


## ORIGINAL ARTICLE

Special Section: From Seed to Pasta IV Congress

# A new wild emmer wheat panel allows to map new loci associated with resistance to stem rust at seedling stage

Anna Maria Mastrangelo<sup>1</sup> | Pablo Roncallo<sup>2</sup> | Oadi Matny<sup>3</sup> | Radim Čegan<sup>4,5</sup> |  
 Brian Steffenson<sup>3</sup> | Viviana Echenique<sup>2</sup> | Jan Šafář<sup>4</sup> | Raffaella Battaglia<sup>6</sup> |  
 Delfina Barabaschi<sup>6</sup> | Luigi Cattivelli<sup>6</sup> | Hakan Özkan<sup>7</sup> | Elisabetta Mazzucotelli<sup>6</sup> 

<sup>1</sup>Research Centre for Cereal and Industrial Crops, Council for Agricultural Research and Economics (CREA), Foggia, Italy

<sup>2</sup>Centro de Recursos Naturales Renovables de la Zona Semiárida (CERZOS), Departamento de Agronomía, Universidad Nacional del Sur (UNS)-CONICET, Bahía Blanca, Argentina

<sup>3</sup>Department of Plant Pathology, University of Minnesota, Minneapolis, Minnesota, USA

<sup>4</sup>Centre of Plant Structural and Functional Genomics, Institute of Experimental Botany of the Czech Academy of Sciences, Olomouc, Czech Republic

<sup>5</sup>Department of Plant Developmental Genetics, Institute of Biophysics of the Czech Academy of Sciences, Brno, Czech Republic

<sup>6</sup>Research Centre for Genomics and Bioinformatics, Council for Agricultural Research and Economics (CREA), Fiorenzuola d'Arda, Italy

<sup>7</sup>Faculty of Agriculture, Department of Field Crops, University of Çukurova, Adana, Turkey

## Correspondence

Elisabetta Mazzucotelli, Research Centre for Genomics and Bioinformatics, Council for Agricultural Research and Economics (CREA), Via San Protaso, 302, I-29017 Fiorenzuola d'Arda (PC), Italy.  
 Email: [elisabetta.mazzucotelli@crea.gov.it](mailto:elisabetta.mazzucotelli@crea.gov.it)

Assigned to Associate Editor Curtis Pozniak.

[Correction added on 19 January, after first online publication: author name/surname is corrected from “Čegan Radim” to “Radim Čegan” and affiliation 5 is added. Affiliation link for Jan Šafář is updated from 5 to 4.]

## Funding information

HORIZON-INFRA-2022-DEV-01, project “Promoting a Plant Genetic Resource

## Abstract

Wheat stem rust, caused by *Puccinia graminis* f. sp. *tritici* (*Pgt*), is a major wheat disease worldwide. A collection of 283 wild emmer wheat [*Triticum turgidum* L. subsp. *dicoccoides* (Körn. ex Asch. & Graebn.) Thell] accessions, representative of the entire Fertile Crescent region where wild emmer naturally occurs, was assembled, genotyped, and characterized for population structure, genetic diversity, and rate of linkage disequilibrium (LD) decay. Then, the collection was employed for mapping *Pgt* resistance genes, as a proof of concept of the effectiveness of genome-wide association studies in wild emmer. The collection was evaluated in controlled conditions for reaction to six common *Pgt* pathotypes (TPMKC, TTTTF, JRCQC, TRTTF, TTKSK/Ug99, and TKTTF). Most resistant accessions originated from the Southern Levant wild emmer lineage, with some showing a resistance reaction toward three to six tested races. Association analysis was conducted considering a 12K polymorphic single-nucleotide polymorphisms dataset, kinship relatedness between accessions,

**Abbreviations:** AMOVA, analysis of molecular variance; BIC, Bayesian information criterion; DAPC, discriminant analysis of principal components; GBS, genotyping by sequencing; GWAS, genome-wide association studies; IT, infection type; LD, linkage disequilibrium; MLM, mixed linear model; MTA, marker-trait association; PA, private alleles; PCA, principal component analysis; *Pgt*, *Puccinia graminis* f. sp. *tritici*; SNP, single-nucleotide polymorphism; SSD, single seed descendance; WEW, wild emmer wheat.

Anna Maria Mastrangelo and Pablo Roncallo contributed equally to this work and shared first authorship.

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs License](https://creativecommons.org/licenses/by-nc-nd/4.0/), which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2023 The Authors. *The Plant Genome* published by Wiley Periodicals LLC on behalf of Crop Science Society of America.

Community for Europe (PRO-GRACE)”; Grantová Agentura České Republiky, Grant/Award Number: 22-00204S; Agritech National Research Center funded by the European Union Next-Generation-EU; PRIMA2019-CerealmEd “Enhancing diversity in Mediterranean cereal farming systems”

and population structure. Eleven significant marker–trait associations (MTA) were identified across the genome, which explained from 17% to up to 49% of phenotypic variance with an average 1.5 additive effect (based on the 1–9 scoring scale). The identified loci were either effective against single or multiple races. Some MTAs colocalized with known *Pgt* resistance genes, while others represent novel resistance loci useful for durum and bread wheat prebreeding. Candidate genes with an annotated function related to plant response to pathogens were identified at the regions linked to the resistance and defined according to the estimated small LD (about 126 kb), as typical of wild species.

## 1 | INTRODUCTION

Wheat is the world’s most widely grown cereal and accounts for ~20% of proteins and calories consumed globally. Wild emmer wheat (WEW) *Triticum turgidum* L. subsp. *dicocoides* (Körn. ex Asch. & Graebn.) Thell, is a progenitor of all forms of cultivated wheat. The loss of the brittle rachis trait was the major domestication event leading to domesticated emmer, *T. turgidum* subsp. *dicoccum* (Schrank ex Schübl.) Thell. Then, the loss of tough glumes and the acquisition of the free-threshing trait ultimately resulted in the durum wheat crop, *T. turgidum* subsp. *durum* (Desf.) Husn. The hybridization of *T. durum* with the diploid goatgrass (*Aegilops tauschii*) gives rise to hexaploid wheat, *T. aestivum* L.

Wild emmer is currently present as native stands across the Fertile Crescent. Two major lineages, geographically and genetically distinct, have been recognized (Luo et al., 2007; Ozkan et al., 2005): (1) the *northern* lineage, with main populations found in north- and east-central Turkey, western Iran, and northern Iraq and corresponding to the slender *horanum* race, and (2) the *southern* lineage found in the Southern Levant, including Syria, Lebanon, Jordan, and Israel, which includes both slender and robust plant types, the latter being typical of the *judaicum* race. Apart from dense and frequent natural populations found in the Upper Jordan Valley catchment area in Israel, and massive stands on the basalt slopes of the Karacadağ mountain range (between Şanlıurfa and Diyarbakır provinces) and of the Kartal-Karadağ hills (between Gaziantep and Kahramanmaraş provinces) in Turkey, wild emmer currently exists in a patchy distribution in the region, with populations being semi-isolated or isolated. Its habitats range in altitude from 100 m below to 1800 m above sea level, with very different climatic conditions, from cool and humid on the Karacadağ mountains to hot and dry in the valleys of Israel (Nevo et al., 2002; Ozkan et al., 2011).

Crop wild relatives have recently gained importance as source of new genes for resistance to emerging new virulence types of crop pathogens. Indeed, domestication and intensive breeding have reduced gene diversity in modern crops (Fu,

2015; Swarup et al., 2021), leading to a potential vulnerability of current breeding pools and varieties to disease outbreaks and environmental stresses in general. In contrast, natural populations of crop relatives are highly diverse and continuously evolving with pathogens endemic of their habitats. This is the case also for wild emmer that possesses broad variation for resistance to wheat pathogens, and thus has enabled the identification and mapping of many qualitative and quantitative resistance loci (Huang et al., 2016). Wild emmer has the same genome formula as durum wheat (AABB), allowing for the easy transfer of valuable genes into cultivated tetraploid wheat and eventually into the hexaploid wheat (Klymiuk et al., 2019). For instance, in wild emmer, the well-known stripe rust (*Puccinia striiformis* f. sp. *tritici*) resistance genes *Yr15* and *Yr36* were cloned (Fu et al., 2009; Klymiuk et al., 2018) and introgressed to durum and common wheat (Hale et al., 2012; Yaniv et al., 2015), similarly to the couple of linked resistance genes for leaf rust (*P. triticina*) and yellow rust, *Lr53* and *Yr35*, respectively, which were mapped and also introgressed to common wheat (Dadkhodaie et al. et al., 2011; Marais et al., 2005). Previously, evaluation of wild emmer populations for resistance to the stem rust pathogen (*P. graminis* f. sp. *Tritici*, *Pgt*) revealed a wide range of variation, with reactions ranging from highly resistance to complete susceptibility (Knott et al., 2005; McVey, 1991; Nevo et al., 1991). Nevertheless, in spite of this variation for reaction to stem rust, no major resistance genes have been mapped in wild emmer so far. The frequency of resistance to single *Pgt* isolates was very low in wild emmer populations from Israel. Indeed, depending on the publications, from 2% (Anikster et al., 2005) to 8% (The et al., 1993) of the accessions were reported to have some level of resistance among about 500 genotypes tested. It has been suggested that the low frequency of stem rust resistance found in wild emmer originated from Israel is due to the low selection pressure imposed by *Pgt* in natural habitats of this country. Stem rust is a warm-weather-adapted fungus, while Israel climate is mainly characterized by frequent hot-dry days that should not be conducive for the development of *Pgt* and high stem rust infection, except some colder/less warm regions.

Indeed, a correlative study, which investigated the relationship between the occurring of stem rust resistance and ecogeographic factors (temperature above all) prevalent in the origin area of the wild emmer populations examined, pointed out that resistant accessions were present only in colder regions where climatic variables enhanced the development of the fungus (Nevo et al., 1991).

To date, the identification of resistance genes from wild emmer has been achieved through linkage mapping of biparental populations (i.e., *Yr15* by Yaniv et al. [2015] and *YrTZ2* by Wang et al. [2018]; recently reviewed in Peng et al. [2021]). Differently from biparental mapping, the genome-wide association study (GWAS) identifies and maps resistance genes based on linkage disequilibrium (LD). In GWAS, the mapping resolution power is a consequence of the number of recombination events historically accumulated in the target population. This condition is obviously enhanced in diverse collections of wild germplasm that have a long history of recombination resulting in rapid LD decay as a function of genetic distance. Provided that genotyping density is sufficiently high to fully cover the genome according to the extent of LD, high mapping precision can be achieved.

A few wild emmer collections have been established to study the genetic evolution of emmer/durum wheat (Luo et al., 2007; Ozkan et al., 2005) and, more recently, to explore the genetic diversity of tetraploid wheats (Yadav et al., 2023) and identify genomic regions that have been historically influenced by domestication and human-driven selection (Maccaferri et al., 2019). Other collections have been used to analyze the genetic differentiation among populations as driven by ecological factors (Ren et al., 2013) or to trace historic gene flow from wild relatives that had substantially contributed to the adaptive diversity of modern bread wheat (He et al., 2019). To date, only a couple of studies exploited a wild emmer collection to discover and map disease-resistant genes, in both cases against wheat stripe rust (Sela et al., 2014; Tene et al., 2022).

This study aims to characterize the population structure and the genetic diversity of a wide *ex situ* collection of WEW accessions, representative of all countries where wild emmer naturally grows, as well as to assess the panel for resistance to several races of the stem rust pathogen with the aim to map resistance loci via a GWAS approach.

## 2 | MATERIALS AND METHODS

### 2.1 | Plant material

A panel of 283 accessions of WEW collected across countries of the Fertile Crescent was organized and is currently maintained at the CREA Research Centre for Genomics and Bioinformatics (WEW-GB: Wild Emmer Wheat at CREA Research

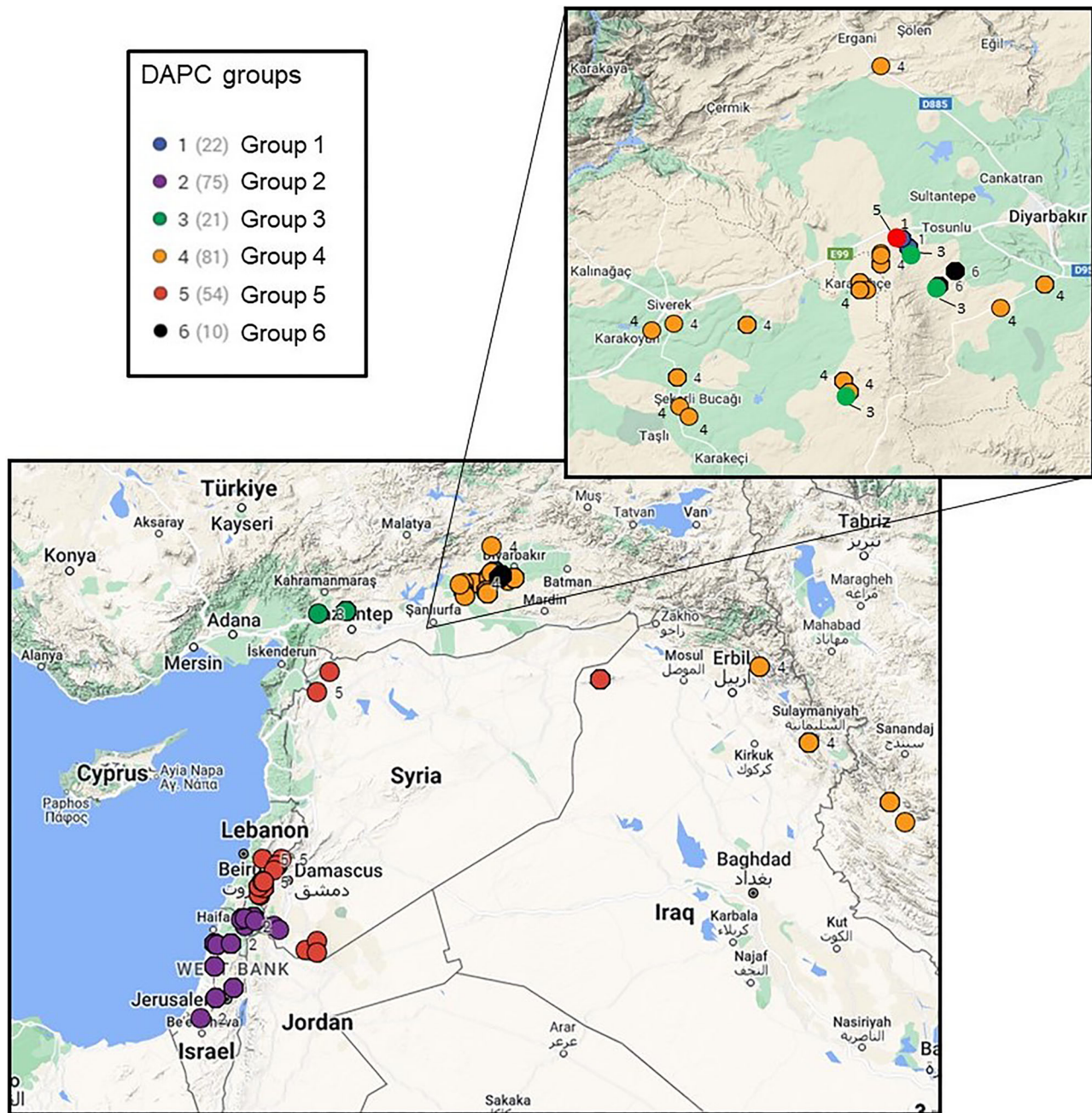
### Core Ideas

- A representative collection of wild emmer wheat was characterized for genetic diversity and resistance to stem rust.
- GWAS for stem rust resistance identified novel marker–trait associations in wild emmer wheat.
- Fast LD decay in wild emmer wheat allowed high-resolution mapping of loci for stem rust resistance.

