

Large-scale structural variations induced by transposable elements promote population-specific divergence in *Triticum turgidum* ssp. *dicoccoides*

Liya Bida

Ben-Gurion University of the Negev

Khalil kashkush

kashkush@bgu.ac.il

Ben-Gurion University of the Negev

Research Article

Keywords: wild emmer wheat, structural variation, transposable elements, population genetics, adaptation, wheat improvement

Posted Date: February 26th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-8854554/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at BMC Plant Biology on June 3rd, 2026. See the published version at <https://doi.org/10.1186/s12870-026-09173-7>.

Abstract

Large-scale structural variations (SVs) are increasingly recognized as important drivers of genome evolution and crop adaptation. In wheat, a young allopolyploid with a history of extensive domestication bottlenecks, SVs remain poorly explored, particularly in its wild progenitors. Here, we analyzed 68 wild emmer wheat (*Triticum turgidum* ssp. *dicoccoides*) accessions representing seven geographically distinct populations across the Fertile Crescent. Comparative genomic analysis between wild emmer and bread wheat (*T. aestivum*) revealed 17 genome-specific large-scale SVs (ranging in size between 1 kb and 1 Mb), many associated with transposable elements (TEs), specifically Gypsy and Copia retrotransposons. PCR-based genotyping validated 16 loci as polymorphic, with some showing strong population specificity. Importantly, most SVs encompassed high-confidence genes, including disease resistance kinases, receptor-like kinases, and stress-response regulators. Our findings demonstrate that wild emmer harbors substantial SV diversity with functional potential for adaptation. These results expand our understanding of wheat genome evolution and highlight population-specific rearrangements as valuable resources for breeding resilient wheat varieties.

Key message

Our findings highlight wild emmer wheat as a reservoir of structural genomic diversity. The population-specific rearrangements identified likely contribute to local adaptation, particularly in stress responses. Structural variants encompassing disease resistance genes or stress regulators represent potential targets for wheat improvement. These results underscore the importance of studying wild progenitors to guide modern breeding strategies.

Introduction

Wheat is one of the most important staple crops globally, providing nearly 20% of the world's caloric intake (Shiferaw et al. 2013). Its significance in global food security cannot be overstated, especially as the global population continues to grow and climate change increasingly impacts agricultural productivity. Bread wheat is a relatively young domesticated allopolyploid species created from two different allopolyploidization events that occurred throughout the evolutionary history of the *Triticum-Aegilops* group of three diploid genomes. The first allopolyploidization event occurred about 0.8 million years ago between a male diploid donor of A subgenome, similar to *Triticum Urartu*, and a female donor of B subgenome related to *Ae. speltoides* led to the creation of the emmer wheat, which is found in nature as a wild species called wild emmer wheat, *Triticum turgidum* ssp. *dicoccoides* ($2n = 4x = 28$, genome composition BBAA). The second allopolyploidization event occurred around 9,000 years ago (Levy and Feldman 2022) between the domesticated emmer as a female genome donor and *Ae. tauschii*, as a male genome donor of the D subgenome, created the bread wheat *Triticum aestivum* ($2n = 6x = 42$, genome BBAAADD) (Petersen et al. 2006).

As wheat allopolyploids are relatively young, they are expected to show limited genetic variation due to the so-called "polyploidy diversity bottleneck." This bottleneck occurs because only a few individuals from the progenitor species were involved in the allopolyploid speciation event. Furthermore, the newly formed allopolyploid quickly becomes reproductively isolated from its parental species. Additionally, there has been insufficient time for accumulating mutations (Feldman and Levy 2012). Nevertheless, allopolyploid species spread faster than their diploid progenitors (Feldman and Levy 2005), as can be seen specifically for the bread wheat, which, soon after its formation and domestication in the South Caspian, was successfully spread worldwide.

The explanation for wheat allopolyploid's success is the mechanism of allopolyploidization. This mechanism is unique in its ability to create new species in a single step, creating two "genetic shocks." The inter-genetic hybridization followed by whole genome duplication led to burst genomic changes, including revolutionary (occurring immediately after allopolyploidization) and evolutionary (occurring during the life of the allopolyploid species) genomic changes in wheat allopolyploids that may not be achieved at the diploid level. These changes include the creation of new traits, new intergenomic interactions, buffering of mutations, epigenetic alteration, gene expression modification, and non-random sequence deletions. Domestication favored allopolyploids in wheat, probably as it allowed for rapid adaptation to the new cultivation environment (Levy and Feldman 2022). Nevertheless, thousands of years of domestication and the involvement of recurrent picking of desirable individuals with desirable traits, such as harvest qualities, led to a significant loss of genetic variation (Bevan et al. 2017; Charmet 2011). The repetitive emphasis on high yields and industrial breeding for standardized characteristics has further diminished the natural resistance of domesticated wheat against biotic (e.g. diseases, pests) and abiotic (e.g. climate change) stress factors (Feldman and Millet 2001; Nevo 2014; Zhou et al. 2020).

Climate change raises the need to investigate the progenitor of the bread wheat, the wild emmer wheat, which has a rich allelic repertoire and can be a potential donor of missing and beneficial traits to the domesticated wheat (Dahan-Meir et al. 2022; Gaurav et al. 2022). Wild emmer wheat (*Triticum dicoccoides*) is the primary progenitor of cultivated tetraploid wheats and contributed the A and B genomes to hexaploid bread wheat (Luo et al. 2007), and the highest biodiversity of wild emmer wheat in Israel, Jordan, southern Syria, and southern Lebanon than in other distribution areas (Nevo and Beiles 1989; Ozbek et al. 2007), it was suggested that wheat was originated in the fertile crescent in the vicinity of Mt. Hermon and the catchment area of the Jordan River (Feldman and Kislev 2007; Özkan et al. 2011). Over the following thousands of years, wild emmer wheat was spread from its original area. Wild emmer can be found and thrive in various environments and can hold habitats with different biotic and abiotic conditions, such as soil types, temperatures, and elevation (from 200m below the sea level in Jordan Valley to 1,600m above sea level in Mt. Hermon in Israel). It can be found in open oak forests, evergreen dwarf shrub formations, pastures, abandoned fields, and on the edge of cultivated fields (Feldman and Kislev 2007; Nevo 2014).

Due to the climatic, topological, and geological diversity existing in Israel, many studies on the genetic diversity of wild emmer populations were done in Israel (Li et al. 2000; Li et al. 2003; Nevo and Beiles 1989; Peleg et al. 2008). Investigating large-scale genomic rearrangements and stability in wheat is crucial for understanding the evolution, diversity, and adaptability of this crop. Such research sheds light on complex genetic structures, including various inversions and translocations, through comparative genomic analyses of different wheat species. These studies reveal critical genomic rearrangements that inform the plant's past adaptations and potential for future resilience. Grasping these genetic intricacies enables the development of wheat varieties with enhanced yields, disease resistance, and stress tolerance, while also preserving genetic diversity. This knowledge is foundational for breeding strategies designed for secure food production in the face of global challenges.

The availability of sequence drafts for wheat species, including bread wheat and wild emmer wheat, facilitated the identification and characterization of large-scale genomic rearrangements (ranging in size from 100 kb to 2 Mb), including insertions/deletions, inversions, translocations, and introgressions. Those rearrangements were associated with allopolyploidization and domestication processes (Bariah et al. 2020a; Dvorak et al. 2018; Huo et al. 2018; Li et al. 2023).

In this study, we have investigated whether large-scale structural variations (SVs) play a role in genetic diversity in 68 accessions of wild emmer wheat, collected from seven different geographically isolated sites, by studying 17

genome-specific large-scale genomic rearrangements. To this end, using PCR screening, we observed polymorphism among wild emmer wheat accessions; in some cases, population-specific patterns were observed. The potential role of these SVs in the adaptation of wild emmer wheat is discussed.

Materials and methods

Plant material

In this study, seeds of 68 wild emmer wheat accessions were grown in a greenhouse at Ben Gurion University of the Negev under common garden conditions. The wild emmer wheat accessions were obtained from four geographically well-defined populations: (1) Mt. Hermon, Israel 8 accessions, (2) Amiad, Israel 10 accessions, (3) Tabgha, Israel 10 accessions, (4) Mt. Amasa, Israel 10 accessions. In addition, seeds were collected from three regional representative collections from (1) Syria, 10 accessions, (2) Lebanon, 10 accessions, and (3) Turkey, 10 accessions. The Syrian, Lebanese, and Turkish accessions were provided by Prof. Hakan Ozkan (Cukurova University) to represent the wider genetic diversity across the northern distribution range of the Fertile Crescent. While the Israeli populations represent specific localized micro-habitats, the regional sets from Syria, Lebanon, and Turkey provide a broader geographical context for identifying SVs that may be absent in the southern Levant. Genomic DNA was extracted from leaves four weeks post-germination using the Geneaid Kit (Geneaid Biotech Ltd., New Taipei City, Taiwan).

The taxonomic classification and nomenclature adopted in this study follow the system where wild emmer wheat is classified as *Triticum turgidum* ssp. *dicoccoides* ($2n = 4x = 28$, genome BBAA) and bread wheat as *Triticum aestivum* ($2n = 6x = 42$, genome BBAADD). This framework is consistent with the genomic assemblies utilized for the comparative analysis, specifically the WEWSeq v1.0 for wild emmer and IWGSC RefSeq v1.0 for bread wheat.

Computer-assisted analysis

Genomic databases:

The genome drafts of two *Triticum* species were used in this study: (1) Wild emmer WEWSeq v1.0 assembly, a whole genome draft of emmer wheat, based on Zavitan accession that was sequenced using whole-genome shotgun sequencing, covering 10.1-Gbp of the genome (87.5%) (Guan et al. 2020) with contigs and scaffold with N50 of 57kb and 7 Mb respectively. (2) Bread wheat IWGSC RefSeq v1.0 assembly, based on *T. aestivum* Chinese Spring accession covering 14.5-Gbp of the genome (97%) (Guan et al. 2020) with contigs and scaffold with N50 of 52kb and 7 Mb, respectively, generated by the International Wheat Genome Sequencing Consortium (IWGSC).

Identification and characterization of large-scale sequence variations in wild emmer wheat vs. bread wheat

Large-scale SVs in bread wheat vs. wild emmer were identified and characterized in a previous study in our lab (see below). In addition, in this study, we utilized publicly accessible databases to identify additional large-scale SVs to have a larger set of SVs cases for PCR screening in wild emmer populations. Here, we used two strategies to identify SVs:

Dyads as genetic markers for large-scale sequence rearrangements: To identify structural variations between bread wheat vs. wild emmer wheat, we used data from publicly available databases (Juery et al. 2021). We specifically focused on pairs of dyads where one copy is in the B subgenome while the other copy is in the D subgenome of the bread wheat genome. Further, ortholog genes were searched in wild emmer wheat using the online database mining tool Ensembl plants Biomart (<http://plants.ensembl.org/biomart/martview>). Groups of genes that were assigned as dyads (two copies) but had an additional homeolog copy assigned by Biomart were excluded from the analysis. Subsequently, dyad groups with ortholog genes in both A and B subgenomes of wild emmer wheat were further analyzed, suggesting an SV within A subgenome of the wild emmer genome. Flanking sequences of the candidate orthologs of wild emmer in A subgenome were aligned to bread wheat chromosome A.

