

Dissecting the Population Structure, Diversity and Genetic Architecture of Disease Resistance in Wild Emmer Wheat (*Triticum turgidum* subsp. *dicoccoides*)

Laxman Adhikari

King Abdullah University of Science and Technology (KAUST)

Pablo Olivera

University of Minnesota

Guotai Yu

King Abdullah University of Science and Technology (KAUST)

John Raupp

Kansas State University

Hanan Sela

University of Haifa

Assaf Distelfeld

University of Haifa

Brande Wulff

King Abdullah University of Science and Technology (KAUST)

Brian Steffenson

University of Minnesota

Jesse Poland

`jesse.poland@kaust.edu.sa`

King Abdullah University of Science and Technology (KAUST)

Research Article

Keywords: Wild Emmer Wheat, *dicoccoides*, whole-genome sequencing, population structure, GWAS, stem rust, leaf rust, stripe rust, judaicum

Posted Date: September 11th, 2024

DOI: <https://doi.org/10.21203/rs.3.rs-4909521/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 Genomics on March 19th, 2026. See the published version at <https://doi.org/10.1186/s12864-026-12751-6>.

Abstract

Background

Wild emmer wheat (WEW) (*Triticum turgidum* subsp. *dicoccoides*) is one of the important crops domesticated in the Neolithic revolution that shifted humanity to agrarian societies. Leveraging whole-genome sequencing (WGS) data from 291 accessions at ~ 9.5x coverage, we identified 3.4 million high-quality SNP markers and utilized them for population genetics and evolutionary studies. We also conducted Genome-wide association studies (GWAS) for seedling-stage resistance to five races of stem, leaf, and stripe/yellow rust pathogens. We investigated candidate genes using ortholog sequence alignment.

Results

Phylogenetic clustering, principal component analysis, and population structure assessment revealed distinct genetic subgroups within WEW, specifically the Northern Population, Southern Levant (SL) population, and a highly distinct subgroup near the Sea of Galilee, known as race *judaicum*. Diversity and pairwise F_{ST} analyses highlighted varying levels of genetic diversity and distances among these subpopulations. The SL accessions exhibited higher resistance to all races of the three rust pathogens compared to Northern and *judaicum* populations. GWAS identified 28, 25, and 32 significant loci associated with stem, leaf, and stripe rust resistance, respectively. Major loci explained up to 60% of phenotypic variation and some loci were linked to multiple races and some were novel. Accessions such as TA11204, TA1041, TA1047, TA11196, TA77, TA93, TA1065, and TA66 demonstrated broad resistance to multiple races.

Conclusions

In summary, this study provides rust resistance WEW germplasms and guides future research on the selection and introgression of rust resistance loci from WEW into cultivated bread and durum wheat, contributing to the stable production of these important food crops.

1. Background

Adverse climate conditions, limited arable land, water scarcity for irrigation, soil degradation, and the prevalence of biotic and abiotic stresses present formidable challenges to sustainable crop production. Diseases such as leaf (caused by *Puccinia triticina*), stem (*Puccinia graminis* f. sp. *tritici*), and stripe rust (*Puccinia striiformis* f. sp. *tritici*) cause substantial economic losses globally to bread wheat (*Triticum aestivum* L.) and durum wheat (*T. turgidum* subsp. *durum*). For instance, a recent probabilistic bio-economic assessment suggests that over 90% of worldwide wheat production is vulnerable to leaf rust [1]. Furthermore, Beddow et al. (2015) highlight a concerning increase in the spread of stripe rust over

the past six decades, leaving 88% of global wheat production vulnerable to infection [2]. Similarly, more than 90% of wheat varieties were found to be susceptible to isolate Ug99 of the stem rust pathogen [3]. Despite ongoing efforts by scientists to breed wheat cultivars resistant to these devastating diseases, the genetic diversity for disease resistance in wheat remains insufficient for the long-term control of these variable rust pathogens. This limitation primarily stems from constraints imposed by the domestication bottleneck and intense selection practices used in breeding. In contrast, wild progenitors and relatives of wheat possess rich genetic diversity, providing a valuable resource to address the inadequate genetic variation observed in cultivated wheat varieties [4–6].

Wild emmer wheat (*T. turgidum* subsp. *dicoccoides*), henceforth referred to as WEW, holds the distinction of being one of the earliest crops to undergo domestication [7]. The tetraploid ($2n = 4x = 28$) WEW shares two sub-genomes (AABB) with bread wheat (AABBDD) and is as the progenitor of both tetraploid and hexaploid wheat [8, 9]. WEW exhibits extensive phenotypic variation for many important agronomic traits, especially disease resistance. For instance, *Pm41* a coiled-coil, nucleotide-binding site and leucine-rich repeat protein (CNL) present in WEW imparts resistance to powdery mildew in wheat [8]. Another powdery mildew gene, *pm36*, has recently been reported and is present a few lines of WEW from the Southern Levant [10]. Additionally, the stripe rust resistance genes *Yr15* and *Yr84* were also cloned in WEW [11, 12]. Other stripe rust genes transferred from WEW to domesticated wheat include *Yr36*, *Yr-SM139*, and additional ASR and APR genes [13, 14]. In previous research, several leaf rust and stem rust resistance genes from WEW were transferred into both durum and bread wheat using conventional breeding methods such as backcrossing [15–17]. Thus, evaluating WEW germplasm for disease resistance holds great promise for identifying and harnessing novel genetic variations that can be used to protect wheat cultivars from important biotic stresses.

The WEW natural existence was reported in the western Fertile Crescent territories of Jordan, Syria, Israel, Lebanon and Palestine, Southeastern Anatolia, and mountain regions of eastern Iraq and western Iran [18]. Understanding the population structure of WEW using a curated whole-genome sequencing (WGS) dataset with genome-wide variants is crucial for effectively leveraging genetic resources that can be utilized in durum breeding programs. Additionally, comprehending the relationships between wild and domesticated emmer wheat in a larger population is invaluable to selecting target breeding germplasms. Previous population studies have primarily relied on morphology-based methods and a limited number of genetic markers, limiting their ability to make definitive conclusions. For instance, despite various claims regarding the domestication site of emmer wheat, including southeastern Turkey, the southern Levant, or multi-regional origins, a clear consensus has yet to be reached [7]. Furthermore, previous disease-related associations and quantitative trait loci (QTL) analyses in WEW have been constrained by small sample sizes and low-coverage sequencing approaches like GBS [14]. Therefore, conducting genome-wide association studies (GWAS) with a larger and more diverse panel and numerous SNPs identified from sequence data across the genome, are invaluable for identifying novel loci. The current study involves a GWAS of leaf, stem and stripe rust resistance in a diverse panel of WEW.

Hence, the primary objectives of this study were to: i) assemble a representative panel of WEW with broader geographical distribution and diversity; ii) conduct population structure and diversity analyses in the assembled panel; iii) perform a comprehensive GWAS to identify genomic regions associated with resistance to the three rust diseases; and iv) provide extensive sequencing and phenotypic data to the research community.

2. Results

2.1 WEW Panel Geographic Distribution

The majority of WEW accessions were sourced from the regions of the Southern Levant (SL) and Southeastern (SE) Anatolia, with only 4 accessions from Iran and 18 from Iraq (Fig. 1, Supplementary Table S1). Additionally, 6 accessions had no passport information and were initially obtained from the other research institutions.

2.2 Genotyping

The average sequencing depth of accessions was 9.5x (Supplementary Figure S1), and the raw paired-end FASTQ files totaled 14 TB data, which were uploaded to the National Center for Biotechnology Information (NCBI) BioProject accession #PRJNA1007489. The number of filtered SNPs totaled 68 million, while in the selected 5% subset, there were 3,410,677 SNPs (Supplementary Table S3). The SNP density plot showed different densities of these variants across the 14 chromosomes (Supplementary Figure S2). For all chromosomes in the genome, there were at least 200K SNPs per chromosome (Supplementary Table S3).

2.3 Passport Data Curation

Allele matching revealed five pairs of WEW accessions with at least one duplicate (Supplementary Table S4). Most of the identified duplicate accessions (TA117 and TA1459; TA11202, TA11203 and TA119; TA131, TA132 and TA133; TA134 and TA135) were collected in nearby zones (Supplementary Table S1), particularly from Golan Heights and nearby regions. We also observed the need to update and ensure consistency in the 'collection sites' between the gene bank and the USDA GRIN database, particularly regarding changes in geopolitical maps.

2.4 Phylogenetic tree and PCA

The unrooted NJ phylogenetic tree revealed three major WEW clades (Fig. 2). Notably, a group of accessions near the Sea of Galilee (Lake Tiberias) formed a distinct clade, indicating significant genetic differentiation within this group, even among accessions from the same region around Lake Tiberias. A comparison of these accessions with the previously recorded spike morphology data of the accessions confirmed that these accessions represent the wide-spiked race *judaicum* [19, 20]. Accessions from Southeastern Anatolia (SE), Iran, and Iraq formed a separate clade, known as the Northern Population[14]. Other than *judaicum*, accessions from the Southern Levant (SL) constituted a single

major clade, with some accessions from Lebanon and Syria forming a small subclade under the SL clade (Fig. 2). The northern and southern populations of WEW have been described elsewhere in the literature [14, 21, 22]; therefore, we named the genetically separated subgroups as the Northern Population, the Southern Levant (SL) Population and the *judaicum*. Additionally, four accessions (TA120, TA1468, TA11175 and TA11193) clustered at the base of the SL clade near the nodes where *judaicum* and the Northern Population separated. Interestingly, the two accessions, TA11208 and TA11218 from northern Syria also clustered with accessions from the northern Lebanon and the southern Syria (Fig. 2).

The division of WEW accessions in the PCA plot reflected their geographic origin. The *judaicum* accessions form a distinct subgroup, clearly differentiating them from the rest. The first principal component (PC1), explaining 29% of the total variation, separated the *judaicum* accessions from the other two groups (Fig. 3). The second principal component (PC2), which accounted for 16% of the variation, distinguished accessions between the Northern Population and the SL. Accessions without passport information were grouped with those from the Northern Population (Fig. 3). As in the phylogenetic tree, four accessions (TA120, TA1468, TA11175 and TA11193) were grouped between the *judaicum* and the other accessions clusters in the PCA plot.

2.5 Population Structure

The population structure analysis further confirmed the differentiation of the WEW collection into various subgroups and admixture groups, aligning well with the phylogenetic tree and PCA grouping. When testing $K = 2$ ancestral subpopulations, only the *judaicum* subgroup was differentiated from the rest of the accessions (Fig. 4). The Northern and the SL population accessions were separated when K was increased to 3 or more (Fig. 4). However, no solid differentiation of the subpopulation was observed after $K = 3$. The four accessions that were separated from the SL and *judaicum* clades in both the phylogenetic tree and PCA were identified as admixtures including ancestry of *judaicum* and SL subgroups (Fig. 4).

To evaluate the relationship between these WEW subpopulations, particularly between *judaicum* and domesticated emmer (*T. turgidum* subsp. *dicoccum*), concerning the point of domestication and the cultivated emmer's immediate ancestor, we utilized 31,277 SNPs generated by genotyping-by-sequencing (GBS) of both wild and cultivated emmer. The genetic clustering based on the GBS data revealed that domesticated emmer is more closely related to wild emmer accessions from the Northern Population (Supplementary Figure S3), suggesting that *judaicum* is not the immediate ancestor of domesticated emmer wheat.

2.6 Diversity and F_{ST}

The Nei's diversity index and the number of segregating alleles computed indicated that the SL population has a higher Nei's index (0.17) compared to the Northern Population (0.15) and *judaicum* (0.06) (Supplementary Table S5). The pairwise F_{ST} calculations among the subpopulations revealed that *judaicum* is more differentiated than both the Northern Population and the SL. All three pairwise F_{ST}

approaches (Weir and Cockerham, beta, and Nei's) displayed a similar pattern of values among subpopulations. Specifically, the Northern and SL populations exhibited the lowest F_{ST} value (Nei's F_{ST} = 0.05), while *judaicum* and the Northern Population showed the highest pairwise F_{ST} values (Nei's F_{ST} = 0.204) [Supplementary Table S6]. Thus, *judaicum* represents the most genetically distinct WEW subgroup compared to the Northern and SL populations.

2.7 Linkage Disequilibrium

In each subpopulation, the linkage decay was rapid over the first 250 Kb (Supplementary Figure S4) with the pattern varying slightly among the subpopulations. Specifically, in the *judaicum* subpopulation, the mean r^2 dropped from 0.8 to 0.4 (half) over this interval, whereas in the Southern Levant and Northern Populations, the mean r^2 dropped from 0.5 to 0.25 and from 0.7 to 0.35, respectively (Supplementary Figure S4).