Centre for Genomics and Bioinformatics). Accessions were kindly provided either from gene banks (the United States Department of Agriculture—Agricultural Research Service National Small Grains Collection [USDA-ARS NSGC], the Leibniz Institute of Plant Genetics and Crop Plant Research [IPK, Germany], the National BioResource Project [NBRP] through KOMUGI, Japan) or from partners (more details in File S1). The plant materials represent the entire Fertile Crescent region where wild emmer naturally occurs, from the Northern East area, encompassing Turkey, Iraq, and Iran, to the Southern region (Southern Levant), which includes Israel, Lebanon, and Syria. File S1 summarizes passport data of accessions used in the present study. Figure 1 shows the geographical origins of all investigated accessions, accompanied by GPS coordinates. Before genotyping, all accessions have been undergone to one cycle of single seed descendance (SSD).

### 2.2 | Genotyping, filtering, and mapping of the polymorphic single-nucleotide polymorphisms on the reference wild emmer genome

Young leaves of a single plant from each SSD were collected, and genomic DNA was extracted using the CTAB method (Doyle & Doyle, 1987). The 283 accessions were genotyped with the Affymetrix 35K Axiom array at the Genomics Facility of University of Bristol. Allele calling was carried out using the Affymetrix proprietary software package Axiom Analysis Suite, according to the Axiom Best Practices Genotyping workflow. It includes a Bayesian-based clustering that was performed by selecting generic priors, as this setting produced less misclustering than when bread wheat-specific priors had been applied. Only single-nucleotide polymorphisms (SNPs) classified as polymorphic at high resolution and as off-target variants by the Affymetrix software were selected. They were then filtered for missing data (max 10%) and for heterozygosity (max 10%).



**FIGURE 1** Geographic origin of wild emmer wheat accessions included in the WEW-GB collection. Accessions were plotted on the geographic map of the Fertile Crescent region, according to GPS coordinates of their sampling area included in passport data. Color code of accession labels is the same of discriminant analysis of principal components (DAPC) clustering.

The probe sequences of the polymorphic SNPs were used as queries in BLASTn similarity searches against the pseudo-molecules of the wild emmer v2 reference genome (accession Zavitan [Zhu et al., 2019]) to obtain marker physical position. Parameters for BLASTn were as follows: word size, 11; gap open, 5; gap extend, 2; penalty, -2; reward, 1. Both allele variants were individually considered for BLAST. Analogous parameters were used to BLASTn sequences of SNP markers against the *T. aestivum* reference genome (IWGSC RefSeqv2.1 [Zhu et al., 2021]) and the *T. durum* genome (Svevo v1 [Maccaferri et al., 2019]).

### 2.3 | Assessment of LD

Intrachromosomal pairwise correlation ( $r^2$  values) between loci was computed on the SNP dataset using TASSEL 5.0 (Bradbury et al., 2007), considering SNP markers that exhibit unique positions in Zavitan genome and have minimum allele frequency (MAF) >0.05. The significant  $r^2$  ( $p < 0.01$ ) values were plotted against the inter-SNP physical distance and fitted to a LOWESS curve (Cleveland, 1979) using MINITAB software both at the single-chromosome and whole-genome level. The critical distance was extrapolated at  $r^2 = 0.2$ , one

of the most commonly used thresholds to define unlinked loci. The smoothing curve was fitted using raw significant  $r^2$  values, but also using average  $r^2$  values calculated on increasing pairwise SNP distances by 5000 pb ranges from zero and till a maximum distance of 10 Mb. Software PLINK version 1.9 (Chang et al., 2015) was used to prune hap-maps at different  $r^2$  threshold for different analyses.

## 2.4 | Analysis of population stratification

Ward's hierarchical clustering of Nei's genetic distances was conducted in DARWIN v6.0 software (Perrier et al., 2003). Population structure analysis was conducted by the Bayesian clustering algorithm implemented in the software STRUCTURE 2.3.4 (Pritchard et al., 2000) using 2444 genome-specific polymorphic loci across the 14 chromosomes. These loci were selected based on low LD, with threshold  $r^2 = 0.3$  and based on MAF  $>0.10$ . The number of populations ( $K$ ) ranging from 2 to 20 was evaluated, with 20 independent runs for each  $k$ . Each run included a burn-in length of 50,000 and a run length of 100,000. The admixture model, which assumes that individuals might have mixed ancestries and correlated allele frequency, was applied. The most likely number of populations ( $K$ ) was determined by using the web-based STRUCTURE HARVESTER software (Earl & Von Holdt, 2012), considering a  $\Delta K$  statistic that depends on the rate of change in log probability [LnP(D)] between consecutive  $K$  values and standard deviations on independent runs for each  $K$ . Individuals were assigned to a population if their maximum ancestry membership coefficient estimated by STRUCTURE ( $Q_{(1^*)}$ ) was  $>0.6$ , or if it was in the range  $0.6-0.5$  and with  $Q_{(1^*)} - Q_{(2^*)} >0.2$ , otherwise they were considered admixed. The 2444 set of genome-specific polymorphic loci was used also in the discriminant analysis of principal components (DAPC). DAPC was performed through the R package "adegenet" v2.0.1 (Jombart et al., 2010) in R studio V 1.3.1056 (R Development Core Team, 2011). The cross-validation method was used to select the number of principal components (PCs) retained (xvalDapc function). The most probable  $K$  was declared based on the first increase in Bayesian information criterion (BIC) value according to the steppingstone genetic model described by Jombart et al. (2010). Finally, the most probable  $K$  based on analysis of the full filtered SNP dataset was selected by usage of ADMIXTURE's cross-validation procedure (Alexander et al., 2009). Analysis of admixture by kinship was performed using ADMIXTURE (version 1.3.0; Alexander et al., 2009) with  $K$  value set to 18.

Genetic diversity among and within populations was assessed by analysis of molecular variance (AMOVA) and by calculating indices as the number of polymorphic loci, Nei's gene diversity (Nei, 1973), and mean number of pairwise dif-

ferences. Population differentiation was assessed based on population pairwise  $F_{ST}$ . All values were derived using the Arlequin 3.5 software (Excoffier & Lischer, 2010), and significance levels for variance components and  $F_{ST}$  statistics were estimated based on 1000 permutations. Alleles found in single populations (as defined by DAPC) of the broader collection, called as private alleles (PAs), were quantified using the GenAlex v6.5 software (Peakall & Smouse, 2012).

## 2.5 | Phenotyping of the WEW-GB collection for stem rust resistance

Within the wild emmer panel, 278 accessions were evaluated for seedling resistance to six races of *Puccinia graminis* f. sp. *tritici* (TPMKC, TTTTF, JRCQC, TRTTF, TTKSK, and TKTTF), in a greenhouse. Race TPMKC (isolate 74MN1409) is virulent on *Sr36* (McVey et al., 2002). Race TTTTF (02MN84A-1-2) is the most widely virulent race reported in the United States but also in Sicily (south of Italy), producing high infection types (ITs) on all lines in the wheat stem rust differential set with the exception of *Sr24* and *Sr31* (Jin et al., 2008; Roelfs & Martens, 1988). Races JRCQC (09ETH80-3) and TRTTF (06YEM34-1), from Ethiopia, are virulent on both resistance genes *Sr9e* and *Sr13*, two genes constituting major components of stem rust resistance in durum cultivars and germplasm (Olivera et al., 2012). Race TTKSK (04KEN156/04, also called Ug99) emerged in 1998 in Uganda and immediately posed concerns for wheat cultivation worldwide because it was virulent for many resistance genes, including the widely used gene *Sr31* (Pretorius et al., 2000). Race TKTTF (13ETH18-1), also from Ethiopia, is virulent on *SrTnp* (Olivera et al., 2015), a gene that confers resistance to Ug99 race. Susceptible controls (cultivar McNair 701 and Morocco) were included in each experiment to monitor the infection level (density of uredinia on leaves) and virulence (maximum uredinial size) of the pathogen races. Additionally, the wheat stem rust differential set also was included in the experiments to confirm the identity and purity of *Pgt* races (Jin et al., 2008; Roelfs & Martens, 1988). Resistant lines within the respective differential wheat sets served as resistant controls, besides two additional lines carrying the resistance genes *Sr24* and *Sr47*. Spores stored at  $-80^\circ\text{C}$  were used for this study (Huang & Steffenson, 2018).

The stem rust evaluations with foreign races were conducted in a Biosafety Level-3 Containment Facility on the St. Paul campus of the University of Minnesota. Three seeds per replicate of each genotype and controls were sown. Sowing methods and substrate, plant growth conditions in greenhouse, and fertilization times and chemicals were as Mamo et al. (2015). Seed dormancy was broken by moving pots at  $4^\circ\text{C}$  for 3 days. Twelve-day-old seedlings (first leaf stage) were inoculated with urediniospores of stem rust races

suspended in a lightweight mineral oil carrier (Soltrol 170; Phillips Petroleum). The spores were heat shocked in the water bath at 45°C for 15 min and placed for 2 h in the humidifier jar with 98%–99% humidity (solution of KOH). The concentration of inoculum used was 15 mg/0.7 mL oil applied at a rate of approximately 0.150 mg per plant (Steffenson et al., 2009). The oil carrier could evaporate before plants were moved into the mist chambers. Plants were then continuously misted for 15 min with ultrasonic humidifiers to establish an initial layer of moisture on the surfaces, then periodically for 2 min on every 15 min for 16 h in dark at 20–22°C, and finally continuously for 2 h with the lights on. Then, doors of the mist chamber were opened to allow plants to dry slightly before being returned to the greenhouse under the same environmental conditions.

Rust ITs were scored on plants 12–14 days postinoculation using the 0–4 scale described by Stakman et al. (1962) and Jin et al., 2007, where IT = 0 represents a highly resistant (incompatible) reaction, 1 = resistant, 2 = moderately resistant, 3 = moderately susceptible, and 4 highly susceptible (compatible) reaction. Plus (+) and minus (–) notations were used with the IT scores to denote uredinia that were larger or smaller than those classically described in Stakman et al. (1962). To convert the raw IT data to a linear scale for association mapping analysis, ITs 0, 1–, 1, 1+, 2–, 2, 2+, 3–, 3, and 3+ were coded as 0, 1, 2, 3, 4, 5, 6, 7, 8, and 9, respectively. IT = 4 was converted to 9. The semicolon symbol for hypersensitive fleck “;” —which means hypersensitive necrotic or chlorotic fleck reactions, typically denoted—was converted to 0. For complex scores that include dual IT, such as 22+, the scale conversion reported by Zhang et al. (2014) and based on weighted average was used.

All the experiments were conducted in a complete random design, with two replicates. Any accessions exhibiting variable reactions across the replicates were repeated in a second test.

## 2.6 | Statistics analysis

Phenotypic data were analyzed using the software Jmp7. Data transformation was attempted to overcome the strong skewed distributions of phenotype data, but none allowed to achieve normal distribution. Therefore, the nonparametric Kruskal–Wallis test (Kruskal & Wallis, 1952) was applied to evaluate genotype differences.

## 2.7 | Genome-wide association scan

TASSEL software, version 5.2.80, was used to carry out mixed linear model (MLM) analyses for association mapping. Different models were tested to select one that best fits the data. These models integrated the kinship matrix (calculated

with the TASSEL software) to account for genotype relatedness with or without components of a principal component analysis (PCA, calculated with the TASSEL software) as a covariate to correct for the effects of population substructure. The model selected for the GWAS was an MLM integrating kinship and the first two PCA components. To reduce the risk of type I error when making multiple comparisons, the Bonferroni threshold for  $p$  value (calculated at  $6.15 \times 10^{-6}$ ) was chosen to select for significant marker–trait associations (MTAs).

## 3 | RESULTS

### 3.1 | Genotyping of the WEW-GB collection

After one SSD cycle, the WEW-GB panel was genotyped with the 35K Axiom Array technology, and 263 genotypes passed the sample call rate quality check. Among the 23,432 SNPs that met the clustering requirements, 12,100 SNPs were polymorphic across the collection. Among them, stepwise data curation discarded 704 SNPs with call rate <90% or residual heterozygosity >10%, and additional 54 SNPs with MAF <0.05. The retained 11,342 polymorphic SNPs were physically assigned to the pseudomolecules of the wild emmer reference genome by BLASTn similarity searches (Zhu et al., 2019). A total of 6931 SNPs had an unambiguous, single BLAST top hit coincident for both allele variant sequences, in addition to chromosome assignment confirmed by at least one genetic map in bread wheat or a BLAST top hit on the durum or bread wheat reference genomes. For those markers showing multiple hits along the genome, the LD with the hypothetical nearby mapped markers was assessed, and eventually a unique position assigned based on the highest  $r^2$  (above a 0.8 threshold). This approach allowed to map other 935 SNPs on the wild emmer genome. For all other SNPs, it was not possible to determine which homoeologous chromosome was the source of the SNP variant since probes aligned with equal/similar scores to two or more pseudomolecule sequences or to unassembled scaffolds of the wild emmer genome. A total of 7866 SNPs resulted physically anchored on the wild emmer reference genome, with an average density of one SNP per 1.26 Mb (Table 1).

### 3.2 | LD decay in the WEW-GB collection

Genome-wide and chromosome-specific LD decay over physical distance was evaluated. Intrachromosomal raw  $r^2$  values, significant at  $p < 0.01$ , allowed fitting LOWESS for the whole genome, which indicated 126 kb as critical distance for a  $r^2$  threshold value of 0.2 (Table 1; Figure S1). In single chromosome analysis, critical values were reached at distances

**TABLE 1** Number of SNP markers, mean values of gene diversity, and LD indices on each chromosome and genome.

Chr	<i>N</i>	Marker coverage (Mb/SNP)	MAF	He	% SNP in LD	LD ( $r^2$ )	% SNP with $r^2 < 0.1$	LD decay (kb)
1A	591	0.99	0.270	0.358	49.3	0.122	60.4	99.8
1B	783	0.87	0.250	0.342	40.1	0.095	70.3	278
2A	558	1.39	0.264	0.352	35.2	0.095	67.1	58 <sup>a</sup>
2B	668	1.18	0.257	0.347	40.6	0.095	70.2	222
3A	488	1.53	0.254	0.343	45.7	0.108	63.7	310 <sup>a</sup>
3B	713	1.17	0.256	0.343	46.2	0.112	63.4	178
4A	384	1.92	0.262	0.351	46.3	0.107	64.1	255
4B	285	2.37	0.277	0.364	52.4	0.138	56.7	95
5A	614	1.09	0.259	0.345	43.1	0.106	65.0	165
5B	656	1.07	0.276	0.358	42.3	0.105	67.4	220
6A	425	1.45	0.284	0.369	45.4	0.133	58.5	190
6B	584	1.20	0.286	0.367	45.6	0.124	61.4	1020
7A	648	1.12	0.261	0.349	44.1	0.114	63.8	390
7B	469	1.54	0.271	0.359	42.8	0.103	66.4	205
A genome	3476	1.40	0.265	0.352	44.2	0.112	63.2	–
B genome	3708	1.38	0.268	0.354	44.3	0.110	65.1	–
Unmapped	4158	–	0.270	0.356	–	–	–	–
Whole genome	11342	1.26	0.266	0.353	44.2	0.111	64.2	126

Note: % SNP in LD, percentage of pairwise SNP in significant LD ( $p < 0.01$ ); LD ( $r^2$ ), mean intrachromosomal LD of significant pairwise SNP ( $p < 0.01$ ); % SNP with  $r^2 < 0.1$ , percentage of unlinked significant pairwise SNP ( $r^2 < 0.1$ , with  $p < 0.01$ ); LD decay (kb), genome-wide and chromosome-specific LD decay critical distance in kilobase pairs extrapolated from the intercept of fitting LOWESS and the  $r^2 = 0.2$  threshold.