Note that some cases of large-scale sequence rearrangements were randomly identified. In this analysis, we utilized genes that exist as dyads in the B and D subgenomes of bread wheat, focusing particularly on their orthologs in wild emmer wheat. For the B subgenome of bread wheat, the dyad member gene was used as a marker; however, the search was not specifically aimed at identifying rearrangements within its orthologous region. Instead, the gene served as a broader marker for a genomic region that includes genes, without excluding the possibility of rearrangements. To thoroughly explore this area, chromosome walking was employed, extending up to 0.5 Mbp on each side of the marker gene, allowing for a detailed examination of the surrounding genomic landscape. In this case, the reason for aligning specifically B subgenome is that comparing the gene order in the A, B, and D subgenomes of bread wheat shows that the B genome appears to be the least organized, with less sequence order conservation (Mirzaghaderi and Mason 2017).

Identification of breakpoints and characterization of large-scale SVs: To polarize the evolutionary direction of the identified SVs (i.e., to distinguish between insertions/introgressions and deletions), we utilized the ancestral state as a reference. By comparing the wild emmer (BBAA) and bread wheat (BBAADD) sequences to the known diploid progenitors and orthologous subgenomes, we categorized events where a sequence was present in the wild emmer but absent in bread wheat as deletions in the bread wheat lineage. Conversely, sequences present in only one species and lacking orthology in the progenitors were classified as potential introgressions or recent TE-mediated insertions. Alignments were performed with BLAST+ stand-alone version 2.9.0, using an e-value less than e^{-100} . In cases where the orthologous locus remained unidentified, a chromosome walking approach was utilized, such that longer flanking sequences of the ortholog gene in wild emmer were aligned to the orthologous chromosome from the bread wheat genome. After identifying the orthologous genome locus in bread wheat, dot plot alignments of the genome loci were performed to create graphical representations and identify structural variations. This is by using UGENE software (Version 49.1, Unipro UGENE, 2023; available at <http://ugene.net/>). using a minimum repeat length of 100bp and 95% repeat identity. The aligned sequences that were identified as an indel (insertion, deletion, or insertion and deletion of a genomic sequence) were further analyzed to identify the breakpoints (the ends of the sequence similarity). Borders between highly similar sequence regions (with 95% sequence identity or higher for a word size of 100) and low similar sequence regions (with lower than 95% sequence identity for a word size of 100) were identified as indel breakpoints by representing suspected large-scale sequence variation in orthologous loci between the wild emmer and bread wheat with a dot plot alignment. Locus lengths were defined as the distances between the 5' and the 3' breakpoints.

Next, to characterize the found SVs, breakpoints and deleted and inserted sequences were annotated to genes and TEs. Gene identifications were performed using EnsemblPlants genome browser for both wild emmer and bread wheat (WEWSeq v.1.0 and IWGSC v.1, respectively) (Bolser et al. 2016). Sequence similarity searches were conducted using the National Center for Biotechnology Information (NCBI) BLASTx tool with standard databases

(<https://blast.ncbi.nlm.nih.gov/Blast.cgi>). TE annotation was performed using CENSOR software tool on the Giri website (Kohany et al. 2006).

PCR analysis – validation of identified SVs and scanning of SVs within wild emmer populations

To validate and scan the identified genomic rearrangements in wild emmer and bread wheat, several steps were undertaken. Primers were designed using the online tool PRIMER3 version 4.1.0, based on the flanking sequences of identified breakpoints for each rearrangement (see Supplemental Table 1 for primer sequences). These primers targeted specific loci in bread wheat accession (Chinese spring 42, CS42) and/or wild emmer accession (Zavitan, Z) to differentiate their genomic DNA amplification through site-specific PCR analysis based on the newly discovered SVs. Following the validation of existing SVs between Zavitan and Chinese Spring 42, these SVs were screened across seven different populations of wild emmer to assess the distribution of SVs within and between populations. This additional step was crucial to determine whether each SV was unique or specific to some populations, or broadly distributed across them, meaning non-specific, thereby providing insights into the evolutionary dynamics and genetic diversity among these populations. The reaction mixture for each PCR included 10µl PCRBIO HS Taq Mix Red, 7µl ultrapure water, 1µl forward primer (10µM), 1µl reverse primer (10µM), and 1µl of genomic DNA template (20 ng/µl), with the conditions set at 95°C for 2 min, followed by 35 cycles of 95°C for 15 sec, the calculated annealing temperature for 15 sec, and 72°C for 15 sec. The PCR products were then visualized using 1% agarose gel. Figures were prepared with GIMP (version 2.10.32; <https://www.gimp.org/>), and Microsoft PowerPoint (www.microsoft.com).

Results

Identification and molecular characterization of large-scale structural variations between bread and wild emmer wheats

To elucidate the genetic diversity and evolutionary dynamics between wild emmer and bread wheat, and within wild emmer wheat accessions, 17 cases of large-scale sequence rearrangements were identified and analyzed (Table 1): (1) Six cases (5B1, 5B2, 5B3, 5B5, 3B2, 3B4) were previously reported by our group (Bariah et al., 2020a), (2) Nine cases (5A11, 5A12, 5A13, 3A14, 3A16, 2A23, 2A24, 7A28, 7A30) were discovered based on dyads markers, and (3) two cases (5B18, 5B19) were identified through a comparative analysis of subgenomes. Note that each locus name is based on chromosome number and genome type, for example, 5B2—the rearrangement found in chromosome 5 of subgenome BB.

Table 1

In silico characterization of large-scale sequence variations identified in bread wheat vs. wild emmer genomes

Wild emmer locus ID ¹	Location in chromosome		Locus length ⁴ (bp)		Type of rearrangement
	Wild emmer ²	Bread wheat ³	Wild emmer	Bread wheat	
5A11	560526756–560575832	565576601–565583682	49076	7081	introgression of new DNA fragment in wild emmer, TE insertional polymorphism
5A12	579641981 – 579641974	584458235–584489770	8	31535	TE insertions
5A13	613959094–613972669	619778803 – 619778795	13575	9	deletion in bread wheat
3A14	14255227–14285277	426135–426139	30050	4	deletion in bread wheat
3A16	651508154–651549596	655634201–655634841	41458	640	deletion in bread wheat
2A23	715408979–715464345	722749722 – 722749560	55366	163	deletion in bread wheat
2A24	53632794–53637564	50957599–50957608	4770	9	deletion in bread wheat
7A28	95718208–95724258	97909675 – 97909673	6050	3	deletion in bread wheat
7A30	448353735–448382339	451518935–451518988	28604	53	deletion in bread wheat
5B18	25591195–26509839	26952358–27151545	918644	199187	introgression of new DNA fragment in wild emmer
5B19	532211441–533072412	527259099–528123920	860971	864821	introgression of new DNA fragment in bread wheat
*5B1	566939353–567135082	561057394–561064945	195730	7552	deletion in bread wheat
*5B2	516702290–516721374	511383608–511385023	19085	1416	deletion in bread wheat
*5B3	363487255–363551431	349934346–349934349	64177	4	deletion in bread wheat
*5B5	610009239–610050693	603942312–603952982	41455	10671	introgression of new DNA fragment
*3B2	284755035–284771490	286353814–286353819	16456	6	deletion in bread wheat
*3B4	538946011–540047920	527682008–527682029	1101910	22	deletion in bread wheat

*Previously found cases by Inbar Bariah (Bariah et al., 2020).

Wild emmer locus ID ¹	Location in chromosome		Locus length ⁴ (bp)		Type of rearrangement
	Wild emmer ²	Bread wheat ³	Wild emmer	Bread wheat	

¹The first number and letter are referred to the chromosome in which the genomic locus is found.

²Coordinates based on WEWSeq v1.0 (<http://wewseq.wix.com/consortium>)

³Coordinates based on RefSeq v1.0 (http://plants.ensembl.org/Triticum_aestivum/Info/Index)

⁴Locus length was determined as the genetic distance between the breakpoints of the sequence variation between wild emmer and bread wheat identified using dot plot alignment with minimum repeat length of 100bp and 95% repeat identity (see Fig. S1a-h)

Identification of SVs using dyads as genetic markers

Since bread wheat originated from three different diploid genomes, it is considered that most of the genes exist in three pairs of homoeologous genes, known as triads. However, triads represent only half of the wheat genes, which exist in various copies, such as dyads— refer to homeologous genes that exist in two out of the three bread wheat subgenomes (AA, BB, and DD). Dyads are mostly located in the distal regions/parts of the chromosomes, making them more prone to rearrangements and, therefore, good markers for detecting rearrangements. Additionally, it was found that dyads are enriched in traits and functions related to adaptations, making them good candidates for gene transfer to improve wheat's survival and success. Overall, 1,738 dyads found in all seven subgenomes (Juery et al. 2021) were analyzed using a chromosome walking approach and dot plot comparison.

To identify structural variations between bread wheat vs. wild emmer wheat, we used publicly available databases (Juery et al. 2021). We specifically focused on pairs of dyads where one copy is in the B subgenome while the other copy is in the D subgenome of the bread wheat genome. Further, ortholog genes were searched in wild emmer wheat using the online database mining tool Ensembl plants Biomart (<http://plants.ensembl.org/biomart/martview>). Subsequently, dyad groups with ortholog genes in both A and B subgenomes of wild emmer wheat were further analyzed, suggesting an SV within the A subgenome of the wild emmer genome. Flanking sequences of the candidate orthologs of wild emmer in A subgenome were aligned to bread wheat chromosome A. We utilized genes that exist as dyads in bread wheat's B and D subgenomes, focusing particularly on their orthologs in wild emmer wheat. For the B subgenome of bread wheat, the dyad member gene was used as a marker, but, the search was not specifically aimed at identifying rearrangements within its orthologous region. Instead, the gene served as a broader marker for a genomic region that includes genes, without excluding the possibility of rearrangements. To thoroughly explore this area, chromosome walking was employed, extending up to 0.5 Mbp on each side of the marker gene, allowing for a detailed examination of the surrounding genomic landscape. In this case, the reason for aligning specifically B subgenome is that comparing the gene order in the A, B, and D subgenomes of bread wheat shows that the B genome appears to be the least organized, with less sequence order conservation (Mirzaghaderi and Mason 2017). To this end, a total of 1,738 dyads were analyzed using a chromosome walking approach and dot plot comparison. After identifying breakpoints, primers were designed from the deletion/insertion (indel) breakpoint regions to be amplified in wild emmer (Zavitan accession) or bread wheat (Chinese Spring). These primers were then used for PCR screening to detect rearrangements between and within populations. Overall, nine cases of large-scale sequence variation were detected and analyzed (5A11, 5A12, 5A13, 3A14, 3A16, 2A23, 2A24, 7A28, 7A30; Table 1).

