



# OPEN Genome-wide association study for resistance to *Macrophomina phaseolina* in maize (*Zea mays* L.)

Gizem Oder<sup>1</sup>, Semiha Yuceer<sup>2</sup>, Canan Can<sup>3</sup>, Muhammed Bahattin Tanyolac<sup>1</sup> & Duygu Ates<sup>1</sup>✉

Maize (*Zea mays* L.) is a frequently used food source in human and animal nutrition. *Macrophomina phaseolina* is a fungal pathogen causing charcoal rot disease in many plants, especially maize. This pathogen causes high yield losses in maize. The development of resistant maize genotypes is of great importance in controlling this disease. In this study, the population structure of 120 different maize genotypes with varying levels of disease resistance was determined and genome-wide association studies were performed. Each genotype was subjected to the pathogen under controlled conditions and their phenotypic responses to the disease were analyzed. Afterwards, single nucleotide polymorphisms were determined by DArT-seq sequencing. After filtering the SNP data, 37,470 clean SNPs were obtained. The population structure was analyzed with STRUCTURE software, and it was determined that the population was divided into two subgroups. The relationship between phenotypic and genotypic data was analyzed using the MLM (Q + K) model in TASSEL software. As a result, seven SNPs markers located on four different chromosomes were associated with disease resistance. The related markers can be used in the future for the development of maize varieties resistant to *M. phaseolina* by marker-assisted selection.

**Keywords** *Zea mays* L., GWAS, Maize, *Macrophomina phaseolina*, Resistance, SNP

Maize (*Zea mays* L. or corn) is a cross-pollinated, annual, and diploid ( $2n=20$ ) member of Poaceae family<sup>1</sup> and has an average genome size of 2.4 Gb<sup>2</sup>. Maize is frequently used in human and animal nutrition due to its high content of vitamins A, B, and C, carbohydrates, protein, and phytochemicals<sup>3,4</sup>. In 2022, maize ranks as the second most produced cereal crop, covering a cultivation area of 43 million hectares globally. It is cultivated in 164 countries, where the annual harvest comes to the amount of about 1.16 billion tons<sup>5</sup>. Türkiye is one of the world's important maize producers with approximately one million hectares and a production of 8.5 million tons annually. Considering the size of maize production worldwide, yield losses can have substantial economic consequences. Nevertheless, several factors negatively affect maize production with some of the major factors being diseases and pests.

Plant diseases caused by fungal infections are the most important pathogenic agents. *Macrophomina phaseolina* (*Mp*) is a broad-host fungi that can infect hundreds of different plants and induce a root cancer called charcoal rot in its host<sup>6</sup>. This pathogen is one of the most destructive diseases that can cause severe yield losses in maize plants and lead to 25–33% or more crop losses. Besides causing plant mortality, it adversely impacts the quality attributes of the product. A recent study observed that high disease severity caused by *Mp* caused a 30–70% decrease in the biophysical and biochemical properties of plants<sup>7</sup>. This highlights the significant adverse economic impact of the pathogen on plants, primarily due to reductions in both yield and quality. *Mp* infection can cause mortality both before and after plant emergence. The most prominent symptom is 'blackleg' due to sclerotia at the plant crown. Infected plants are characterized by lesions with dark margins and light grey centres<sup>6</sup>, wilted appearance and discoloured leaves<sup>8</sup>. Furthermore, the phytotoxic metabolites released by *Mp* cause vascular obstructions that impede nutrient transport, leading to plant death. The ability of this pathogen to persist in soil for long periods and its wide genetic diversity contributes to its widespread in different climatic conditions<sup>9</sup>.

The development of resistant cultivars to control *Mp* has become a critical requirement due to the limitations of current control methods. Chemical control can lead to resistance development of the pathogen and environmental problems; in addition, the effectiveness of systemic fungicides is limited due to their inability to reach the roots. Physical control methods require high cost and labour, which negatively affects agricultural

<sup>1</sup>Department of Bioengineering, Ege University, Izmir, Turkey. <sup>2</sup>Phytopathology Department, Biological Control Research Institute, Adana, Turkey. <sup>3</sup>Department of Biology, Gaziantep University, Gaziantep, Turkey. ✉email: duygu.ates@ege.edu.tr

productivity<sup>10</sup>. Although agronomic practices aim to reduce the density of microsclerotia, some methods such as irrigation are not effective<sup>11,12</sup>. Biological control varies depending on environmental conditions and does not always give reliable results. Although plant defense stimulants trigger defense mechanisms, their effects vary depending on genotype and environmental conditions, and in some cases may negatively affect other plant traits<sup>13</sup>. Consequently, due to the limitations of current methods, the development of resistant cultivars appears to be the most effective and sustainable strategy for permanent control of *Mp*.

Conventional breeding methods require many years to develop disease-resistant varieties in plants. However, modern molecular techniques such as genome-wide association studies (GWAS) accelerate this process and make significant contributions to breeding studies. GWAS reveals the relationship between genetic variants and plant traits using genotypic and phenotypic data. Since this technique directly identifies the genetic factors affecting phenotypic variation, it provides more precise and efficient results compared to traditional methods<sup>14</sup>. In particular, in previous studies, diseases such as Fusarium-induced ear rot<sup>15</sup>, *Aspergillus flavus* infection<sup>16</sup>, *Phytophthora blight*-induced stalk rot<sup>17</sup> and maize rough dwarfism<sup>18</sup> were investigated by GWAS and resistance genes were identified. In addition, GWAS has also been used in the development of resistant varieties against the western corn rootworm<sup>19</sup>. Thus, it enables a better understanding the genetic background of complex traits such as disease resistance. By measuring the contribution of specific genetic variants to a trait, GWAS accelerates the identification of genetic markers to be used in breeding programs. This allows the development of resistant varieties in a shorter time.

One of the technologies used to provide genotypic data for GWAS is Diversity Array Technology (DARt). DARt detects thousands of single nucleotide polymorphism (SNP) markers quickly and efficiently without the need for whole genome sequencing<sup>20</sup>. SNP markers produced by DARt are used for mapping gene loci that control phenotypic traits and for the identification of genetic variants. Therefore, DARt and similar technologies make breeding programs more effective by providing genetic data to GWAS studies.

This study aimed to perform GWAS analysis to identify molecular markers for *Mp* resistance using 120 different maize genotypes. The research aims to identify genetic variations contributing to *Mp* resistance among different genotypes and specific SNPs associated with these variations. The identification of these molecular markers will facilitate future studies by providing an understanding of the genetic basis of *Mp* resistance. Furthermore, the findings will contribute to the development of breeding programs for the development of resistant varieties, ultimately strengthening plant protection strategies against *Mp* infection and contributing to sustainable agricultural practices.