Abbreviations: He, expected heterozygosity (Nei's gene diversity); MAF, minor allele frequency; *N*, number of SNPs; LD, linkage disequilibrium.

<sup>a</sup>For chromosome 2A and 3A, critical distances were obtained using a LOWESS curve based on averaging significant  $r^2$  value on classes of inter-SNP distances with amplitudes increasing by 5000 bp steps.

in the range of 95 kbp to 1 Mb (Figure 2; Table 1). Exceptions were chromosomes 2A and 3A where low LD even at short distances impaired fitting to a nonparametric curve, and thus the option of averaging  $r^2$  values on classes of inter-SNP distances was adopted. Unlinked markers ( $r^2 < 0.1$ ) were an average of around 64% on the whole genome, with a minimum of 56.7% for chromosome 4B and similar values in the two subgenomes (Table 1).

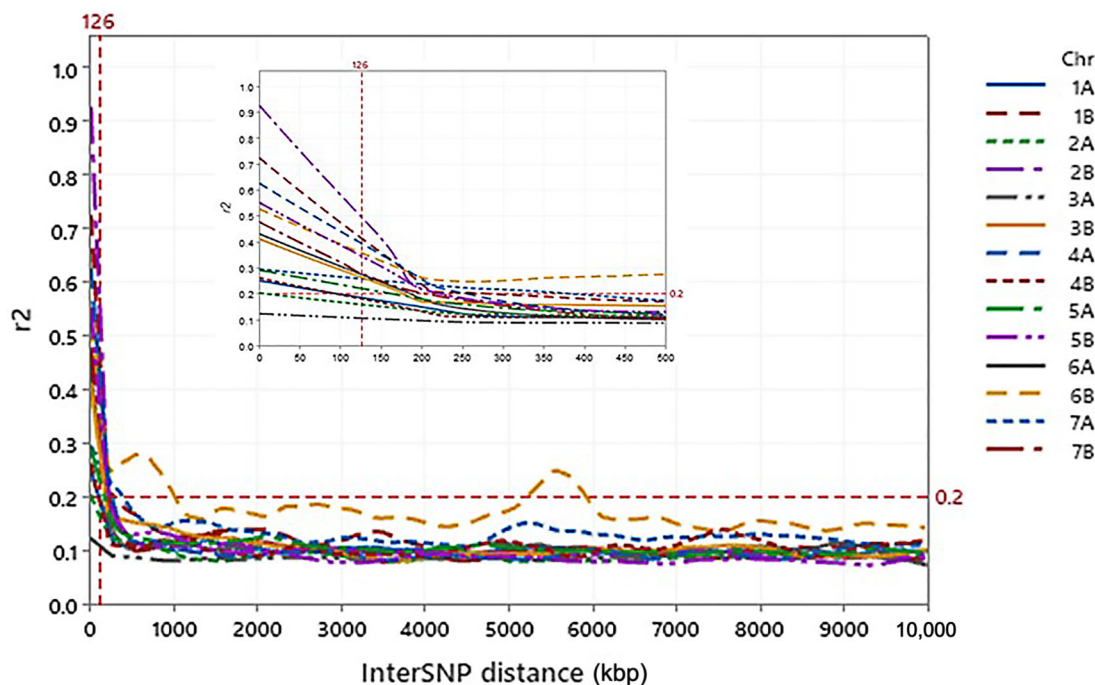
### 3.3 | Genetic stratification of the WEW-GB collection

To assess the pattern of population stratification within the WEW-GB collection, kinship/clustering analyses were applied on a selected set of 2444 SNP markers in low LD across the wild emmer genome ( $r^2 < 0.3$ ) and having MAF  $> 0.10$ .

The phylogenetic relationship among accessions obtained using Ward's hierarchical clustering clearly distinguished two main wild emmer lineages that mostly correspond to two different geographic areas. The *northern* lineage comprehends accessions from Turkey, Iran, and Iraq, while the *southern*

lineage includes accessions from Israel, Syria, and Lebanon (Figure 3A). The only exception is represented by the Kartal's population in Turkey that phylogenetically located in the *southern* lineage.

The Bayesian clustering approach implemented by STRUCTURE was run from  $K = 1$  to 20. The estimated log probability  $\text{Ln}(\text{Pr})$  increased continuously with increasing  $K$  (Figure S2A); however, after  $K = 7$ , it showed reduced variation with an increase in  $K$ , and  $\Delta K$  peaked at  $K = 4/6$  (Figure S2B). As for number of admixed genotypes, the maximum (36) was at  $K = 4$ , while 23 to 26 admixed genotypes were found at  $K5$ – $K9$ . File S1 associates  $Q$  membership to each accession at  $K$  values from 4 to 9, also represented in Figure 3B, and subpopulation names, as defined by clustering analyses and geographic origin. At  $K = 4$ , two clusters involving two populations each were defined within the *northern* and *southern* lineages defined by Ward's clustering, and no accession of one lineage was found within the clusters defined for the other lineage. With respect to the *southern* lineage, the major group included 99 accessions mostly originating from Israel (74), in addition to some accessions from Lebanon (13) and Syria (12), while the minor group encompassed 15 genotypes all coming from



**FIGURE 2** Chromosome-specific linkage disequilibrium (LD) decay as a function of single-nucleotide polymorphism (SNP) physical distance in the WEW-GB collection. Pairwise  $r^2$  values were calculated up to a maximum inter-SNP distance of 10 Mb, and then significant values ( $p < 0.01$ ) were plotted against physical distance. The threshold of critical  $r^2$  value is shown. The box shows a zoom in at max 0.5 Mb SNP distance.

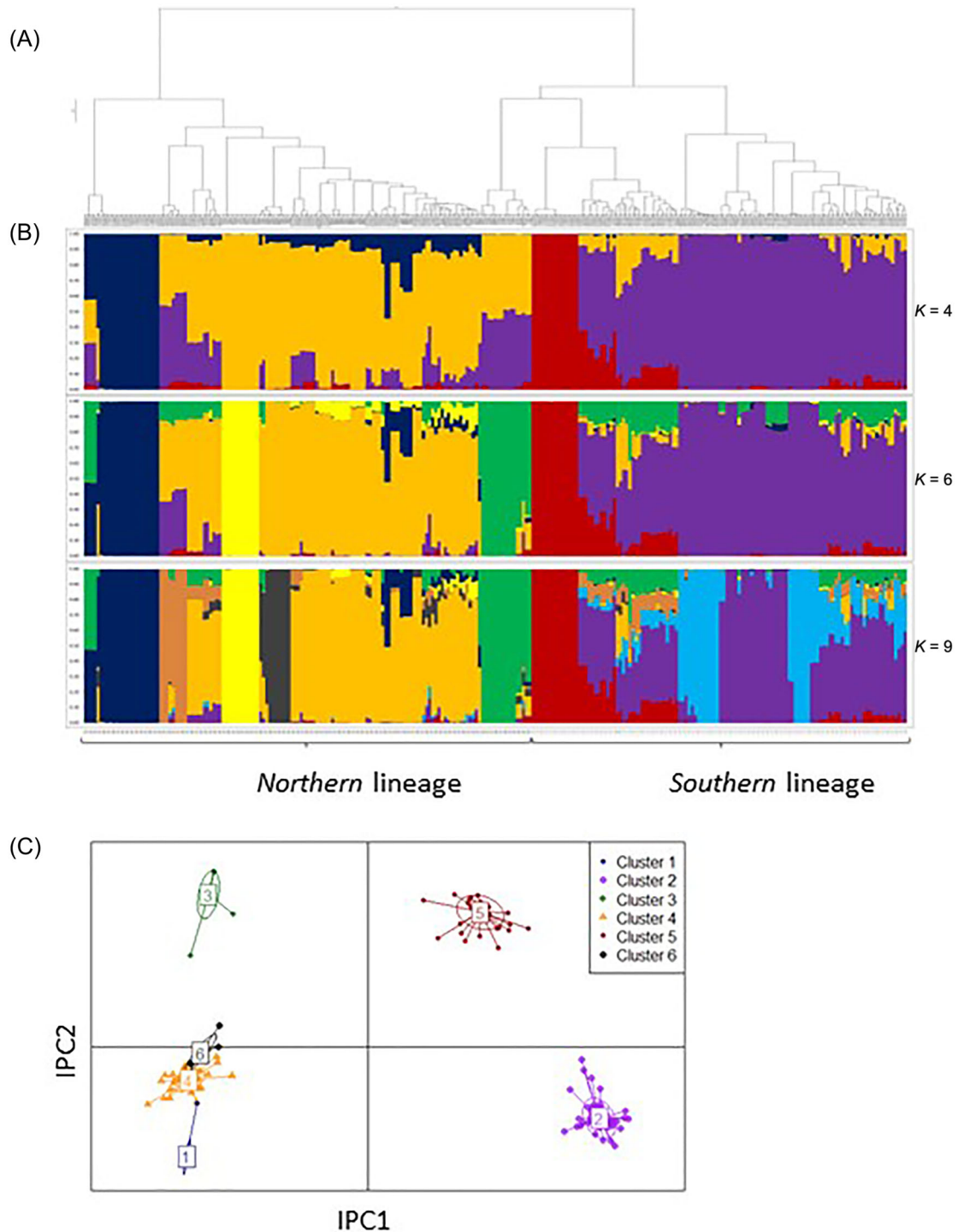
Lebanon, and four accessions were admixed between the two. For the *northern* lineage, at  $K = 4$ , the few separated accessions (22) were a subset among those originating from Karacadağ mountains (TUR-Diyarbakir population), which contrasted a major cluster including accessions from different regions in Turkey (80), but also from Iraq (7) and Iran (4). No admixed genotypes were found among the two *northern* clusters. Instead, many admixed genotypes (28) were found among the major groups of the two main lineages, and they were mostly originating from a specific region of Iraq (IQ-Sinjar) and Turkey (TUR-Kartal), the latter already evidenced by Ward's clustering. The two *southern* clusters defined at  $K = 4$  were stable across increasing  $K$ , until  $K = 7$ , when an additional group was defined by two subsets of accessions originating from North Israel (SYR/ISR-Golan and ISR-Afula). Within the *northern* lineage, one additional cluster was defined starting from  $K = 5$  and comprised lines from the E-Siverek region in Turkey. The TUR-Kartal population shows a peculiar behavior: admixed at  $K = 4$ , it was included in the major cluster of the *northern* lineage at  $K = 5$ , and then defined as an individual group from  $K = 6$ . Lastly,  $K = 8$  can separate genotypes originating from Iraq in two groups, the first one made of lines constituting the IQ-Sinjar population, previously considered as admixed, and the second one including five genotypes from the IN/IQ-Zagros region.

To complement the model-based clustering method, a multivariate DAPC analysis was performed. The BIC analysis

upon DAPC supported an optimum number of subpopulations equal to 6 (Figure S2C; File S1). A scatterplot of the first two principal components of the DAPC, which accounted for 82.2% of the total variance, describes the relationships among the six clusters (Figure 3C).

The first component (IPC1) separated accessions mostly based on the two known wild emmer lineages, with the *northern* lineage on the left side of the DAPC graph, and the *southern* lineage on the right side. The second component (IPC2) identified clusters related to more specific geographic origins. The DAPC patterning almost corresponded to the one highlighted by STRUCTURE at  $K = 6$  (Table 2); however, DAPC clusters were more coherent with the geographic separation. With respect to the *southern* lineage, DAPC divided accessions in two populations (C2 and C5). The largest one (C2) was composed of 75 lines mostly originating from Israel, while C5 included 54 genotypes mostly originating from Lebanon and Syria. C5 also included 10 accessions of the *northern* lineage, mostly from the IQ-Sinjar mountain range, and thus it is the only cluster gathering accessions from both lineages, mostly identified as admixed by STRUCTURE. In addition, it includes accessions from Lebanon that STRUCTURE grouped with the Israel main cluster till  $K = 8$ . Within the *northern* lineage, four populations were identified by DAPC. Clusters C1 (22 accessions) and C3 (21 accessions) corresponded to TUR-Diyarbakir and TUR-Kartal populations defined by STRUCTURE at  $K = 4$  and  $K = 6$ , respectively. Cluster C4 included most of the





**FIGURE 3** Population structure in the wild emmer collection according to Ward's clustering (A), the STRUCTURE model-based clustering with a box plot representation of accession membership at  $K = 4$ ,  $K = 6$ , and  $K = 9$  (B), and the multivariate discriminant analysis of principal components (DAPC) with a dot plots representation of principal component 1 (IPC1) and 2 (IPC2) distances for six subpopulations (C). Correspondence between STRUCTURE populations and DPCA clusters is evidenced by using the same colors for bar plot and dot plot graphs, respectively. The only exceptions are the group 5 from STRUCTURE K6 (TUR\_E-Siverek), which is included in DAPC cluster 4, and the DAPC cluster 6 (TUR-Pirinlick), which is included in group 3 for STRUCTURE K6,

**TABLE 2** Correspondences between clustering of the WEW-GB collection based on two Bayesian admixture models as calculated by STRUCTURE (group from 1 to 6, plus admixed accessions, based on the LD and MAF-filtered 2444 polymorphic SNP set) and by ADMIXTURE (population ID from Pop 1 to Pop 18, based on the MAF-filtered 11,342 polymorphic SNP set), or on the multivariate analysis DAPC (cluster from C1 to C6, based on the LD and MAF-filtered 2444 polymorphic SNP set).

DAPC	STRUCTURE K6	ADMIXTURE K18	Subpopulation origin
C1	4	Pop 15	TUR_36Km W Diyarbakir
C1	Admixed	Pop 1	TUR_36Km W Diyarbakir
C2	2	Pop 2	ISR_Afula
C2	2	Pop 3	ISR_Rosh Pinna
C2	2	Pop 6	SYR-ISR_Golan
C2	2	Pop 10	ISR_North/Central
C2	2	Pop 16	ISR_Bat Shelomo
C2	2	Admixed	ISR_admixed
C3	6	Pop 14	TUR_Kartal
C4	3	Pop 1	TUR_Karacadag
C4	5	Pop 4	TUR_S-Siverek
C4	3	Pop 7	TUR_52Km Diyarbakir
C4	3	Pop 8	TUR_Karabahce
C4	3	Pop 9	TUR_24Km SW Diyarbakir
C4	3	Pop 12	INIQ_Zagros
C4	3	Pop 13	TUR_E-Siverek
C4	3	Admixed	TUR_admixed
C5	2	Admixed	SYR_admixed
C5	Admixed	Pop 11	IQ_Sinjar
C5	2	Pop 17	LBN_SYR
C5	1	Pop 5	LBN_Aiha-Kfarkouk
C6	3	Pop 18	TUR_Pirinclik

Abbreviations: DAPC, discriminant analysis of principal components; LD, linkage disequilibrium; MAF, minimum allele frequency; SNPs, single-nucleotide polymorphisms.