2.8 Genetic Resources for Disease Resistance in WEW

With the stem rust evaluations, we found two accessions (TA11165 and TA11196) resistant or moderately resistant to four races of the stem rust pathogen (Supplementary Table S1), whereas four (1.4%) other accessions (TA1042, TA1045, TA11208, and TA1464) were resistant or moderately resistant to three races. Fourteen other accessions (4.8%) were resistant or moderately resistant to two races, and 28 accessions (9.6%) were resistant or moderately resistant to only one race (Supplementary Table S1). The largest number of accessions resistant to an individual stem rust pathogen race was 26 (9%) to race TTKSK, followed by 17 (5.8%) to TKTTF, 14 (4.8%) to TTTTF, 10 (3.4%) to JRCQC, and 9 (3.1%) to TTRTF.

For the leaf rust evaluations, we found four (1.4%) accessions (TA66, TA1014, TA1434, and TA1465) that were either resistant or moderately resistant to all five races of leaf rust pathogen (Supplementary Table S1 and Supplementary Figure S6). The accession TA11156, which was resistant across all races, exhibited no symptoms (score '0') for any leaf rust pathogen. Unfortunately, TA11156 is not included in our genotyping list and could not be displayed in the phylogenetic tree. The accessions TA11204 and TA1474 were resistant or moderately resistant to four races of the leaf rust pathogen. Four (1.4%) other accessions (TA1389, TA1470, TA10667, and TA11208) were resistant or moderately resistant to three races. Similarly, 14 accessions (4.8%) were resistant or moderately resistant to two of the five races; and 33 accessions (11.3%) were resistant/moderately resistant to only one race (Supplementary Table S1 and Supplementary Figure S6). The largest number of accessions resistant/moderately resistant to an individual race was 41 (14.1%) to race TBBGS, followed by 19 (6.6%) to TNBGJ, 18 (6.2%) to BBBQD, 12 (4.2%) to TFBGQ, and 11 (3.8%) to MFPSB.

With the stripe rust evaluations, we found 59 accessions (20.3%) accessions were resistant or moderately resistant to all five races; 28 accessions (9.6%) were resistant or moderately resistant to four of the five races; 23 accessions (8%) were resistant or moderately resistant to three of the five races; 24 accessions (8.2%) were resistant to two of the five races; and 50 accessions (17.2%) were resistant to only one race (Supplementary Table S1 and Supplementary Figure S7). The largest number of

accessions resistant to an individual race was 150 (51.5%) to race PSTv-37, followed by 117 (40.2%) to race PST5006, 106 (36.4%) to race PSTv-143, 105 (36.1%) to race PSTv14, and 96 (33%) to race PSTv-221.

We also identified accessions that were resistant to all three rust pathogens. Notably, accession TA11204 was resistant or moderately resistant to 10 out of 15 tested races and was resistant or moderately resistant to at least two races of all three rust pathogens (Supplementary Table S1). Two other accessions, TA1474 and TA11196, were resistant to nine races together of three pathogens. Interestingly, TA1474 was not resistant to any stem rust pathogen races, whereas TA11196 was not resistant to any leaf rust races, but both accessions were resistant or moderately resistant to all five races of stripe rust (Supplementary Table S1). Two other accessions, TA11165 and TA1470, were resistant or moderately resistant to at least eight races. Thirteen (4.5%) other accessions were resistant to seven of 15 tested races, whereas 17 (5.8%) other accessions were resistant or moderately resistant to six races.

2.9 Distributions of Resistant Accessions Across Geography

We compared the rust reaction of accessions to their geographic origin to discern patterns of concentration for this trait. Comparison of the rust reactions with their sites of origin revealed that the SL population has a higher proportion of resistant accessions than the northern population and *judaicum* (Table 1, Fig. 5–6, Supplementary Figure S6 and S7). Additionally, the response of each rust pathogen varied across each subpopulation.

Table 1
Percentage of resistant or moderately resistant accessions per population group. Total accessions that were genetically grouped under SL,

Disease	Pathogen Races	% Resistant accessions per subgroup		
		SL	Northern Population	<i>judaicum</i>
Stem rust	TTKSK	14.4	0	5.7
	JRCQC	6.1	0	0
	TKTTF	7.8	0	8.6
	TTRTF	4.2	0	2.9
	TTTTF	7.2	0	5.7
Leaf Rust	TBBGS	19.2	4.1	8.6
	BBBQD	7.2	0	11.4
	TFBGQ	3.6	4.1	0
	MFPSB	3.6	2.6	0
	TNBGJ	7.8	2.6	5.7
Stripe Rust	PSTv-221	51.3	1.3	29.4
	PSTv14	51.8	14.3	17.6
	PSTv-143	47.6	11.7	40
	PST5006	51.2	15.6	48.6
	PSTv-37	65.9	21.8	54.3

2.9.1 Geographical distributions of stem rust resistance

For all stem rust races, the majority of resistant and moderately resistant accessions originated from the SL subpopulation (Table 1), particularly from the Haifa zone (Supplementary Table 1). Within the SL subgroup, the highest percentage of resistant accessions (14.4%) was observed against the race TTKSK, whereas the lowest percentage of resistant or moderately resistant accessions (4.2%) was observed against the race TTRTF (Table 1). Surprisingly, no resistant or moderately resistant accessions were observed in the Northern Population for any races of the stem rust pathogen. Within the *judaicum* subpopulation, the highest percentage of resistant accessions (8.6%) was observed against the race TKTTF, while no accessions were resistant to the race JRCQC (Table 1, Fig. 6).

2.9.2 Geographical distributions of leaf rust resistance

For all leaf rust pathogen races, the majority of resistant or moderately resistant accessions originated from the SL subpopulation (Table 1, Supplementary Table 1). Within the SL subgroup, the highest percentage of resistant accessions (19.2%) was observed against the race TBBGS, whereas the lowest percentage of resistant accessions (3.6%) was observed against the races TFBGQ and MFPSB (Table 1). Within the Northern Population, the highest percentage of resistant or moderately resistant accessions (4.2%) was observed against races like TBBGS and TFBGQ (Table 1), whereas no accessions resistant to BBBQD were observed in this subpopulation. Within the *judaicum* subpopulation, we observed resistant or moderately resistant accessions as follows: 11.4% to BBBQD, 8.6% to TBBGS, 5.7% to TNBGJ, and no resistant accessions to TFBGQ and MFPSB (Table 1, Supplementary Table 1).

2.9.3 Geographical distributions of stripe rust resistance

For the stripe rust pathogens, a higher number of resistant or moderately resistant accessions originated from the Southern Levant (SL) subpopulation (Table 1, Supplementary Table S1, and Supplementary Figure S7). Within the SL subgroup, the highest percentage of resistant accessions (65.9%) was observed against race PSTv-37, whereas the lowest percentage of resistant accessions (47.6%) was observed against race PSTv-143 (Table 1). This indicates that about 50% of accessions in the SL group are resistant to all five races of the stripe rust pathogen. Within the Northern Population, the highest percentage of resistant or moderately resistant accessions (21.8%) was observed against race PSTv-37, whereas the lowest percentage (1.3%) of resistant accessions was against PSTv-221 (Table 1). Within the *judaicum* subpopulation, the highest percentage (54.3%) of resistant or moderately resistant accessions was observed against race PSTv-37, while the lowest percentage (17.6%) of resistant accessions was observed against race PSTv-14 (Table 1, Supplementary Table S1, and Figure S7).

2.10 Genome-Wide Association Study (GWAS)

The GWAS revealed highly significant novel and some known loci that explain higher phenotypic variances (up to 60%) (Table 2; Supplementary Tables S7 and S8). Most of the GWAS models identified common causal regions whether we ran an analysis in the entire or only in the SL population.

Table 2

The genomic regions that are statistically significant ($p < 0.05$) after Bonferroni correction for the marker-trait association for stem rust (SR) resistance in the WEW. The phenotypic variance explained (PVE) in % and P-value was reported for the GWAS model represented by the bold characters (G, V, F, b). The Zavitan genome (v2.1) positions were used to track the annotated genes for the position. The loci naming at the second column follows chromosome_position.

Pathogen Races	Peak loci and Position	GWAS model	PVE %	P-value	Pop Group	Stem Rust QTL and functional genes on Zavitan [WEW_v2.1] around the GWAS peak marker region, and the literature cited for the known genes
TTKSK (Ug99)	1B_54289580	G, V, F, b	54	$4.5e^{-12}$	To, SL	Trichome birefringence-like 10 (<i>TBL</i>)[23, 44] (54.29 Mb), and QTL at 1BS in durum [68] and WEW [23]
	4B_672753066	G, b	9.4	$7.5e^{-10}$	To	Gibberellin 2-beta-dioxygenase 6-like at 672.9 Mb and zinc finger protein KNUCKLES-like (672.5 Mb)
	1B_94672107	G, V , F	7.5	$1.5e^{-9}$	To, SL	Probable sugar phosphate/phosphate translocator At3g11320 (94.674 Mb)
	3B_380702044	F	3	$2.4e^{-12}$	SL	Uncharacterized protein at 380.7 Mb; putative L-cysteine desulfhydrase 1, transcript variant X1 at 381.2 Mb
	5A_567843744	G	3	$2.5e^{-10}$	SL	BTB/POZ domain-containing protein At4g01160-like (567.85 Mb)
JRCQC	1B_54289580	G, V, F, b	26.5	$2.2e^{-14}$	To, SL	*** location overlapped with the loci for TTKSK at 1BS***
	3A_55126148	G, b , F	25	$7.9e^{-11}$	SL	JRCQC QTL at 3A of durum[69]; uncharacterized protein at 55.12 Mb; histone H2B.4-like at 55.14 Mb and probable LRR receptor-like serine/threonine-protein kinase At3g47570 at 54.95
G = rrBLUP; V = CMLM; F = FarmCPU and b = Blink						
SL = Southern Levant population except for <i>judaicum</i>						
Pop = Population						
To = Total population in the panel						

Pathogen Races	Peak loci and Position	GWAS model	PVE %	P-value	Pop Group	Stem Rust QTL and functional genes on Zavitan [WEW_v2.1] around the GWAS peak marker region, and the literature cited for the known genes
	5A_390655074	G, b	18.8	$6.12e^{-10}$	To	QTL reported at 5AL of the WEW and durum wheat[68, 70]; myb-related protein 306-like at 390.55 Mb
	6B_709557414	G, F, b	12.1	$3.1e^{-12}$	To, SL	i) <i>Sr13</i>[71] ortholog [MW033594.1 <i>T. turgidum</i> subsp. <i>durum</i> cultivar CAT-A1 <i>Sr13</i>] aligned (91%) at 717.02 Mb: RGA5-like, transcript variant X1 in the annotated file ii) Lysine-specific histone demethylase 1 homolog 3-like at 709.59 Mb iii) Putative disease resistance protein <i>RGA3</i> and <i>RGA4</i> at 715.7 Mb iv) Disease resistance protein PIK6-NP-like and G-type lectin S-receptor-like serine/threonine-protein kinase SD2-5 at 716 Mb v) Reported QTL at 6BL of durum and WEW [68, 70]
TKTTF	2B_714517390	G, V, b	14.7	$1.1e^{-14}$	To, SL	Importin subunit beta-1-like, transcript variant X1 and QTL for the race in durum 2BL [72]; Wheat ortholog <i>SR9E_H2</i> at vicinity (699.23 Mb)
	5A_597691058	G, b	12	$1.e^{-16}$	To, SL	Probable <i>WRKY</i> transcription factor 57 at 597.69 Mb, role of <i>WRKY</i> in plant defense [73] ***
	1B_56521356	b	6.7	$5.13e^{-16}$	SL	5'-methylthioadenosine/S-adenosylhomocysteine nucleosidase-like, transcript variant X2 (56.52 Mb)
G = rrBLUP; V = CMLM; F = FarmCPU and b = Blink						
SL = Southern Levant population except for <i>judaicum</i>						
Pop = Population						
To = Total population in the panel						

Pathogen Races	Peak loci and Position	GWAS model	PVE %	P-value	Pop Group	Stem Rust QTL and functional genes on Zavitan [WEW_v2.1] around the GWAS peak marker region, and the literature cited for the known genes
	3B_127728931	F, b	5	6.25e ⁻²¹	SL	Acyl transferase 9-like at 127.64 Mb; glutamine-tRNA ligase-like at 127.5 Mb and berberine bridge enzyme-like 26 at 127.49 Mb
	3B_116165318	G, b	3	2.3e ⁻¹⁷	To, SL	Serine/threonine-protein kinase Aurora-2 at 116.33 Mb; barley <i>Rpg1</i> candidate gene [74]
	6B_165520546	G, b	3	5.8e ⁻¹²	To	BOI-related E3 ubiquitin-protein ligase 1-like (165.41 Mb); putative homeobox-leucine zipper protein HO at 165.30 Mb and probable C-terminal domain small phosphatase at 165.77 Mb: plant immunity-related gene [75]
TTRTF	6A_534202576	G, V	41.3	1.1e ⁻⁸	To, SL	QTL for TTRTF reported at 6AL in durum wheat[76]; bZIP transcription factor 2-like at 534.23 Mb and uncharacterized protein at 534.20 Mb; and eukaryotic translation initiation factor 5-like, transcript variant X1 at 534.23; <i>Sr13</i> ortholog from durum mapped on 633.6 Mb
	3A_673490015	F, b	29.4	5.2e ⁻⁰⁹	SL, To	QTL at 3AL for this race in durum wheat[76]; putative disease resistance protein RGA3, transcript variant X3 at 673.56 Mb and exocyst complex component EXO70A1-like at 673.55 Mb
	5A_597691058	b	10	5.7e ⁻¹⁴	To	*** location overlapped with the loci for TKTTF at 5AL ***
G = rrBLUP; V = CMLM; F = FarmCPU and b = Blink						
SL = Southern Levant population except for <i>judaicum</i>						
Pop = Population						
To = Total population in the panel						