In two of the nine loci (5A11 and 5A12, Table 1), a high percentage of TE insertions between the 5' and 3' ends of the indel was detected in wild emmer and bread wheat. Dot plot comparison of genomic locus 5A12 from bread wheat chromosome 5A and wild emmer chromosome 5A (Fig. S1a) revealed a 31.5kbp segment consisting of 92% of transposable elements, with ~ 85% belonging to the Copia superfamily in bread wheat. The flanking sequences of the 5A12 locus in wild emmer wheat were aligned with the corresponding orthologous sequence in bread wheat, showing an interruption of a high confidence (HC) gene coding segment (TRIDC5AG056530) that shows a high sequence identity (> 80%) when only exons are blasted for rust resistance kinase LR10-like from *Triticum urartu* in wild emmer wheat due to TE insertions in bread wheat. LR10 is a gene that confers resistance to leaf rust, caused by the fungus *Puccinia triticina*. This disease is a significant concern for farmers because it can drastically reduce both the yield and quality of wheat harvests, causing yield losses sometimes approaching > 50% (Riaz and Wong 2017). The gene encodes a protein with a CC-NBS-LRR structure, which is a common motif in plant resistance proteins (Feuillet et al. 2003). In this case, the 31.5kbp segment of the bread wheat is flanked by 25bp inverted sequences, annotated as DNA-TA-1_TM TE, which is part of the Mariner superfamily (Class II, 'cut & paste' mechanism) (Fig. 1). Further analysis and BLAST results indicate that these are part of terminal inverted repeats (TIRs). This may suggest a mechanism of action similar to what is described in (Gray 2000), where two parts of a transposon are associated with initiating the insertion of a large sequence into a genome. The presence of only half of a TIR at each end suggests potential mutations or alterations in the typical Mariner TIR structure, which may disable the normal transposase action of 'cut and paste' but allow for alternative transpositional events that result in complex rearrangements.

Additionally, dot plot comparison of the genomic locus 5A11 from bread wheat chromosome 5A and the orthologous locus in wild emmer (Fig. S1b) chromosome 5A revealed a 7kbp segment consisting of 81.5% transposable elements, with ~ 94% of the TE belonging to the Gypsy superfamily in bread wheat (this calculation includes an unread fragment inside one of the Gypsy TEs). This indel involved the replacement of a ~ 49kbp segment in wild emmer, which consists of 79% TE, with the 7kbp segment in bread wheat. The ~ 49kbp segment in wild emmer includes a high confidence (HC) protein-coding gene (TRIDC5AG053440) that shows a high sequence identity (> 93%) to a wall-associated receptor kinase 5-like gene from *Triticum aestivum* (Fig. S2a). Wall-associated receptor kinases are involved in various processes, including pathogen defense and cell wall metabolism, which are crucial for plant survival. Blast analysis showed that the 7kbp segment existing in bread wheat also exists in *Triticum turgidum* Svevo, domesticated emmer wheat, and may indicate the event occurring during the evolution of the tetraploid wheat, perchance during the domestication. SVs, such as 5A11 and 5A12, that are enriched with TEs represent more than simple 'cut and paste' transposition; the presence of inverted repeats (TIRs) and motifs associated with MMEJ (microhomology-mediated end joining) suggests that these TEs acted as 'anchors' or triggers for larger-scale genomic rearrangements. This demonstrates that TEs in wild emmer are not merely parasitic elements but are active drivers of large-scale structural diversity.

In a previous work (Bariah et al. 2020a), we obtained similar patterns at 3B4, 5B1, and 5B5 loci (Table 1). Analysis of 7 different loci (5A13, 3A14, 3A16, 2A23, 2A24, 7A28, and 7A30, Table 2) showed large-scale genomic rearrangements with a missing segment in bread wheat vs. wild emmer wheat. A ~ 13.5kb segment consisted of 42.5% of TEs in wild emmer chromosome 5A, locus 5A13 was absent in the orthologous genomic locus in the bread wheat genome (Fig. S1c). The orthologous locus in bread wheat was identified by the flanking sequences alignment. The missing 13.5kb sequence includes an HC gene for an uncharacterized protein (TRIDC5AG062850). Additionally, a dot plot comparison revealed that the 13.5kb segment was flanked by a 3bp motif 'AGA' whereas in bread wheat, this motif existed in only one copy between the conserved sequences flanking the 5A13 locus (Fig. S2b). This sequence signature is typical for double-strand break (DSB) repair via microhomology-mediated end

joining (MMEJ) and may indicate that a DSB occurred within the wild emmer sequence. The signature suggests repair of the DSB involved endonuclease activity utilizing these repeats as micro-homology (Khodaverdian et al. 2017; Ranjha et al. 2018). Also, 5B3 loci showed the same DSB repair mechanism with 'A' mononucleotide appearing at both the 5' and 3' ends of 64kb segment in wild emmer while the orthologous locus, missing that 64kb segment, in bread wheat showed a single copy of the 'A' mononucleotide (Bariah et al. 2020a).

Table 2

PCR – based screening of large-scale sequence variations wild emmer wheat accessions from seven populations

Wild emmer locus ID	Species ²	Mt. Hermon	Amiad	Mt. Amasa	Tabgha	Turky	Syria	Lebanon	Status of the rearrangement ¹
5A11	bread wheat	0	0	0	0	2	2	0	absent in some populations and polymorphic in others-population specific
5A12	wild emmer	2	1	1	1	2	2	2	polymorphic in some populations
	bread wheat	2	0	0	0	2	2	2	absent in some populations and polymorphic in others
5A13	wild emmer	1	1	1	1	1	1	1	monomorphic in all populations
	bread wheat	0	0	0	0	0	0	0	absent in all populations
3A14	wild emmer	2	2	2	2	2	2	0	absent in one population and present in others
	bread wheat	0	0	0	0	0	0	0	absent in all populations
3A16	wild emmer	0	2	0	2	2	2	0	absent in some populations and polymorphic in others
	bread wheat	0	0	0	0	0	0	0	absent in all populations
2A23	wild emmer	2	2	1	2	2	2	2	monomorphic in one population
	bread wheat	0	0	0	0	2	2	0	absent in some populations and

¹The specificity of a rearrangement is determined based on its occurrence in PCR analysis within different populations.

²Primers designed to amplify the specific species locus. A population-specific rearrangement refers to a sequence that appears in at most 2 populations. Cell value of the matrix: 0- No amplification observed in any accession of the population. 1- Amplification observed in all accessions of the population. 2- Partial amplification was observed in some accessions of the population.

Wild emmer locus ID	Species ²	Mt. Hermon	Amiad	Mt. Amasa	Tabgha	Turky	Syria	Lebanon	Status of the rearrangement ¹
									polymorphic in others- population specific
2A24	wild emmer	1	1	1	1	2	1	1	polymorphic in one population
	bread wheat	0	0	2	0	2	2	0	absent in some populations and polymorphic in others
7A28	wild emmer	1	1	2	1	2	2	2	polymorphic in some populations
	bread wheat	0	0	0	0	2	2	0	absent in some populations and polymorphic in others- population specific
7A30	wild emmer	1	1	1	2	1	1	2	polymorphic in two populations
	bread wheat	0	0	0	0	0	0	0	absent in all populations
5B18	wild emmer	1	2	2	2	2	2	2	monomorphic in one population
	bread wheat	1	2	1	2	1	2	2	polymorphic in some populations
5B19	wild emmer	0	2	0	2	0	0	0	absent in some populations and polymorphic in others- population specific
	bread wheat	0	2	2	0	2	2	2	absent in some populations and

¹The specificity of a rearrangement is determined based on its occurrence in PCR analysis within different populations.

²Primers designed to amplify the specific species locus. A population-specific rearrangement refers to a sequence that appears in at most 2 populations. Cell value of the matrix: 0- No amplification observed in any accession of the population. 1- Amplification observed in all accessions of the population. 2- Partial amplification was observed in some accessions of the population.

Wild emmer locus ID	Species ²	Mt. Hermon	Amiad	Mt. Amasa	Tabgha	Turky	Syria	Lebanon	Status of the rearrangement ¹
									polymorphic in others
5B1	wild emmer	0	2	0	2	2	0	1	absent in some populations and polymorphic in others
5B2	wild emmer	1	2	2	1	2	2	2	polymorphic in some populations
5B3	wild emmer	1	1	1	1	2	2	2	polymorphic in some populations
	bread wheat	2	0	2	0	2	2	2	absent in some populations and polymorphic in others
3B4	bread wheat	1	2	1	2	1	2	2	polymorphic in some populations
3B2	wild emmer	1	1	1	1	1	2	2	polymorphic in some populations
5B5	wild emmer	0	2	0	1	0	0	0	absent in some populations and polymorphic in others
	bread wheat	1	2	1	2	1	1	2	polymorphic in some populations
¹ The specificity of a rearrangement is determined based on its occurrence in PCR analysis within different populations.									
² Primers designed to amplify the specific species locus. A population-specific rearrangement refers to a sequence that appears in at most 2 populations. Cell value of the matrix: 0- No amplification observed in any accession of the population. 1- Amplification observed in all accessions of the population. 2- Partial amplification was observed in some accessions of the population.									

An additional ~ 30kb in wild emmer wheat, chromosome 3A, locus 3A14 was missing in the orthologous genomic locus in bread wheat (Fig. S1d). The indel includes two HC protein-coding genes within the wild emmer genome. TRIDC3AG003040 shows high sequence similarity for Predicted *Triticum dicoccoides* protein EXECUTER 2, chloroplastic-like (LOC119266574), mRNA (GO term shows singlet oxygen-mediated programmed cell death), and TRIDC3AG003030 that show high sequence similarity for homeobox-leucine zipper protein HOX15-like (Fig. S2C).

~ 41kb fragment include ~ 62% of TE with ~ 53% Gypsy superfamily exists in wild emmer chromosome 3A' locus 3A16 was missing in bread wheat (Fig. S1e). 560bp upstream to the 3' breakpoint, there is a 11.8kb TE from Gypsy superfamily. A common fragment of 605bp was found in both wild emmer and bread wheat genomes between the breakpoints. The 41kb missing fragment contains two HC coding genes, the first, TRIDC3AG058490, showing high sequence identity (> 90%) for anthranilate O-methyltransferase 3-like protein from *Triticum dicoccoides* and the second, TRIDC3AG058500, showing high sequence identity (100%) for 2'-deoxymugineic-acid 2'-dioxygenase-like protein (Fig. 2).