## Results

### Phenotypic variations of the population

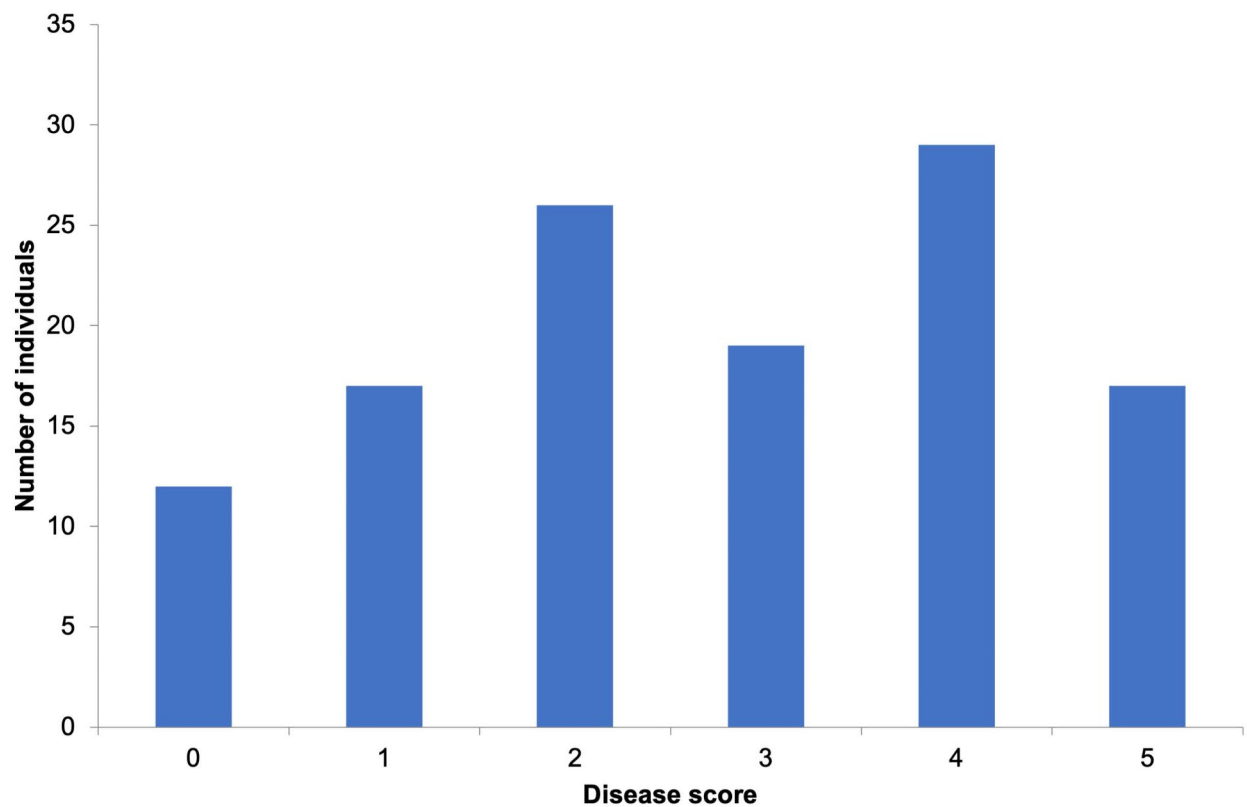
The resistance to *Mp* for 120 different maize genotypes under controlled conditions was evaluated and disease symptoms were scored. The average disease score was calculated as 4.2. Out of 120 genotypes, 12 were classified as resistant, 43 as moderately resistant/tolerant, 48 as moderately susceptible/susceptible, and 17 as highly susceptible. As a result of the examination of the disease response distribution, it was determined that the population used exhibited a normal distribution suitable for the GWAS study. The normal distribution of genotypes is depicted in the histogram graph in Fig. 1. Analysis of variance of *Mp* disease scores reveals significant effects of maize genotypes on disease severity. The results of the analysis of variance are presented in Table 1. Analysis of variance showed that the resistance reactions to *Mp* among the 120 maize genotypes were statistically significant at  $P \leq 0.05$  level which shows that there are differences among the genotypes in terms of reaction to the pathogen infection. ANOVA assumptions, including homogeneity of variances and normality, were tested and confirmed to be accurate. In addition, the broad sense heritability score was calculated as 0.80.

### Population structure analysis

A total of 79,166 SNPs were obtained through DARt analysis. The initial SNP data were subjected to strict filtration criteria allowing only SNPs with 95% call rate and 5% minor allele frequency (MAF) to remain, resulting in 37,470 high-quality SNPs. This process was performed to ensure the reliability of subsequent analyses. SNPs are uniformly distributed among chromosomes, with an average of 2,500 SNPs per chr. The highest SNP number was found in chr 1 with 4,129 SNPs and the lowest SNP number was found in chr 10 with 1,639 SNPs. Furthermore, the polymorphism information (PIC) content values of SNPs ranged from 0.01 to 0.5, with an average of 0.31. This indicates a moderate level of genetic diversity within the population. Following filtration, clean SNP data were utilized in STRUCTURE analyses with 50,000 iterations. The  $\Delta K$  value was determined by conducting calculations for each group ranging from 1 to 10 using STRUCTURE HARVESTER software, with the peak observed at  $K = 2$  (Fig. 2). The population was divided into two groups: population 1 (green) and population 2 (red), as shown in Fig. 3.

### Association analysis and candidate genes

GWAS analysis was carried out in TASSEL software. From the option, the MLM (Q+K) model was chosen to analyze associations between marker and trait data. The MLM (Q+K) model decreases the false positive rate in analysis by accounting for both population structure and genetic linkages. This model adjusts the population structure using the Q matrix and genetic relationships through the K matrix, thereby yielding more precise results<sup>21</sup>. According to the GWAS analysis, seven SNPs were detected along the four different chromosomes exceeding the FDR corrections. FDR corrections ( $-\log_{10}P$  value is  $\geq 3.28$ ) were utilized to filter the most likely SNPs to avoid false positives. The Manhattan plot illustrates the distribution of SNPs across various chromosomes and identifies significant SNPs concentrated in specific regions (Fig. 4). After the filtering process, seven SNP markers associated with resistance were identified. The  $-\log_{10}P$ ,  $R^2$ , and PIC values of the SNPs found were used to evaluate the significance and effect of these markers (Table 2). The  $-\log_{10}P$  value indicates the