Pathogen Races	Peak loci and Position	GWAS model	PVE %	P-value	Pop Group	Stem Rust QTL and functional genes on Zavitan [WEW_v2.1] around the GWAS peak marker region, and the literature cited for the known genes
	2B_529637064	G, b	9	$6.2e^{-13}$	To	Peptide methionine sulfoxide reductase A2-1-like at 529.62 Mb [77]; NAC transcription factor NAM-B2-like; and BTB/POZ and TAZ domain-containing protein 2-like at 529.9 Mb
	5B_359661332	F, b	5	$9.9e^{-10}$	To	aminopeptidase M1-C-like at 360.63 Mb
TTTTF	4B_452141975	G, F, b	28.7	$1.2e^{-8}$	To	QTL at 4BL for this race in durum panel [68]; putative clathrin assembly protein At2g25430 at 452.14 Mb; palmitoyltransferase akr1-like, transcript variant X1 at 452.15 Mb and E3 ubiquitin-protein ligase XB3-like, transcript variant X2 at 453.06 Mb
	1B_469154488	b	15.2	$4.2e^{-10}$	To	Several QTL for stem rust race TTTTTF at 1B in durum [68]; probable glycosyltransferase STELLO2 at 469.16 Mb and probable membrane-associated kinase regulator 1, transcript variant X1 at 469.01 Mb
	2A_608467259	b	14	$4.6e^{-18}$	SL	CBS domain-containing protein CBSCBSPB3-like, transcript variant X6 at 608.37 Mb and putative E3 ubiquitin-protein ligase SINA-like 6 at 608.55 Mb
	5B_539977539	b	13.6	$1.2e^{-19}$	SL	Cortical cell-delineating protein-like and T-complex protein 1 subunit beta-like at 539.95 at Mb
G = rrBLUP; V = CMLM; F = FarmCPU and b = Blink						
SL = Southern Levant population except for <i>judaicum</i>						
Pop = Population						
To = Total population in the panel						

Pathogen Races	Peak loci and Position	GWAS model	PVE %	P-value	Pop Group	Stem Rust QTL and functional genes on Zavitan [WEW_v2.1] around the GWAS peak marker region, and the literature cited for the known genes
	5A_4511589	F, b	12.6	$7.1e^{-22}$	To	Pentatricopeptide repeat-containing protein At1g18900-like at 4.51 Mb
	2A_288803558	b	7.4	$4.6e^{-18}$	SL	Probable ribose-5-phosphate isomerase 3, chloroplastic at 289.0 Mb
	5A_263785125	F, b	4	$1.9e^{-11}$	SL	Serine/threonine-protein phosphatase 7 long form homolog at 263.93 Mb and basic leucine zipper 19-like at 263.50 Mb, which is plant defense gene [78–80]
	4B_570338558	F, b	3	$1.5e^{-10}$	SL	Protein IQ-DOMAIN 1-like at 570.32 Mb basic blue protein-like at 570.31 Mb and QTL at 4BL in durum [68]
G = rrBLUP; V = CMLM; F = FarmCPU and b = Blink						
SL = Southern Levant population except for <i>judaicum</i>						
Pop = Population						
To = Total population in the panel						

2.10.1 GWAS for the stem rust

Across all five races of stem rust evaluated, we identified 28 genomic regions significantly associated with resistance to *P. graminis* f. sp. *tritici* based on Bonferroni-corrected P-values (Table 2). In response to race TTKSK, we found a highly significant locus at 1BS, at 54.289 Mb, explaining 54.3% of the phenotype variance. This region was identified by various GWAS models such as Blink, FarmCPU, CMLM, and rrBLUP (Table 2 and Fig. 7). Additional loci associated with resistance to this race were detected on chromosomes 3B (380 Mb), 4B (672 Mb), and 5A (567 Mb) (Table 2). While many of these associations are novel, some have been reported in earlier studies [23]. For the JRCQC stem rust race, we found a major association at the same location at 1BS (PVE = 26.5%) as with resistance to TTKSK (54.28 Mb) (Table 2). Other significant loci for this race were identified at 3A, 5A and 6B (Table 2).

For the TKTF race, we observed two major associations at 2BL (PVE = 14.7%) and 5AL (PVE = 12%). Additional significant loci for resistance to this race were detected at 1BS, 3BS, and 6BS (Table 2). Similarly, to race TTRTF, the major genomic region associated with resistance was identified at 3AL, 5AL,

2BL, and 5BL (Table 2). Likewise, for the TTTTF race, a major QTL was detected at 4BL. Other significantly associated loci for this race were observed at 1BL, 2AL, 4BL, 5AS, and 5BL (Table 2).

2.10.2 GWAS for the leaf rust

Twenty-five (25) loci were found associated with resistance to the five races of *Puccinia triticina*. Major effect loci for resistance to race TBBGS were found at 1AS (PVE = 10%), 2A, 5A, 4B, 4A, and 3A (Supplementary Table S7). With race BBBQD, significant associations were found at 3BS, 6BL, 4BL, 3BL, 2AS, and 1AS (Supplementary Table S7). Perhaps due to the limited number of resistant accessions identified to race TFBGQ (Supplementary Figure S6), only two loci were identified: on chromosomes 3B and 6B, exclusively within the SL group (Supplementary Table S7). For race MFPSB, four different genomic regions were associated with resistance at 2AL, 2BL, 3BS, and 7BS (Supplementary Table S7). For race TNBGJ, significant associations (PVE = 46%) were observed around 257.9 Mb on chromosome 2A, as well as on 3BS (87 Mb), 2BL (498 Mb), 1BS (7 Mb), 2AS (31 Mb), and 2AL (733 Mb) (Supplementary Table S7).

2.10.3 GWAS for the stripe/yellow rust

For the five yellow/stripe rust races, we detected a total of 32 loci associated with resistance, ranging from 4 to 7 loci per race (Supplementary Table S8). A highly significant major effect locus for resistance to race PSTv-221 was positioned at 99 Mb on chromosome 1BS, explaining about 16% of the phenotypic variance (Supplementary Table S8). Other significant loci associated with resistance to this race were detected at 1A (321 Mb), 3A (633 Mb), 6B (444 Mb), 5A (563 Mb), and 7A (221 Mb) (Supplementary Table S8). For race PSTv-14, significantly associated loci were observed on chromosomes 1A, 1B, 2A, 4B, 5B, 7A, and 7B, contributing from 3–16.7% of phenotypic variance (Supplementary Table S8). For stripe rust race PSTv-143, significant marker-trait associations were found at 1A, 1B, 2A, 3B, 4A, and 5B, contributing up to 22.2% of phenotypic variance. Similarly, for race PST5006, the most significant association was observed at 6B at 155 Mb, while other loci detected for this race were on 1A, 1B, 2A, 4B, and 7A of WEW (Supplementary Table S8). For another race, PSTv-37, we obtained a highly influential loci at 3AL that contributed nearly 60% of PVE and was located around 716 Mb of the Zavitan (v2.1) genome. Another important association for this race was observed at 1BS, and other minor loci were also found at 1A, 2A, 4B, 6A, and 6B (Supplementary Table S8).

2.11 Known Candidate Genes and Novel Loci Identified by GWAS

The highly significant locus at 1BS associated with resistance to *P. graminis* f. sp. *tritiraces* TTKS and JRCQC was discovered near the ortholog of the trichome birefringence-like (*TBL*) family gene (Supplementary Table S9) on the Zavitan genome. The locus consistently emerged in GWAS conducted with the entire population or the SL, across multiple GWAS models, affirming the reliability of the identified loci. Another locus, '6B_709557414', found on chromosome 6B could be *Sr13*. Notably, when aligning the 2Kb coding sequence of durum *Sr13*, a high alignment rate (96%) was observed at 6AL, the expected position of *Sr13*, but also a substantial alignment (91%) at 709 Mb of 6B, suggesting a

potential additional copy of *Sr13* at 6BL (Supplementary Table S9). This could indicate the introgression of durum gene *Sr13* into the Zavitan, given the significant introgression from durum observed in the wild emmer wheat [24]. Alternatively, the gene at 6B could be a novel gene. The region around 709 Mb at 6BL harbors several functional genes, including the disease-related gene RGA3 at 715 Mb (Table 2).

For leaf rust, the GWAS-identified loci were associated with several potential candidate genes. For instance, the locus identified for the TBBGS race on 4BS at 70 Mb likely represents the receptor protein kinase-like protein *ZAR1*, a gene associated with leaf rust resistance [25]. Another gene for the same race, located on the long arm of 4A (748 Mb), likely represents the ortholog of a leaf rust-related kinase, known as the Lr10-like gene (Supplementary Table S7). Additionally, a locus for MFPSB at 2A (2A_733147249) was found to be collocated with an ortholog of the RING-H2 finger protein ATL43-like gene, which plays a role in plant defense [26]. Similarly, another orthologous gene, 'NF-X1-type zinc finger NFXL2-like', collocated with the GWAS-identified locus at 3B (159 Mb) for TBBGS, has also been reported as related to leaf rust resistance [27]. Some other GWAS-identified loci are collocated with other functional genes and most likely represent novel loci (Supplementary Table S7).

For stripe rust resistance, some GWAS peaks we identified were collocated with the ortholog sequences of previously reported stripe rust resistance genes, such as *Yr15*, *YrU1*, and *Yr36* (Supplementary Table S8-S9). A GWAS peak loci for races PSTv-221 and PSTv-14 (1B_99349879), located at 99 Mb on chromosome 1BS, was identified near the alignment (93%) position of the *Yr15* at 97.9 Mb, while the GWAS locus for PSTv-14 (5A_563398244) was closer to the alignment (99%) position (Supplementary Table S9) of the *urartu* stripe rust gene *YrU1* at 554 Mb of 5AL. Similarly, the GWAS peak at 6B (6B_155315249) was in the vicinity of the alignment (98%) position of an ortholog of the wheat stripe rust gene *Yr36* (Supplementary Table S8). Although *Yr36* is an adult plant resistance gene, when pyramiding with other adult plant resistance genes such as *Yr18* and *Yr28* also conferred seedling stage resistance[28]. Additionally, a GWAS-identified locus for PSTv-14 (7A_638642386) was found to collocate with an annotated gene, *Ethylene-responsive transcription factor 3-like* at 638.77 Mb on 7A. Similarly, the *WRKY* transcription factor was found to collocate with the GWAS loci at 1BS (1B_123847300) for the races PSTv-143 and PSTv-37. Both of these genes have been previously reported as related to stripe rust resistance [29, 30]. Some other previously known stripe rust QTL and the loci we identified here using GWAS were collocated or resided in the vicinity of the same chromosomes as described in Supplementary Table S8.

3. Discussions

3.1 WEW Population Structure and *judaicum* Race

The genetic separation of a unique *judaicum* subgroup, distinct from the rest of the Southern Levant (SL) population, is a notable observation in this study (Figs. 1 and 2). Furthermore, the clustering of two northern Syrian lines with Lebanese and other southern Syrian accessions likely suggests the dispersal of WEW along the Mediterranean coastal region. This is supported by several studies identifying

southeastern Anatolia and northern Syria as centers of emmer domestication [31] [32]. Since we lack accessions from the region between northern Lebanon and northern Syria, it would be beneficial to include accessions from this area in future studies. The four admixture accessions (Fig. 2, 3, and 4) observed with shared ancestry between SL and northern populations indicate that the extent of admixture in wild populations could be larger if a more extensive population, from southeastern Anatolia to the southern Levant, is surveyed.

In previous studies, some authors recognized *judaicum* [19] as a separate subspecies of *T. turgidum*, while other recognized the group as a race and analyzed its morphological differences with another race *horanum* [20] of the SL population. In this study, we found that *judaicum* has a stronger population differentiation from the rest of the WEW accessions and could represent one of the diverse genetic groups (Fig. 2–4). Blumler (1998) suggested the origin of *judaicum* stems from the hybridization of wild emmer and durum wheat [22, 24]. However, further investigation into this diverse genetic subgroup is necessary to confirm its unique genetic and evolutionary differentiation. In this study, with the GBS data, we observed that *judaicum* is not an immediate ancestor of cultivated emmer, using thousands of genetic variants. Based on supplemental GBS data containing thousands of SNPs, our study concludes that *judaicum* is a naturally occurring genetically distinct subgroup within wild emmer wheat. It is confirmed not to be an immediate ancestor of domesticated emmer nor a feral emmer subgroup (Fig. 4).