Dot plot comparison of genomic locus 2A23 (Fig. S1f). from bread wheat chromosome 2A and wild emmer chromosome 2A revealed ~ 55kb segment, missing in bread wheat, which consisted of 63.4% of TEs while ~ 54% belong to Gypsy superfamily and ~ 25% belong to NonLTR/L1 superfamily in wild emmer wheat. The ~ 55kb segment includes two HC genes; TRIDC2AG068840 that show high sequence similarity (100% of identity with 73% coverage, when blast only exons) to senescence-specific cysteine protease SAG39-like from *Triticum dicoccoides* and TRIDC2AG068810 that show high sequence similarity (90.66% of identity with 91% coverage, when blast only exons) mitochondrial transcription factor 1 from *Triticum aestivum* (Fig. S2d).

Dot plot comparison of genomic locus 2A24 (Fig. S1g) from bread wheat chromosome 2A and wild emmer chromosome 2A revealed ~ 4.7kbp segment, missing in bread wheat, that consisted of ~ 14.4% of transposable elements, while ~ 31% belonged to Mariner superfamily in wild emmer. This segment contains a non-coding protein gene (ENSRNA050007490), tRNA-Gly for anticodon UCC.

Dot plot comparison of genomic locus 7A28 (Fig. S1h) from bread wheat chromosome 7A and wild emmer chromosome 7A revealed ~6kbp segment, missing in bread wheat, that consisted of ~ 27.2% of transposable elements while ~ 36.7% belonged to Helitron superfamily and ~ 34% to NonLTR/L1 superfamily in wild emmer. No genes were found.

Dot plot comparison of genomic locus 7A30 (Fig. S1i) from bread wheat chromosome 7A and wild emmer chromosome 7A revealed a ~ 28.6kbp segment, consisting of ~ 29% transposable elements, while ~ 64% belonged to the Copia superfamily in wild emmer wheat. One HC protein-coding gene is found within the 28.9kb fragment, TRIDC7AG044240, which shows high sequence similarity to four isoforms of vacuolar protein sorting-associated protein 13b-like in *Triticum dicoccoides*. This gene covers 96.5% of the 28.9kbp segment, with a portion of the gene remaining in the 5' region of the indel in bread wheat (Fig. S2e).

Identification of SVs using comparative analysis

Here, two different cases of large-scale genomic rearrangements were discovered through comparative analysis. A dot plot comparison of the genomic locus 5B18 (Fig. S1j) from the bread wheat chromosome 5B and the wild emmer chromosome 5B revealed the replacement of a ~ 918.6kb segment from wild emmer, which consisted of 75.9% transposable elements, while ~ 61.7% belong to the Gypsy superfamily and ~ 13.9% to the EnSpm superfamily in the wild emmer wheat genome. This segment was replaced with a ~ 199.1kb segment in the bread wheat genome, consisting of 77% transposable elements. Within this segment, about 64.7% belong to the Gypsy superfamily and ~ 21.4% to the Copia superfamily. The ~ 918.6kb segment belongs to wild emmer wheat, contains one HC protein-coding gene, TRIDC5BG004390, that shows high sequence similarity (100% of identity with 97% coverage) to a predicted *Triticum dicoccoides* (wild emmer) F-box/FBD/LRR-repeat protein At1g13570-like protein. The ~ 199.1kb segment belonging to bread wheat contains an HC protein-coding gene, TraesCS5B02G028100, that also show high sequence similarity (100% of identity with 77% coverage, when blast only exons) to F-box/FBD/LRR-

repeat protein At1g13570-like in *Triticum aestivum* (bread wheat). In addition, this locus contains an HC protein-coding gene, TraesCS5B02G02800, that shows high sequence similarity (100% of identity with 64% coverage, when blast only exons) to acyl-coenzyme A thioesterase 13-like from *Triticum aestivum* (Fig. 3).

Dot plot comparison of the genomic locus 5B19 (Fig. S1k) from the bread wheat chromosome 5B and the wild emmer chromosome 5B revealed the replacement of a ~ 860 kb segment from wild emmer, which consisted of 75% transposable elements, while 48.3% belong to the Gypsy superfamily and ~ 21.4% to the Copia superfamily in the wild emmer wheat genome. This segment was replaced with a ~ 864 kb segment in the bread wheat genome, consisting of 77.33% transposable elements. Within this segment, about 54% belonged to the Gypsy superfamily and ~ 18.1% to the Copia superfamily. The ~ 860kb segment belonging to wild emmer wheat, contains an HC protein-coding gene (TRIDC5BG054370) that show a high sequence identity (100% identity with 67% coverage, based on exon sequences) with disease resistance protein RGA4-like (resistance gene analogs 4) from *Triticum dicoccoides*. An additional HC protein-coding gene, TRIDC5BG054400, that shows high sequence similarity to uncharacterized protein LOC119311448 isoform X1 in *Triticum dicoccoides* (100% identity with 94% coverage, based on exon sequences) and disease resistance protein RGA5-like isoform X2 from *Triticum dicoccoides* (100% of identity with 81% coverage, based on exon sequences). Another HC protein-coding gene, TRIDC5BG054470 shows high sequence similarity (92% of identity with 85% coverage) to auxin-responsive protein SAUR36-like from *Triticum dicoccoides*. Additionally, to other HC protein-coding proteins that showed high sequence similarity to uncharacterized proteins; TRIDC5BG054430 to LOC119307060 and TRIDC5BG054460 to LOC119306066 in *Triticum dicoccoides*. In contrast to wild emmer segment, the ~ 864kb segment belonging to bread wheat, contains three HC protein-coding genes that show high sequence similarity to auxin-responsive protein SAUR36-like from *Triticum aestivum*; TraesCS5B02G345000 (100% identity with 99% coverage), TraesCS5B02G344700 (100% identity with 68% coverage) and TraesCS5B02G344800 (99% identity with 76% coverage). Similar to the orthologous locus in wild emmer, this locus also contains an HC protein-coding gene, TraesCS5B02G344300, that shows high sequence similarity (100% of identity with 69% coverage, based on exon sequences) to disease resistance protein RGA4-like in *Triticum aestivum*. Also, similar to the wild emmer locus, this locus contains an HC protein-coding gene, TraesCS5B02G344100, that shows high sequence similarity to disease resistance protein RGA5-like isoform X2 from *Triticum aestivum*, and additional HC protein-coding gene, TraesCS5B02G344200, that has a sequence similarity to predicted *Triticum aestivum* disease resistance protein RGA5-like (LOC123112960). Also, it contains hypothetical protein CFC21_073049 TraesCS5B02G344500, hypothetical protein CFC21_073044, TraesCS5B02G344000, and three uncharacterized proteins: TraesCS5B02G344600, TraesCS5B02G344900, and TraesCS5B02G345100 (Fig. S2f).

Validation and PCR screening of the identified large-scale sequence rearrangement

PCR-based analysis of the observed 17 SVs between wild emmer vs. bread wheat was performed. The PCR analysis included validation of the indel breakpoints, screening the 68 wild emmer wheat accessions collected from seven different wild emmer populations. Details on primer sequencing and PCR conditions can be seen in supplement Table 1. Note that all PCR reactions were also performed in bread wheat and wild emmer accession Zavitan for validation. Table 2 summarizes the PCR screening in all wild emmer wheat populations.

Locus 5A11

An orthologous locus corresponding to the introgression of a new DNA fragment in wild emmer wheat. The PCR primers were designed from the introgression breakpoints to discriminate between wild emmer wheat accessions with and without the introgression. Amplification (Fig. 4) was detected only in nine accessions from the Turkish population and two accessions from the Syrian population (Table 2), meaning it is polymorphic only in these populations. These results indicate that this rearrangement is population-specific to both the Turkish and Syrian populations.

Locus 5A12

PCR reaction using primers specifically designed from the indel breakpoint was performed in all wild emmer wheat accessions. The populations from Mt. Amasa, Amiad, and Tabgha show monomorphism of this locus, while the other populations exhibit polymorphism. Amplification was observed in five accessions from Mt. Hermon, Syria, and Turkey, and seven of Lebanon populations (Fig. S3a1). Additionally, a second PCR reaction was conducted using primers specifically designed to fit bread wheat, one primer targeting the breakpoint and the other the indel fragment. This revealed polymorphism at the same locus in the populations from Mt. Hermon, Turkey, Syria, and Lebanon. Amplification was observed in three accessions from Mt. Hermon, five accessions each from Turkey and Syria, and two accessions from Lebanon. In contrast, the other three populations showed no amplification, probably indicating the absence of this locus (Fig. S3a2). The results from both sets of primers are complementary to each other in all populations except for Lebanon, where the patterns differed.

Locus 5A13

PCR reaction using primers specifically designed from the indel breakpoint was performed in all wild emmer wheat accessions. The forward primer was designed to pair with the flanking sequence at the 5' end of the breakpoint, and the reverse primer was designed to target the ~ 13kb segment of wild emmer wheat. Amplification was observed in all accessions across all seven populations, indicating monomorphism at this locus (Fig. 5). Moreover, a separate PCR reaction was carried out using primers specifically designed to amplify the bread wheat allele. No amplification was detected in any accession across all populations, likely indicating the absence of this sequence in all tested accessions. These results align with the findings from the wild emmer-specific primers, further supporting the monomorphism observed at this locus.

Locus 3A14

PCR reaction using primers specifically designed from the indel breakpoint was performed in all wild emmer wheat accessions. The forward primer was designed to pair with the flanking sequence at the 5' end of the breakpoint, and the reverse primer was designed to target the ~ 30kb segment of wild emmer wheat. Here, the Lebanese population shows monomorphism (lack of PCR amplification) in all accessions, while the remaining populations show polymorphism of this locus. Amplification was observed in one accession from Amiad and one from Mt. Amasa, nine accessions in Tabgha and Turkey, five in Mt. Hermon, and two in Syria (Fig. S3b1). Additionally, a second PCR reaction using primers specifically designed to amplify the bread wheat allele based on the breakpoints was conducted. In this case, no amplification was observed in any accession across all populations (Fig. S3b2).

Locus 3A16

PCR reaction using primers specifically designed from the indel breakpoint was performed in all wild emmer wheat accessions. The forward primer was designed to target the ~ 41kb segment of wild emmer wheat, and the reverse

primer was designed to pair with the flanking sequence at the 3' end of the breakpoint. The populations from Mt. Hermon, Mt. Amasa, and Lebanon showed monomorphism for this locus, with no amplification observed in any accessions. Amplification was observed in seven accessions from Tabgha, two from Syria and Amiad, and three from Turkey (Fig. S3c1). Furthermore, a subsequent PCR reaction was carried out using primers based on breakpoints designed to fit the bread wheat allele. In this case, no amplification was observed in any accession across all populations, suggesting the possible absence of the bread wheat allele in these wild emmer accessions (Fig. S3c2).