**Fig. 1.** Disease score distribution for the 120 maize genotypes based on 0–5 scores.

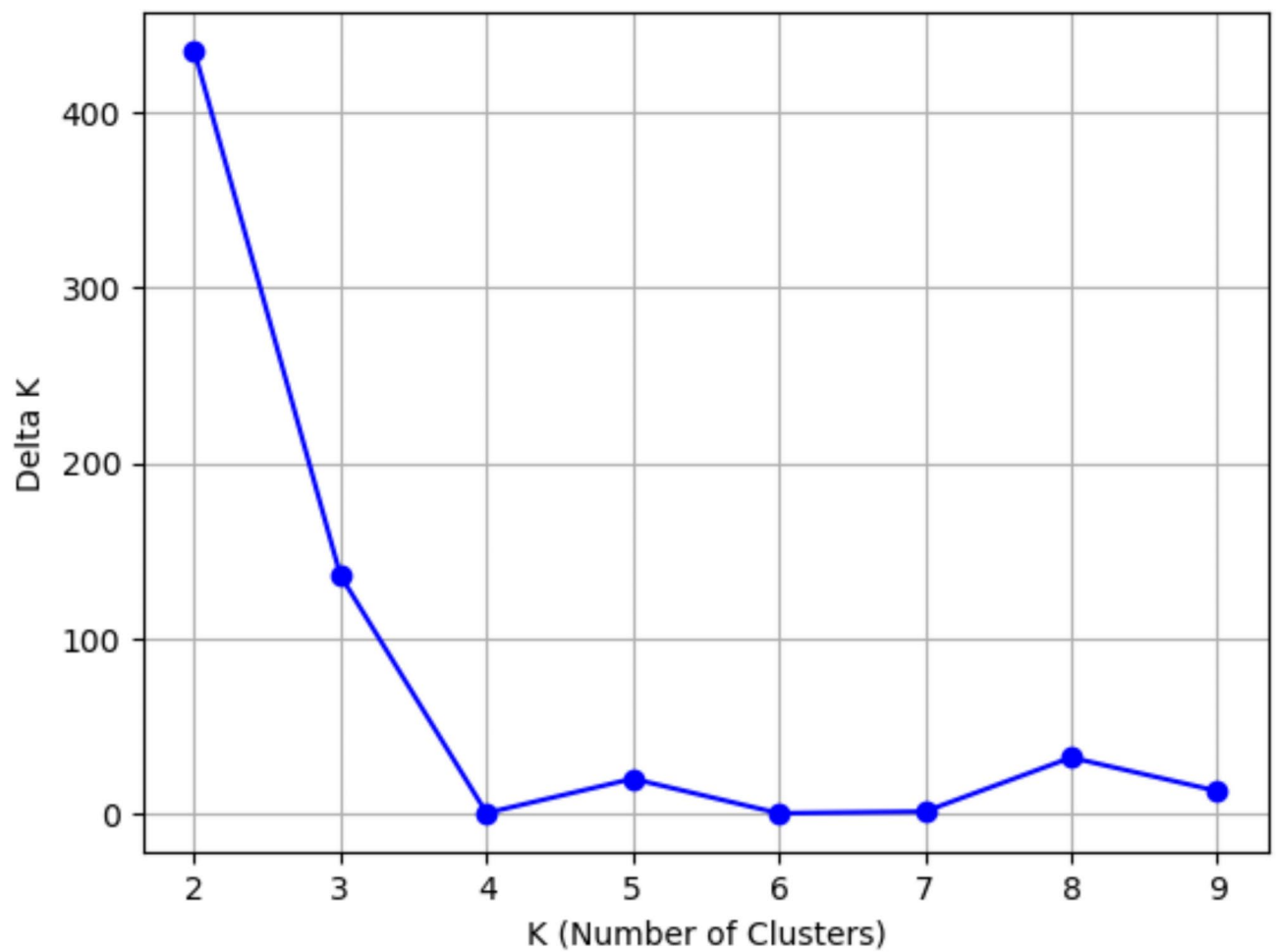
	Sum of squares	df	Mean square	F	Sig.
Genotype	972.62	119	8.17	13.31	1.29
Error	147.33	240	0.61		
Total	1119.95	359			

**Table 1.** Analysis of variance results of disease severity and genotype. *df*: degree of freedom; F: F-value; Sig. significance.

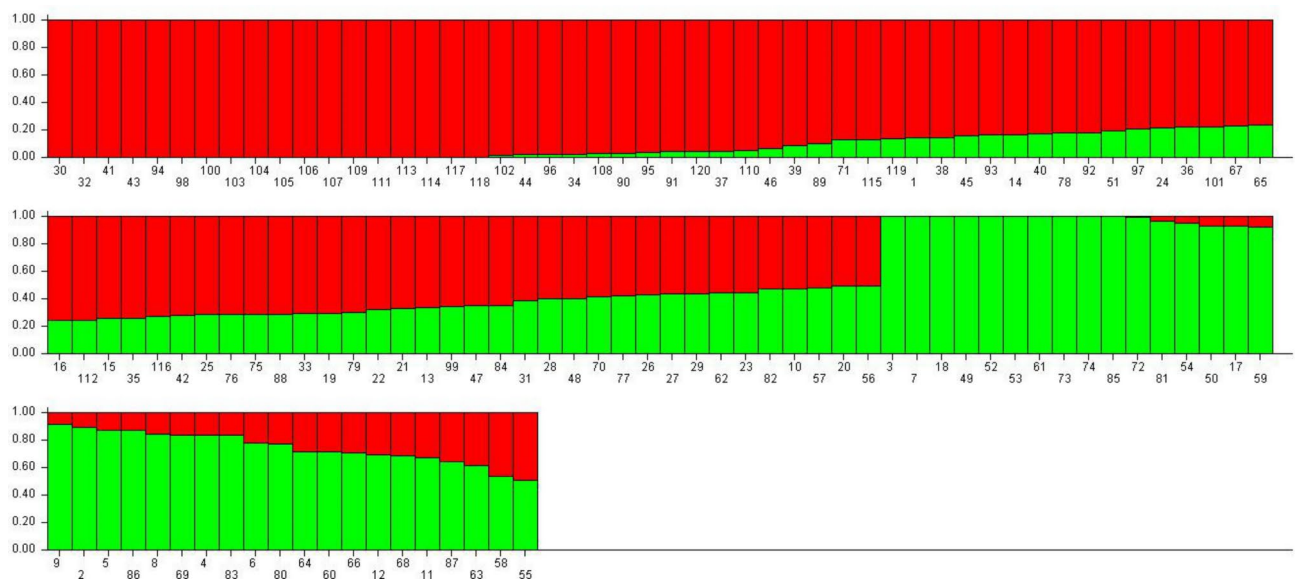
statistical significance of the SNP's association with the phenotype, while the  $R^2$  value reflects the proportion of phenotypic variation explained by this SNP. The PIC value expresses the SNP's capacity to reflect genetic variation. SNP6999, located on chr 2, was found to be the most associated marker with the highest  $-\log_{10} P$  value (3.70). It was followed by SNP12080 and SNP34600 on chr 8 ( $-\log_{10} P$ , 3.61 and 3.60, respectively). As seen in the Q-Q plot, most data points are positioned on the diagonal line, indicating that the observed  $P$  values show a good fit to the expected distribution. This fit suggests that the  $P$  values were appropriately adjusted for multiple testing and that the analysis was robust to control for potential false positives (Fig. 5). Candidate genes were identified by screening up and down 100,000 base pairs of SNP locations, and 33 different genes with various activities were found. These genes are involved in disease resistance and play important roles in plant defence mechanisms such as cell wall synthesis, phytohormone signalling and activation of transcription factors. They also show the potential to alter cell wall composition, regulate stress responses and increase resistance to fungal pathogens. The candidate genes associated with the SNP markers and their functions are given in Supplementary Table S2.

## Discussion

*Mp* is one of the most destructive pathogens of maize production, causing major losses in maize yields. It causes yield loss not only in maize but also in soya bean<sup>22,23</sup>, canola<sup>24</sup>, sorghum<sup>25,26</sup> and other plants<sup>27–29</sup>. For this reason, many studies have been carried out to determine the genetic factors that provide resistance in various plants. Statistical analysis of the disease resistance data obtained because of phenotypic observations revealed that there was significant diversity among the genotypes used in the study and that the data were normally distributed. However, environmental factors and growth conditions were reported to have a significant effect on resistance



**Fig. 2.**  $\Delta K$  plot obtained from STRUCTURE HARVESTER software.



**Fig. 3.** The population structure of the 120 maize genotypes was analyzed based on 37,470 DArT-seq markers using STRUCTURE.