Turkish accessions and some genotypes from Iraq and Iran and corresponded to the major group identified at  $K = 4$  within the *northern* lineage. However, it also included the E-Siverek population, which was instead individually clustered by STRUCTURE since  $K = 5$ , while it excluded accessions from Pirinclik area. Indeed, this area hosts cluster C6, a population that separated from the largest group of the *northern* lineage only from  $K = 9$  model-based clustering.

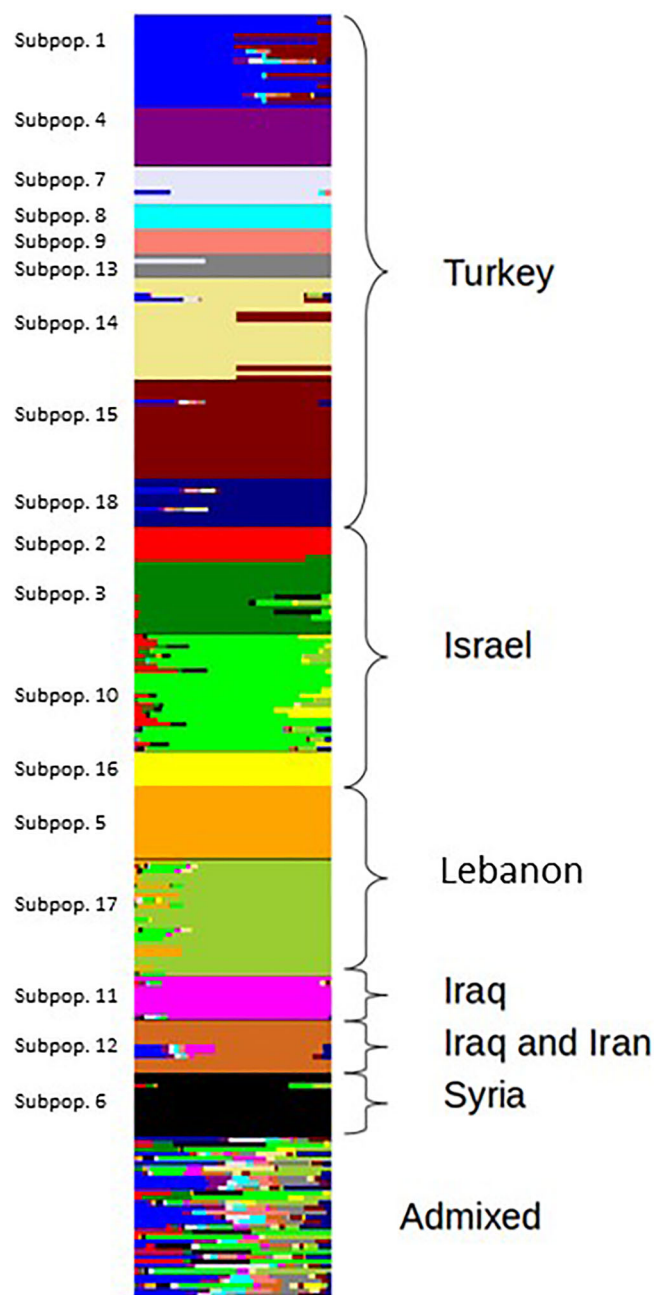
To test if the population structure could be affected by the number of assessed SNPs, the Bayesian admixture model was recalculated through the ADMIXTURE procedure considering the full SNP dataset (11,342 SNPs). The pattern of the cross-validation error indicated  $K = 18$  as the most likely number of populations (Figure S2D), which largely reflected the specific geographic area of origin at high resolution (Figure 4). Populations 1, 4, 7, 8, 9, 13, 14, 15, and 18 comprise samples from Turkey. Populations 2, 3, 10, and 16 include accessions from Israel. Populations 5 and 17 are composed of accessions from Lebanon. Populations 11 and 12 contain samples from Iraq and Iran, and Population 6

from Syria. The number of accessions in each of the 18 populations ranged from 5 to 24. ADMIXTURE was able to identify additional subgroups within those already recognized by STRUCTURE and DPCA, and one population only was split between two DAPC clusters (only two accessions).

Thus, ADMIXTURE allowed reaching a higher clustering resolution while keeping consistency between the recognized subpopulations and larger clusters (Table 2).

### 3.4 | Gene diversity within populations and population differentiation

Considering the whole SNP dataset of 11,342 SNPs, we described the pattern of genetic diversity between the two lineages, as defined according to the most frequent population structure and passport data (*northern* lineage for accessions originated from Turkey, Iran, and Iraq; *southern* lineage for accessions from Israel, Syria, and Lebanon). Therefore, the Kartal's population from Turkey was included in the *northern*



**FIGURE 4** Bar plot of ADMIXTURE results showing admixture proportions for each genotype, and clustering of genotypes in 18 subpopulations,

lineage according to DAPC and Bayesian clustering outputs, despite Ward's clustering had defined it as a branch of the *southern* lineage. AMOVA highlighted a moderate level (32%) of genetic variance distinguishing the two lineages and the largest portion of variance within lineages (68%). The *northern* lineage was the most diverse with an average of 3666 pairwise differences within population compared to 3112 in the *southern* lineage. The same evidence was obtained by calculating Nei's gene diversity index, which was 0.274 for the *southern* lineage and 0.32 for the *northern* lineage, while the

two lineages have a similar number of polymorphic markers (10,765 against 10,987, respectively). Populations as defined by ADMIXTURE were also assessed for intrapopulation gene diversity to identify the most interesting population, upon excluding admixed accessions ( $n = 33$ ) (Table 3). Five subpopulations (Pop 1, 10, 12, 14, and 17) showed higher gene diversity based on Nei's index, number of pairwise differences, and number of polymorphic markers. Among them, Pop 1, 10, and 17 include accessions with various ecogeographic origins, from Karacadag region in Turkey to locations ranging from southern Lebanon and Syria to central Israel. The two remaining populations (Pop 12 and 14) correspond to a small and well-defined area next to Sulaymaniyah in Iraq (Zagros region) and on the Kartal-Karadag hills in Turkey, respectively. Other subpopulations showed a remarkable low level of gene diversity, although constituted by a considerable number of accessions, for instance, the Aiha-Kfarkouk's population in Lebanon and the 36Km W. Diyarbakir's population in Turkey, including 15 and 20 genotypes, respectively.

With respect to among-population comparisons, high differentiation  $F_{ST}$  values were found, with only a few exceptions (Figure 5), which suggested that population structure as identified by ADMIXTURE explained most of the genetic variation. Indeed,  $F_{ST}$  values higher than 0.9 were notably frequent for populations originating from well-defined specific geographic areas, as in the cases of Afula's and Bat Shelomo subpopulations in Israel (Pop 2 and 16) and S-Siverek subpopulation in Turkey (Pop 4). Notably, some subpopulations showed higher  $F_{ST}$  values when evaluated in pair-wise comparison with other subpopulations both from the same lineage and the other one (i.e., Pop 4 and 15). Relatively lower values were generally exhibited by those subpopulations that included accessions originating from less delimited areas (i.e., Pop 1, 10, and 17). Interestingly, the Kartal's population (Pop 14) exhibited moderate  $F_{ST}$  values against all other subpopulations, with a minimum against the Turkish Pop 1. The latter was also the closest subpopulation to the Zagros population in Iraqi/Iran (Pop 12), while the Iraqi population from the Sinjar mountains (Pop 11) had its  $F_{ST}$  lowest value when assessed against the Israeli Pop 10.

Finally, considering the whole SNP dataset of 11,342 SNPs, DAPC clusters were analyzed for private/unique alleles (SNPs occurring in a specific group only). A total of 225 PAs were identified that occurred in four out of the six DAPC clusters (File S1). The frequency of these alleles in their respective groups is generally high; indeed, 141 PAs have a frequency higher than 25% and five alleles have a frequency higher than 90%, while only four alleles have a frequency lower than 10%. Out of the 225 unique alleles, 125 were mapped on the wild emmer genome, with a prevalence on chromosomes 3B and 5B. Notably, three SNPs with higher frequency were located close to each other at 527–528 Mb

Subpop ID	Pop 7_TUR	Pop 18_TUR	Pop 3_ISR	Pop 6_SYR	Pop 2_ISR	Pop 4_TUR	Pop 15_TUR	Pop 8_TUR	Pop 9_TUR	Pop 13_TUR	Pop 5_LEB	Pop 16_ISR	Pop 11_IN/IQ	Pop 12_INQ	Pop 1_TUR	Pop 14_TUR	Pop 17_LEB
Pop 18_TUR	0.591																
Pop 3_ISR	0.646	0.701															
Pop 6_SYR	0.743	0.785	0.528														
Pop 2_ISR	0.766	0.813	0.564	0.706													
Pop 4_TUR	0.807	0.830	0.829	0.910	0.973												
Pop 15_TUR	0.824	0.843	0.850	0.915	0.960	0.971											
Pop 8_TUR	0.703	0.736	0.774	0.875	0.956	0.998	0.959										
Pop 9_TUR	0.703	0.744	0.770	0.875	0.955	0.997	0.962	0.998									
Pop 13_TUR	0.644	0.709	0.751	0.855	0.932	0.971	0.941	0.947	0.945								
Pop 5_LEB	0.817	0.850	0.681	0.809	0.880	0.959	0.953	0.945	0.944	0.929							
Pop 16_ISR	0.766	0.499	0.588	0.727	0.854	0.970	0.958	0.950	0.950	0.926	0.881						
Pop 11_INQ	0.610	0.687	0.606	0.706	0.734	0.851	0.866	0.779	0.774	0.747	0.790	0.737					
Pop 12_IN/IQ	0.486	0.528	0.614	0.688	0.688	0.700	0.747	0.594	0.610	0.574	0.755	0.684	0.551				
Pop 1_TUR	0.354	0.409	0.580	0.643	0.631	0.535	0.497	0.412	0.438	0.364	0.690	0.632	0.510	0.353			
Pop 14_TUR	0.452	0.805	0.522	0.595	0.586	0.639	0.668	0.564	0.556	0.532	0.641	0.587	0.495	0.446	0.381		
Pop 17_LEB	0.508	0.583	0.312	0.427	0.423	0.690	0.723	0.629	0.623	0.601	0.404	0.409	0.446	0.508	0.490	0.431	
Pop 10_ISR	0.490	0.561	0.238	0.325	0.322	0.670	0.706	0.608	0.601	0.580	0.482	0.348	0.433	0.485	0.476	0.403	0.177

FIGURE 5 Population differentiation calculated as pairwise  $F_{ST}$ . Subpopulation IDs are the same as in Figure 4.

on chromosome 5A, thus co-locating with the *CBF* locus, the gene cluster of C-repeat Binding Factors that have a key role in plant cold acclimation (Guo et al., 2018). Cluster 2 had the highest mean number of unique alleles (33), followed by clusters 4 and 5 with 16 and 20, respectively, up to an average of five PAs within cluster 3. Three specific geographic regions appear to harbor accessions that were significantly enriched in PAs. The first are Golan Heights where 10 accessions with more than 60 PAs originated; other eight Israeli accessions with around 40 unique alleles were all originated from the Bat Shelomo moshav. Lastly, in cluster 5, accessions originating from Rashaya district have an average of 30 PAs.

### 3.5 | Stem rust resistance in the wild emmer collection

Out of 283 genotyped wild emmer accessions, 278 lines were evaluated for reaction to stem rust infection in controlled conditions at seedling stage. A high susceptibility scoring on susceptible checks in all experiments demonstrated that optimal infection levels were achieved, allowing for the clear scoring of IT. The IT score = 6 in the linear scale was chosen as threshold for resistance, since it corresponds to small to medium uredinia, often surrounded by chlorosis or necrosis, which are symptoms of a resistance reaction. The frequency distribution of IT scoring values highlighted a prevalence of susceptible genotypes (Figures 6A and S3). Focusing only on resistant lines (IT < 6), the distribution plot of IT values (Figure 6B) showed a higher frequency of moderate resistance (IT > 5) against races JRCQC and TRTTF, as opposed to a slightly higher frequency of strong resistance (IT < 1) for TTKSK. A further different distribution profile was shown by resistant IT values against TKTTF and TTTTF with a homogeneous distribution in the range 0–6.

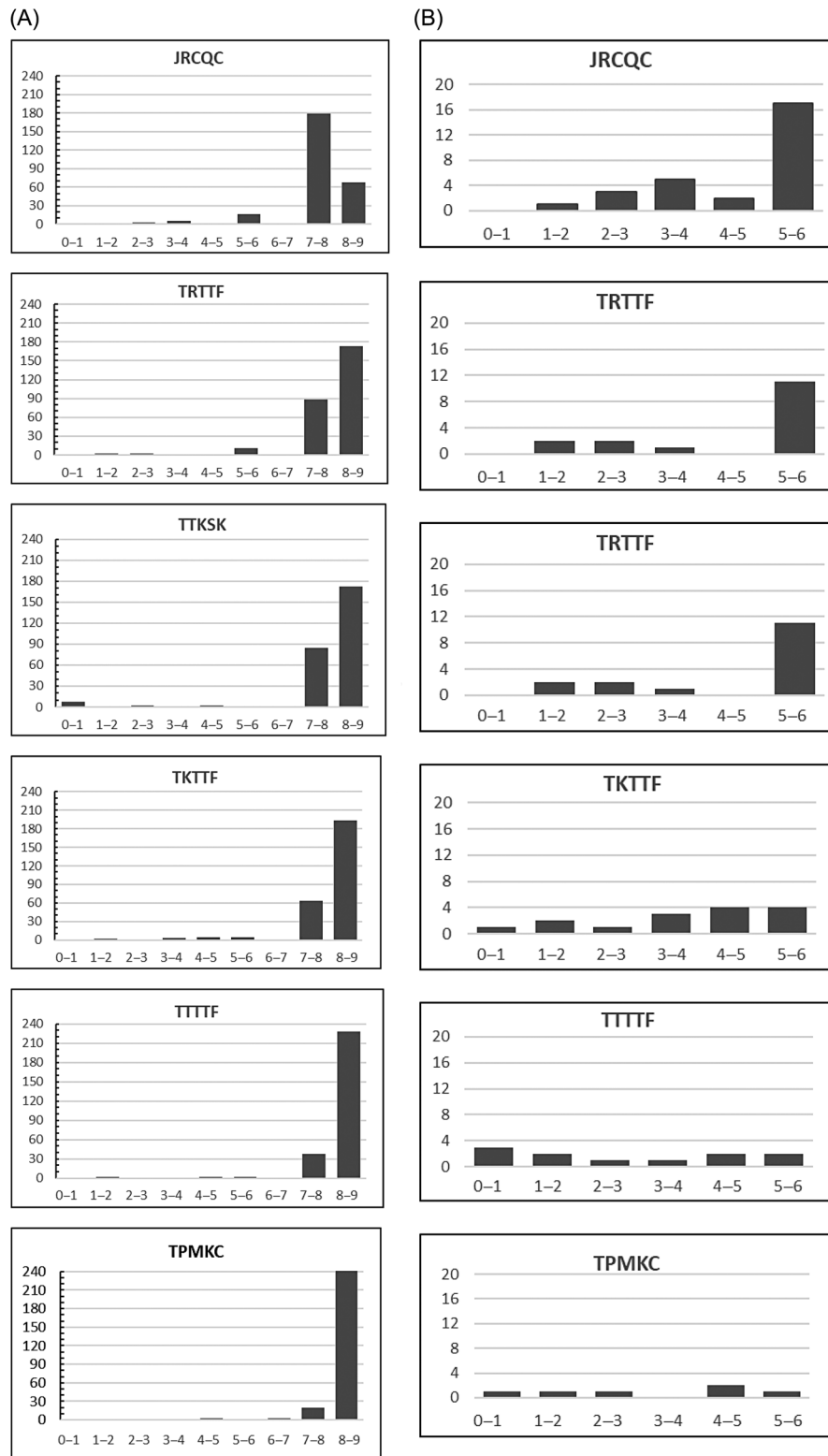
The highest number of resistant accessions was found in response to infection with the race JRCQC (28 accessions out of 278 scored, 10%), followed by TRTTF (16 accessions, 6%), while for race TPMKC, only six accessions were resistant.