Locus 5B18

PCR reaction using primers specifically designed from the indel breakpoint was performed in all wild emmer wheat accessions. The forward primer was designed to align with the flanking sequence at the 5' end of the breakpoint, while the reverse primer was designed to target the ~ 918kb segment of wild emmer wheat. The Mt. Hermon population showed monomorphism at this locus, with amplification observed in all accessions. The rest of the populations show polymorphism with amplification detected in nine accessions from Tabgha and Mt. Amasa, eight from Turkey, seven from Lebanon, three from Syria, and two from Amiad (Fig. S3d1). A subsequent PCR reaction was carried out to amplify the bread wheat allele. The forward primer was designed to align with the flanking sequence at the 5' end of the breakpoint, while the reverse primer targeted the ~ 199kb segment of bread wheat. Amplification was observed in all accessions from Mt. Hermon, Mt. Amasa, and Turkey, as well as in nine accessions from Amiad, five from Tabgha, and three from Lebanon. Interestingly, it appears that many accessions carry both alleles, except for one accession from Lebanon (Lebanon 4) and one from Syria (Syria 3), where neither allele was detected (Fig. S3d2).

Locus 5B19

PCR reaction using primers specifically designed from the indel breakpoint was performed in all wild emmer wheat accessions. The forward primer was designed to align with the flanking sequence at the 5' end of the breakpoint, while the reverse primer was designed to target the ~ 860kb segment of wild emmer wheat. Amplification was detected only in nine accessions from the Tabgha population and two accessions from the Amiad population (Table 2), meaning it was polymorphic only in these populations (Fig. S3e1). These results indicate that this rearrangement is population-specific for both the Tabgha and Amiad populations. Additionally, a PCR reaction was conducted to amplify the bread wheat allele. The forward primer was designed to align with the flanking sequence at the 5' end of the breakpoint, while the reverse primer targeted the ~ 864kb segment of bread wheat. Amplification was observed in nine accessions from Mt. Amasa, two from Amiad, and three from Syria, as well as in three from Turkey and one from Lebanon. Interestingly, no amplification was detected in Tabgha or Mt. Hermon populations (Fig. S3e2).

Locus 2B23: PCR reaction using primers specifically designed from the indel breakpoint was performed in all wild emmer wheat accessions. The forward primer was designed to align with the flanking sequence at the 5' end of the breakpoint, while the reverse primer was designed to target the ~ 55kb segment of wild emmer wheat. In Mt. Amasa, the locus showed monomorphism, with amplification observed in all accessions. Amplification results for other populations were as follows: seven accessions in Mt. Hermon, eight in both Tabgha and Turkey, four in Syria, three in Lebanon, and two in Amiad (Fig. S3f1). A subsequent PCR reaction was conducted using primers based on the breakpoints to amplify the bread wheat allele. Amplification was observed only in two accessions from Turkey and one from Syria, while the remaining accessions showed no amplification (Fig. S3f2).

Locus 2A24

PCR reaction using primers specifically designed from the indel breakpoint was performed in all wild emmer wheat accessions. The forward primer was designed to the flanking sequence at the 5' end of the breakpoint, whereas the reverse primer was located at the ~4.7kb segment of wild emmer wheat. This locus was monomorphic in six populations, while the Turkey population showed polymorphism with amplification seen in seven accessions. A follow-up PCR reaction was carried out using primers based on the breakpoints to amplify the bread wheat allele. Amplification was detected in one accession from Mt. Amasa, four from Turkey, and three from Syria, while the remaining accessions showed no amplification.

Locus 7A28

PCR reaction using primers specifically designed from the indel breakpoint was performed in all wild emmer wheat accessions. The forward primer was designed from the flanking sequence at the 5' end of the breakpoint, whereas the reverse primer was located in the ~6kb segment of wild emmer wheat. Monomorphism was observed with amplification present in all accessions from Mt. Hermon, Tabgha, and Amiad, while the remaining populations show polymorphism with amplification detected in three accessions from Mt. Amasa, five accessions from Turkey, five from Syria, and nine from Lebanon (Fig. S3g1). Subsequently, a PCR reaction was conducted using primers based on the breakpoints to amplify the bread wheat allele. Amplification was observed in five accessions each from Syria and Turkey, while the other populations showed no amplification (Fig. S3g2).

Locus 7A30

PCR reaction using primers specifically designed from the indel breakpoint was performed in all wild emmer wheat accessions. The forward primer was designed to the flanking sequence at the 5' end of the breakpoint, while the reverse primer was located at the ~28kb segment of wild emmer wheat. Polymorphism was observed only in the Tabgha and Lebanon populations, with amplification detected in 9 out of 10 accessions. In contrast, the remaining populations exhibited monomorphism, with amplification noted in all accessions. Note that some accessions exhibited a lower PCR product size than expected (Fig. S3h) (approximately 500bp instead of 534bp). Sequencing and subsequent analysis of this shorter product revealed the absence of a TE from the Helitron superfamily in the smaller fragment (Fig. S3h1). A follow-up PCR reaction was conducted using primers based on the breakpoints to amplify the bread wheat allele. However, no amplification was observed in any of the accessions across all populations, suggesting the possible absence of this allele in these wild emmer accessions (Fig. S3h2).

Locus 5B1

PCR reaction using primers specifically designed from the indel breakpoint was performed in all wild emmer wheat accessions. The forward primer was located on a sequence absent in bread wheat, and a reverse primer specific to the wild emmer 3' flanking sequence of the SV (Bariah et al. 2020a). Monomorphism at this locus was observed in the populations from Mt. Hermon, Mt. Amasa, and Syria, with no amplification detected in any accessions. However, amplification was seen in nine accessions from Tabgha, seven from Amiad, six from Turkey, and three from Lebanon (Fig. S3i).

Locus 5B2

PCR reaction using primers specifically designed from the indel breakpoint was performed in all wild emmer wheat accessions. The forward primer was located on a sequence absent in bread wheat, and a reverse primer was

located at the 3' flanking sequence of the SV (Bariah et al. 2020a). Monomorphism at this locus was seen in the populations from Mt. Hermon and Tabgha, with amplification detected in all accessions, while polymorphism was seen in seven accessions from Turkey and Syria, six from Lebanon, five from Amiad, and one from Mt. Amasa (Fig. S3j).

Locus 5B3: PCR reaction using primers specifically designed from the indel breakpoint was performed in all wild emmer wheat accessions. The forward primer was designed to be located on a sequence ~ 64kb in length that is absent in bread wheat, and the reverse primer was located at the indel's 3' flanking region. In the populations from Mt. Hermon, Amiad, Mt. Amasa, and Tabgha, the locus exhibited monomorphism, with amplification observed in all accessions, while in Turkey, amplification was seen in nine accessions: eight in Syria, and six in Lebanon (Fig. S3k1). Another PCR reaction was conducted using primers based on the breakpoints to amplify the bread wheat allele. Amplification was observed in six accessions from Lebanon, three from Mt. Hermon, three from Syria, two from Turkey, and one from Mt. Amasa (Fig. S3k2).

Locus 3B4

PCR reaction using primers specifically designed from the indel breakpoint was performed in all wild emmer wheat accessions. The forward primer was located on the 5' flanking sequence of the SV, and the reverse primer was located on the 3' flanking sequence. The populations from Mt. Hermon, Mt. Amasa, Lebanon, and Turkey were monomorphic, with amplification observed in all accessions. In contrast, amplification was seen in one accession from Tabgha, eight from Amiad, and nine from Syria (Fig. S3l).

Locus 3B2

PCR reaction using primers specifically designed from the indel breakpoint was performed in all wild emmer wheat accessions. The forward primer was designed to pair with the flanking sequence at the 5' end of the breakpoint, while the reverse primer was located at a ~ 16kb segment of wild emmer wheat. Except for the populations from Syria and Lebanon, which showed amplification in 9 out of 10 accessions, all populations exhibited monomorphism with amplification in all accessions (Fig. S2m).

Locus 5B5

PCR reaction using primers specifically designed from the indel breakpoint was performed in all wild emmer wheat accessions. The forward primer was located at the ~ 41kb segment present in wild emmer wheat, and the reverse primer was located at the 3' flanking area of the SV. Monomorphism was observed in the Tabgha population, with amplification detected in all accessions, while the populations from Mt. Hermon, Turkey, Syria, and Lebanon also showed monomorphism but lacked the PCR product in all accessions. Amplification was shown in two accessions from Amiad and one from Mt. Amasa (Fig. S3n1). These results indicate that this rearrangement is population-specific, particularly to the Tabgha and Amiad populations. The lower bands in the bread wheat accession (CS42) were sequenced and identified in the D subgenome of the bread wheat. A subsequent PCR reaction using primers based on the breakpoints to amplify the bread wheat allele was performed. Amplification was observed in all accessions from Mt. Hermon, Mt. Amasa, Syria, and Turkey, as well as in eight accessions from Amiad, one from Tabgha, and nine from Lebanon. Notably, some of the accessions exhibited a higher band (~ 1210bp) than expected (787bp). Sequencing of these higher bands revealed an insertion of a transposable element (TE) within the amplified fragment, identified as HARB-N25_SBi from the Harbinger family (Fig. S3n2).

Discussion

In this study, we conducted a detailed comparative genomic analysis to explore large-scale genomic rearrangements between wild emmer wheat and bread wheat. Utilizing dyads as genomic markers, we identified significant genomic differences between these two species. We addressed these and previously identified SVs from our lab and screened them across seven isolated wild emmer wheat populations in the Fertile Crescent. Four of these populations are from Israel (Mt. Hermon, Amiad, Tabgha, and Mt. Amasa), in addition to accessions collected from Syria, Turkey, and Lebanon. The results of this study indicate significant structural differences that may contribute to phenotypic diversity and adaptation in these species.

The identification and characterization of SVs, such as deletions, insertions, and introgressions, highlight the dynamic nature of the wild emmer wheat genome. These SVs are not merely genomic curiosities; the potential genes and regulatory sequences they contain are likely crucial for the ecological and evolutionary resilience of wild emmer wheat and can be used as a potential donor of lost and valuable traits in bread wheat and other crops, potentially improving agricultural productivity in areas facing climatic challenges (Dahan-Meir et al. 2022; Gaurav et al. 2022; Schiessl et al. 2019).

In this study, we identified and characterized 11 cases of large-scale genomic rearrangements between wild emmer and bread wheat, including 7 deletions, 3 introgressions of new sequences, and TE insertional polymorphism. These sequences showed high sequence similarity to genes associated with disease resistance (biotic stress), such as the rust resistance kinase (TRIDC5AG056530). Leaf rust resistance genes are a family of genes found in plants that provide resistance against leaf rust, a common fungal disease caused by *Puccinia* species. This disease affects leaves, stems, and fruits and is one of the most damaging fungal pathogens threatening global wheat cultivations (Wu et al., 2019). Lr10, when used in combination with other resistance genes like Lr17, Lr27, and Lr31, contributes to a robust defense mechanism against leaf rust (Rasheed et al. 2012).

Additionally, another gene, TRIDC5AG053440, which shows high sequence similarity to wall-associated receptor kinase 5, is involved in resistance to diseases caused by the necrotrophic fungus *Rhizoctonia cerealis*. It also facilitates communication between the extracellular matrix and the cytoplasm (Yang et al. 2014). Another significant finding involves the gene TRIDC3AG003030 from the HOX family of proteins. Homeobox-leucine zipper proteins, a class of transcription factors, play a pivotal role in plant growth and stress responses. These proteins are integral to the transcriptional regulation of gene expression involved in developmental processes and adaptations to environmental challenges.