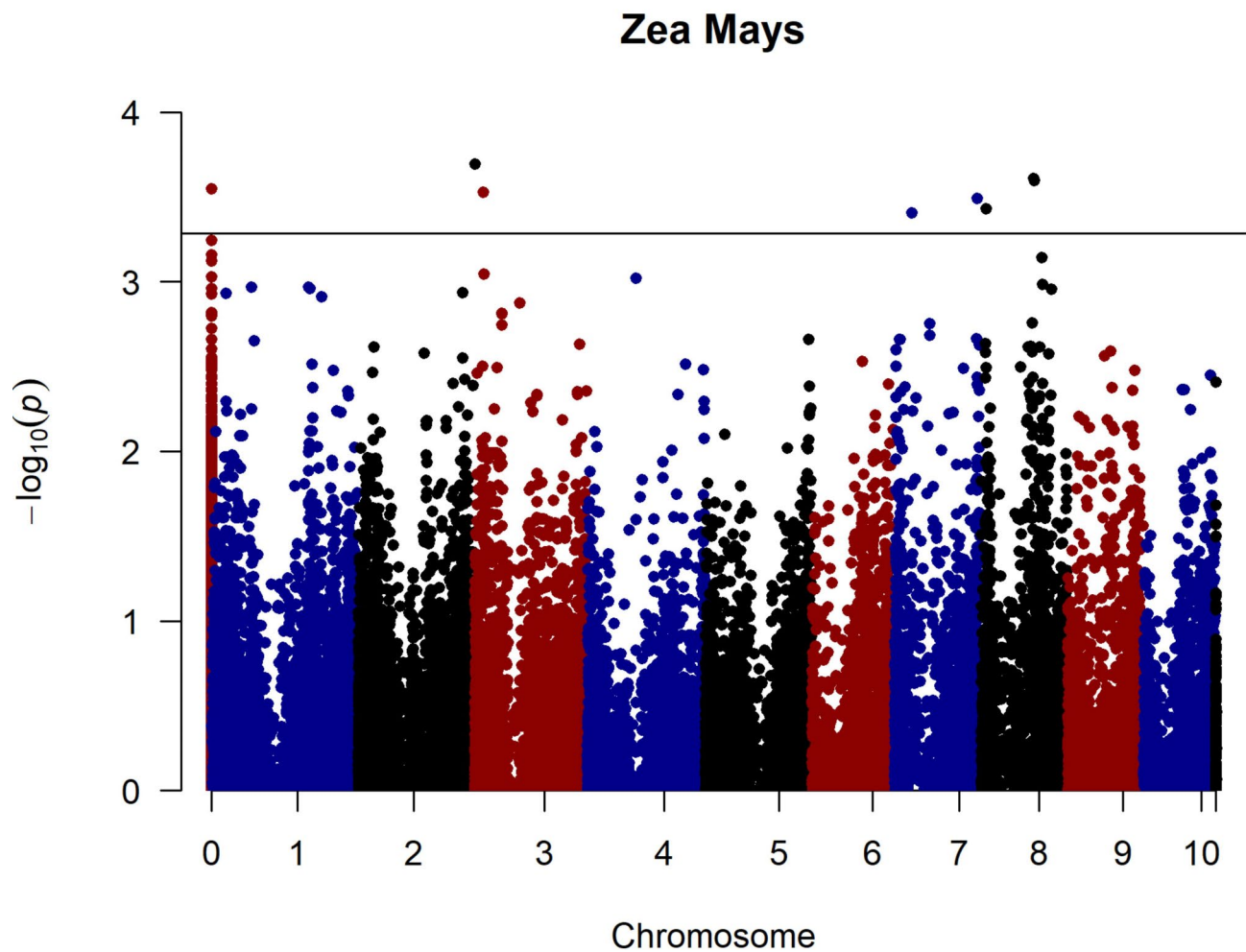


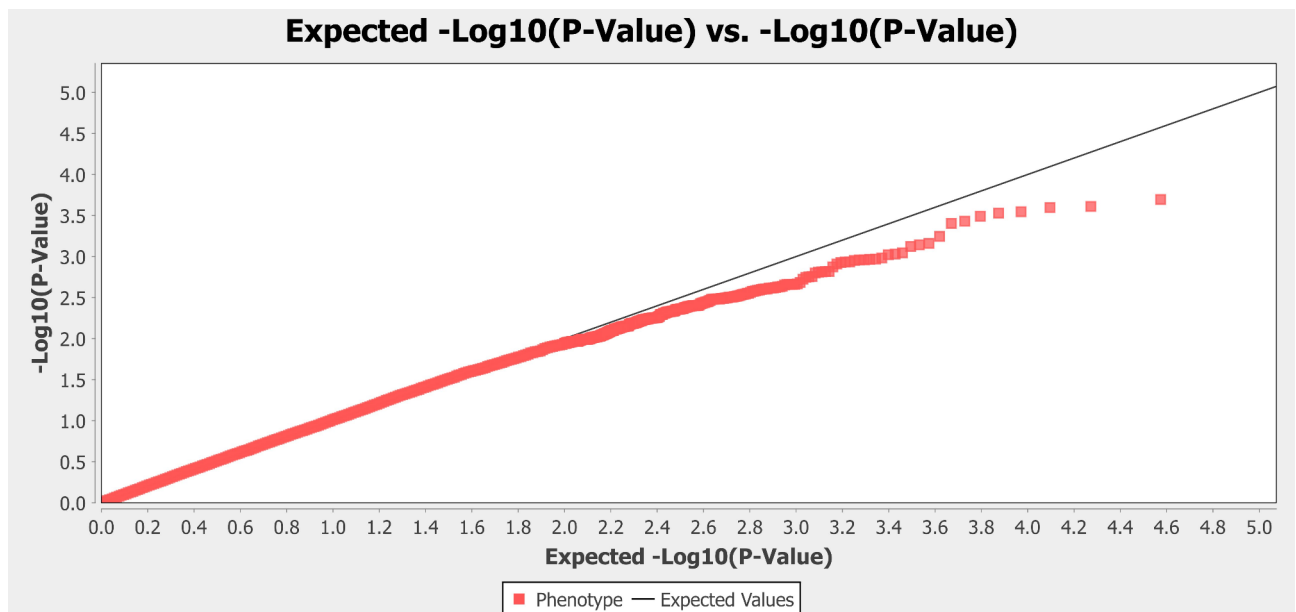
Fig. 4. Manhattan plot of *Mp* resistance in 120 maize genotypes.

Marker	Chromosome	Position	$-\log_{10}P$	$R^2$	PIC values
SNP6999	2	244,407,872	3.70	0.12	0.27
SNP12080	8	109,468,297	3.61	0.12	0.26
SNP34600	8	110,434,136	3.60	0.11	0.27
SNP22096	3	17,631,976	3.53	0.11	0.25
SNP892	7	173,715,255	3.49	0.11	0.32
SNP23394	8	9,998,386	3.43	0.11	0.24
SNP26238	7	36,306,357	3.40	0.11	0.49

Table 2. The list of SNP markers associated with *Mp* resistance.

to fungal infections<sup>30</sup>. In this study, variation between genotypes was analyzed by ANOVA on phenotypic data and this variation was attributed to genetic factors. Broad sense heritability indicates the proportion of total phenotypic variation attributable to genetic factors and was found to be 0.80 in this study; therefore, it can be concluded that broad sense heritability demonstrates that the disease resistance trait is generally inherited genetically. This value was higher than those reported in the literature. The heritability value was determined as 0.54–0.67 in maize<sup>31</sup> and approximately 0.6 in soybean<sup>32</sup> in GWAS studies for resistance to *Mp*. These values indicated that resistance to *Mp* is highly heritable. Therefore, the high heritability of disease resistance is an indication of the reproducibility of the study and the reliability of its use in breeding studies.