These resistant genotypes belonged to both lineages, although with a majority from the *southern* lineage. Resistant accessions against the other three races were all from the *southern* lineage: 15 accessions over 275 scored (6%) showed resistance to TKTTF, 13 on 270 (5%) against TTKSK, and 11 on 278 (4%) against TTTTF.

One accession (DIC199) was resistant to all tested races, two accessions (DIC061 and DIC101) were resistant to five races, but both susceptible to TPMKC, three accessions to four races (DIC193, DIC256, DIC288), four accessions to three races (DIC130, DIC255, DIC289, DIC353), and six accessions to two races (DIC058, DIC088, DIC107, DIC111, DIC202, DIC326). All these accessions belong to the *southern* lineage, except genotypes DIC288, DIC289, and DIC353. With respect to these last three accessions, morphological observations of plants together with failed crossing attempts and failed genotyping suggested that they were misclassified as *T. dicoccoides*, while they most likely belong to *Triticum timopheevii*. It is noteworthy that the six accessions showing resistance to TPMKC also showed resistance to TTTTF, with only one exception.

All subpopulations of the *southern* lineage defined by ADMIXTURE showed stem rust disease resistance; however, the distribution of reaction specificity to *Pgt* race was not homogeneous. Indeed, the subpopulations Pop 3 and 10 (ISR-RoshPinna and ISR-North/Central, respectively) carry resistance genes to all *Pgt* races, as opposite to the Pop 2 (ISR-Afula), Pop 5 (LBN\_SYR), and Pop 17 (LBN\_Aiha-Kfarkouk) that contain resistance sources to one *Pgt* race only (TTKSK, TKTTF, and JRCQC, respectively). Besides the mentioned subpopulations and corresponding geographic areas, other resistant lines originated mainly from Bat Shelomo in Israel (Pop 10 and 16) and Kazrin in Syria (Pop 6).

The strong skewness for the distribution of phenotypic data impaired the application of analysis of variance to assess genotypic differences. Therefore, the nonparametric Kruskal–Wallis test was performed that highlighted statistically significant differences among genotypes for resistance to all *Pgt* races (Table 4).



**FIGURE 6** Frequency distribution of infection type (IT) scoring values in response to six stem rust races, calculated for the whole WEW-GB collection (A) or limited to resistant accessions (IT ≤ 6) only (B).

TABLE 3 Gene diversity within subpopulations as defined by ADMIXTURE.

Pop ID	POP 1	POP 2	POP 3	POP 4	POP 5	POP 6	POP 7	POP 8	POP 9	POP 10	POP 11	POP 12	POP 13	POP 14	POP 15	POP 16	POP 17	POP 18	Mean
Country	TUR	ISR	ISR	TUR	TUR	ISR	TUR	TUR	TUR	ISR	ISR	ISR	TUR	TUR	TUR	ISR	LBN/ SYR	TUR	-
Location	Kara- caddag	Afula	Rosh- Pinna	S Siver.	Aiha- Kfark.	Golan	52Km W Diyarb	Kara- balce	24Km SW Diyarb	North- Central	Sinjar	Zagros	E Siver.	Kartal	36Km W Diyarb.	Bat Shel.	LBN- SYR	Pirin- clik	-
Number	19	7	15	12	15	13	8	5	5	24	9	11	5	21	20	7	24	10	nr
Net's gene diversity	0.23	0.03	0.14	0.001	0.03	0.08	0.15	0.001	0.001	0.24	0.14	0.21	0.03	0.22	0.013	0.04	0.22	0.12	0.11
Number of polymorphic loci	6222	1187	3561	31	684	2231	3358	10	14	7085	3664	5288	661	6647	1370	1105	6649	3619	2966
Mean number of pairwise differences	2401	349	1466	8	340	831	1590	5	6	2459	1469	2250	261	2325	140	391	2286	1256	1102

Note: Origin country, main geographic location, and the number of accessions for each subpopulation are indicated, besides indices for gene diversity. Subpopulation IDs are the same as Figure 4.

TABLE 4 Kruskal–Wallis test of IT scoring values in response to six stem rust races of the WEW-GB collection.

Pgt race	H	p value	CV	Mean
JRCQC	790.3	<0.0001	14.52	7.78
TRTTF	713.97	<0.0001	13.75	8.16
TTKSK	698.35	<0.0001	18.87	8.12
TKTTF	685.08	<0.0001	15.99	8.24
TTTTF	697	<0.0001	14.85	8.36
TPMKC	413.13	<0.0001	14.82	8.48

Note: H is the statistic test for the Kruskal–Wallis analysis.

Abbreviations: CV, coefficient of variation; IT, infection type; Pgt, *Puccinia graminis* f. sp. *tritici*.

### 3.6 | Association mapping of stem rust reaction on the wild emmer panel

Considering the successfully genotyped panel of 263 accessions, the frequency of resistant lines was 9% and 5% for JRCQC and TRTTF, but with only 5 and 4 accessions on a total of 144 genotypes for the *northern* lineage, respectively. Such a low frequency of resistant lines might impair finding of associated alleles with MAF >0.05. For the races TKTTF, TTKSK, and TTTTTF, resistant accessions were all belonging to the *southern* wild emmer lineage. The frequency of resistant accessions within the *southern* genotyped and phenotyped panel (118 accessions in total) was as follow: 16% ( $n = 19$  accessions) for JRCQC, 8.5% ( $n = 10$ ) for TRTTF, 12% ( $n = 14$ ) for TKTTF, 8.5% ( $n = 10$ ) for TTKSK, and 6% ( $n = 7$ ) for TTTTTF. GWAS was therefore conducted only on the *Southern* subset for these five Pgt races. For this analysis, only molecular markers polymorphic on the *southern* subset were used for a total of 8125 SNPs, including 5013 physically anchored SNPs and 3112 unmapped. With respect to data for race TPMKC, only three among the six resistant accessions were successfully genotyped; therefore, association analysis was not conducted.

The GWAS analysis was carried out with an MLM, including kinship and the first two PCA components to account for population structure and genotype relatedness. On the basis of the Bonferroni-corrected threshold, 19 SNPs were significantly associated to the resistance, out of which six were in response to race JRCQC, seven to TRTTF, three to TTKSK, and three to TTTTTF, with some of these markers in common among the races. No markers were significantly associated to TKTTF (File S2); however, the SNP AX-94616911 was considered because it was close to the threshold and found associated with resistance to TTTTTF. To identify MTAs with a strong and significant effect on the resistance level, we analyzed the difference in resistance to Pgt between classes of genotypes carrying the alternative alleles at each MTA, and we focused on those MTAs for which the mean IT value of one genotypic class was lower than 6. On the basis of this criteria,

11 MTAs were identified for four races (Table 5; File S2): one in response to race TTTTF, two to TTKSK and TRTTF, and six to JRCQC, for a total of eight different SNPs. No MTAs related to TKTTF passed this filter. For these MTAs, the most probable position on WEW reference genome was provided (Table 5), although chromosome assignment remained ambiguous for some markers. For each MTA, the chromosome region surrounding the peak marker within the limits of LD decay was inspected for other supporting markers significantly associated based on a relaxed  $p$ -value threshold. As expected, for all MTAs, the minor alternative was the allele conferring resistance phenotypes. To exclude an effect of the population structure, the occurrence of the associated alleles in the different subpopulations was counted for each MTA.

The first locus for resistance against race TRTTF (SNP AX-94637067) mapped to the long arm of chromosome 2A and explained 17% of the observed phenotypic variation. The additive allelic effect was 1.7 with a positive role on resistance for the T allele. The association of this chromosome region might be corroborated by another SNP, AX-94749848, in high LD ( $r^2 = 0.74$ ) and 2 kb distant with respect to AX-94637067, although with a  $p$ -value above the threshold ( $2.69 \times 10^{-4}$ ). Interestingly, the second region for resistance to TRTTF corresponded to the SNP AX-94530330 located on the short arm of chromosome 5B, which was also the unique marker significantly associated with reaction to TTTTF. This SNP explained 17% and 21% of the observed variation for reaction to races TRTTF and TTTTF, with IT additive effects of 1.4 and 1.5, respectively.

Two MTAs were identified for *Pgt* race TTKSK, AX-94444583 and AX-94661492, which explained 20% and 29% of the phenotypic variation, respectively. The IT additive allelic effect calculated on the two genotypic classes was 1.65 and 1.9, respectively. The same markers were significantly associated also with the reaction to *Pgt* race JRCQC. In this case, the two markers explained 36% and 49% of the observed phenotypic variation, respectively, with an additive effect of 2.05 for both markers. The two SNPs are located on chromosome group 1, but BLAST results could not provide a conclusive assignment to a specific subgenome of the Zavitan v2 genome, neither by attempting an LD-based positioning. Analogously, when CS RefSeqv2.1 and Svevo v1 genomes were searched with the sequence of the two markers as queries, hits on chromosome 1B and 1A were retrieved for both (Table 5; File S2), with a small difference between Blast score and E values. On all three genome assemblies, the most likely chromosome assignment is 1B for AX-94661492 and 1A for AX-94444583, although their moderate LD ( $r^2 = 0.69$ ) would suggest a co-location.

Four other significant MTAs were identified for race JRCQC. AX-94479993 explained 22% phenotypic variation. For this SNPs, the best chromosome assignment by BLAST similarity searches was 2A; however, the second hit on chro-

mosome 1B at 3 Mb from AX-94444583 cannot be excluded (Table 5; File S2), also based on a moderate LD of about  $r^2 = 0.51$  with this latter SNP, which listed chromosome 1B as one possible assignment based on all three genome assemblies (Zavitan v2, IWGSC RefSeqv2.1, and Svevo v1) considered. Notably, on chromosome 1B, the three markers AX-94444583, AX-94661492, and AX-94479993 are located in the same order and at the same relative roughly distance in the three genomes.

The remaining three markers (AX-94766528, AX-94430011, and AX-94879817) all have the same  $p$ -value and explained 19% of the variance. The strong LD among them ( $r^2 = 0.95$ ) would suggest a close physical linkage; however, their chromosome position on the Zavitan v2 genome is doubtful since different indications were obtained by BLAST hits of SNP sequences (Table 5; File S2). Indeed, compelling evidence for a position on chromosome 7A was obtained for AX-94766528 only, while a very low Blast score was found for chromosome 4A (at around 683 Mb). The strong LD ( $r^2 = 0.95$ ) among the three markers would suggest a close physical linkage; however, mapping the other two SNPs on chromosome 7A based on LD with AX-94766528 seems not reliable. Indeed, the position for AX-94879817 remained uncertain among chromosomes 6B, 2B, and 7B, but with no hit on 7A and 4A, while AX-94430011 was totally unassigned. When the sequence of the three markers was blasted against the durum Svevo v1 genome, chromosome 4A was indicated as the best hit for two of them and among the best hits for AX-94879817. Analogous results were obtained on the IWGSC RefSeqV2 genome assembly (File S2). Although with a higher  $p$ -value, AX-94427726, a SNP in high LD with the latter three SNPs, can provide further evidence for the location of such genomic region on chromosome 4A.

### 3.7 | Candidate genes at MTA chromosome regions

Putative candidate genes for resistance to *Pgt* races were searched within the most probable genomic regions of MTAs considering as interval the extension of LD as calculated for each chromosome (Table 1) on the right and left side of the MTA. The number of high confidence genes annotated in each region of the Zavitan v2 genome varied from 1, for markers AX-94661492 on chromosome 1B and for marker AX-94766528 on chromosome 7A (intervals of  $\pm 278$  and  $\pm 390$  kb, respectively), to 16, for marker AX-94879817 on chromosome 6B (interval of  $\pm 1020$  kb). Genes with an annotated function related to plant response to pathogens were found in some cases. A region of  $\pm 58$  kb around the marker AX-94479993 on chromosome 2A contained six high confidence genes: five showed a disease-related annotation, as *LRR receptor-like serine threonine-kinase*, probable

**TABLE 5** SNPs significantly associated to reaction to stem rust infection in the southern wild emmer subset, according to Bonferroni's test at  $p = 0.05$  ( $-\text{Log}_{10} = 5.2$  for 8125 SNPs).