The involvement of HOX proteins in stress response mechanisms is particularly notable in the context of their potential role in enhancing disease resistance through the modulation of plant immune responses and stress signaling pathways. The upregulation of such genes in response to stressors, including pathogen attack, could be part of a larger genomic strategy to bolster the plant's resilience to environmental and biotic stresses (Li et al. 2019).

Moreover, proteins like TRIDC7AG044240, which show high sequence similarity to vacuolar protein sorting-associated protein, highlight the complex nature of cellular responses to stress. In *Citrus sinensis*, for instance, the expression of genes related to cell transport, including a putative vacuolar protein sorting-associated protein 13B-like isoform, is discussed as part of the plant's response to boron deficiency (Lu et al. 2015).

Enhancing genetic resistance against biotic stress remains a principal aim within many wheat breeding programs. However, contemporary wheat varieties exhibit limited genetic diversity for pest and disease resistance, with the ongoing threat of emergent diseases and pests that might overcome established resistance genes. These findings can be a great contribution to breeding programs as those genes can be potential donors for beneficial and crucial traits.

Using genetic markers to identify and understand large-scale genomic rearrangements is a crucial tool in plant genetics (Kordrostami and Rahimi 2015). These markers, such as TEs (Bariah et al. 2020a), SNPs (Balla et al. 2022), microsatellites (Peleg et al. 2008) are instrumental in detecting and characterizing genetic diversity, which is essential for crop improvement. The results, which largely found gene-containing SVs, demonstrate that this approach is particularly effective in discerning the genomic structures of wild emmer wheat and bread wheat, facilitating precise identification of SVs that are likely crucial for ecological and evolutionary resilience, and potentially enhancing agricultural productivity under adverse conditions. In contrast to other studies, such as Inbar Bariah's research (Bariah et al. 2020a), which utilized Fatima, a Gypsy long-terminal repeat retrotransposon, my use of dyadic markers offers a distinct advantage.

Understanding the types of structural variations (SVs) and their genetic content and screening them within and between populations to assess their distribution is crucial for future insights into their potential contributions to phenotypes and a species' adaptability to various conditions and environmental changes (Zhao et al. 2023). Different polymorphism patterns were exhibited in the wild emmer populations. The observed polymorphism of large-scale SVs within and among populations may seem at odds with classical models of heterozygote disadvantage, which typically lead to the fixation or loss of such variants. However, wheat is an allopolyploid, and the unique 'genetic shock' of its formation provides a buffering mechanism where the presence of multiple homoeologous genomes (A and B) allows the plant to tolerate significant genomic changes that would be deleterious in diploids. These SVs are subgenome-specific; because the PCR markers were designed specifically for one subgenome (e.g., A), the variations do not impact the stability of the entire chromosomal set. Furthermore, as wild emmer is a predominantly self-pollinating species, most individuals are naturally homozygous at these loci. Therefore, the polymorphism we report reflects the presence or absence of these variants among different individuals (accessions) in the population, rather than unstable heterozygous states within individuals." The wild emmer accessions collected from Turkey, Syria, and Lebanon showed the most polymorphic patterns. Among the loci identified, some were found to be population-specific. For example, the bread wheat allele of the 5A11 locus (Table 2), is specific to Turkish and Syrian populations. This locus in wild emmer contains a protein-coding gene with high sequence similarity to the WAK protein. It is possible that this important gene, shown to be associated with disease resistance, is present in all other examined accessions and absent only in some Turkish and Syrian accessions. The wild emmer allele of locus 5B19 (Table 2), showed amplification in the Amiad and Tabgha populations. This locus underwent introgression in bread wheat, and both orthologous loci have the same genes, except for two additional uncharacterized proteins and some hypothetical proteins existing only in the bread wheat locus. Among the previously identified loci, the wild emmer allele of the 5B5 locus (Table 2) is also specific to the Amiad and Tabgha populations and contains no genes (Bariah et al. 2020b). The bread wheat allele shows complementary PCR results, except in one accession in Tabgha and one in Lebanon. This suggests that the introgression likely occurred during the early stages of bread wheat development, probably around the time or shortly after the introgression event. The only locus that showed monomorphic amplifications in all accessions in 5A13. This large-scale rearrangement could be triggered by an allopolyploidization event, which triggers genomic rearrangements, thus existing in all examined wild emmer populations but not in the CS42 (bread wheat), potentially conferring some evolutionary or adaptive advantages. Additionally, the sole gene found in this locus is

uncharacterized and may be associated with stress resistance, growth, or other critical functions that are preserved despite the genomic rearrangements typically associated with allopolyploidy.

In summary, the divergent polymorphism patterns across the populations provide deeper insights into the historical and geographical influences on genetic diversity. These findings not only increase our understanding of plant genetics and evolution but also can assist future researchers and attempting to improve agricultural practices by informing breeding programs aimed at introducing resilient traits into crop varieties.

Declarations

Competing Interests:

The authors declare no financial or non-financial interests, direct or indirect, related to the work submitted for publication.

Data statement:

All data generated or analyzed during this study are included in this published article [and its supplementary information files].

Funding:

This work was supported by a grant from the Israel Science Foundation (grant# 1311/21) to K. K.

Author Contribution

L.B.—Designed the project, generated data, analyzed data, and wrote the manuscript
K.K.— Designed the project, analyzed data, wrote, edited, and submitted the manuscript.

Acknowledgement

We would like to thank Vadim Khasdan for his assistance with the manuscript preparation. This work was supported by a grant from the Israel Science Foundation (grant# 1311/21) to K. K.

Data Availability

All data generated or analyzed during this study are included in this published article [and its supplementary information files].

References

1. Balla MY, Gorafi YSA, Kamal NM, Abdalla MGA, Tahir ISA, Tsujimoto H. Exploiting Wild Emmer Wheat Diversity to Improve Wheat A and B Genomes in Breeding for Heat Stress Adaptation. *Front Plant Sci.* 2022;13:895742.

2. Bariah I, Keidar-Friedman D, Kashkush K. Identification and characterization of large-scale genomic rearrangements during wheat evolution. *PLoS ONE*. 2020a;15:e0231323.
3. Bariah I, Keidar-Friedman D, Kashkush K. Where the wild things are: transposable elements as drivers of structural and functional variations in the wheat genome. *Front Plant Sci*. 2020b;11:585515.
4. Bevan MW, Uauy C, Wulff BB, Zhou J, Krasileva K, Clark MD. Genomic innovation for crop improvement. *Nature*. 2017;543:346–54.
5. Bolser DM, Staines DM, Perry E, Kersey PJ. Ensembl plants: integrating tools for visualizing, mining, and analyzing plant genomic data. *Plant Genomics Databases: Methods and Protocols*. Springer; 2016. pp. 1–31.
6. Charmet G. Wheat domestication: lessons for the future. *CR Biol*. 2011;334:212–20.
7. Dahan-Meir T, Ellis TJ, Sela H, Mafessoni F. (2022) The genetic structure of a wild wheat population has remained associated with microhabitats over 36 years. *bioRxiv*.
8. Dvorak J, Wang L, Zhu T, Jorgensen CM, Luo M-C, Deal KR, Gu YQ, Gill BS, Distelfeld A, Devos KM. Reassessment of the evolution of wheat chromosomes 4A, 5A, and 7B. *Theor Appl Genet*. 2018;131:2451–62.
9. Feldman M, Kislev ME. Domestication of emmer wheat and evolution of free-threshing tetraploid wheat. *Isr J Plant Sci*. 2007;55:207–21.
10. Feldman M, Levy A. Allopolyploidy—a shaping force in the evolution of wheat genomes. *Cytogenet Genome Res*. 2005;109:250–8.
11. Feldman M, Levy AA. Genome evolution due to allopolyploidization in wheat. *Genetics*. 2012;192:763–74.
12. Feldman M, Millet E. The contribution of the discovery of wild emmer to an understanding of wheat evolution and domestication and to wheat improvement. *Isr J Plant Sci*. 2001;49:25–36.
13. Feuillet C, Travella S, Stein N, Albar L, Nublat A, Keller B. (2003) Map-based isolation of the leaf rust disease resistance gene Lr10 from the hexaploid wheat (*Triticum aestivum* L.) genome. *Proceedings of the National Academy of Sciences* 100:15253–15258.
14. Gaurav K, Arora S, Silva P, Sanchez-Martin J, Horsnell R, Gao L, Brar GS, Widrig V, John Raupp W, Singh N, Wu S, Kale SM, Chinoy C, Nicholson P, Quiroz-Chavez J, Simmonds J, Hayta S, Smedley MA, Harwood W, Pearce S, Gilbert D, Kangara N, Gardener C, Forner-Martinez M, Liu J, Yu G, Boden SA, Pascucci A, Ghosh S, Hafeez AN, O'Hara T, Waites J, Cheema J, Steuernagel B, Patpour M, Justesen AF, Liu S, Rudd JC, Avni R, Sharon A, Steiner B, Kirana RP, Buerstmayr H, Mehrabi AA, Nasyrova FY, Chayut N, Matny O, Steffenson BJ, Sandhu N, Chhuneja P, Lagudah E, Elkot AF, Tyrrell S, Bian X, Davey RP, Simonsen M, Schauser L, Tiwari VK, Randy Kutcher H, Hucl P, Li A, Liu DC, Mao L, Xu S, Brown-Guedira G, Faris J, Dvorak J, Luo MC, Krasileva K, Lux T, Artmeier S, Mayer KFX, Uauy C, Mascher M, Bentley AR, Keller B, Poland J, Wulff BBH. Population genomic analysis of *Aegilops tauschii* identifies targets for bread wheat improvement. *Nat Biotechnol*. 2022;40:422–31.
15. Gray YH. It takes two transposons to tango: transposable-element-mediated chromosomal rearrangements. *Trends Genet*. 2000;16:461–8.
16. Guan J, Garcia DF, Zhou Y, Appels R, Li A, Mao L. The battle to sequence the bread wheat genome: a tale of the three kingdoms. *Genom Proteom Bioinform*. 2020;18:221–9.
17. Huo N, Zhang S, Zhu T, Dong L, Wang Y, Mohr T, Hu T, Liu Z, Dvorak J, Luo M-C. Gene duplication and evolution dynamics in the homeologous regions harboring multiple prolamin and resistance gene families in hexaploid wheat. *Front Plant Sci*. 2018;9:673.
18. Juery C, Concia L, De Oliveira R, Papon N, Ramírez-González R, Benhamed M, Uauy C, Choulet F, Paux E. New insights into homoeologous copy number variations in the hexaploid wheat genome. *Plant Genome*.

2021;14:e20069.