In this study, the MLM (Q + K) model is preferred to correct the population structure. This model effectively eliminates potential false positives by taking into account random and fixed effects according to Henderson's notation and provides precision in analysing complex traits<sup>33</sup>. Previous research has revealed that between the observed and expected  $-\log_{10}P$  values shown by the Q-Q plot, the G-test model exhibits the highest deviation, followed by the general linear model (GLM or G + Q model). However, the MLM model shows the least bias



**Fig. 5.** Quantile-quantile plot for the *Mp* resistance.

as it corrects for both population structure (Q) and relatedness (K)<sup>31</sup>. Furthermore, FDR (False Discovery Rate) correction was applied to reduce false positive associations. These methods are critical for improving the reliability of results in large-scale GWAS studies.

In the present study, a GWAS study was performed with a total of 37,470 SNPs by filtering SNPs obtained by the DArT-seq. As a result of the analyses, seven SNPs were identified with  $-\log_{10}P$  values ranging from 3.70 to 3.40, indicating the statistical significance of the effect of SNPs on phenotype. In addition,  $R^2$  values, which indicate how much of the phenotypic variance is explained by genotypes, were found to vary between 0.12 and 0.11. Considering the genetic basis of complex characteristics such as disease resistance, a single SNP is not expected to explain a large proportion of the phenotypic variance. Therefore,  $R^2$  values around 0.1 can be predicted and the results obtained in this study are under the polygenic nature of complex traits. According to a study conducted in soybeans, 19 out of 35,683 clean SNPs obtained with Illumina Infinium SoySNP50K BeadChip were detected to be associated with the *Mp*<sup>32</sup>. In a different study, 3,780 clean SNP data in soybeans were also obtained using the same method. Among the SNPs obtained, six SNPs with  $R^2$  values ranging between 9.5 and 11.8% were reported to be associated with resistance<sup>34</sup>. These  $R^2$  values align with the findings of the current study, supporting the consistency and reliability of the results. However, only one GWAS study in maize has been identified so far. In the study in which 396 maize varieties adapted to Asia were used, filtering was performed according to call rate  $> 0.7$  and MAF  $> 0.05$ , resulting in 296 K clean SNPs obtained by Genotyping by Sequencing (GBS) method<sup>31</sup>. In the present study, a more stringent call rate filter ( $> 0.95$ ) was used to minimize false positives. A total of 19 SNP markers associated with disease resistance, with  $-\log_{10}P$  values ranging from 5.23 to 4.32, were identified. While the reported  $-\log_{10}P$  values in that study were statistically more significant than those observed here, lower heritability values (ranging from 0.67 to 0.54) were reported, likely due to the field conditions under which the study was conducted. Environmental factors beyond control may have negatively impacted the heritability estimates<sup>30</sup>. In contrast, the controlled conditions used in this study resulted in higher heritability values. Similar to the current study, most SNPs were identified on chromosomes 3, 6, and 8<sup>31</sup>. Additionally, for the first time, genomic regions associated with *Mp* resistance on chromosome 7 in maize were identified in this study.

Gene loci likely to be associated with disease resistance were identified through association mapping. However, candidate genes can only be predicted by association mapping studies, and further research is needed to determine the role of these genes in the resistance mechanism<sup>35</sup>. SNP6999, located on chr 2, is localized between several gene regions such as protein kinases, transcription factors, and ATP-DNA binders. The defense mechanisms of plants against pathogens are quite diverse. Among these, phytohormones such as salicylic acid and ethylene play an important role. The gene LOC100502174 has not yet been characterized in the maize genome, but the protein it expresses has an AP2/ERF domain. Ethylene-containing elements, also known as Ethylene-Responsive Elements (ERE), induce the expression of some plant defense-related genes<sup>36</sup>. ERE binding factors (ERF) are transcription factors that recognize ERE sites. In previous studies, the association of ERF sites with resistance to *Mp* has been demonstrated<sup>37–39</sup>. Moreover, the most basic defense mechanism of plants is the cell wall. Cell walls are composed of cellulose, hemicellulose, and pectins. The LOC103648302 gene is involved in the synthesis of xyloglucan (hemicellulose), an important component of the cell wall. In this regard, it may be related to disease resistance.

Only a small number of genes are physically close to SNP 22,096. The gene LOC100272653 is the most important of these. This region affects the morphology and physiology of the cell wall and contains domains

of pectin methyltransferase. Cell respiration and photosynthetic efficiency can be changed by pectin methyltransferase. It can also change the plasticity of cell walls<sup>40,41</sup>. This gene is likely to contribute to fungal pathogen resistance by modifying the cell wall<sup>40</sup>.

SNP 892 and SNP 26,238 were found to be close to nine different gene regions, including various protein kinases, transmembrane proteins, and transcription factors. Prior research has indicated that ubiquitinated proteins<sup>42</sup> and protein kinases<sup>39</sup> are involved in host defense. Moreover, pectin methylesterase activity was detected in LOC100274042, which hardens and consolidates the cell wall against various pathogens and generates damage signals to activate defense mechanisms<sup>43</sup>.

The chr 8 was found to have a significant association with resistance to *Mp*. This region included three different SNPs that indicate ten different gene regions. SNP 12,080 was found to be close to the *mbd101* gene, which expresses zinc-finger regions with proven activity in plant defense<sup>44,45</sup>, and to the cytochrome P450 gene<sup>46</sup>, which is involved in secondary metabolite synthesis and oxidation and reduction reactions. SNP 23,394 was located generally close to genomic regions containing various kinases such as serine/threonine, cysteine, and rust resistance. All these kinases are associated with polysaccharides in the cell wall and thus contribute to the cell wall defense mechanism<sup>47</sup>.

## Conclusion

The genetic background of *Mp* resistance in maize was investigated by GWAS. In this study, SNP markers were generated using DArT-seq technology for the first time among the association studies for *Mp*. In GWAS analyses, seven SNPs were found to be significantly associated. Several candidate gene regions were identified on chromosomes 2, 3, 7, and 8 that contribute to disease resistance in different ways. The SNP markers associated with resistance to *Mp* identified in this study could be used in marker-assisted selection (MAS), including MAS backcrossing and MAS for quantitative traits, to select lines resistant to *Mp* disease in future maize breeding programs. Moreover, 12 maize genotypes found to be resistant to the disease have the potential could be used as donor parents in breeding programs.

## Materials and methods

### Plant material and DNA isolation

The plant materials were kindly received from Polen Seed Co. in Türkiye. A list of 120 genotypes that were used listed in Supplementary Table S1. Liquid nitrogen was used to freeze fresh leaves that were obtained from the genotypes and weighed 0.1 g. After that, they were grounded to a fine powder using a tissue lyser (Technogene, Türkiye). The Qiagen DNA isolation kit (Catalogue No./ID: 69181) was applied to extract DNA from the powdered leaves. The quality of the extracted DNA was examined on an agarose gel, and the quantity was measured using a Nanodrop One (Thermo Sci. Co.). The isolates of the *Mp* were collected from previous studies and pathogenicity was determined, and then the most aggressive pathogen was used for the determination of genotype reaction to *Mp* resistance. Following a 7-day incubation period of the fungus in potato dextrose agar medium, ten agar disks were added to each bottle containing oat medium. The bottles were then incubated at  $25 \pm 2$  °C for 15–20 days. After the completion of incubation, a mixture of composted animal manure, sterilized (45 min at 121 °C) soil, and fine sand (2:1:1) was combined with 100 g of inoculum per 1 kg of soil and the mixture was then incubated for five days.