Trait	SNP Id	On WEWSeq v2.0 (Zavitan)			On Svevo v1			On IWGSC RefSeqv2.1 (CS)			-Log <sub>10</sub> (p)	R <sup>2a</sup>	SNP alleles <sup>b</sup>	Average IT of accessions carrying		Additive effect	MAF (%)	Number of new accessions with the R allele in the GWAS panel	Subpopulations including accessions with the associated allele	LD (r <sup>2</sup> )
		Chr	bp	Chr	bp	Chr	bp	Chr	bp	Major allele				Minor allele						
TRTF	AX-94530330	5B	4,581,777	unk	nd	5A	LC	5A	LC	5.2	0.17	G/C	8.4	5.6	1.4	7	8	Pop 3(1), Pop 10(4), Pop 17(1), Adm(2)	-	
TRTF	AX-94637067	2A	646,378,206	2A	637,819,187	2A	646,607,353	2A	646,607,353	5.2	0.17	G/T	8.4	5	1.7	5	6	Pop6(3), Adm(3)	0.74	
TRTF	AX-94749848	2A	646,380,114	2A	637,060,799	2A	646,609,254	2A	646,609,254	3.6	0.11	G/A	8.4	5.8	1.3	7	8	Pop6(3), Pop 10(2), Adm(3)	-	
TTTTF	AX-94530330	5B	4,581,777	unk	nd	5A	LC	5A	LC	6.5	0.21	G/C	8.6	5	1.8	8	8	Pop3(1), Pop 10(4), Pop 17(1), Adm(2)	-	
TTKSK	AX-94661492	1B*	54,292,926*	1B*	37,821,758*	1B*	47,606,593*	1B*	47,606,593*	9.1	0.29	A/G	8.3	4.5	1.9	9	10	Pop3(1), Pop 10(4), Pop 16(1), Adm(1)	0.69	
TTKSK	AX-94444583	1A*	43,104,857*	1A*	29,659,368*	1A*	35,073,917*	1A*	35,073,917*	6.4	0.20	A/G	8.2	4.9	1.65	8	9	Pop3(1), Pop 10(4), Pop 16(1), Adm(1)	-	
JRCQC	AX-94661492	1B*	54,292,926*	1B*	45,661,342*	1B*	47,606,593*	1B*	47,606,593*	17.5	0.49	A/G	7.8	3.7	2.05	9	10	Pop3(1), Pop 10(4), Pop 16(1), Adm(1)	0.59– 0.51	
JRCQC	AX-94444583	1A*	43,104,857*	1A*	34,782,400*	1A*	35,073,917*	1A*	35,073,917*	12.3	0.36	A/G	7.8	3.7	2.05	8	9	Pop3(1), Pop 10(4), Pop 16(1), Adm(1)	-	
JRCQC	AX-94479993	2A*	31,561,505*	2A*	27,631,080*	2A*	28,869,214*	2A*	28,869,214*	7.0	0.22	G/C	7.8	4.7	1.55	11	13	Pop3(4), Pop 10(4), Pop 16(1), Pop 17(4)	-	
JRCQC	AX-94766528	7A*	45,544,188*	4A	670,278,881	4A	653,860,028	4A	653,860,028	6.1	0.19	C/T	7.8	5.5	1.15	16	19	Pop3(5), Pop 10(4), Pop 16(7), Pop 17(1), Adm(2)	0.94–1	
JRCQC	AX-94430011	nd	nd	4A	670,051,203	4A	654,083,434	4A	654,083,434	6.1	0.19	G/C	7.8	5.5	1.15	16	19	Pop3(5), Pop 10(4), Pop 16(7), Pop 17(1), Adm(2)	-	
JRCQC	AX-94879817	6B	35,385,985	4A	670,069,103	4A	654,066,068	4A	654,066,068	6.1	0.19	C/T	7.8	5.5	1.15	16	19	Pop3(5), Pop 10(4), Pop 16(7), Pop 17(1), Adm(2)	-	
JRCQC	AX-94427226	7A	37,303,228	4A	672,313,841	4A	652,385,669	4A	652,385,669	4.7	0.15	C/T	7.9	5.6	1.15	19	20	Pop3(5), Pop 10(4), Pop 16(7), Pop 17(1), Adm(3)	-	

*Note:* Only SNPs associated to average IT values <6 on accessions carrying the associated allele are reported. SNP IDs in bold are associated to reaction to different *Pgt* races. SNP IDs in italics are not significant according to Bonferroni; however, they are located next to significantly associated SNPs. The most probable chromosome is provided as mapping information on the three genome assemblies: wild emmer wheat (Zhu et al., 2019), durum wheat (Maccacferri et al., 2019), and bread wheat (Zhu et al., 2021). Additional information is reported in File S2. Asterisk (\*) indicates that the position reported is on the most probable chromosome, since for these SNPs blasting provided ambiguous results between homoeologous chromosomes; SNP sequence alignments for these SNPs are reported in File S2.

<sup>a</sup>Variance explained by the locus.

<sup>b</sup>Major allele is written in bold letters.



*long-chain-alcohol O-fatty-acyltransferase 4, fruit bromelain, xylanase inhibitor 1*, and *shock SRC2*. Notably, marker AX-94479993 falls nearly 60 bp upstream of the gene coding for a xylanase inhibitor. Two disease resistance-related genes are present in the region around the marker AX-94530330, annotated as *pentatricopeptide repeat-containing At1g74750-like* and *plasma membrane ATPase 1*, the latter starting nearly 200 bp downstream of the MTA. Three genes corresponding to *indole-2-monooxygenase* and an *LRR receptor-like serine threonine-kinase* are nearby the marker AX-94879817 (chromosome 6B). The SNP AX-94637067 (chromosome 2A) falls within a *putative serine threonine-kinase D6PK*, while the SNP AX-94444583 (chromosome 1A) is nearby a *kinase-interacting 1-like* gene. Within the confidence interval of AX-94661492 on chromosome 1B, only one high confidence gene is annotated as *trichome birefringence like*, a family of proteins mediating xylan acetylation, and the MTA falls within its sequence.

For markers with ambiguous mapping, regions surrounding their alternative genomic positions were also inspected for candidate genes. For the latter two markers, AX-94444583 and AX-94661492, the region surrounding their position on chromosomes 1B and 1A, respectively, was also inspected. In the first case, one gene with a kinase and interacting domain is present near the marker, while only one annotated gene was present in the interval of marker AX-94661492 on chromosome 1A, and it corresponds to an *O-acetyltransferase* involved in cell wall organization or biogenesis, a kind of gene for which a role in response to diseases has been previously documented (Gao et al., 2017). The four SNPs markers (AX-94766528, AX-94430011, AX-94879817, and AX-94427726) associated with resistance to race JRCQC were consistently located at 670 Mb on chromosome 4A in both durum and bread wheat genomes; therefore, this region was also inspected to identify possible *T. durum* homologous genes of candidates linked to stem rust resistance in WEW. Between 668 and 673 Mb, several genes related to plant reaction to diseases were found. Among them, some genes annotated as *Disease resistance protein RPM1* (669.5 Mb), *Calcium-binding EF-hand family protein* (669.9 Mb), and *Mitogen-activated protein kinase* (670.1 Mb) are located very near the position of the markers.

## 4 | DISCUSSION

### 4.1 | A wide and representative WEW collection

Wild emmer grows at many different latitudes and altitudes in a discontinuous arc, from Israel to Western Iran, thus coping with a wide range of ecological conditions of temperature and water availability. A large ex situ collection has been estab-

lished that both encompasses the two phylogenetic lineages of wild emmer previously recognized and includes accessions originating from all countries of the Fertile Crescent (WEW-GB collection). This work reports about its characterization for genetic diversity based on a high-density genotyping SNP dataset and its effective use for mapping stem rust resistance loci by GWAS.

In the literature, other complete wild emmer collections have been reported in studies related to wheat origin and domestication (i.e., Luo et al., 2007; Oliveira et al., 2020; Ozkan et al., 2005), while Ren et al. (2013) assessed 25 Turkish and Israeli wild emmer populations for genetic diversity and its relationship with ecological factors. Recently, a collection of wild and related tetraploid wheats of the Wheat Genetics Resource Center (WGRC) gene bank including both *T. turgidum* and *T. timopheevii* species was characterized for genetic diversity based on high-density genotyping by sequencing (GBS) technology to create a reference core set (Yadav et al., 2023). Although being quite large (around 900 accessions), this collection included only one accession from Iraq and none from Iran, two countries that showed to host wild emmer populations with specific genetic characteristics. Then, a small collection of Israeli accessions was used to prove effectiveness of GWAS mapping for resistance loci to yellow rust in WEW (Sela et al., 2014). Tene et al. (2022) conducted a GWAS analysis for adult plant resistance using GBS markers on 188 Israeli accessions of WEW. These collections were therefore either complete but characterized through a low number, although multi-allelic, molecular markers or extensively genotyped but not complete. Indeed, panels including only Israeli accessions or without an effective number of accessions originating from Iran and Iraq cannot be considered representative of all countries hosting WEW. The WEW-GB collection did not include accessions from Jordan; however, wild emmers from this country were generally reported to cluster or to be very similar to Syrian and Israeli accessions (Mazzucotelli et al., 2020; Ozkan et al., 2005; Yadav et al., 2023), which are well represented in the present collection. Therefore, the WEW-GB panel can be considered one of the most complete collections of wild emmer characterized so far.

Both low- and high-resolution stratification analyses were conducted on the WEW-GB collection. Insights confirmed that wild emmer genetic patterning is related to a first subdivision in the two known *northern* and *southern* lineages and further to the specific geographic origin of the accessions, as indeed previously shown by genetic structure analysis based on AFLP and RFLP markers (Luo et al., 2007; Ozkan et al., 2005), and more recently on SNP (Ren et al., 2013). Despite similar size and composition, the number of populations identified by genetic analysis in the different studied collections ranged from four or five (Ozkan et al., 2005; Ren et al., 2013) to nine (Luo et al., 2007). This is in

general agreement with the six to 11 populations recognized by our low-resolution stratification analyses through multivariate and Bayesian clustering methods (DPCA and STRUCTURE, respectively). Indeed, despite different statistics, these methods could not clearly identify a critical  $K$  value to define the number of populations (Figure S2), as also reported by Ren et al. (2013). Nevertheless, our high-resolution analysis through ADMIXTURE clustering was able to identify up to 18 populations, which in most cases have well-defined and small geographic boundaries.

The clusters defined by Ozkan et al. (2005) and Luo et al. (2007) mostly correspond to the populations and clusters we defined based on STRUCTURE and DPCA. Thus, the two Turkish mountain regions of Kartal-Karadağ and Karacadağ host two populations genetically distinguished by both our analysis (cluster 3 and cluster 4, respectively) and by Ozkan et al., 2005 (group I and II, respectively). Similarly, our separation of the Iraqi Sinjar population (Population 7 at  $K = 8$ , and ADMIXTURE pop 11) from the Iraqi and Iranian accessions originating from Zagros mountain range (Population 4 at  $K = 7$ , and ADMIXTURE pop 12—IN/IQ-Zagros) confirmed the distinction found by Ozkan et al. (2005, and further confirmed in Ozkan et al., 2011) between groups III' and III, respectively, and also the higher genetic similarity of the IN/IQ-Zagros population with the Turkish Karacadağ clusters than with the Iraqi population from the Sinjar mountain range, despite a higher geographic distance. These results were instead in contrast with findings from Luo et al. (2007), which showed more close relations between the two Iraqi populations and a greater distance with Turkish accessions. Noteworthy, our results also indicated an admixture of the Sinjar's and Kartal's populations between the two wild emmer lineages, never reported in previous analysis of wild emmer diversity. In other cases, differences between stratification of the different collections can be related to the different occurrence of accessions from some specific subpopulations or geographic areas. For instance, Luo et al. (2007) defined a cluster of 13 accessions showing spike morphology traits corresponding to those of the *judaicum* wild emmer race. In our case, only one accession in addition to Zavitan was included in the present collection, thus impairing to define an individual cluster.

Notably, high genetic differentiation between populations, more evident in the *northern* lineage (i.e., Pop 4, 8, and 6), combined with low (i.e., Pop 2 and 15) or very low (i.e., Pop 5 and 13) gene diversity within the populations is a specific feature of the WEW-GB collection. This could suggest that adaptation to specific environmental conditions, combined with limited contacts between populations, might have shaped the structure of natural wild emmer populations. A sharp local differentiation associated to ecological, climate factors and irrespective of geographic separation was already reported for wild emmer, with even microenvironmental vari-

ations between populations at the same site (Li et al., 2000; Ren et al., 2013; Zhang et al., 2018). Consequently, stratification in wild emmer collections strictly depends on microscale level of accession sampling within its patchy distribution, beside the resolution achievable by the specific clustering method. This evidence has strong implications for ex situ conservation of wild emmer diversity. As for this evidence, the WEW-GB panel being genetically characterized and widely representative of wild emmer geographic distribution is an interesting resource for its effective conservation as well as for its utilization in prebreeding programs.

## 4.2 | Resistance to stem rust in WEW

A low frequency of stem rust resistance in wild emmer was previously reported, in contrast with the higher frequency of yellow rust resistance, and mostly localized in the northern part of Israel (Huang et al., 2016). Similarly, from 4% to 10% of the accessions of the WEW-GB collection were moderate to highly resistant depending on the *Pgt* race. This indicates that rare but valuable resistance genes to stem rust can be identified in wild emmer. Low ITs were mostly found against TTKSK, while moderate resistance was common against races JRCQC and TRTTF, as already shown in a previous characterization of a tetraploid wheat collection against stem rust races (Saccomanno et al., 2018). Of great interest are accessions that carried resistance to different *Pgt* races. GWAS allowed to map most of their resistance loci. For instance, in the accession DIC199, which showed resistance phenotype against all six tested *Pgt* races, four resistance loci were mapped (two for TRTTF and JRCQC, and one each for the other two races tested in GWAS, TTTTF, and TTKSK). Analogously, accession DIC101, which was resistant to all *Pgt* races but TPMKC, carried the resistant alleles to all MTAs associated to JRCQC and TTKSK, and at the locus AX-94530330 for resistance against TKTTF and TTTTF.

Most of resistant lines belong to the *southern* lineage, suggesting that environmental conditions of the geographic areas where this wild emmer type naturally occurs are more favorable to stem rust pathogen and thus promoting the evolution and maintenance of plant resistance. In a similar vein, the distribution of resistance genes in relationship with the *Pgt* race specificity was not homogeneous among the different populations defined by ADMIXTURE, with some populations carrying resistance to different *Pgt* races. Most of the resistant lines originated from Rosh Pinna and Kazrin in northern Israel and from Bat Shelomo, which is closer to the Mediterranean shore. The nonhomogenous geographic distribution of plant resistance genes could be related to differential adaptation to specific local environmental factors of Israeli *Pgt* races present in the different wild emmer sampling sites and having similarities with *Pgt* race tested in the present study.

For instance, no resistance to JRCQC race was found in lines originated from Kazrin, while low ITs were scored against JRCQC only in Pop 17. However, more research for characterizing local Israeli *Pgt* races would be necessary to confirm this hypothesis.

Having effective sources of resistance in adapted genetic backgrounds represents a great advantage in breeding. Indeed, a resistance locus from wild accessions could be linked to deleterious alleles or might be ineffective upon introgression into a very different genomic background. Nevertheless, many effective resistance genes are needed by breeders to counter the continuing evolution of new virulence determinants in the pathogen, and wild together with domesticated gene pools represent a wider source of new and effective resistant genes/alleles compared to elite cultivars. Further studies are needed to elucidate the genetic basis of the resistance found in WEW accessions. Crossing them to a completely durum susceptible genotype represents a first step in the strategy to dissect this trait.

### 4.3 | Effective association mapping in WEW and novel loci for resistance to *Pgt*

The fast LD decay of wild species might facilitate high mapping resolution of loci, but it can also affect the effectiveness of genome-wide association mapping to retrieve useful variability from crop wild relatives. Indeed, a low-range LD requires high genotyping density up to genome resequencing to fully mine genome variation in the wild species. Thus, only three studies, two with yellow rust resistance (Sela et al., 2014; Tene et al., 2022) and the other on heading date (Strejckova et al., 2023), have demonstrated association genetics to be an effective strategy for mapping traits in wild emmer. Our collection was assessed for LD decay, as this is a critical parameter affecting the resolution achievable in GWAS. Fast LD decay was expected as previously showed for relatives of other self-pollinated crop species such as *Hordeum vulgare* subsp. *spontaneum* (Morrell et al., 2005), *Oryza rufipogon* (Li et al., 2009), and *Glycine soja* (Zhou et al., 2015). In our collection, LD decayed within 126 kb, on a genome-wide average, a value close to the 195 kb previously reported for wild emmer based on a dataset of 17,340 nonredundant, genetically, and physically mapped SNP (Maccaferri et al., 2019). However, at the intralocus level, the analysis of the wild emmer homologous to the bread wheat disease resistance gene *Lr10* in a small panel of wild emmer populations from Israel showed even faster LD decay (1–2 kb) (Sela et al., 2011). Genome resequencing of a large representative panel might allow to fully assess the LD pattern of this wheat relative.