3.2 Geographical Origin and Disease Resistance

In this study, we identified significant regional variation in the genetic resistance of WEW accessions to rust diseases. Specifically, the SL population has more resistant accessions, followed by the *judaicum* subgroup (Table 2, Supplementary Table S1). Regarding stem rust, we did not observe any resistant accessions in the Northern Population, highlighting the importance of considering this finding when selecting breeding materials in the future. These findings are consistent with previous studies that reported higher stem rust resistance within the SL population and more susceptibility in the Northern Population [23]. However, unlike the patterns observed for stem rust, the Northern Population exhibited some resistant accessions to at least four races of leaf rust and all five races of the stripe rust pathogen (Table 1, Supplementary Table S1, and Supplementary Figures S6-S7). This indicates that there is not a complete absence of an R gene pool in this population type. Similarly, the *judaicum* group also failed to provide resistant accessions for two races of leaf rust and one race of stem rust, highlighting that the SL population is the only subgroup with accessions resistant to all 15 tested races (Table 1). Yet, the performance of SL group accessions against stem and leaf rust pathogen races is not as proportionate as against the stripe rust races (Table 1). This observation aligns with a previous study indicating a relatively lower abundance of leaf rust resistance in the WEW panel compared to rust caused by Pst [33].

On the other hand, the Northern Population may also harbor resistance genomic regions (R genes) more similar to those in domesticated wheat, where there is much more selective pressure due to population abundance. Consequently, the R genes present in the Northern Population were much more likely to be overcome. Additionally, the overall genetic diversity in the Southern Population is higher (Supplementary Table S5), which may also be reflected in the R gene pool. Over time, natural selection in response to

local environmental conditions, including temperature, could have shaped the genetic diversity and effectiveness of R genes in WEW, providing regional variation. Several temperature-sensitive R genes in WEW were discussed earlier [34].

3.3 Seedling Stage Disease Resistance

In this study, we observed that the rust pathogens' genetic resistance at the seedling stage is associated with 4–6 loci, exhibiting a range of resistance levels from highly resistant to highly susceptible. Multiple quantitative trait loci (QTL) for seedling resistance are commonly found in both cultivated wheat and wild wheat. For instance, D Lhamo, Q Sun, Q Zhang, X Li, JD Fiedler, G Xia, JD Faris, Y-Q Gu, U Gill, X Cai, et al. [35] reported 15 QTL for seedling resistance against a single leaf rust race, BBBQD, in cultivated emmer wheat. Similarly, Tene et al. (2022) identified several minor QTL for seedling-stage stripe rust resistance in WEW, while Megerssa et al. (2021) identified multiple QTL for seedling resistance against various stem rust races like TTKSK, TKTTF, JRCQC, and TTRTF in a durum wheat panel. This complexity underscores the necessity of considering multiple genetic factors and environmental variables in breeding strategies aimed at enhancing rust resistance in wheat and its relatives. This is most probably the largest panel screened so far with the WGS-based millions of variants. Therefore, the genetic resources we selected are robust and can also be screened for other disease genes for gene stacking.

3.4 Disease Resistance Accessions- Valuable Genetic Resources

This study represents one of the most robust population genomics studies of wild emmer wheat (WEW) in terms of accession diversity and geographical distribution coverage. To our knowledge, this is the first genetic analysis screening many pathogen races (15) on WEW together using WGS-based SNPs. Notably, races such as Pst-221 and Pst-143 of the stripe rust pathogen have not been previously tested in this wild wheat species, or the information about them is not publicly available. Similarly, there have been limited reports on genetic association studies of leaf rust races in WEW. Therefore, the rust-resistant accessions identified in our study represent valuable genetic resources, offering an excellent source of novel genes. For instance, TA66, highly resistant to all leaf rust races (Supplementary Table S1), could be a valuable germplasm for enhancing bread wheat resistance against leaf rust. Accessions such as TA11204, which was resistant or moderately resistant to many races (10 out of 15 tested), along with other rust-specific resistant accessions like TA1041, TA1047, and TA11196, hold promise as key sources of resistance for developing advanced breeding lines with broad-spectrum resistance.

3.5 Abundance Stripe Rust Resistance in SL Subpopulation

Forty-seven to sixty-six percent of total accessions under the SL group were resistant or moderately resistant to all five races of the stripe rust pathogen (Table 2). This is very promising and suggests that this group of populations most likely represents greater genetic resources for other stripe rust pathogen races not tested here as well. A previous study also reported the superiority of the southern population over the northern population in terms of stripe rust resistance [14]. Nevertheless, all population subgroups showed higher resistance to stripe rust pathogens than leaf rust and stem rust, suggesting

that the entire WEW population is a crucial reservoir of genetic diversity for stripe rust resistance (Table 2, Supplementary Table S1).

3.6 Cloned Genes from the SL Subgroup

To date, the majority of cloned R genes from WEW have originated from the southern population. This explains why more resistant lines are found in the Southern Levant (SL) subgroup. For example, an Israeli accession from Mount Hermon, H52, was identified as the source of the stripe rust resistance gene *YrH52* [36]. The broad-spectrum R-gene *Yr15*, which encodes a putative kinase-pseudokinase protein, was identified in the Israeli accession G25 (TA1474) from Rosh Pinna [37]. *Yr15* has proven effective against more than 3000 genetically and geographically diverse Pst isolates [38]. Our disease screening analysis revealed that TA1474 was resistant or moderately resistant to four races of leaf rust and all races of stripe rust pathogens (Supplementary Table 1), demonstrating the accuracy of our phenotyping screening. Additionally, the powdery mildew resistance gene *PMG25* was discovered in the same accession [39]. Another powdery mildew resistance gene, *PmG3M*, was identified in the Southern Levant accession G-305-M from Israel, and it is effective against 41 powdery mildew isolates from diverse wheat species and locations across Israel [40–42]. Similarly, the *Pm36* gene is present in only a few WEW lines of the SL subgroup [10]. Therefore, it is important to investigate the effectiveness and abundance of the R gene pool in the SL subgroup compared to the Northern Population.

3.7 GWAS and Novel Loci

Employing comparative genomics, leveraging sequence homology, and comparing data on published resistance genes, we discerned that some of the identified genes are likely candidate genes characterized earlier. In this study, we also identified several new loci through GWAS (Table 2, Supplementary Table S7 and S8). The identified major loci present promising targets for breeding efforts aimed at improving resistance in cultivated wheat varieties. For instance, upon investigating a highly significant GWAS-identified region for stem rust race TTKSK (Ug99) at 1BS, we identified a candidate gene, Trichome birefringence-like 10 (*TBL*), in the region and confirmed the presence of a gene with wheat *TBL* orthologs alignment (Supplementary Table S9) using nucleotide BLAST. In Arabidopsis, the *TBL* genes contribute to cellulose synthesis and deposition required for the secondary cell wall, most likely through the esterification state of pectic polymers [43]. There are several previous reports which observed that cytomembrane-related genes are associated with the R genes. R Sharma Poudel, AF Al-Hashel, T Gross, P Gross and R Brueggeman [44] reported that out of 7 *Rrr1* candidate genes for stem rust resistance in barley, 3 belong to the *TBL* family, which functions as both positive and negative regulators of plant pathogen defense, depending on the pathosystem. There are other instances of defense genes, which are located in the cytomembrane, involved in the cell wall matrix, and transform signals into the cytoplasm, such as wall-associated kinase (WAK) [45]. Therefore, further investigation of *TBL*, taking the gene into account as a strong candidate gene for stem rust resistance, is essential.

In another example, we identified the *YR36* gene, previously reported for adult plant stripe rust resistance [46], as one of the causative loci against stripe rust races PST5006 and PSTv-37 at 6BS near 149 Mb

(Supplementary Table S8). A recent study also indicated that pyramiding *YR36* with other adult plant resistance genes such as *Yr18* and *Yr28* conferred seedling stage resistance [28]. Thus, it is plausible that *Yr36* alone could also be effective against races like PST5006 and PSTv-37.

3.8 WGS data for the research community

The study provides a substantial amount (14 TB) of whole-genome sequencing (WGS) data for 291 accessions of WEW on the NCBI SRA, aiming to contribute to the research community by granting access to the sequencing data and enabling further exploration and analysis. This high-coverage sequencing (~ 9.5x) data has great value to researchers, especially those gene banks with duplicated WEW germplasms [47], the sequence data can be easily utilized for trait mapping on their plants, eliminating the need for additional sequencing. Thus, the finding observed in the study and the data resources generated certainly adds new information to WEW population genetics studies and highlight the potential of utilizing their genetic diversity to improve cultivated wheat.

4. Conclusion

In conclusion, this study leveraged a massive whole-genome sequencing (WGS) dataset of wild emmer wheat (WEW) to explore the population structure, genetic diversity, and loci associated with rust disease resistance, highlighting WEW as a potential genetic resource for improving wheat. The analysis identified distinct genetic subgroups, namely 'SL, Northern Population, and *judaicum*,' within WEW, along with several genetic regions linked to resistance against various rust diseases. Additionally, several novel candidate genomic regions for rust resistance were revealed. The screening of some rust races for the first time in WEW added further knowledge about WEW's genetic response to these pathogens. By making this sequencing data (14 TB) publicly available, the research opens doors for further exploration into disease resistance genetics and eventually aids in the development of wheat varieties with enhanced rust resistance.

5. Methods

5.1 Plant Materials and Pathogen Isolates

We developed a genetically diverse panel of WEW comprising 291 accessions from four gene banks: 137 from the Wheat Genetics Resource Center (WGRC) at Kansas State University (KSU) in Manhattan KS, USA; 76 from the USDA-ARS National Small Grains Collection (NSGC) in Aberdeen, ID, USA; 58 accessions from the Council for Agricultural Research and Agricultural Economics Analysis (CREA) in Fiorenzuola d'Arda, Italy which were originally obtained from the National BioResource Project (NBRP) in Komugi, Kyoto, Japan and selfed to maintain. Further, we obtained additional 20 accessions directly from NBRP. The selection of accessions aimed to cover the broadest geographic distribution of the species, encompassing Southeastern (SE) Anatolia, Iraq, Iran and the Southern Levant (SL). All accessions were increased from single plants grown in the greenhouse. As a result, the collection represents a refined germplasm collection with a high level of homozygosity suitable for various

experiments. The curated passport data includes detailed information on all of the accessions in the panel (Supplementary Table S1). Accessions from the NSGC were chosen based on previous evaluations for stem rust resistance. The panel was evaluated for resistance at the seedling stage to five diverse races each of *Puccinia graminis* f. sp. *tritici*, *P. triticina* and *P. striiformis* f. sp. *tritici*. Details regarding the rust isolates, inoculation protocol and disease assessment methods are given in Supplementary Note S1 and Supplementary Table S1.

5.2 Genotyping WGS Panel

For whole-genome sequencing (WGS), we generated PCR-free libraries using the TruSeq DNA PCR-FREE LT Sample Prep Kit (Illumina, Inc.), followed by paired-end sequencing at Novogene (Novogene Co., Ltd., Beijing, China). Raw sequencing files were trimmed using fastp [48] to eliminate adapters and any undesired sequences. High-quality sequences were aligned to a WEW genome assembly available for the 'Zavitan' accession (Genome assembly WEW_v2.1) [49], using the HISAT2 aligner [50], and the resulting alignment files were utilized for variant calling using bcftools [51]. The raw variants in the VCF file were subjected to an initial filtration process based on quality (≥ 30), read depth (≥ 20), allele frequency (AF) ($0.01 < AF < 0.99$), and missing data (≤ 0.2) using the *bcftool filter* command in bcftools [51]. Additional filtering was conducted for minor allele frequency (MAF) and heterozygosity. Loci with less than 0.05 MAF and those exhibiting more than 10% heterozygosity were removed. For computational efficiency in subsequent analyses and preservation of representative genomic information, we employed the 'SelectVariants' option in GATK tool as detailed by [52], to select a 5% subset of the filtered variants from each of the 14 chromosomes. Subsequently, the selected variants from the 14 chromosomes were merged into a single VCF file.

5.2.1 GBS Data and Relationship with Cultivated and Wild Emmer Wheat

In a separate analysis, we also used genotyping-by-sequencing (GBS) data to assess the genetic relationship between the subgroups of *T. turgidum* subsp. *dicoccoides*, especially the *judaicum* race, and cultivated emmer (*T. turgidum* subsp. *dicoccum* L.). For this analysis, we investigated 483 accessions and 31,277 GBS SNPs using WGRG wild and domesticated emmer wheat collection. The SNP matrix and accession information are available in the Dryad deposition link provided in the data availability section.

5.3 Passport Data Curation

We conducted allele matching within the panel accessions, employing a threshold of $\geq 99\%$, and documented any accessions that exhibited genetic identity [4, 47]. Furthermore, we cross-referenced passport data provided by the different gene banks with the USDA-ARS Germplasm Resources Information Network (GRIN) and Genesys (<https://www.genesys-pgr.org/>) databases to identify any existing discrepancies. Any questionable passport information, particularly concerning the geographical coordinates of collection sites, was verified with the assistance of gene bank curators.