### Inoculum preparation and phenotypic analysis

For the genotype reaction studies, the fungal inoculum was first developed by filling 500 ml glass bottles with a mixture of sand-water-cornmeal (9:2:1) or bran. Bottles with media were sterilized in an autoclave at 121 °C for 1 h on 2 consecutive days. Fungal isolates and sclerotia, developed in PDA, were placed in each bottle in the form of 10 agar discs, 5 mm in diameter, and incubated at  $30 \pm 2$  °C for 15–20 days. Once the fungal incubation was completed, a mixture of sterilized garden soil (121 °C for 45 min), peat, perlite, and fine sand (2:1:1) was filled into trays. The inoculum, prepared in the bottles, was then mixed into the soil at a rate of 100 g inoculum per 1 kg of soil and incubated for 4–5 days. After the sterilization with 2% sodium hypochlorite 6 seeds of each genotype were sown in 21 cm diameter pots filled with a mixture of sterile soil, fine sand, peat (2:1:1) containing inoculum. The cultivar P31G98 was utilized as a positive control due to its being known to be sensitive to *Mp*<sup>48</sup>. Negative control plants were planted in disease-free pots.

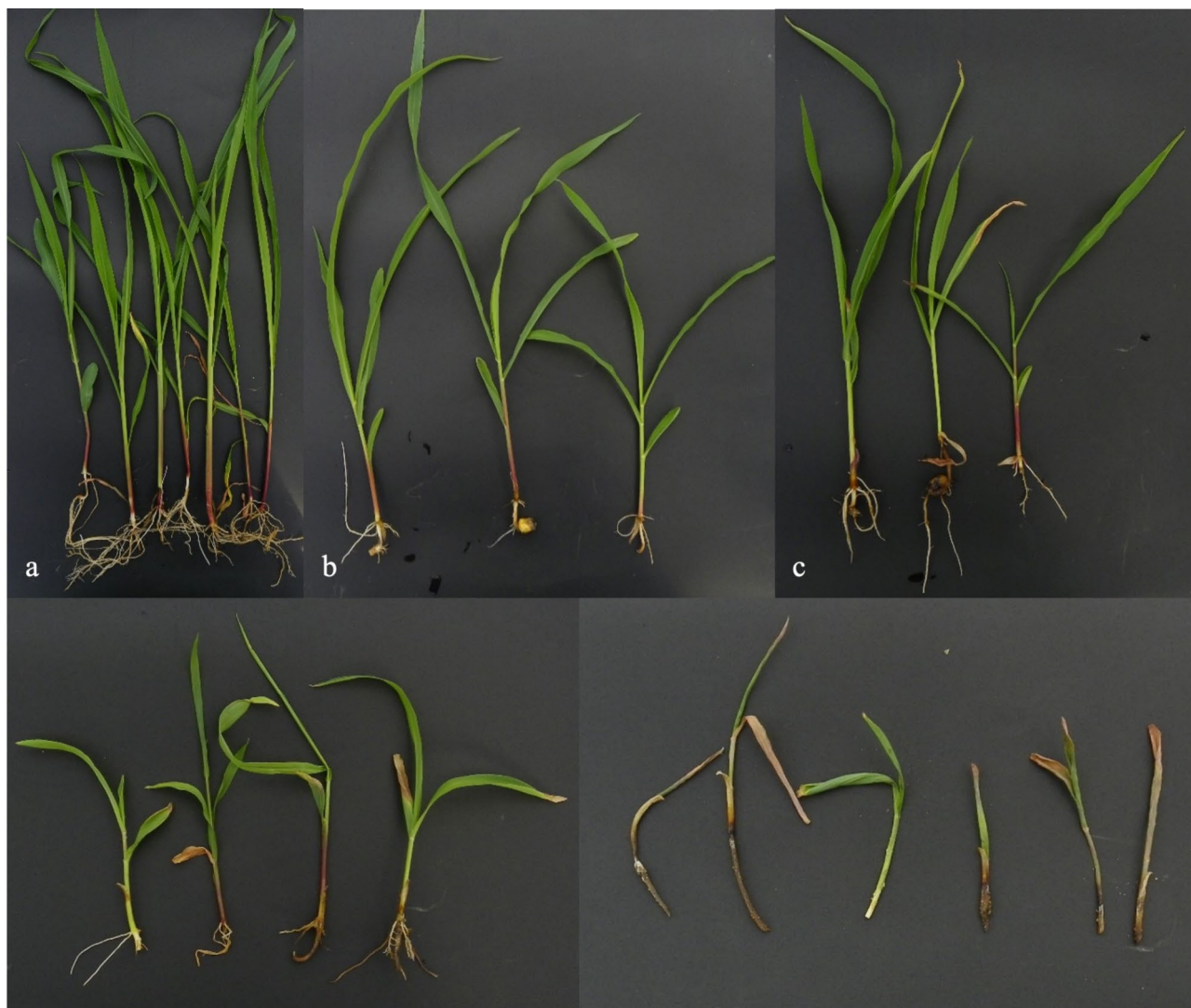
The pots were maintained in climate chambers at 30 °C, 30–45% humidity, and a 16-hour light/8-hour dark cycle. Plants were watered as necessary, ensuring that the soil moisture was not too high<sup>49</sup>. The pathogenicity trials were set up in a randomized complete block design with 3 replicates. Each pot was considered one replicate. The disease severity in the plants was assessed using a 0–5 scale<sup>48</sup> (Fig. 6). After 40 days under the controlled conditions, 40-day-old plants were uprooted, disease symptoms were scored and the disease severity index (DSI) on roots and stems was calculated using the Townsend-Heuberger formula ( $S$  = score,  $P$  = number of plants,  $TP$  = total number of plants,  $HS$  = highest score) (Table 3)<sup>50</sup>.

$$\text{Disease Severity \%} = \sum \frac{S \times P}{TP \times HS} \times 100$$

Genotype reactions were defined according to DSI and phenotypes were arranged in 6 groups (Table 3).

### Statistical analysis of phenotypic data

Variance analyses (ANOVA) were conducted using the SPSS software package (version 12.0 J; SPSS Inc., Chicago, IL, USA). At the  $p \leq 0.05$  significance level, the variances according to the disease severity and genotype interactions were identified. Broad sense heritability score calculated as described by<sup>51</sup>.



**Fig. 6.** Plants representing the 0–5 scale used for scoring corn genotypes. 0: No disease symptoms observed in the plant (a); 1: General growth retardation in the plant (b); 2: Growth retardation in the plant, chlorosis in the leaves, and mild root rot (c); 3: Growth retardation in the plant, severe root rot, wilting, and intense chlorosis (d); 4: Significant growth retardation, complete root rot in the plant (e); 5: No germination observed.

Scale	Description	Phenotype	DSI
0	No disease symptoms observed in the plant	Resistant	0–20
1	General growth retardation in the plant	Moderately resistant	20–40
2	Growth retardation in the plant, chlorosis in the leaves, and mild root rot	Tolerant	40–60
3	Growth retardation in the plant, severe root rot, wilting, and intense chlorosis	Moderately susceptible	60–80
4	Significant growth retardation, complete root rot in the plant	Susceptible	80–99
5	No germination observed	Highly susceptible	100

**Table 3.** Criteria used in phenotyping and corresponding values<sup>50</sup>.

### Molecular marker analysis

SNP markers were determined using DArT analysis (Diversity Arrays Technology Pty. Ltd., Canberra, Australia). Extracted DNA was cut with *Pst* I-*Mse* I enzymes and ligated to adapters compatible with these enzymes as described previously<sup>52</sup>. PCR and adapter ligation were performed following established protocols<sup>53</sup>. Sequencing was conducted using Illumina HiSeq 2000 (Illumina Inc., USA).