The genotyping density of the WEW-GB collection is not perfectly suitable with respect to the low range of LD decay;

thus, a mapping strategy based on GWAS could not be exhaustive of the full genome variation. However, the significance of MTAs of single-locus association models should not be affected. Thus, in this work, we were able to identify eight significant MTAs based on a high stringent threshold of statistical significance. Since resistant genotypes were rare in the *northern* lineage, association analysis was conducted only on the *southern* genotypes, which also removed possible effects of population structure. The eight significant MTAs were on chromosomes 1A, 1B, 2A, 4A, 5B, 6B, and 7A and three of them are associated with resistance to two different races (Table 3). The comparison between these loci and previously reported ones was done by using Svevo v1 genome as common framework. Notably, markers AX-94444583 and AX-94661492 were both associated with resistance to two races, TTKSK and JRCQC, but with higher percentage of explained phenotypic variation against the latter race. The most probable locations of these markers were on chromosomes 1A (AX-94444583) and 1B (AX-94661492) based on BLAST searches of marker sequences against the Zavitan v2 genome (File S2). On the basis of the position on the Svevo v1 genome, the region corresponding to AX-94444583 (43.1 Mb on chromosome 1A) may be novel, since no resistant loci were previously mapped in the proximity. Also considering the high percentage of explained variation, this marker might represent a useful tool to transfer the *Pgt* resistance gene to prebreeding lines. The marker AX-94661492 (37.8 Mb on 1B of the Svevo genome) maps very close to an MTA previously identified at 36.9 Mb based on field phenotyping upon natural inoculum of a collection of *T. durum* genotypes (Letta et al., 2013). Marker AX-94479993 mapped to the short arm of chromosome 2A, at 23.4 Mb on Svevo genome, in a region where *Pgt* resistant loci were previously identified at the seedling stage for races TRTTF, TTTTF, and TTKSK (Letta et al., 2014—21.3 Mb) and at the adult stage in field with a mix of five *Pgt* races (TTKSK, TKTTF, JRCQC, TTTTF, and TRTTF) in a durum wheat collection (Megerssa et al., 2020—21.0 Mb). This region could therefore contain different race-specific resistance genes/alleles, or loci conferring broad resistance. This locus, if unique in this chromosome region, could have originated very early during wheat evolution, as it has been identified in both wild and durum wheat.

Other markers corresponding to putatively novel resistant regions are AX-94637067, mapped to chromosome 2A (646.4 Mb on Zavitan v2 and 637.8 Mb on Svevo v1 genomes), and AX-94430011, for which only the BLAST against the Svevo genome gave results with the first hit at 670.1 Mb. For marker AX-94879817, associated with resistance to race JRCQC on chromosome 6B (35.4 and 28.8 Mb on Zavitan v2 and Svevo v1 genomes, respectively), some correspondence was found with MTAs previously identified at 30.6 and 31.3 Mb and reported again by Megerssa et al. (2020). More complicated is the analysis of marker

AX-94766528. It was mapped to chromosome 7A of Zavitan v2 genome, while two different hits were found on Svevo v1 genome: one on chromosomes 4A (670.3 Mb, *E*-value  $1 \times 10^{-28}$  and percentage of identity 98.59; File S2), and the other on 7A (43.7 Mb, *E*-value  $4 \times 10^{-25}$  and percentage of identity 95.77), with the former chromosome being the best hit. This BLAST result is due to the presence of a chromosomal translocation between these regions of long arm of 7A and short arm of 4A, which is well documented also at the genomic level in both durum wheat and WEW (Avni et al., 2017; Maccaferri et al., 2019). The different result between Svevo v1 and Zavitan v2 genome regions could be due to small differences in the extension of the translocation in the two species. Associations with *Pgt* reaction have been previously mapped to chromosome 7A at 40–54 Mb in durum wheat following infection with races TRTTF and TTKSK in controlled conditions (Letta et al., 2014) and artificial inoculation in the field with a mix of five races (Megerssa et al., 2020). Given the consistency of the position of the four SNPs markers (AX-94766528, AX-94430011, AX-94879817, and AX-94427726) in durum and bread wheat, the region at 670 Mb on chromosome 4A was also considered, and no correspondence was identified with resistance loci previously published. If confirmed on chromosome 4A, such a locus could be therefore novel.

## 5 | CONCLUSIONS

This work characterized the genetic diversity of a comprehensive ex situ collection of WEW (WEW-GB) that represents a worth germplasm resource to mine the genetic diversity for important genetic traits. As a proof of concept, the collection was proven to be effective to identify genetic determinants of major genes for stem rust resistance by association mapping. Future resequencing initiatives of wild emmer germplasm will maximize the potential of mapping of association genetics and hasten the cloning of agriculturally valuable genes.

### AUTHOR CONTRIBUTIONS

**Anna Maria Mastrangelo:** Formal analysis; funding acquisition; investigation; methodology; writing—original draft; writing—review and editing. **Pablo Roncallo:** Formal analysis; investigation; methodology; visualization; writing—original draft; writing—review and editing. **Oadi Matny:** Investigation; methodology; writing—review and editing. **Radim Čegan:** Formal analysis; writing—review and editing. **Brian Steffenson:** Funding acquisition; methodology; writing—review and editing. **Viviana Echenique:** Resources; writing—review and editing. **Jan Šafář:** Resources; writing—review and editing. **Raffaella Battaglia:** Formal analysis; writing—review and editing.

**Delfina Barabaschi:** Methodology; writing—review and editing. **Luigi Cattivelli:** Conceptualization; funding acquisition; supervision; writing—review and editing. **Hakan Özkan:** Conceptualization; resources; writing—review and editing. **Elisabetta Mazzucotelli:** Conceptualization; data curation; formal analysis; funding acquisition; investigation; methodology; supervision; visualization; writing—original draft; writing—review and editing.

### ACKNOWLEDGMENTS

This research was supported by project PRIMA2019-Cerealm “Enhancing diversity in Mediterranean cereal farming systems”; the Czech Science Foundation, grant number 22–00204S; the Horizon Europe program, project “Promoting a Plant Genetic Resource Community for Europe (PRO-GRACE),” project number 101094738; and the Agritech National Research Center funded by the European Union Next-Generation-EU (PIANO NAZIONALE DI RIPRESA E RESILIENZA (PNRR)—MISSIONE 4 COMPONENTE 2, INVESTIMENTO 1.4—D.D. 1032 17/06/2022, CN00000022). This manuscript reflects only the authors’ views and opinions; neither the European Union nor the European Commission can be considered responsible for them.

### CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

### DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories: Graingenes ([https://wheat.pw.usda.gov/GG3/global\\_durum\\_genomic\\_resources](https://wheat.pw.usda.gov/GG3/global_durum_genomic_resources)) and Dryad (<https://doi.org/10.5061/dryad.pzgmbsctb>).

### ORCID

Elisabetta Mazzucotelli  <https://orcid.org/0000-0002-8578-6502>

### REFERENCES

- Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, 19, 1655–1664. <https://doi.org/10.1101/gr.094052.109>
- Anikster, Y., Manisterski, J., Long, D. L., & Leonard, K. J. (2005). Leaf rust and stem rust resistance in *Triticum dicoccoides* populations in Israel. *Plant Disease*, 89(1), 55–62. <https://doi.org/10.1094/PD-89-0055>
- Avni, R., Nave, M., Barad, O., Baruch, K., Twardziok, S. O., Gundlach, H., Hale, I., Mascher, M., Spannagl, M., Wiebe, K., Jordan, K. W., Golan, G., Deek, J., Ben-Zvi, B., Ben-Zvi, G., Himmelbach, A., MacLachlan, R. P., Sharpe, A. G., Fritz, A., ... Distelfeld, A. (2017). Wild emmer genome architecture and diversity elucidate wheat evolution and domestication. *Science*, 357(6346), 93–97. <https://doi.org/10.1126/science.aan0032>

- Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y., & Buckler, E. S. (2007). TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinformatics*, *23*(19), 2633–2635. <https://doi.org/10.1093/bioinformatics/btm308>
- Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., & Lee, J. J. (2015). Second-generation PLINK: Rising to the challenge of larger and richer datasets. *GigaScience*, *4*(1). <https://doi.org/10.1186/s13742-015-0047-8>
- Cleveland, W. S. (1979). Robust locally weighted regression and smoothing scatterplots. *The Journal of the American Statistical Association*, *74*(368), 829–836. <https://doi.org/10.2307/2286407>
- Dadhodaie, N. A., Karaoglou, H., Wellings, C. R., & Parks, R. F. (2011). Mapping genes Lr53 and Yr35 on the short arm of chromosome 6B of common wheat with microsatellite markers and studies of their association with Lr36. *Theoretical Applied Genetics*, *122*, 479–487. <https://doi.org/10.1007/s00122-010-1462-y>
- Doyle, J. J., & Doyle, J. L. (1987). A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin*, *19*, 11–15. [https://doi.org/10.1016/0031-9422\(88\)92985-0](https://doi.org/10.1016/0031-9422(88)92985-0)
- Earl, D. A., & Von Holdt, B. M. (2012). STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, *4*(2), 359–361. <https://doi.org/10.1007/s12686-011-9548-7>
- Excoffier, L., & Lischer, H. E. L. (2010). Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, *10*, 564–567. <https://doi.org/10.1111/j.1755-0998.2010.02847.x>
- Fu, D., Uauy, C., Distelfeld, A., Blechl, A., Epstein, L., Chen, X., Sela, H., Fahima, T., & Dubcovsky, J. (2009). A kinase-START gene confers temperature-dependent resistance to wheat stripe rust. *Science*, *323*, 1357–1360. <https://doi.org/10.1126/science.1166289>
- Fu, Y. B. (2015). Understanding crop genetic diversity under modern plant breeding. *Theoretical Applied Genetics*, *128*, 2131–2142. <https://doi.org/10.1007/s00122-015-2585-y>
- Gao, Y., He, C., Zhang, D., Liu, X., Xu, Z., Tian, Y., Liu, X. H., Zang, S., Pauly, M., Zhou, Y., & Zhang, B. (2017). Two trichome birefringence-like proteins mediate xylan acetylation, which is essential for leaf blight resistance in rice. *Plant Physiology*, *173*(1), 470–481. <https://doi.org/10.1104/pp.16.01618>
- Guo, X., Liu, D., & Chong, K. (2018). Cold signaling in plants: insights into mechanisms and regulation. *Journal of Integrative Plant Biology*, *60*, 745–756. <https://doi.org/10.1111/jipb.12706>
- Hale, I., Zhang, X., Fu, D., & Dubcovsky, J. (2012). Registration of wheat lines carrying the partial stripe rust resistance gene Yr36 without the Gpc-B1 high grain protein content allele. *Journal of Plant Registrations*, *7*(1), 108–112. <https://doi.org/10.3198/jpr2012.03.0150crg>
- He, F., Pasam, R., Shi, F., Kant, S., Keeble-Gagnere, G., Kay, P., Forrest, K., Fritz, A., Hucl, P., Wiebe, K., Knox, R., Cuthbert, R., Pozniak, C., Akhunova, A., Morrell, P. L., Davies, J. P., Webb, S. R., Spangenberg, G., Hayes, B., ... Akhunov, E. (2019). Exome sequencing highlights the role of wild-relative introgression in shaping the adaptive landscape of the wheat genome. *Nature Genetics*, *51*(5), 896–904. <https://doi.org/10.1038/s41588-019-0382-2>
- Huang, L., Raats, D., Sela, H., Klymiuk, V., Lidzbarsky, G., Feng, L., Krugman, T., & Fahima, T. (2016). Evolution and adaptation of wild emmer wheat populations to biotic and abiotic stresses. *Annual Review of Phytopathology*, *54*(1), 279–301. <https://doi.org/10.1146/annurev-phyto-080614-120254>
- Huang, S., & Steffenson, B. (2018). Resistance of *Aegilops longissima* to the rusts of wheat. *Plant Disease*, *102*, 1124–1135. <https://doi.org/10.1094/PDIS-06-17-0880-RE>
- Jin, Y., Singh, R. P., Ward, R. W., Wanyera, R., Kinyua, M., Njau, P., Fetch, T., Pretorius, Z. A., & Yahyaoui, A. (2007). Characterization of seedling infection types and adult plant infection responses of monogenic Sr gene lines to race TTKS of *Puccinia graminis* f. sp. *tritici*. *Plant Disease*, *91*(9), 1096–1099. <https://doi.org/10.1094/PDIS-91-9-1096>
- Jin, Y., Szabo, L. J., Pretorius, Z. A., Singh, R. P., Ward, R., & Fetch, T. (2008). Detection of virulence to resistance gene Sr24 within race TTKS of *Puccinia graminis* f. sp. *tritici*. *Plant Disease*, *92*, 923–926. <https://doi.org/10.1094/PDIS-92-6-0923>
- Jombart, T., Devillard, S., & Balloux, F. (2010). Discriminant analysis of principal components: A new method for the analysis of genetically structured populations. *BMC Genetics*, *11*(1), 94. <https://doi.org/10.1186/1471-2156-11-94>
- Klymiuk, V., Fatiukha, A., Huang, L., Wei, Z., Kis-Papo, T., Saranga, Y., Krugman, T., & Fahima, T. (2019). Durum wheat as a bridge between wild emmer wheat genetic resources and bread wheat in applications of genetic and genomic research in cereals. In T. Miedaner & V. Korzun (Eds.), *Applications of genetic and genomic research in cereals* (pp. 201–230). Woodhead Publishing Series in Food Science, Technology and Nutrition. <https://doi.org/10.1016/B978-0-08-102163-7.00010-7>
- Klymiuk, V., Yaniv, E., Huang, L., Raats, D., Fatiukha, A., Chen, S., Feng, L., Frenkel, Z., Krugman, T., Lidzbarsky, G., Chang, W., Jääskeläinen, M. J., Schudoma, C., Paulin, L., Laine, P., Bariana, H., Sela, H., Saleem, K., Sørensen, C. K., ... Fahima, T. (2018). Cloning of the wheat Yr15 resistance gene sheds light on the plant tandem kinase-pseudokinase family. *Nature Communications*, *9*(1), 3735. <https://doi.org/10.1038/s41467-018-06138-9>
- Knott, D. R., Bai, D., & Zale, J. (2005). The transfer of leaf and stem rust resistance from wild emmer wheats to durum and common wheat. *Canadian Journal of Plant Science*, *85*(1), 49–57. <https://doi.org/10.4141/P03-212>
- Kruskal, W. H., & Wallis, W. A. (1952). Use of ranks in one-criterion variance analysis. *The Journal of the American Statistical Association*, *47*(260), 583–621.
- Letta, T., Maccaferri, M., Badebo, A., Ammar, K., Ricci, A., Crossa, J., & Tuberosa, R. (2013). Searching for novel sources of field resistance to Ug99 and Ethiopian stem rust races in durum wheat via association mapping. *Theoretical Applied Genetics*, *126*, 1237–1256. <https://doi.org/10.1007/s00122-013-2050-8>
- Letta, T., Olivera, P., Maccaferri, M., Jin, Y., Ammar, K., Badebo, A., Salvi, S., Noli, E., Crossa, J., & Tuberosa, R. (2014). Association mapping reveals novel stem rust resistance loci in durum wheat at the seedling stage. *The Plant Genome*, *7*. <https://doi.org/10.3835/plantgenome2013.08.0026>
- Li, X., Tan, L., Zhu, Z., Huang, H., Liu, Y., Hu, S., & Sun, C. (2009). Patterns of nucleotide diversity in wild and cultivated rice. *Plant Systematics and Evolution*, *281*, 97–106. <https://doi.org/10.1007/s00606-009-0191-7>
- Li, Y. C., Röder, M. S., Fahima, T., Kirzhner, V. M., Beiles, A., Korol, A. B., & Nevo, E. (2000). Natural selection causing microsatellite divergence in wild emmer wheat at the ecologically variable microsite at Ammiad, Israel. *Theoretical Applied Genetics*, *100*, 985–999. <https://doi.org/10.1007/s001220051380>