5.4 Population Structure Analysis

We explored the WEW population composition by phylogenetic clustering, principal component analysis (PCA) and population structure analysis as described [4, 47]. Phylogenetic clustering uses the 'dist' function to compute the genetic distance. We used the R packages of 'ape' [53] and 'phyclust' [54] to convert the genetic distance to a phylo object and for plotting the tree, respectively. The branches of an unrooted neighbor-joining (NJ) tree were colored based on their sites of origin. Additionally, a PCA plot was also generated with eigenvalues and eigenvectors computed using the 'e' function using the A.mat computed with rrBlup in R [47, 55]. The first two eigenvectors were plotted and colored the subgroups using a corresponding color pattern as was done for the NJ tree.

The population structure analysis was done in R [56] using the Landscape and Ecological Association Studies (LEA) package [57]. In this approach, we computed least-squares estimates of ancestry proportions and ancestral allelic frequency using the function 'snmf' assuming $K = 6$ for ancestral populations. The numbers of population subgroups (K) describing the genotypic data were defined using the cross-entropy value based on the maximal local contribution of ancestry. Then, we plotted the Q matrix value for the population ancestry proportions using the R package Pophelper [58].

5.5 Diversity and Pairwise F_{ST}

We computed Nei's diversity index [59] and the total segregating alleles per subpopulation as previously described [4]. The diversity indices were computed for all accessions together and the subpopulations separately. The Weir and Cockerham F_{ST} [60], betas [61] F_{ST} and the Nei's pairwise F_{ST} [62] were computed between the subgroups using 'hierfstat' [<https://github.com/jgx65/hierfstat>] package in R [56].

5.6 Linkage Disequilibrium and GWAS

To study the linkage disequilibrium (LD) and its decay in the WEW panel, we selected subsets of 20K SNPs for each of the Northern Population, *judaicum* and the SL population. For each subpopulation, pairwise correlations (r^2) between alleles at two loci for all SNP pairs was computed using VCFtools [63]. The LD decay curves were plotted using the ggplot2 [64] package in R (version 4.3.2) [56], where only the means of pairwise LD (r^2), computed using a 100 Kb window, were plotted.

GWAS was carried out using various models including rrBLUP, CMLM, BLINK, and FarmCPU to validate the outputs of one another and to avoid algorithm bias. The latter three methods were implemented using GAPIT software [65, 66]. For GWAS, we used the Q matrix as covariates, which was calculated using PCA, and the kinship matrix as individual relatedness along with the genotypic data. Since some of the WEW accessions from the Northern Population exhibited no or very low resistance to the tested rust races, we also independently ran the GWAS only in the SL population and compared the output obtained for the entire population.

5.7 Comparative Genomics and Potential Candidate Genes

We conducted a comparison between the GWAS-identified loci and previously reported seedling resistance-related loci and candidate genes by examining their relative positions on the genome. Disease-related candidate genes with ortholog sequences were directly aligned using BLASTn[67] on the Zavitan genome (v2.1), while other loci were queried within the updated annotated file (unpublished). We searched for annotated functional genes in the vicinity of the GWAS peak loci (approximately 1 Mb) and reviewed the literature for any previous reports or evidence of their involvement in disease response.

Declarations

Data and Code Availability

A variant call format (VCF) file, hapmap file with the SNP matrix, and Linux script and statistical codes used in the study, were deposited in the Dryad public repository under doi.org/10.5061/dryad.zcrjdfnmb. The whole genome sequencing (WGS) data of the wild emmer wheat (WEW) panel investigated in the study has been deposited in the NCBI public repository sequence read archive (SRA) under the BioProject accession # PRJNA1007489.

Acknowledgement

The authors acknowledge the funding support (Award No. 2020-67013-30871) received from the National Institute of Food and Agriculture (NIFA) for conducting this study. We thank the Wheat Genetics Resource Center at Kansas State University for providing germplasms. The ‘Beocat’ supercomputer facilities of the Kansas State University and KAUST supercomputer facility, ‘lbex’ were used for computational analysis and we are grateful for these facilities. The Department of Plant Pathology at the University of Minnesota, and the Center for Desert Agriculture, King Abdullah University of Science and Technology (KAUST).

Authors Contributions

B.S., J.P., and B.W. conceptualized the study, secured grants, and provided ongoing guidance throughout the project. P.O. conducted phenotyping for the GWAS and analyzed phenotype data, while L.A. curated the initial sequencing data obtained in different batches, genotyped the population, and performed downstream population analysis. Both L.A. and P.O., with the assistance of B.S., J.P., and B.W., prepared the initial draft of the manuscript. G.Y. conducted GWAS and contributed to manuscript preparation. J.R. provided germplasms, and both J.R. and H.S. curated the passport information of the germplasms. A.D. and H.S. assisted in population analysis and provided updated Zavitan genome annotation. B.S., J.P., and B.W. revised the manuscript, and all authors read and reviewed the final version.

References

1. Chai Y, Pardey PG, Hurley TM, Senay SD, Beddow JM. A Probabilistic Bio-Economic Assessment of the Global Consequences of Wheat Leaf Rust. *Phytopathology*® 2020, 110(12):1886–1896.

2. Beddow JM, Pardey PG, Chai Y, Hurley TM, Kriticos DJ, Braun H-J, Park RF, Cuddy WS, Yonow T. Research investment implications of shifts in the global geography of wheat stripe rust. *Nat Plants*. 2015;1(10):15132.
3. Singh RP, Hodson DP, Huerta-Espino J, Jin Y, Bhavani S, Njau P, Herrera-Foessel S, Singh PK, Singh S, Govindan V. The emergence of Ug99 races of the stem rust fungus is a threat to world wheat production. *Annu Rev Phytopathol*. 2011;49:465–81.
4. Adhikari L, Raupp J, Wu S, Koo D-H, Friebe B, Poland J. Genomic characterization and gene bank curation of *Aegilops*: the wild relatives of wheat. *Front Plant Sci* 2023, 14.
5. Balla MY, Gorafi YSA, Kamal NM, Abdalla MGA, Tahir ISA, Tsujimoto H. Harnessing the diversity of wild emmer wheat for genetic improvement of durum wheat. *Theor Appl Genet*. 2022;135(5):1671–84.
6. Ahmed HI, Heuberger M, Schoen A, Koo D-H, Quiroz-Chavez J, Adhikari L, Raupp J, Cauet S, Rodde N, Cravero C, et al. Einkorn genomics sheds light on history of the oldest domesticated wheat. *Nature*. 2023;620(7975):830–8.
7. Oliveira HR, Jacocks L, Czajkowska BI, Kennedy SL, Brown TA. Multiregional origins of the domesticated tetraploid wheats. *PLoS ONE*. 2020;15(1):e0227148.
8. Li M, Dong L, Li B, Wang Z, Xie J, Qiu D, Li Y, Shi W, Yang L, Wu Q, et al. A CNL protein in wild emmer wheat confers powdery mildew resistance. *New Phytol*. 2020;228(3):1027–37.
9. Gioia T, Nagel KA, Beleggia R, Fragasso M, Ficco DBM, Pieruschka R, De Vita P, Fiorani F, Papa R. Impact of domestication on the phenotypic architecture of durum wheat under contrasting nitrogen fertilization. *J Exp Bot*. 2015;66(18):5519–30.
10. Li M, Zhang H, Xiao H, Zhu K, Shi W, Zhang D, Wang Y, Yang L, Wu Q, Xie J, et al. A membrane associated tandem kinase from wild emmer wheat confers broad-spectrum resistance to powdery mildew. *Nat Commun*. 2024;15(1):3124.
11. Klymiuk V, Chawla HS, Wiebe K, Ens J, Fatiukha A, Govta L, Fahima T, Pozniak CJ. Discovery of stripe rust resistance with incomplete dominance in wild emmer wheat using bulked segregant analysis sequencing. *Commun Biol*. 2022;5(1):826.
12. Klymiuk V, Yaniv E, Huang L, Raats D, Fatiukha A, Chen S, Feng L, Frenkel Z, Krugman T, Lidzbarsky G, et al. Cloning of the wheat Yr15 resistance gene sheds light on the plant tandem kinase-pseudokinase family. *Nat Commun*. 2018;9(1):3735.
13. Elkot AF, Singh R, Kaur S, Kaur J, Chhuneja P. Mapping novel sources of leaf rust and stripe rust resistance introgressed from *Triticum dicoccoides* in cultivated tetraploid wheat background. *J Plant Biochem Biotechnol*. 2021;30(2):336–42.
14. Tene M, Adhikari E, Cobo N, Jordan KW, Matny O, del Blanco IA, Roter J, Ezrati S, Govta L, Manisterski J, et al. GWAS for Stripe Rust Resistance in Wild Emmer Wheat (*Triticum dicoccoides*) Population: Obstacles and Solutions. *Crops*. 2022;2(1):42–61.
15. Bai D, Knott DR. Genetic studies of leaf and stem rust resistance in six accessions of *Triticum turgidum* var. *dicoccoides*. *Genome*. 1994;37(3):405–9.

16. Knott DR, Bai D, Zale J. The transfer of leaf and stem rust resistance from wild emmer wheats to durum and common wheat. *Can J Plant Sci.* 2005;85(1):49–57.
17. Bai D. Transfer of rust resistance genes to tetraploid and hexaploid wheat from some related *Triticum* species. Saskatoon, SK: University of Saskatchewan; 1992.
18. 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(1):11–53.
19. Poyarkova H. Morphology, geography and infraspecific taxonomics of *Triticum dicoccoides* Körn. A retrospective of 80 years of research. *Euphytica.* 1988;38(1):11–23.
20. Poyarkova H, Gerechter-Amitai ZK, Genizi A. Two variants of wild emmer (*Triticum dicoccoides*) native to Israel: morphology and distribution. *Can J Bot.* 1991;69(12):2772–89.
21. Jorgensen C, Luo M-C, Ramasamy R, Dawson M, Gill BS, Korol AB, Distelfeld A, Dvorak J. A High-Density Genetic Map of Wild Emmer Wheat from the Karaca Dağ Region Provides New Evidence on the Structure and Evolution of Wheat Chromosomes. *Front Plant Sci* 2017, 8.
22. Luo MC, Yang ZL, You FM, Kawahara T, Waines JG, 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(6):947–59.
23. Mastrangelo AM, Roncallo P, Matny O, Čegan R, Steffenson B, Echenique V, Šafář J, Battaglia R, Barabaschi D, Cattivelli L et al. A new wild emmer wheat panel allows to map new loci associated with resistance to stem rust at seedling stage. *The Plant Genome* 2023, n/a(n/a):e20413.
24. Blumler MA. Introgression of durum into wild emmer and the agricultural origin question. In: *The origins of agriculture and crop domestication Proc Harlan Symp: 1998; Aleppo, Syria*. Published jointly by ICARDA, IPGRI, FAO, and UC/GRCP; 1998: 252–268.
25. Gu J, Sun J, Liu N, Sun X, Liu C, Wu L, Liu G, Zeng F, Hou C, Han S, et al. A novel cysteine-rich receptor-like kinase gene, TaCRK2, contributes to leaf rust resistance in wheat. *Mol Plant Pathol.* 2020;21(5):732–46.
26. Aguilar-Hernández V, Aguilar-Henonin L, Guzmán P. Diversity in the architecture of ATLs, a family of plant ubiquitin-ligases, leads to recognition and targeting of substrates in different cellular environments. *PLoS ONE.* 2011;6(8):e23934.
27. Satapathy L, Singh D, Ranjan P, Kumar D, Kumar M, Prabhu KV, Mukhopadhyay K. Transcriptome-wide analysis of WRKY transcription factors in wheat and their leaf rust responsive expression profiling. *Mol Genet Genomics.* 2014;289(6):1289–306.
28. Wang F, Zhang M, Hu Y, Gan M, Jiang B, Hao M, Ning S, Yuan Z, Chen X, Chen X, et al. Pyramiding of Adult-Plant Resistance Genes Enhances All-Stage Resistance to Wheat Stripe Rust. *Plant Dis.* 2023;107(3):879–85.
29. Wang J, Tao F, Tian W, Guo Z, Chen X, Xu X, Shang H, Hu X. The wheat WRKY transcription factors TaWRKY49 and TaWRKY62 confer differential high-temperature seedling-plant resistance to *Puccinia striiformis* f. sp. *tritici*. *PLoS ONE.* 2017;12(7):e0181963.