## Population structure and genome-wide association mapping

The raw SNP data obtained from sequencing were filtered based on a 95% call rate and a minor allele frequency (MAF) greater than 5%. Subsequently, the population structure was determined using STRUCTURE software to identify groups within the population that exhibited discriminant allele frequencies. The number of groups in the population is indicated by the K value, which is defined as the allele frequencies at each locus. For every group, a range of 1 to 10 was calculated to find the most favorable K value. Ten replicates were analyzed for every K value, and 50,000 Markov Chain Monte Carlo (MCMC) replicates were performed for the burn-in length and burn-in period. The STRUCTURE HARVESTER<sup>54</sup> software was utilized to determine the  $\Delta K$  value with the highest probability for every K value that was obtained.

The association between SNP markers and resistance of maize genotypes to *Mp* was determined in the TASSEL software<sup>55</sup>. For this purpose, the significance levels of the relationship between phenotypic data, marker data, and Q matrix, which represents the population structure, were analyzed with TASSEL software. The Mixed Linear Model (MLM) model was employed in this study to minimize false positives by integrating population structure and genetic relationships. It utilizes the Q matrix for population adjustments and the K matrix for genetic linkage, resulting in enhanced accuracy in the analysis<sup>56</sup>. FDR (False Discovery Rate) is a method used to control the proportion of false positive results in statistical testing<sup>57</sup>. The FDR correction threshold was calculated to minimize the likelihood of detecting false positive SNPs<sup>58</sup>. This approach enables a more accurate identification of significant SNPs while maintaining the integrity of the analysis. When FDR Correction was exceeded, SNP markers were considered statistically significant. The visual representation of these markers exceeding the FDR threshold line was drawn using Manhattan blocks. In order to visualize the relationship between observed and expected phenotypic data and P values, quantile-quantile (Q-Q) plots were also drawn<sup>56</sup>.

## Candidate genes

SNP markers linked to *Mp* in maize were examined using Sequence Viewer 3.49.0 on the National Center for Biotechnology (NCBI) database to identify potential genes. Analysis of candidate genes was carried out 100 K upstream and 100 K downstream of the marker's positions.

## Data availability

The data generated from the study will be provided by the corresponding author upon request.

Received: 24 July 2024; Accepted: 22 January 2025

Published online: 06 March 2025

## References

- del Molina, M., Naranjo, C. A. & C. & Cytogenetic studies in the genus *Zea*. *Theor. Appl. Genet.* **73**, 542–550 (1987).
- Haberer, G. et al. Structure and architecture of the maize genome. *Plant Physiol.* **139**, 1612–1624 (2005).
- Kehinde, A. D., Ojo, T. O. & Ogundej, A. A. Impact of participation in social capital networks on the technical efficiency of maize producers in Southwest Nigeria. *Agric. Food Secur.* **13**, 12 (2024).
- Mishra, M., Sharma, S. & Sharma, R. Corn (*Zea mays*) as a nutrient source and diet: a review. *J. Pharm. Res. Int.* **33**, 299–303 (2021).
- FAO Statistical database - Food and Agriculture Organization of the United Nations, Rome Italy (2024).
- Kaur, S. et al. Emerging phytopathogen *Macrophomina phaseolina*: Biology, economic importance and current diagnostic trends. *Crit. Rev. Microbiol.* **38**, 136–151 (2012).
- Ahmed, M. J. & Shoaib, A. Unveiling the biocontrol potential of *Pseudomonas syringae* through seed biopriming against charcoal rot disease in maize. *Physiol. Mol. Plant Pathol.* **133**, 102370 (2024).
- Baird, R. E., Watson, C. E. & Scruggs, M. Relative longevity of *Macrophomina phaseolina* and associated mycobiota on residual soybean roots in soil. *Plant Dis.* **87**, 563–566 (2003).
- Khalili, E. et al. Evaluation of Trichoderma isolates as potential biological control agent against soybean charcoal rot disease caused by *Macrophomina phaseolina*. *Biotechnol. Biotechnol. Equip.* **30**, 479–488 (2016).
- Marquez, N., Giachero, M. L., Declerck, S. & Ducasse, D. A. *Macrophomina phaseolina*: General characteristics of pathogenicity and methods of control. *Front. Plant Sci.* **12**, 634397 (2021).
- Jordaan, E., van der Waals, J. E. & McLaren, N. W. Effect of irrigation on charcoal rot severity, yield loss and colonization of soybean and sunflower. *Crop Prot.* **122**, 63–69 (2019).
- Kendig, S. R., Rupe, J. C. & Scott, H. D. Effect of irrigation and soil water stress on densities of *Macrophomina phaseolina* in soil and roots of two soybean cultivars. *Plant Dis.* **84**, 895–900 (2000).
- Pawlowski, M. L., Bowen, C. R., Hill, C. B. & Hartman, G. L. Responses of soybean genotypes to pathogen infection after the application of elicitors. *Crop Prot.* **87**, 78–84 (2016).
- Lango Allen, H. et al. Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature* **467**, 832–838 (2010).
- Guo, Z. et al. Complex Genetic System involved in Fusarium ear rot resistance in Maize as revealed by GWAS, Bulk Sample Analysis, and genomic prediction. *Plant Dis.* **104**, 1725–1735 (2020).
- Warburton, M. L. et al. Comparative analysis of multiple GWAS results identifies metabolic pathways associated with resistance to *A. flavus* infection and aflatoxin accumulation in maize. *Toxins* **14**, 738 (2022).
- Hou, M. et al. Genome-wide association study of maize resistance to *Pythium aristosporum* stalk rot. *Front. Plant Sci.* **14**, 1239635 (2023).
- Zhao, M. et al. Identification of genetic loci associated with rough dwarf disease resistance in maize by integrating GWAS and linkage mapping. *Plant Sci.* **315**, 111100 (2022).
- Washburn, J. D. et al. GWAS analysis of maize host plant resistance to western corn rootworm (Coleoptera: Chrysomelidae) reveals candidate small effect loci for resistance breeding. *J. Econ. Entomol.* **116**, 2184–2192 (2023).
- Jaccoud, D., Peng, K., Feinstein, D. & Kilian, A. Diversity arrays: A solid state technology for sequence information independent genotyping. *Nucleic Acids Res.* **29**, e25–e25 (2001).
- Wen, J. et al. Association mapping of major economic traits and exploration of elite alleles in *Prunus sibirica*. *Euphytica* **219**, 1–13 (2023).
- Mengistu, A., Smith, J. R., Ray, J. D. & Bellaloui, N. Seasonal progress of charcoal rot and its impact on soybean productivity. *Plant Dis.* **95**, 1159–1166 (2011).