19. Khodaverdian VY, Hanscom T, Yu AM, Yu TL, Mak V, Brown AJ, Roberts SA, McVey M. Secondary structure forming sequences drive SD-MMEJ repair of DNA double-strand breaks. *Nucleic Acids Res.* 2017;45:12848–61.
20. Kohany O, Gentles AJ, Hankus L, Jurka J. Annotation, submission and screening of repetitive elements in Repbase: RepbaseSubmitter and Censor. *BMC Bioinformatics.* 2006;7:474.
21. Kordrostami M, Rahimi M. Molecular markers in plants: concepts and applications. *Genet 3rd Millenn.* 2015;13:4024–31.
22. Levy AA, Feldman M. Evolution and origin of bread wheat. *Plant Cell.* 2022;34:2549–67.
23. Li C, Fan R, Ma C, Zhang Z, Li Z, Zhu L, Nie F, Li Y, Liu X, Xie J. Reciprocal translocations hidden by phenotype and genotype within the same wheat cultivar. *Crop Sci.* 2023;63:2727–39.
24. Li Y, Fahima T, Korol AB, Peng J, Kirzhner RMS, Beiles V, Nevo A E. Microsatellite diversity correlated with ecological-edaphic and genetic factors in three microsites of wild emmer wheat in North Israel. *Mol Biol Evol.* 2000;17:851–62.
25. Li Y, Fahima T, Röder M, Kirzhner V, Beiles A, Korol A, Nevo E. Genetic effects on microsatellite diversity in wild emmer wheat (*Triticum dicoccoides*) at the Yehudiyya microsite. *Isr Heredity.* 2003;90:150–6.
26. Li Z, Pan X, Guo X, Fan K, Lin W. Physiological and transcriptome analyses of early leaf senescence for ospls1 mutant rice (*Oryza sativa* L.) during the grain-filling stage. *Int J Mol Sci.* 2019;20:1098.
27. Lu Y-B, Qi Y-P, Yang L-T, Lee J, Guo P, Ye X, Jia M-Y, Li M-L, Chen L-S. Long-term boron-deficiency-responsive genes revealed by cDNA-AFLP differ between *Citrus sinensis* roots and leaves. *Front Plant Sci.* 2015;6:585.
28. Luo M-C, Yang Z-L, You F, Kawahara T, Waines J, Dvorak J. The structure of wild and domesticated emmer wheat populations, gene flow between them, and the site of emmer domestication. *Theor Appl Genet.* 2007;114:947–59.
29. Mirzaghaderi G, Mason AS. Revisiting pivotal-differential genome evolution in wheat. *Trends Plant Sci.* 2017;22:674–84.
30. Nevo E. Evolution of wild emmer wheat and crop improvement. *J Syst Evol.* 2014;52:673–96.
31. Nevo E, Beiles A. Genetic diversity of wild emmer wheat in Israel and Turkey: structure, evolution, and application in breeding. *Theor Appl Genet.* 1989;77:421–55.
32. Ozbek O, Millet E, Anikster Y, Arslan O, Feldman M. Comparison of the genetic structure of populations of wild emmer wheat, *Triticum turgidum* ssp. *dicoccoides*, from Israel and Turkey revealed by AFLP analysis. *Genet Resour Crop Evol.* 2007;54:1587–98.
33. Özkan H, Willcox G, Graner A, Salamini F, Kilian B. Geographic distribution and domestication of wild emmer wheat (*Triticum dicoccoides*). *Genet Resour Crop Evol.* 2011;58:11–53.
34. Peleg Z, Saranga Y, Krugman T, Abbo S, Nevo E, Fahima T. Allelic diversity associated with aridity gradient in wild emmer wheat populations. *Plant Cell Environ.* 2008;31:39–49.
35. Petersen G, Seberg O, Yde M, Berthelsen K. 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.* 2006;39:70–82.
36. Ranjha L, Howard SM, Cejka P. Main steps in DNA double-strand break repair: an introduction to homologous recombination and related processes. *Chromosoma.* 2018;127:187–214.

37. Rasheed A, Mumtaz AS, Shinwari ZK. Genetic characterization of novel Lr gene stack in spring wheat variety Chakwal86 and its effectiveness against leaf. Pak J Bot. 2012;44:507–10.
38. Riaz M, Wong Y. (2017) Estimation of Yield Losses Due to Leaf Rust and Late Seeding on Wheat (*Triticum aestivum* L) Variety Seher-06 in District Faisalabad, Punjab, Pakistan. Adv Biotech Micro 4.
39. Schiessl S-V, Katche E, Ihien E, Chawla HS, Mason AS. The role of genomic structural variation in the genetic improvement of polyploid crops. Crop J. 2019;7:127–40.
40. Shiferaw B, Smale M, Braun H-J, Duveiller E, Reynolds M, Muricho G. Crops that feed the world 10. Past successes and future challenges to the role played by wheat in global food security. Food Secur. 2013;5:291–317.
41. Yang K, Qi L, Zhang Z. Isolation and characterization of a novel wall-associated kinase gene TaWAK5 in wheat (*Triticum aestivum*). Crop J. 2014;2:255–66.
42. Zhao J, Li X, Qiao L, Zheng X, Wu B, Guo M, Feng M, Qi Z, Yang W, Zheng J. Identification of structural variations related to drought tolerance in wheat (*Triticum aestivum* L). Theor Appl Genet. 2023;136:37.
43. Zhou Y, Zhao X, Li Y, Xu J, Bi A, Kang L, Xu D, Chen H, Wang Y, Wang YG, Liu S, Jiao C, Lu H, Wang J, Yin C, Jiao Y, Lu F. Triticum population sequencing provides insights into wheat adaptation. Nat Genet. 2020;52:1412–22.

Figures

Wild emmer

Bread wheat

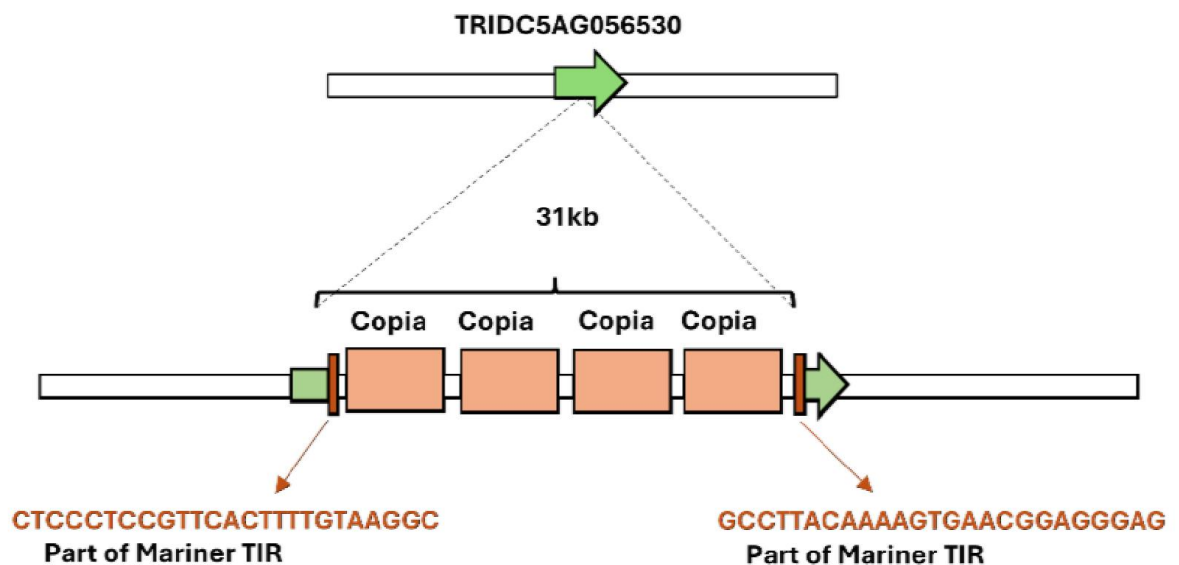


Figure 1

Schematic representation of locus 5A12 in wild emmer and bread wheat genomes. An insertion of a 31kb segment in bread wheat compared to wild emmer resulted in significant structural differences between the two genomes. Sequence lengths are not drawn to scale. The insertion includes transposable elements (TEs), particularly from the Copia family, represented by orange boxes, although the number of boxes does not accurately reflect the total number of inserted TEs. Brown boxes represent parts of TIRs. A gene (TRIDC5AG056530) in wild emmer, represented by a green arrow, shows high sequence similarity to rust resistance kinase LR10-like. Dashed lines

indicate the alignment between orthologous sequence segments at the indel borders and the terminal ends of the represented sequences.

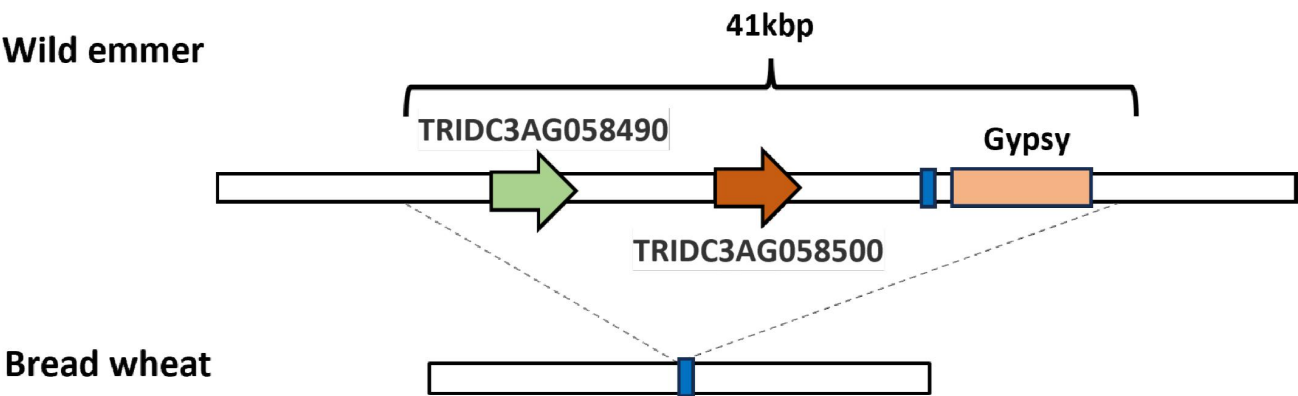


Figure 2

Schematic representation of the locus containing 3A16 in the wild emmer (top) and bread wheat genomes (bottom). A missing 41kb segment in bread wheat compared to wild emmer resulted in a large-scale structural difference between the two genomes. Sequence lengths are not drawn to scale. A gene (TRIDC3AG058490) in wild emmer, represented by a green arrow, shows high sequence identity to anthranilate O-methyltransferase 3-like protein. Another gene (TRIDC3AG058500), represented by a brown arrow, shows sequence identity to 2'-deoxymugineic-acid 2'-dioxxygenase-like protein. An orange box represents a Gypsy transposable element, and a blue box indicates the common fragment of 605bp shared between the genomes. Dashed lines indicate the alignment between orthologous sequence segments at the indel borders and the terminal ends of the represented sequences.