30. Hawku MD, Goher F, Islam MA, Guo J, He F, Bai X, Yuan P, Kang Z, Guo J. TaAP2-15, An AP2/ERF Transcription Factor, Is Positively Involved in Wheat Resistance to *Puccinia striiformis* f. sp. tritici. *Int J Mol Sci* 2021, 22(4).
31. Özkan H, Brandolini A, Schäfer-Pregl R, Salamini F. AFLP analysis of a collection of tetraploid wheats indicates the origin of emmer and hard wheat domestication in southeast Turkey. *Mol Biol Evol.* 2002;19(10):1797–801.
32. Lev-Yadun S, Gopher A, Abbo S. The cradle of agriculture. *Science.* 2000;288(5471):1602–3.
33. Huang L, Raats D, Sela H, Klymiuk V, Lidzbarsky G, Feng L, Krugman T, Fahima T. Evolution and Adaptation of Wild Emmer Wheat Populations to Biotic and Abiotic Stresses. *Annu Rev Phytopathol.* 2016;54(1):279–301.
34. Gerechter-Amitai ZK, van Silfhout CH. Race-specificity of temperature-sensitive genes for resistance to *Puccinia striiformis* in *Triticum dicoccoides*. *Euphytica.* 1989;43(1):7–14.
35. Lhamo D, Sun Q, Zhang Q, Li X, Fiedler JD, Xia G, Faris JD, Gu Y-Q, Gill U, Cai X, et al. Genome-wide association analyses of leaf rust resistance in cultivated emmer wheat. *Theor Appl Genet.* 2023;136(1):20.
36. Peng JH, Fahima T, Röder, Huang QY, Dahan A, Li YC, Grama A, Nevo E. High-density molecular map of chromosome region harboring stripe-rust resistance genes YrH52 and Yr15 derived from wild emmer wheat, *Triticum dicoccoides*. *Genetica.* 2000;109(3):199–210.
37. Klymiuk V, Fatiukha A, Fahima T. Wheat tandem kinases provide insights on disease-resistance gene flow and host-parasite co-evolution. *Plant J.* 2019;98(4):667–79.
38. Wan A, Wang X, Kang Z, Chen X. Variability of the Stripe Rust Pathogen. In: *Stripe Rust*. Edited by Chen X, Kang Z. Dordrecht: Springer Netherlands; 2017: 35–154.
39. Alam MA, Mandal M, Wang C, Ji W. Chromosomal location and SSR markers of a powdery mildew resistance gene in common wheat line N0308. *Afr J Microbiol Res.* 2013;7(6):477–82.
40. Xie CJ, Ni ZF, Sun QX, Yang ZM, Liu BS, Wei YL. [Molecular tagging of a major powdery mildew resistance gene MIG in wheat derived from wild Emmer by using microsatellite marker]. *Yi Chuan Xue Bao.* 2001;28(11):1034–9.
41. Wei ZZ, Klymiuk V, Bocharova V, Pozniak C, Fahima T. A Post-Haustorial Defense Mechanism is Mediated by the Powdery Mildew Resistance Gene, PmG3M, Derived from Wild Emmer Wheat. *Pathogens* 2020, 9(6).
42. Xie W, Nevo E. Wild emmer: genetic resources, gene mapping and potential for wheat improvement. *Euphytica.* 2008;164(3):603–14.
43. Bischoff V, Nita S, Neumetzler L, Schindelasch D, Urbain A, Eshed R, Persson S, Delmer D, Scheible WR. TRICHOME BIREFRINGENCE and its homolog AT5G01360 encode plant-specific DUF231 proteins required for cellulose biosynthesis in Arabidopsis. *Plant Physiol.* 2010;153(2):590–602.
44. Sharma Poudel R, Al-Hashel AF, Gross T, Gross P, Brueggeman R. Pyramiding rpg4- and Rpg1- Mediated Stem Rust Resistance in Barley Requires the Rrr1 Gene for Both to Function. *Front Plant Sci* 2018, 9.

45. Zhang B, Su T, Xin X, Li P, Wang J, Wang W, Yu Y, Zhao X, Zhang D, Li D, et al. Wall-associated kinase BrWAK1 confers resistance to downy mildew in Brassica rapa. *Plant Biotechnol J*. 2023;21(10):2125–39.
46. Uauy C, Brevis JC, Chen X, Khan I, Jackson L, Chicaiza O, Distelfeld A, Fahima T, Dubcovsky J. High-temperature adult-plant (HTAP) stripe rust resistance gene Yr36 from *Triticum turgidum* ssp. *dicoccoides* is closely linked to the grain protein content locus Gpc-B1. *Theor Appl Genet*. 2005;112(1):97–105.
47. Adhikari L, Raupp J, Wu S, Wilson D, Evers B, Koo DH, Singh N, Friebe B, Poland J. Genetic characterization and curation of diploid A-genome wheat species. *Plant Physiol*. 2022;188(4):2101–14.
48. Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*. 2018;34(17):i884–90.
49. Zhu T, Wang L, Rodriguez JC, Deal KR, Avni R, Distelfeld A, McGuire PE, Dvorak J, Luo M-C. Improved genome sequence of wild emmer wheat Zavitan with the aid of optical maps. *G3: Genes Genomes Genet*. 2019;9(3):619–24.
50. Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol*. 2019;37(8):907–15.
51. Li H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics*. 2011;27(21):2987–93.
52. Van der Auwera GA, O'Connor BD. *Genomics in the cloud: using Docker, GATK, and WDL in Terra*. O'Reilly Media; 2020.
53. Paradis E, Schliep K. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics*. 2019;35(3):526–8.
54. Chen W-C. Overlapping codon model, phylogenetic clustering, and alternative partial expectation conditional maximization algorithm. Iowa State University; 2011.
55. Endelman JB. Ridge Regression and Other Kernels for Genomic Selection with R Package rrBLUP. *Plant Genome*. 2011;4(3):250–5.
56. Team RC. R: A language and environment for statistical computing. In.; 2023.
57. Frichot E, François O. LEA: An R package for landscape and ecological association studies. *Methods Ecol Evol*. 2015;6(8):925–9.
58. Francis RM. pophelper: an R package and web app to analyse and visualize population structure. *Mol Ecol Resour*. 2017;17(1):27–32.
59. Nei M. Analysis of gene diversity in subdivided populations. *Proceedings of the national academy of sciences* 1973, 70(12):3321–3323.
60. Weir BS, Cockerham CC. Estimating F-statistics for the analysis of population structure. *Evolution* 1984:1358–70.

61. Weir BS, Goudet J. A Unified Characterization of Population Structure and Relatedness. *Genetics*. 2017;206(4):2085–103.
62. Nei M. *Molecular evolutionary genetics*. Columbia university; 1987.
63. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, et al. The variant call format and VCFtools. *Bioinformatics*. 2011;27(15):2156–8.
64. Kassambara A. *Ggplot2: guide to create beautiful graphics in R*. Alboukadel KASSAMBARA; 2013.
65. Lipka AE, Tian F, Wang Q, Peiffer J, Li M, Bradbury PJ, Gore MA, Buckler ES, Zhang Z. GAPIT: genome association and prediction integrated tool. *Bioinformatics*. 2012;28(18):2397–9.
66. Wang J, Zhang Z. GAPIT Version 3: Boosting Power and Accuracy for Genomic Association and Prediction. *Genom Proteom Bioinform*. 2021;19(4):629–40.
67. Sayers EW, Bolton EE, Brister JR, Canese K, Chan J, Comeau DC, Connor R, Funk K, Kelly C, Kim S, et al. Database resources of the national center for biotechnology information. *Nucleic Acids Res*. 2022;50(D1):D20–6.
68. Letta T, Olivera P, Maccaferri M, Jin Y, Ammar K, Badebo A, Salvi S, Noli E, Crossa J, Tuberosa R. Association Mapping Reveals Novel Stem Rust Resistance Loci in Durum Wheat at the Seedling Stage. *The Plant Genome* 2014, 7(1):plantgenome2013.2008.0026.
69. Megerssa SH, Ammar K, Acevedo M, Bergstrom GC, Dreisigacker S, Randhawa M, Brown-Guedira G, Ward B, Sorrells ME. QTL mapping of seedling and field resistance to stem rust in DAKIYE/Reichenbachii durum wheat population. *PLoS ONE*. 2022;17(10):e0273993.
70. Saccomanno A, Matny O, Marone D, Laidò G, Petruzzino G, Mazzucotelli E, Desiderio F, Blanco A, Gadaleta A, Pecchioni N et al. Genetic Mapping of Loci for Resistance to Stem Rust in a Tetraploid Wheat Collection. *Int J Mol Sci* 2018, 19(12).
71. Yu L-X, Barbier H, Rouse MN, Singh S, Singh RP, Bhavani S, Huerta-Espino J, Sorrells ME. A consensus map for Ug99 stem rust resistance loci in wheat. *Theor Appl Genet*. 2014;127(7):1561–81.
72. Aoun M, Kolmer JA, Rouse MN, Elias EM, Breiland M, Bulbula WD, Chao S, Acevedo M. Mapping of Novel Leaf Rust and Stem Rust Resistance Genes in the Portuguese Durum Wheat Landrace PI 192051. *G3 (Bethesda)* 2019, 9(8):2535–47.
73. Chen S, Zhang W, Bolus S, Rouse MN, Dubcovsky J. Identification and characterization of wheat stem rust resistance gene Sr21 effective against the Ug99 race group at high temperature. *PLoS Genet*. 2018;14(4):e1007287.
74. Brueggeman R, Drader T, Kleinhofs A. The barley serine/threonine kinase gene Rpg1 providing resistance to stem rust belongs to a gene family with five other members encoding kinase domains. *Theor Appl Genet*. 2006;113(6):1147–58.
75. Wei J, Sun W, Zheng X, Qiu S, Jiao S, Babilonia K, Koiwa H, He P, Shan L, Sun W, et al. Arabidopsis RNA polymerase II C-terminal domain phosphatase-like 1 targets mitogen-activated protein kinase cascades to suppress plant immunity. *J Integr Plant Biol*. 2023;65(10):2380–94.

76. Megerssa SH, Sorrells ME, Ammar K, Acevedo M, Bergstrom GC, Olivera P, Brown-Guedira G, Ward B, Degete AG, Abeyo B. Genome-wide association mapping of seedling and adult plant response to stem rust in a durum wheat panel. *Plant Genome*. 2021;14(2):e20105.
77. Oh S-K, Baek K-H, Seong ES, Joung YH, Choi G-J, Park JM, Cho HS, Kim EA, Lee S, Choi D. CaMsrb2, Pepper Methionine Sulfoxide Reductase B2, Is a Novel Defense Regulator against Oxidative Stress and Pathogen Attack. *Plant Physiol*. 2010;154(1):245–61.
78. Mapuranga J, Zhang N, Zhang L, Liu W, Chang J, Yang W. Harnessing genetic resistance to rusts in wheat and integrated rust management methods to develop more durable resistant cultivars. *Front Plant Sci*. 2022;13:951095.
79. Thurow C, Schiermeyer A, Krawczyk S, Butterbrodt T, Nickolov K, Gatz C. Tobacco bZIP transcription factor TGA2.2 and related factor TGA2.1 have distinct roles in plant defense responses and plant development. *Plant J*. 2005;44(1):100–13.
80. et al: bZIP10-LSD1 antagonism modulates basal defense and cell death in *Arabidopsis* following infection. *The EMBO Journal* 2006, 25(18):4400-4411-4411.

Figures

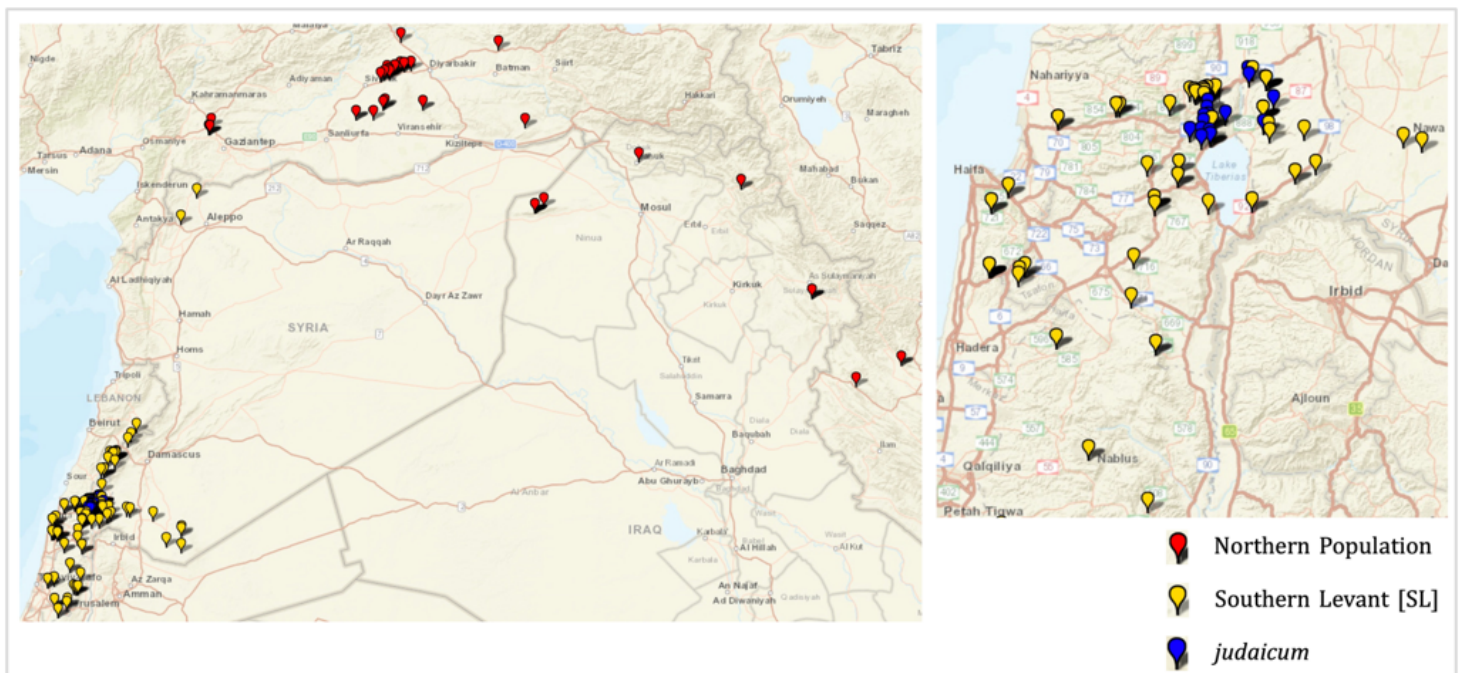


Figure 1

Geographic distributions of the WEW accessions in the WGS panel. The WEW accessions genetic grouping separated the panel into different subgroups, each marked by a unique color on the map. (Left) the distribution of all accessions in the panel, and (Right) a magnified view of the Southern Levant (SL) and the *judaicum* subgroups.