- Luo, M. C., Yang, Z. L., You, F. M., Kawahara, T., Waines, J. G., & Dvorak, J. (2007). The structure of wild and domesticated emmer wheat populations, gene flow between them, and the site of emmer domestication. *Theoretical Applied Genetics*, *114*, 947–959. <https://doi.org/10.1007/s00122-006-0474-0>
- Maccaferri, M., Harris, N. S., Twardziok, S. O., Pasam, R. K., Gundlach, H., Spannagl, M., Ormanbekova, D., Lux, T., Prade, V. M., Milner, S. G., Himmelbach, A., Mascher, M., Bagnaresi, P., Faccioli, P., Cozzi, P., Lauria, M., Lazzari, B., Stella, A., Manconi, A., ... Cattivelli, L. (2019). Durum wheat genome highlights past domestication signatures and future improvement targets. *Nature Genetics*, *51*(5), 885–895. <https://doi.org/10.1038/s41588-019-0381-3>
- Mamo, B. E., Kevin, P. S., Brueggeman, R. S., & Steffenson, B. J. (2015). Genetic characterization of resistance to wheat stem rust race TTKSK in landrace and wild barley accessions identifies the Rpg4/Rpg5 locus. *Phytopathology*, *105*, 99–109. <https://doi.org/10.1094/PHYTO-12-13-0340-R>
- Marais, G. F., Pretorius, Z. A., Wellings, C. R., McCallum, B., & Marais, A. S. (2005). Leaf rust and stripe rust resistance genes transferred to common wheat from *Triticum dicoccoides*. *Euphytica*, *143*, 115–123. <https://doi.org/10.1007/s10681-005-2911-6>
- Mazzucotelli, E., Sciara, G., Mastrangelo, A. M., Desiderio, F., Xu, S. S., Faris, J., Hayden, M. J., Tricker, P. J., Ozkan, H., Echenique, V., Steffenson, B. J., Knox, R., Niane, A. A., Udupa, S. M., Longin, F. C. H., Marone, D., Petruzzino, G., Corneti, S., Ormanbekova, D., ... Bassi, F. M. (2020). The Global Durum Wheat Panel (GDP): An international platform to identify and exchange beneficial alleles. *Frontiers in Plant Science*, *11*, 569905. <https://doi.org/10.3389/fpls.2020.569905>
- McVey, D. V. (1991). Reaction of a group of related wheat species (AABB genome and an AABBDD) to stem rust. *Crop Science*, *31*, 1145–1149. <https://doi.org/10.2135/cropsci1991.0011183X003100050012x>
- McVey, D. V., Long, D. L., & Roberts, J. J. (2002). Races of *Puccinia graminis* in the United States during 1997 and 1998. *Plant Disease*, *86*, 568–572. <https://doi.org/10.1094/PDIS.2002.86.6.568>
- Megerssa, S. H., Ammar, K., Acevedo, M., Brown-Guedira, G., Ward, B., Degete, A. G., Randhawa, M. S., & Sorrells, M. E. (2020). Multiple-race stem rust resistance loci identified in durum wheat using genome-wide association mapping. *Frontiers in Plant Science*, *11*, 598509. <https://doi.org/10.3389/fpls.2020.598509>
- Morrell, P. L., Toleno, D. M., Lundy, K. E., & Clegg, M. T. (2005). Low levels of linkage disequilibrium in wild barley (*Hordeum vulgare* ssp. *spontaneum*) despite high rates of self-fertilization. *Proceedings of the National Academy of Sciences of the United States of America*, *102*, 2442–2447. <https://doi.org/10.1073/pnas.0409804102>
- Nei, M. (1973). Analysis of gene diversity in subdivided populations. *Proceedings of the National Academy of Sciences of the United States of America*, *70*, 3321–3323.
- Nevo, E., Gerechter-Amitai, Z., & Beiles, A. (1991). Resistance of wild emmer wheat to stem rust: Ecological, pathological and allozyme associations. *Euphytica*, *53*, 121–130. <https://doi.org/10.1007/BF00023792>
- Nevo, E., Korol, A. B., Beiles, A., & Fahima, T. (2002). *Evolution of wild emmer and wheat improvement*. Springer.
- Oliveira, H. R., Jacocks, L., Czajkowska, B. I., Kennedy, S. L., & Brown, T. A. (2020). Multiregional origins of the domesticated tetraploid wheats. *PLoS ONE*, *15*(1), e0227148. <https://doi.org/10.1371/journal.pone.0227148>
- Olivera, P., Newcomb, M., Szabo, L. J., Rouse, M., Johnson, J., Gale, S., Luster, D. G., Hodson, D., Cox, J. A., Burgin, L., Hort, M., Gilligan, C. A., Patpour, M., Justesen, A. F., Hovmøller, M. S., Woldeab, G., Hailu, E., Hundie, B., Tadesse, K., ... Jin, Y. (2015). Phenotypic and Genotypic characterization of race TKTF of *Puccinia graminis* f. sp. *tritici* that caused a wheat stem rust epidemic in southern Ethiopia in 2013–2014. *Phytopathology*, *105*, 917–928. <https://doi.org/10.1094/PHYTO-11-14-0302-FI>
- Olivera, P. D., Jin, Y., Rouse, M., Badebo, A., Fetch, T., Jr., Singh, R. P., & Yahyaoui, A. (2012). Races of *Puccinia graminis* f. sp. *tritici* with combined virulence to Sr13 and Sr9e in a field stem rust screening nursery in Ethiopia. *Plant Disease*, *96*, 623–628. <https://doi.org/10.1094/pdis-09-11-0793>
- Ozkan, H., Brandolini, A., Pozzi, C., Effgen, S., Wunder, J., & Salamini, F. (2005). A reconsideration of the domestication geography of tetraploid wheat. *Theoretical Applied Genetics*, *110*, 1052–1060. <https://doi.org/10.1007/s00122-005-1925-8>
- Özkan, H., Willcox, G., Graner, A., Salamini, F., & Kilian, B. (2011). Geographic distribution and domestication of wild emmer wheat (*Triticum dicoccoides*). *Genetic Resources and Crop Evolution*, *58*, 11–53. <https://doi.org/10.1007/s10722-010-9581-5>
- Peakall, R., & Smouse, P. E. (2012). GenAlEx 6.5: Genetic analysis in Excel. Population genetic software for teaching and research—an update. *Bioinformatics*, *28*(19), 2537–2539. <https://doi.org/10.1093/bioinformatics/bts460>
- Peng, J., Zhiyong, L., Xiongjun, L., Yan, J., Sun, D., & Nevo, E. (2021). Evolutionary agriculture domestication of wild emmer wheat. In S. P. Wasser & M. Frenkel-Morgenstern (Eds.), *New horizons in evolution* (pp. 193–255). Academic Press. <https://doi.org/10.1016/B978-0-323-90752-1.00007-9>
- Perrier, X., Flori, A., & Bonnot, F. (2003). Data analysis methods. In P. Hamon, M. Seguin, X. Perrier, & J. C. Glaszmann (Eds.), *Genetic diversity of cultivated tropical plants* (pp. 43–76). Science Publishers.
- Pretorius, Z. A., Singh, R. P., Wagoire, W. W., & Payne, T. S. (2000). Detection of virulence to wheat stem rust resistance gene Sr31 in *Puccinia graminis* f. sp. *tritici* in Uganda. *Plant Disease*, *84*(2), 203. <https://doi.org/10.1094/pdis.2000.84.2.203b>
- Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, *155*, 945–959. <https://doi.org/10.1093/genetics/155.2.945>
- R Development Core Team. (2011). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.
- Ren, J., Chen, L., Sun, D., You, F. M., Wang, J., Peng, Y., Nevo, E., Beiles, A., Sun, D., Luo, M. C., & Peng, J. (2013). SNP-revealed genetic diversity in wild emmer wheat correlates with ecological factors. *BMC Evolutionary Biology*, *13*, 169. <https://doi.org/10.1186/1471-2148-13-169>
- Roelfs, A. P., & Martens, J. W. (1988). An international system of nomenclature for *Puccinia graminis* f. sp. *tritici*. *Phytopathology*, *78*, 526–533.
- Saccomanno, A., Matny, O., Marone, D., Laidò, G., Petruzzino, G., Mazzucotelli, E., Desiderio, F., Blanco, A., Gadaleta, A., Pecchioni, N., De Vita, P., Steffenson, B., & Mastrangelo, A. M. (2018). Genetic mapping of loci for resistance to stem rust in a tetraploid wheat collection. *International Journal of Molecular Sciences*, *19*, 3907. <https://doi.org/10.3390/ijms19123907>
- Sela, H., Ezrati, S., Ben-Yehuda, P., Manisterski, J., Akhunov, E., Dvorak, J., Breiman, A., & Korol, A. (2014). Linkage disequilibrium and association analysis of stripe rust resistance in wild emmer wheat

- (*Triticum turgidum* ssp. *dicoccoides*) population in Israel. *Theoretical Applied Genetics*, 127, 2453–2463. <https://doi.org/10.1007/s00122-014-2389-5>
- Sela, H., Loutre, C., Keller, B., Schulman, A., Nevo, E., Korol, A., & Fahima, T. (2011). Rapid linkage disequilibrium decay in the Lr10 gene in wild emmer wheat (*Triticum dicoccoides*) populations. *Theoretical Applied Genetics*, 122(1), 175–187. <https://doi.org/10.1007/s00122-010-1434-2>
- Stakman, E. C., Stewart, P. M., & Loegering, W. (1962). *Identification of physiologic races of Puccinia graminis var. tritici*. US Department of Agriculture, Agricultural Research Service.
- Steffenson, B. J., Jin, Y., Brueggeman, R. S., Kleinhofs, A., & Sun, Y. (2009). Resistance to stem rust race TTKSK maps to the rpg4/Rpg5 complex of chromosome 5H of barley. *Phytopathology*, 99, 1135–1141. <https://doi.org/10.1094/PHYTO-99-10-1135>
- Strejčková, B., Mazzucotelli, E., Čegan, R., Milec, Z., Brus, J., Çakır, E., Mastrangelo, A. M., Özkan, H., & Šafář, J. (2023). Wild emmer wheat, the progenitor of modern bread wheat, exhibits great diversity in the VERNALIZATION1 gene. *Frontiers in Plant Science*, 13, 1106164. <https://doi.org/10.3389/fpls.2022.1106164>
- Swarup, S., Cargill, E. J., Crosby, K., Flagel, L., Kniskern, J., & Glenn, K. C. (2021). Genetic diversity is indispensable for plant breeding to improve crops. *Crop Science*, 61, 839–852. <https://doi.org/10.1002/csc2.20377>
- Tene, M., Adhikari, E., Cobo, N., Jordan, K. W., Matny, O., del Blanco, I. A., Roter, J., Ezrati, S., Govta, L., Manisterski, J., Yehuda, P. B., Chen, X., Steffenson, B., Akhunov, E., & Sela, H. (2022). GWAS for stripe rust resistance in wild emmer wheat (*Triticum dicoccoides*) population: Obstacles and solutions. *Crops*, 2(1), 42–61. <https://doi.org/10.3390/crops2010005>
- The, T. T., Nevo, E., & McIntosh, R. A. (1993). Responses of Israeli wild emmers to selected Australian pathotypes of *Puccinia* species. *Euphytica*, 71, 75–81.
- Wang, Z., Xie, J., Guo, L., Zhang, D., Li, G., Fang, T., Chen, Y., Li, J., Wu, Q., Lu, P., Li, M., Wu, H., Zhang, H., Zhang, Y., Yang, W., Luo, M., Tzion, F., & Liu, Z. (2018). Molecular mapping of YrTZ2, a stripe rust resistance gene in wild emmer accession TZ-2 and its comparative analyses with *Aegilops tauschii*. *Journal of Integrative Agriculture*, 17(6), 1267–1275. [https://doi.org/10.1016/S2095-3119\(17\)61846-X](https://doi.org/10.1016/S2095-3119(17)61846-X)
- Yadav, I. S., Singh, N., Wu, S., Raupp, J., Wilson, D. L., Rawat, N., Gill, B. S., Poland, J., & Tiwari, V. K. (2023). Exploring genetic diversity of wild and related tetraploid wheat species *Triticum turgidum* and *Triticum timopheevii*. *Journal of advanced research*, 48, 47–60. <https://doi.org/10.1016/j.jare.2022.08.020>
- Yaniv, E., Raats, D., Ronin, Y., Korol, A. B., Grama, A., Bariana, H., Dubcovsky, J., & Schulman, A. H. (2015). Evaluation of marker-assisted selection for the stripe rust resistance gene Yr15, introgressed from wild emmer wheat. *Molecular Breeding*, 35, 43. <https://doi.org/10.1007/s11032-015-0238-0>
- Zhang, D., Bowden, R. L., Yu, J., Carver, B. F., & Bai, G. (2014). Association analysis of stem rust resistance in U.S. winter wheat. *PLoS ONE*, 9(7), e103747. <https://doi.org/10.1371/journal.pone.0103747>
- Zhang, Y., Hu, X., Islam, S., She, M., Peng, Y., Yu, Z., Wylie, S., Juhasz, A., Dowla, M., Yang, R., Zhang, J., Wang, X., Dell, B., Chen, X., Nevo, E., Sun, D., & Ma, W. (2018). New insights into the evolution of wheat avenin-like proteins in wild emmer wheat (*Triticum dicoccoides*). *Proceedings of the National Academy of Sciences of the United States of America*, 115(52), 13312–13317. <https://doi.org/10.1073/pnas.1812855115>
- Zhou, Z., Jiang, Y., Wang, Z., Gou, Z., Lyu, J., Li, W., Yu, Y., Shu, L., Zhao, Y., Ma, Y., Fang, C., Shen, Y., Liu, T., Li, C., Li, Q., Wu, M., Wang, M., Wu, Y., Dong, Y., ... Tian, Z. (2015). Resequencing 302 wild and cultivated accessions identifies genes related to domestication and improvement in soybean. *Nature Biotechnology*, 33, 408–414. <https://doi.org/10.1038/nbt.3096>
- Zhu, T., Wang, L., Rimbart, H., Rodriguez, J. C., Deal, K. R., & Oliveira, D. (2021). Optical maps refine the bread wheat *Triticum aestivum* cv. Chinese Spring genome assembly. *The Plant Journal*, 107, 303–314. <https://doi.org/10.1111/tj.15289>
- Zhu, T., Wang, L., Rodriguez, J. C., Deal, K. R., Avni, R., Distelfeld, A., McGuire, P. E., Dvorak, J., & Luo, M. C. (2019). Improved genome sequence of wild emmer wheat Zavitan with the aid of optical maps. *G3: Genes, Genomes, Genetics*, 9(3), 619–624. <https://doi.org/10.1534/g3.118.200902>

## SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Mastrangelo, A. M., Roncallo, P., Matny, O., Čegan, R., Steffenson, B., Echenique, V., Šafář, J., Battaglia, R., Barabaschi, D., Cattivelli, L., Özkan, H., & Mazzucotelli, E. (2024). A new wild emmer wheat panel allows to map new loci associated with resistance to stem rust at seedling stage. *The Plant Genome*, e20413. <https://doi.org/10.1002/tpg2.20413>