23. Ruzmetov, D. R., Sherimbetov, A. G. & Adilov, B. S. First report of *Macrophomina phaseolina* causing charcoal rot in soybean (*Glycine max* (L.) Merr.) plants in Uzbekistan. *Plant Dis.* **7**, 108 (2024).
24. Gaetán, S. A., Fernandez, L. & Madia, M. Occurrence of charcoal rot caused by *Macrophomina phaseolina* on canola in Argentina. *Plant Dis.* **90**, 524 (2006).
25. Mahmoud, A. F., Abou-Elwafa, S. F. & Shehzad, T. Identification of charcoal rot resistance QTLs in sorghum using association and in silico analyses. *J. Appl. Genet.* **59**, 243–251 (2018).
26. Mughogho, L. K. & Pande, S. Charcoal rot of sorghum (1983).
27. Mayek-Pérez, N., García-Espinosa, R., López-Castañeda, C., Acosta-Gallegos, J. A. & Simpson, J. Water relations, histopathology and growth of common bean (*Phaseolus vulgaris* L.) during pathogenesis of *Macrophomina phaseolina* under drought stress. *Physiol. Mol. Plant Pathol.* **60**, 185–195 (2002).
28. Zveibil, A., Mor, N., Gnyem, N. & Freeman, S. Survival, host-pathogen interaction, and management of *Macrophomina phaseolina* on strawberry in Israel. *Plant Dis.* **96**, 265–272 (2012).
29. Fuhlbohmer, M. J., Ryley, M. J. & Aitken, E. A. B. Infection of mungbean seed by *Macrophomina phaseolina* is more likely to result from localized pod infection than from systemic plant infection. *Plant Pathol.* **62**, 1271–1284 (2013).
30. Adeyanju, A., Little, C., Yu, J. & Tesso, T. Genome-wide association study on resistance to stalk rot diseases in grain sorghum. *G3 Genes Genomes Genetics* **5**, 1165–1175 (2015).
31. Rashid, Z. et al. Identification and validation of genomic regions associated with charcoal rot resistance in tropical maize by genome-wide association and linkage mapping. *Front. Plant Sci.* **12**, 1–13 (2021).
32. Coser, S. M. et al. Genetic architecture of charcoal rot (*Macrophomina phaseolina*) resistance in soybean revealed using a diverse panel. *Front. Plant Sci.* **8**, 1–12 (2017).
33. Cappa, E. P. et al. Impacts of population structure and analytical models in genome-wide association studies of complex traits in forest trees: a case study in *Eucalyptus globulus*. *PLoS ONE* **8**, e81267 (2013).
34. Vinholes, P., Rosado, R., Roberts, P., Borém, A. & Schuster, I. Single nucleotide polymorphism-based haplotypes associated with charcoal rot resistance in Brazilian soybean germplasm. *Agron. J.* **111**, 182–192 (2019).
35. Gu, X. et al. Genome-wide association of single nucleotide polymorphism loci and candidate genes for frogeye leaf spot (*Cercospora sojina*) resistance in soybean. *BMC Plant Biol.* **21**, 588 (2021).
36. Broglie, K. E., Biddle, P., Cressman, R. & Broglie, R. Functional analysis of DNA sequences responsible for ethylene regulation of a bean chitinase gene in transgenic tobacco. *Plant Cell.* **1**, 599–607 (1989).
37. Chowdhury, S., Basu, A. & Kundu, S. Biotrophy-necrotrophy switch in pathogen evoke differential response in resistant and susceptible sesame involving multiple signaling pathways at different phases. *Sci. Rep.* **7**, 17251 (2017).
38. Gaige, A. R., Ayella, A. & Shuai, B. Methyl jasmonate and ethylene induce partial resistance in *Medicago truncatula* against the charcoal rot pathogen *Macrophomina phaseolina*. *Physiol. Mol. Plant Pathol.* **74**, 412–418 (2010).
39. Shirai, M. & Eulgem, T. Molecular interactions between the soilborne pathogenic fungus *Macrophomina phaseolina* and its host plants. *Front. Plant Sci.* **14**, 1264569 (2023).
40. Kim, S. J., Held, M. A., Zemelis, S., Wilkerson, C. & Brandizzi, F. CGR2 and CGR3 have critical overlapping roles in pectin methylesterification and plant growth in *Arabidopsis thaliana*. *Plant J.* **82**, 208–220 (2015).
41. Weraduwa, S. M. et al. Pectin methylesterification impacts the relationship between photosynthesis and plant growth. *Plant Physiol.* **171**, 833–848 (2016).
42. Marino, D., Peeters, N. & Rivas, S. Ubiquitination during plant immune signaling. *Plant Physiol.* **160**, 15–27 (2012).
43. Coculo, D. et al. Arabidopsis subtilases promote defense-related pectin methylesterase activity and robust immune responses to botrytis infection. *Plant Physiol. Biochem.* **201**, 107865 (2023).
44. Guo, Y. H. et al. GhZFP1, a novel CCCH-type zinc finger protein from cotton, enhances salt stress tolerance and fungal disease resistance in transgenic tobacco by interacting with GZIRD21A and GZIPR5. *New Phytol.* **183**, 62–75 (2009).
45. Gupta, S. K., Rai, A. K., Kanwar, S. S. & Sharma, T. R. Comparative analysis of zinc finger proteins involved in plant disease resistance. *PLoS ONE* **7**, e42578 (2012).
46. Schuler, M. A. & Werck-Reichhart, D. Functional genomics of P450s. *Annu. Rev. Plant Biol.* **54**, 629–667 (2003).
47. Lannoo, N. & Van Damme, E. J. M. Lectin domains at the frontiers of plant defense. *Front. Plant Sci.* **5**, 397 (2014).
48. Talapov, T. Determination of sunflower (*Helianthus annuus*) ve soybean (*Glycine max*) variety reactions to *Macrophomina phaseolina* and characterization of the agent. *Master's Thesis, Gaziantep University* (2020).
49. Twizeyimana, M., Hill, C. B., Pawlowski, M., Paul, C. & Hartman, G. L. A cut-stem inoculation technique to evaluate soybean for resistance to *Macrophomina Phaseolina*. *Plant Dis.* **96**, 1210–1215 (2012).
50. Townsend, G. K. & Heuberger, J. W. Methods for estimating losses caused by diseases in fungicide experiments. *Plant Dis. Rep.* **27**, 340–343 (1943).
51. Ma, L. et al. GWAS with a PCA uncovers candidate genes for accumulations of microelements in maize seedlings. *Physiol. Plant* **172**, 2170–2180 (2021).
52. Erdogmus, S. et al. Genome-wide association studies of ca and mn in the seeds of the common bean (*Phaseolus vulgaris* L.). *Genomics* **112**, 4536–4546 (2020).
53. Kilian, A. et al. Diversity arrays technology: A generic genome profiling technology on open platforms. *Methods Mol. Biol.* **888**, 67–89 (2012).
54. Earl, D. A. & VonHoldt, B. M. Structure harvester: A website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv. Genet. Resour.* **4**, 359–361 (2012).
55. Bradbury, P. J. et al. TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinformatics* **23**, 2633–2635 (2007).
56. Karaca, N. et al. Association mapping of magnesium and manganese concentrations in the seeds of *C. Arietinum* and *C. Reticulatum*. *Genomics* **112**, 1633–1642 (2020).
57. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B* **57**, 289–300 (1995).
58. Karaca, N. et al. Genome-wide association studies of protein, lutein, vitamin c, and fructose concentration in wild and cultivated chickpea seeds. *Crop Sci.* **59**, 2652–2666 (2019).

## Acknowledgements

We gratefully acknowledge the Scientific and Technological Research Council of Türkiye for their support, under grant number 121O537.

## Author contributions

G.O. Bioinformatic analysis, writing manuscript; S.Y. collection of Mp pathogens and determination of resistance reaction of genotypes to Mp; C.C. Collection of pathogens; M.B.T. GWAS analysis, principal investigator; D.A. Corresponding author, GWAS analysis, Bioinformatic analysis.

## Declarations

### Competing interests

The authors declare no competing interests.

### Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-025-87798-8>.

**Correspondence** and requests for materials should be addressed to D.A.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2025