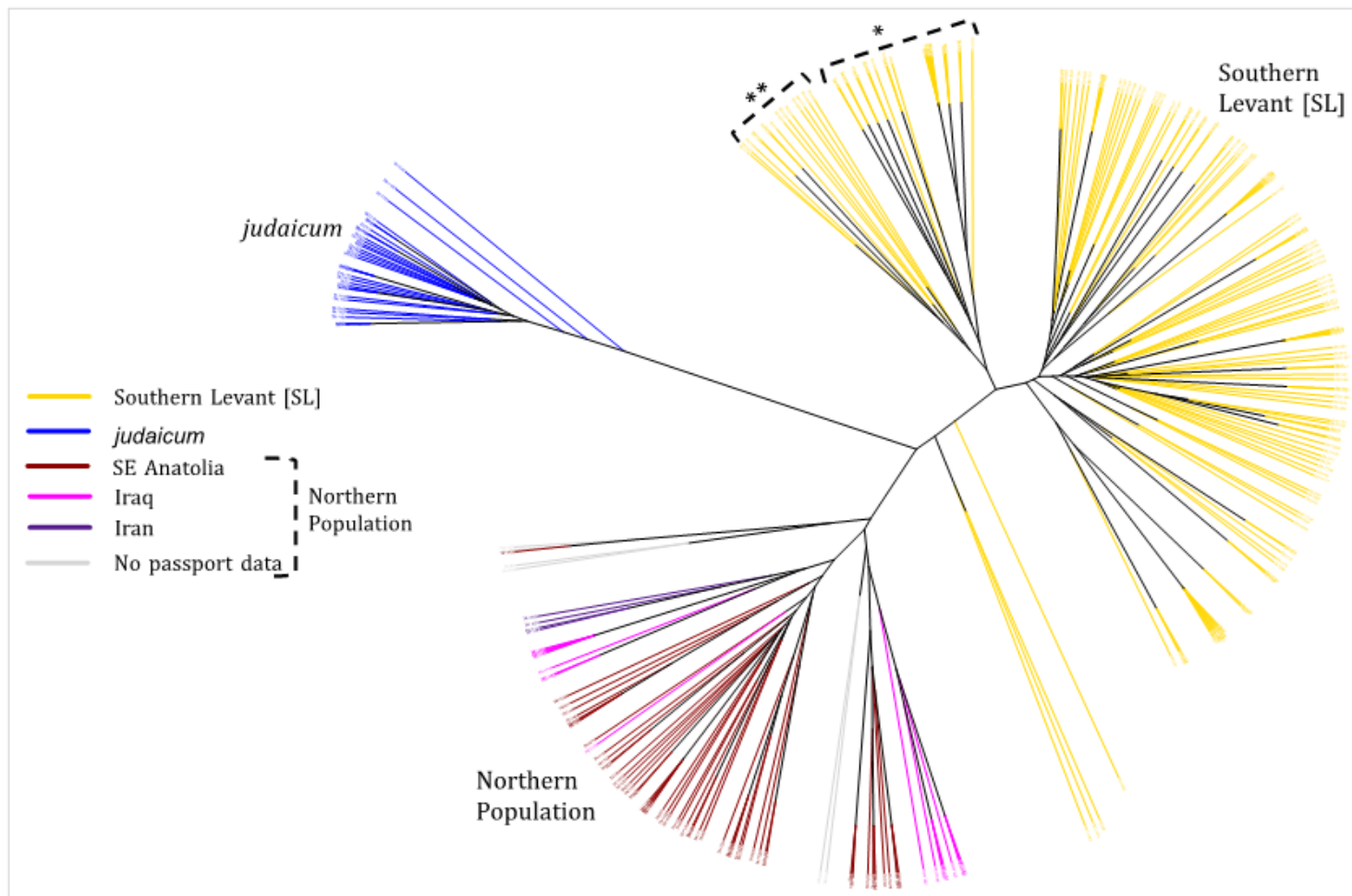


Figure 2

An unrooted Neighbor-Joining tree of wild emmer wheat in the WGS panel. The tree branches were colored based on the accessions' origination. The three major clades included the accessions from the Northern Population, Southern Levant (SL) and the WEW race *judaicum*. The smaller subclades under the SL, indicated by '*', included most of the accessions from Lebanon, while another subclade, indicated by '**', represented most of the accessions from Syria. Four accessions were observed at the base of the SL clade near the node where *judaicum* and Northern Population separated.

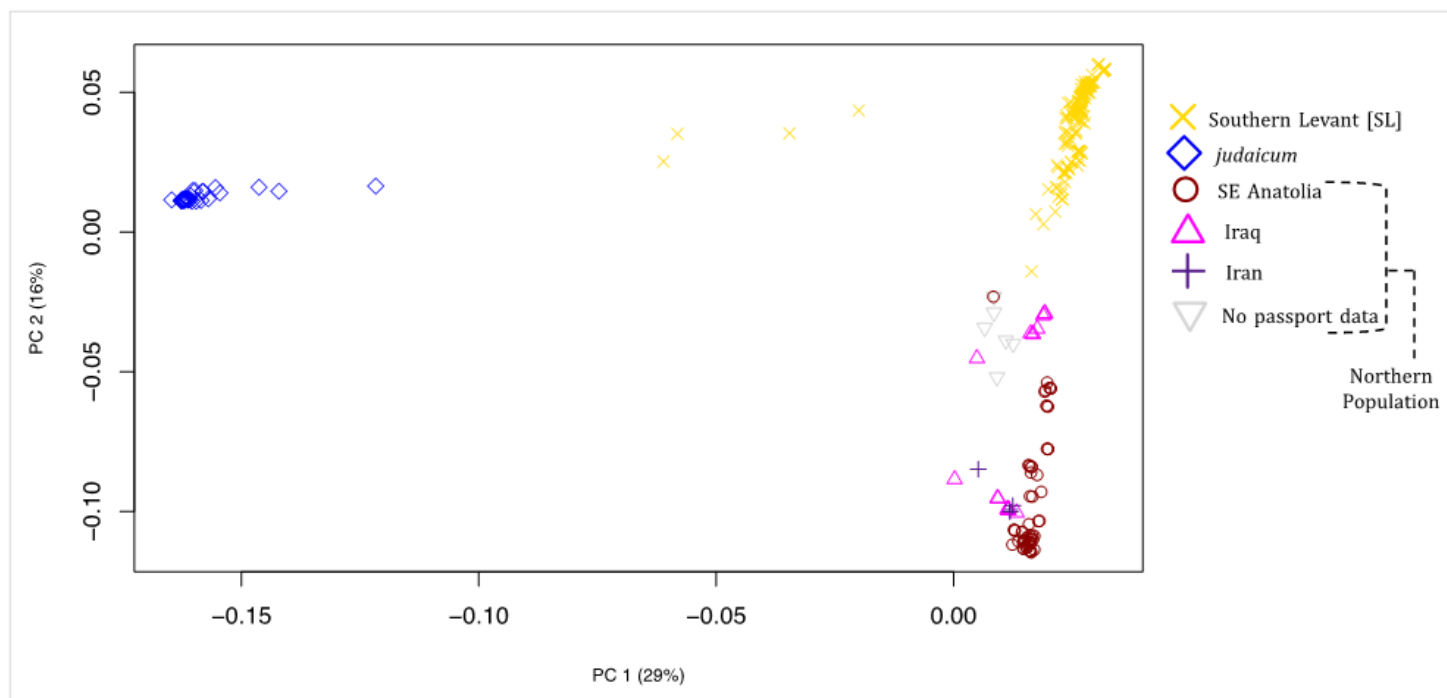


Figure 3

Principal component analysis (PCA) plot for 291 WEW accessions with the first two principal components (PCs).

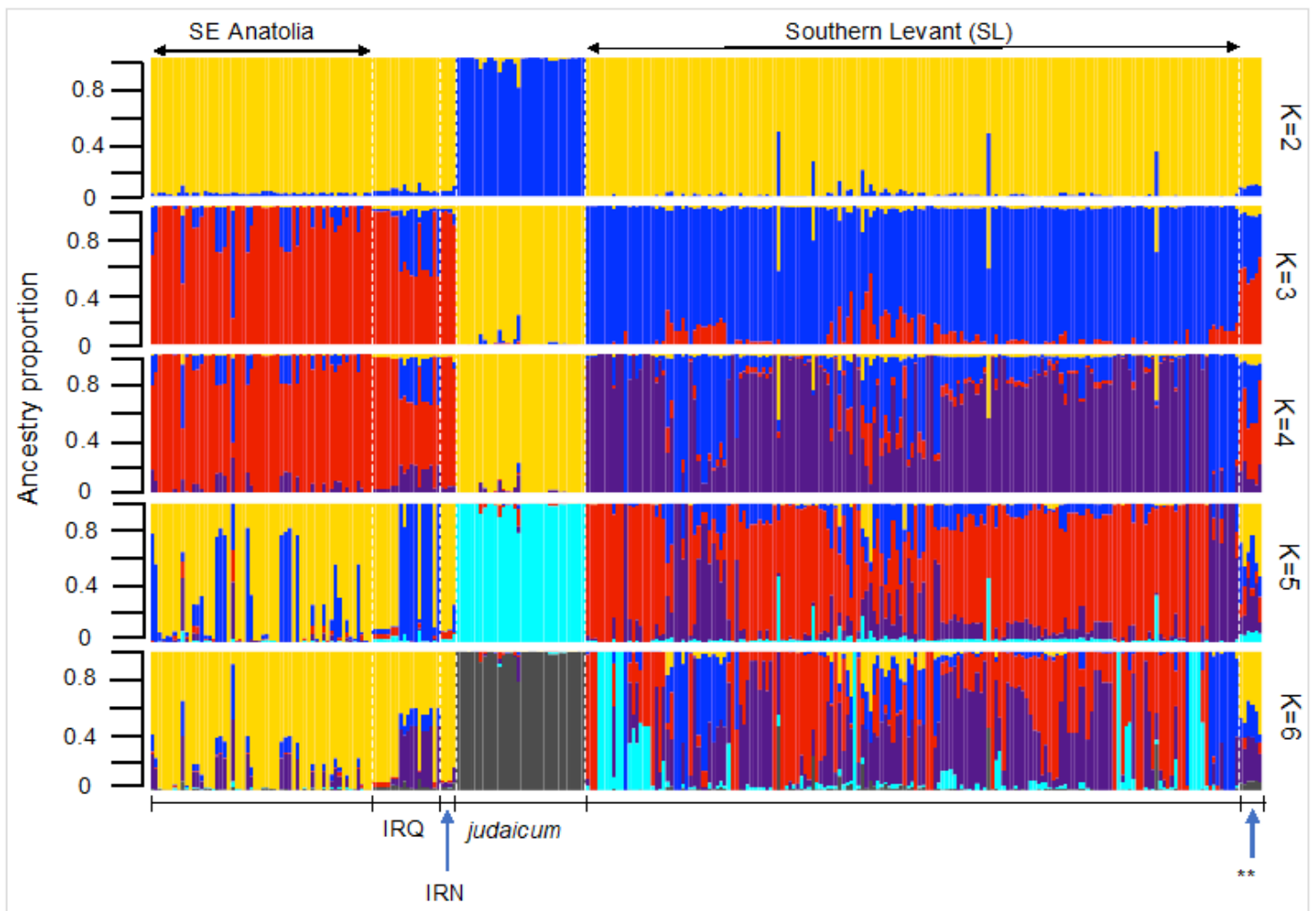


Figure 4

Population structure of the wild emmer wheat (WEW) panel subjected to whole genome sequencing-based markers. Each vertical bar indicates an accession and individual colors represent the ancestry proportion assigned to differentiated subpopulations for population structure from K=2 to K=6 (K refers to the number of ancestral populations). The geographical sites of each subgroup were provided, where the abbreviations 'IRQ' indicates Iraq, 'IRN' indicates Iran, and '**' represents the accessions without passport information.