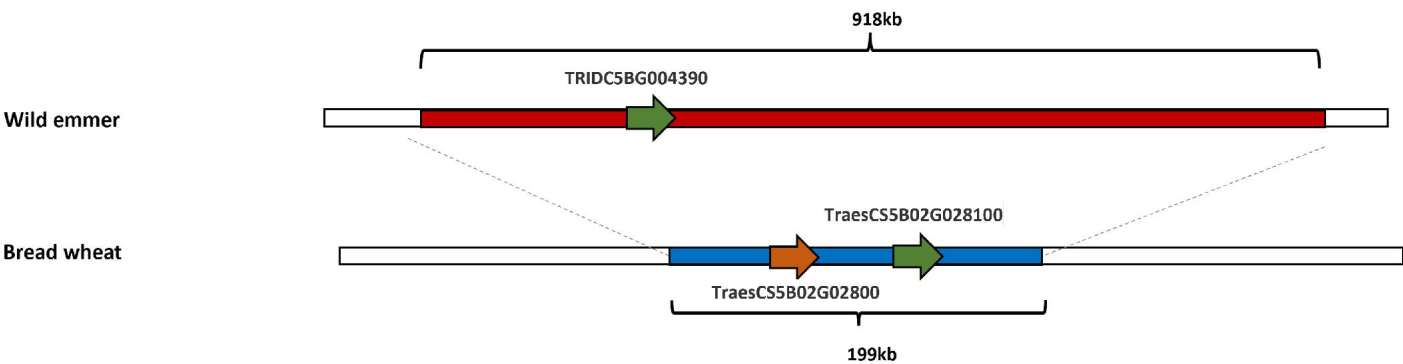


Figure 3

Schematic representation of locus 5B18 in wild emmer (top) and bread wheat (bottom). The sequence length is unscaled. Genes are represented by arrows, with each color representing similar genes. Green arrows represent genes similar to F-box/FBD/LRR-repeat protein At1g13570-like protein: TRIDC5BG004390 in wild emmer and TraesCS5B02G028100 in bread wheat. Brown arrows represent TraesCS5B02G028000, which shows similarity to

acyl-coenzyme A thioesterase 13-like. Red lines represent wild emmer-specific sequences, while blue lines represent bread wheat-specific sequences. Dashed lines indicate the alignment between orthologous sequence segments at the indel borders and the terminal ends of the represented sequences.

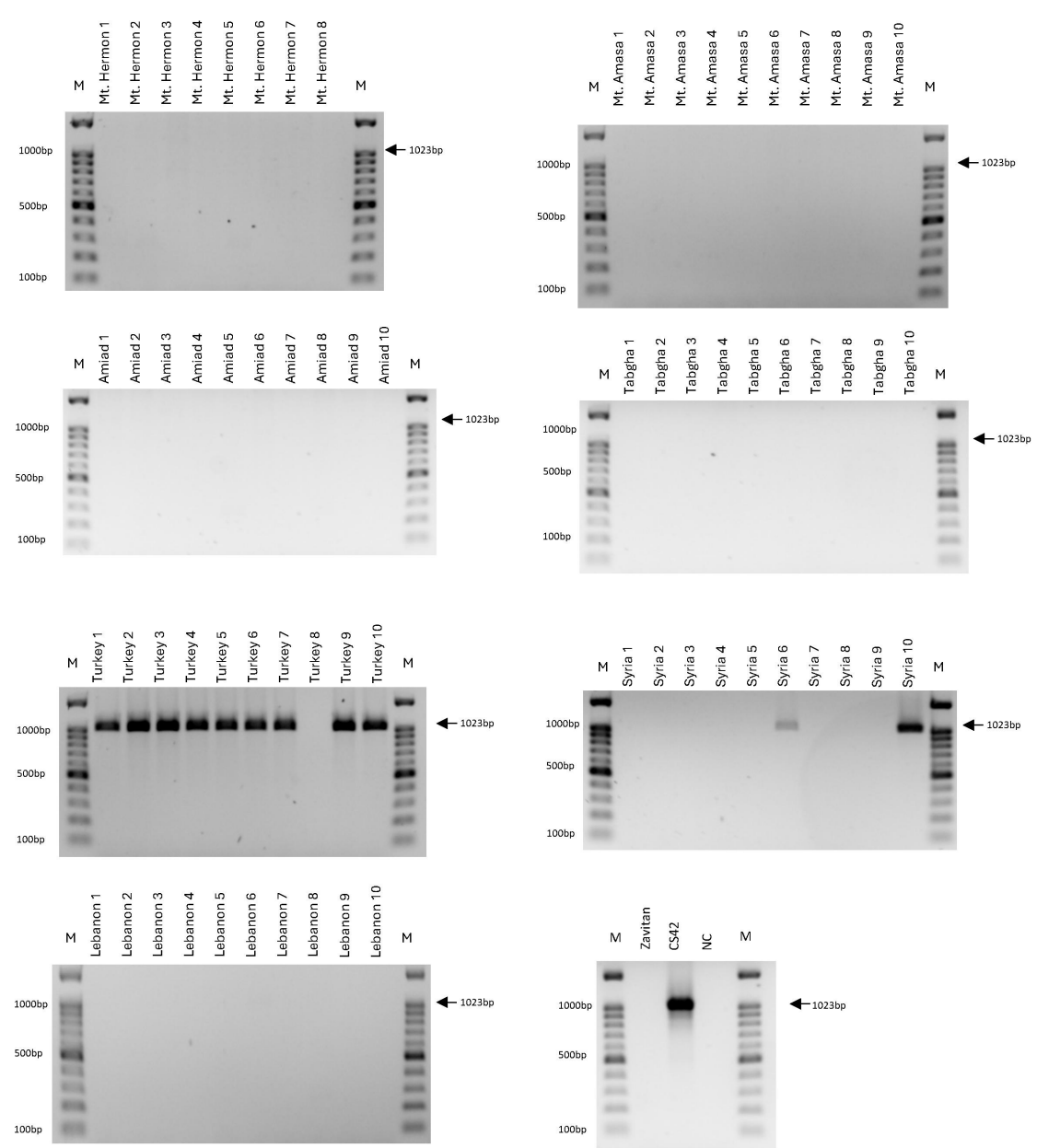


Figure 4

PCR analysis using primers designed based on introgression identified in locus 5A11. The forward primer was designed based on 5' flanking sequence of 5A11 locus and reverse primer was designed from 7kb bread wheat

specific segment to amplify the bread wheat locus. 'M' refers to the size marker, 'NC' refers to the negative control (ddH2O was used as the template in its PCR reactions). The black arrow indicates the expected product size.

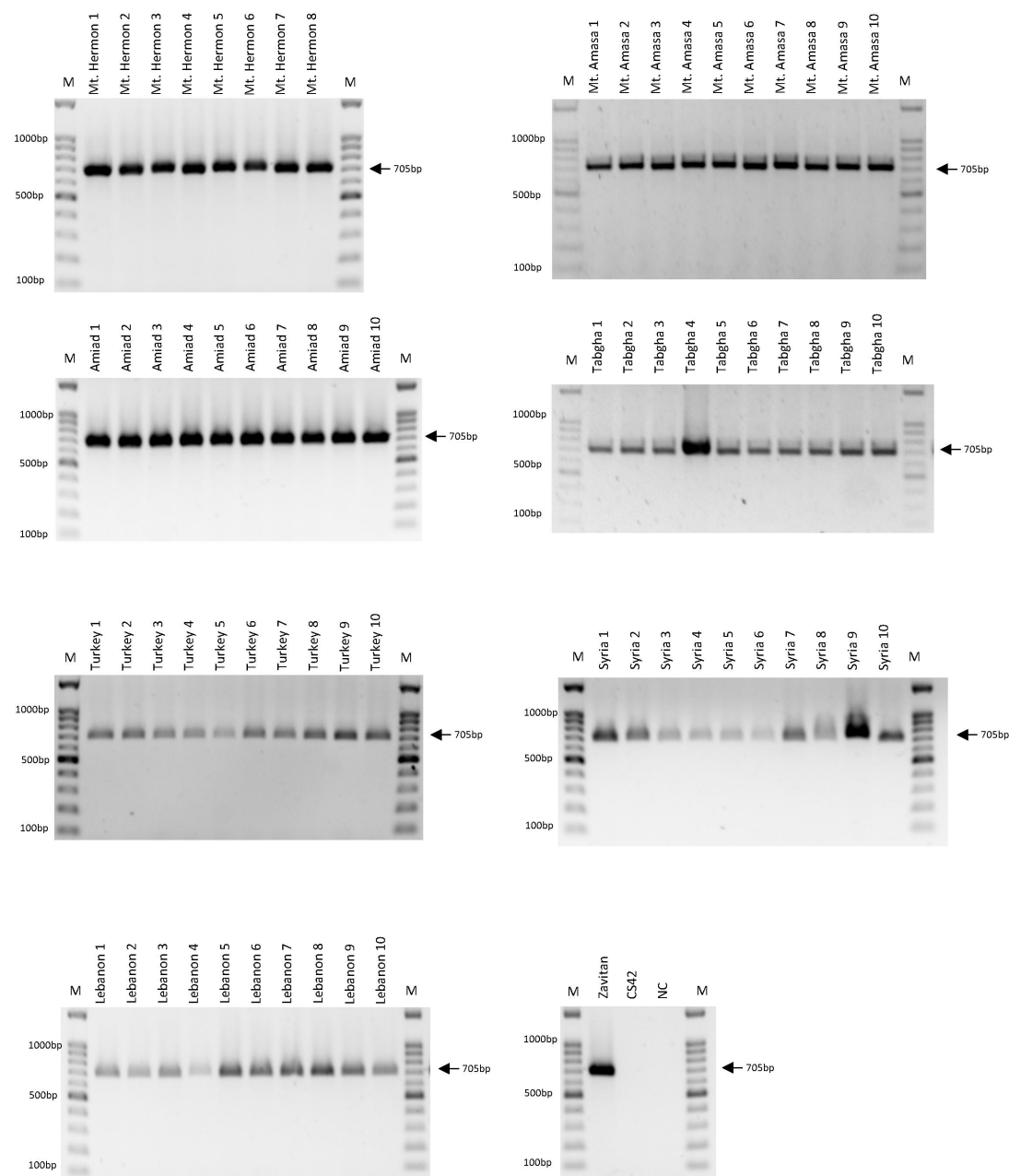


Figure 5

PCR analysis using primers designed based on the deletion identified in the orthologous locus 5A13 in bread wheat. The forward primer was designed based on 5' flanking sequence of 5A13 locus and the reverse primer was designed from the 13.5kb deleted sequence to amplify the wild emmer wheat allele. 'M' refers to the size marker,

'NC' refers to the negative control (ddH₂O was used as the template in its PCR reactions). The black arrow indicates the expected product size.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementTable1.pdf](#)
- [Supplementfigure1.pdf](#)
- [Supplementalfigure2.pdf](#)
- [Supplementfigure3.pdf](#)