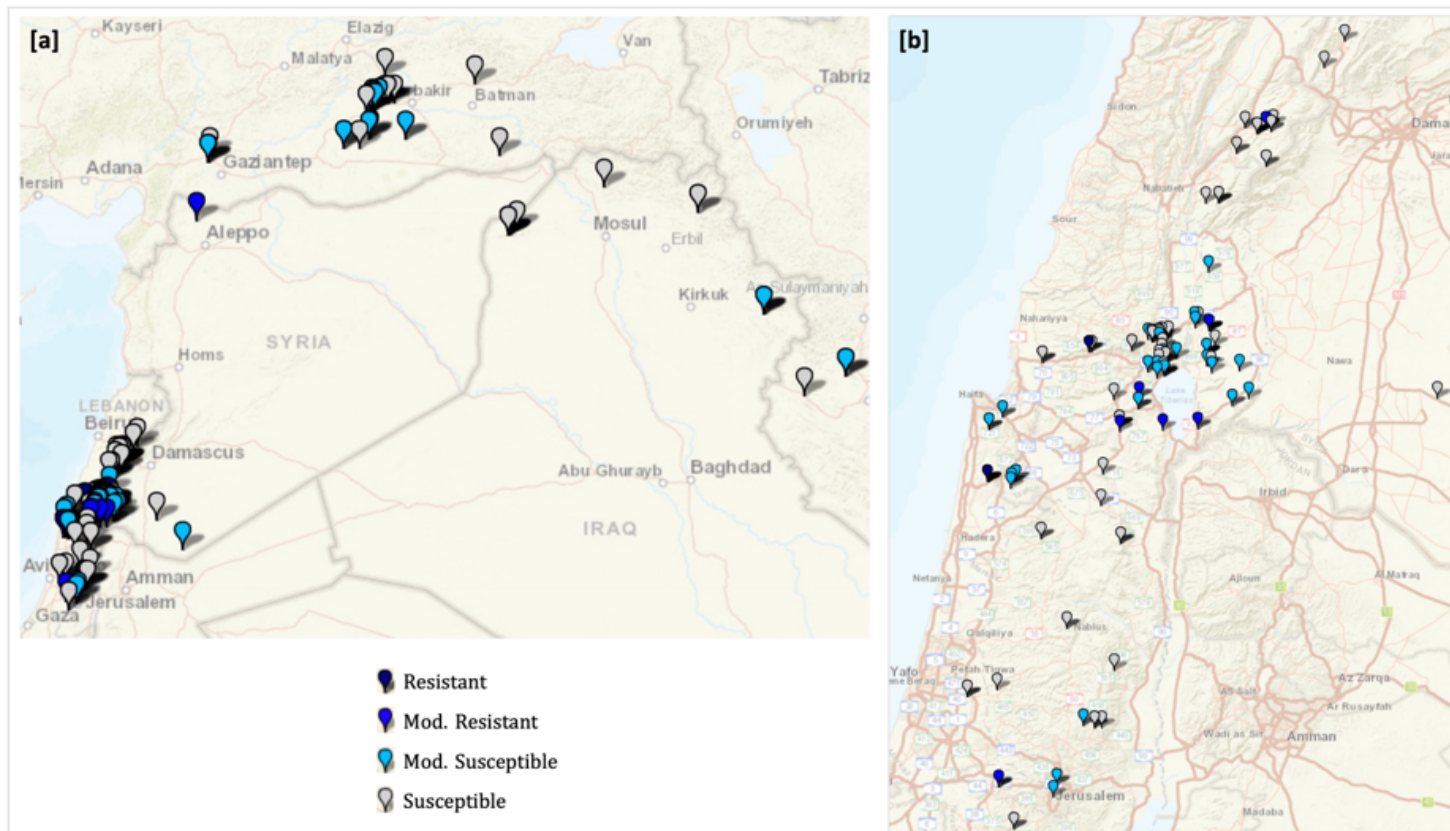


Figure 5

Seedling resistance levels of Wild Emmer Wheat (WEW) accessions to *P. graminis* f. sp. tritici-race TTKSK in; [a] the entire panel, [b] a magnified view of the Southern Levant and the *judaicum* subgroups.

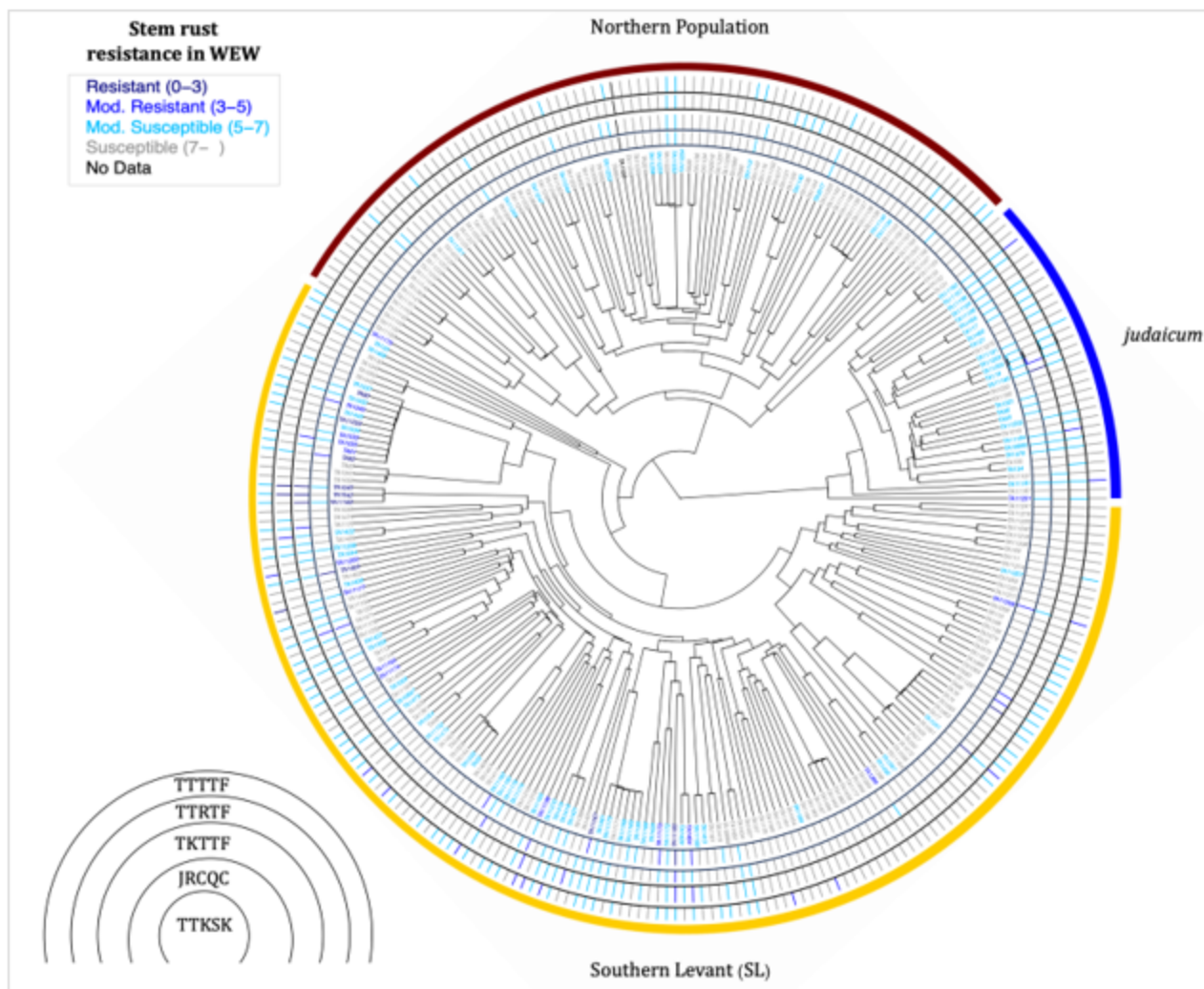


Figure 6

Clustering of WEW accessions within a neighbor-joining (NJ) tree, showing their geographic origins and levels of resistance to *P. graminisf. sp. tritici* races at the seedling stage. The resistance levels were differentiated based on disease scores, with corresponding colors on the tree matching the legend color on the top left. The innermost tree features labels of accessions as TA#, while the outer tree branches are represented by colors only. The outer colored circle's segments within the NJ tree delineate subgroups: SL, northern population, and *judaicum*.

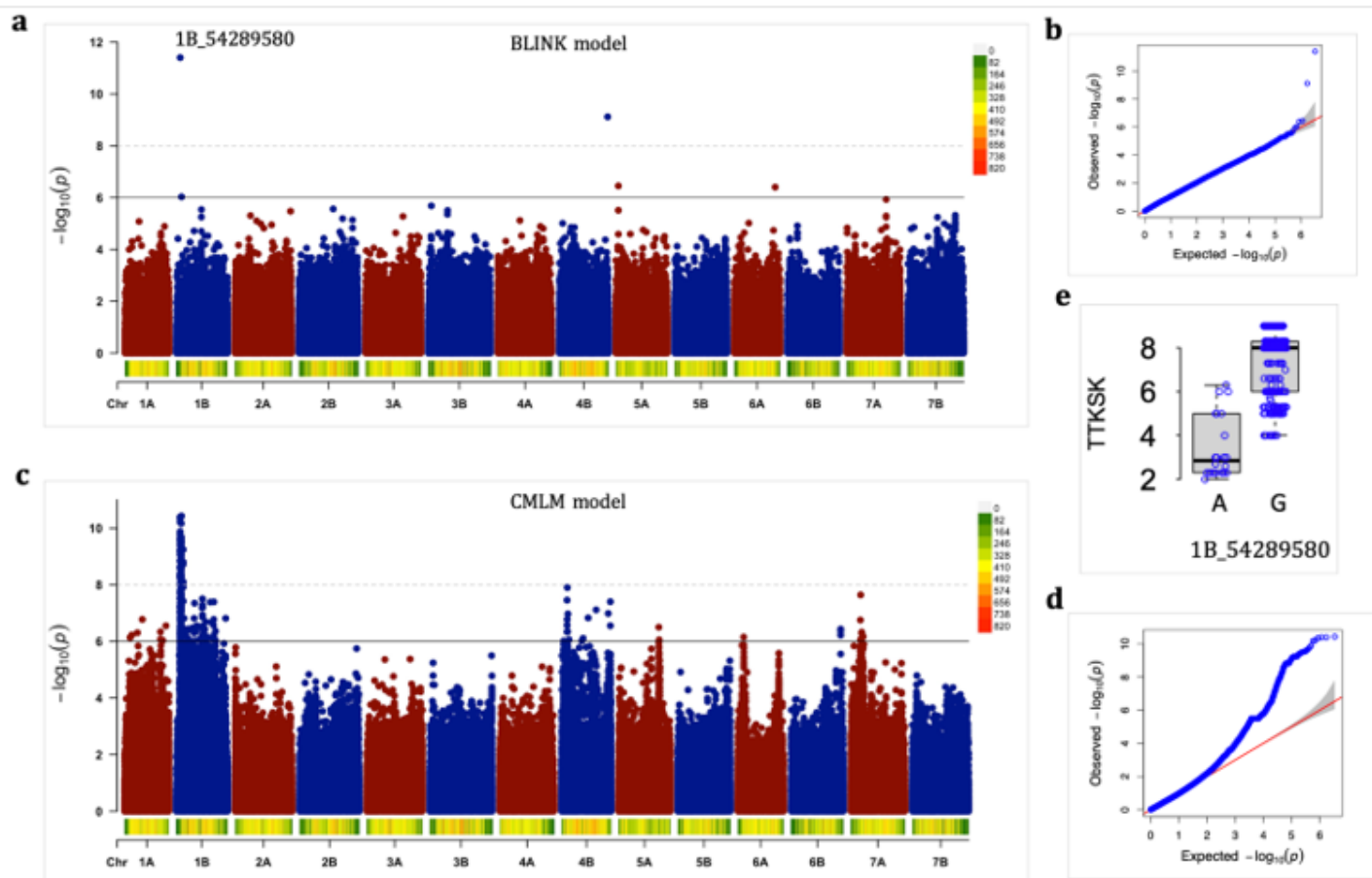


Figure 7

GWAS of stem rust resistance for *P. graminis* f. sp. *tritici*-race TTKSK, where we detected a candidate genomic region at the short arm of 1B with CMLM and BLINK methods. **[a]** TTKSK Manhattan plot with Blink model; **[b]** QQ-plot (right) showing the deviation of the observed P-values from the null hypothesis for the Blink method; **[c]** TTKSK Manhattan plot with CMLM method; **[d]** the P-values QQ plot for the CMLM method; and **[e]** the phenotype distribution for the alleles of the most significant loci at 54289580 at 1BS. The dotted gray-colored horizontal line in both Manhattan plots represents the Bonferroni-adjusted P-value and the solid black-colored horizontal line represents the FDR-adjusted P-value. For both methods, the threshold cutoff was set as 0.05 [$p_{adj} = 0.05$] as default.

Supplementary Files

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

- [SupplementaryTableS1.xlsx](#)
- [WEWSupplementaryFiles20240813.docx](#)