Heliyon 10 (2024) e38939

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

# Identifying SSR/InDel loci related to tobacco bacterial wilt resistance using association mapping

Ruiqiang Lai [a,b,1], Yanshi Xia [a,1], Ronghua Li [a,1], Qinghua Yuan [c], Weicai Zhao [d,**], Kadambot H.M. Siddique [e], Peiguo Guo [a,*]

[a] Guangdong Provincial Key Laboratory of Plant Adaptation and Molecular Design, School of Life Sciences, Guangzhou University, Guangzhou, 510006, China
[b] College of Agriculture & Biology, Zhongkai University of Agriculture and Engineering, Guangzhou, 510225, China
[c] Crop Research Institute, Guangdong Academy of Agriculture, Guangzhou, 510640, China
[d] Guangdong Research Institute of Tobacco Science, Shaoguan, 512029, China
[e] The UWA Institute of Agriculture, The University of Western Australia, Perth, WA, 6001, Australia

ARTICLE INFO

*Keywords:*
*Nicotiana tabacum*
Bacterial wilt
SSR
InDel
Association mapping

ABSTRACT

Identifying molecular markers linked to tobacco bacterial wilt resistance is crucial for developing resistant tobacco varieties, thereby enhancing tobacco production and quality. In this two-year study, we evaluated the tobacco bacterial wilt disease index (TBWDI) in a mapping population of 78 tobacco accessions using SSR/InDel markers across 1377 marker loci. Two association models, GLM_Q and MLM_Q + Kinship, were used for association analysis. By considering multiple environments, selection thresholds, and phenotype values, we identified 19 reliable marker loci (P = 7.07-E05–4.98-E02) that explained 5.37–19.10 % of the phenotypic variation. Among these, we selected 11 accessions with high resistance to tobacco bacterial wilt, each containing at least one excellent locus. The models predicted five hybrid combinations whose offspring aggregated additional excellent loci. Additionally, one locus was identified during quantitative trait loci mapping, with accessions carrying this locus exhibiting significantly different TBWDI values from those without it, confirming its reliability and stability. A candidate gene (LOC107795335) was also identified downstream at 17.58 Kbp of the locus. Our study pinpointed reliable loci associated with bacterial wilt resistance, identified important materials for breeding resistant tobacco varieities, and predicted the best hybrid combinations for low disease indexes. The results of our study offer invaluable theoretical insights for breeding varieties with high and stable resistance.

## 1. Introduction

Tobacco (*Nicotiana tabacum* L.) is a globally important economic crop, with breeders continually striving to improve its yield and quality. However, tobacco is prone to bacterial wilt during growth, significantly decreasing yield and quality [1]. Current control strategies, such as agricultural control [2], biological control [3], and chemical control [4], have limited effectiveness [5]. Developing

* Corresponding author.
** Corresponding author.
*E-mail addresses:* zhaoweicai@126.com (W. Zhao), guopg@gzhu.edu.cn (P. Guo).
[1] Both authors contributed equally to this work.

https://doi.org/10.1016/j.heliyon.2024.e38939
Received 7 May 2024; Received in revised form 25 September 2024; Accepted 2 October 2024
Available online 3 October 2024

bacterial wilt-resistant tobacco varieties offers a promising solution to mitigate the damage caused by this disease. Moreover, identifying molecular markers associated with tobacco bacterial wilt (TBW) resistance can expedite the selection of resistant materials, accelerating the breeding of resistant varieties.

Two main approaches for identifying molecular markers linked to TBW resistance are quantitative trait loci (QTL) mapping using pedigree populations and linkage disequilibrium (LD) mapping using natural populations [6]. Nishi et al. [7] pioneered QTL mapping for TBW resistance using a double haploid (DH) population and identified a QTL associated with resistance. Later studies by Lan et al. [1], Qian et al. [8], and Drake-Stowe et al. [9] also explored the genetics of TBW resistance through genetic linkage mapping, identifying eight, four, and three QTL, respectively. While QTL mapping has advanced the understanding of TBW resistance, it has limitations. Using pedigree populations in linkage analysis only involves two alleles at the same locus, resulting in a narrow genetic background and limited recombination events, compromising mapping accuracy and resolution [10]. In contrast, association analysis leverages natural populations with greater genetic variation, eliminating the need to construct mapping populations, and is more time-efficient [11]. By using multiple generations of recombination events, association analysis offers broader genetic backgrounds and higher resolutions [10]. Consequently, association analysis has gained prominence in mapping loci related to TBW resistance [12].

Despite the advances in identifying QTL or molecular markers linked to TBW resistance through linkage and LD analyses, the body of literature remains limited and practical applications of these findings are scarce. Resistance to TBW is a complex quantitative trait [13] influenced by multiple genes and environmental factors, rendering it impractical to await the identification of adequate gene loci before aggregating them. Therefore, identifying stable and reliable representative molecular markers and applying these findings to assess TBW resistance is essential. This approach facilitates the selection of materials with multiple excellent variant loci, improving the likelihood of selecting tobacco materials with high TBW resistance.

In this study, we used SSR/InDel markers developed in previous research [14–16] and in-house [17] to genotype 78 tobacco accessions. We conducted association analysis with a multi-year TBW disease index (TBWDI) to achieve three key objectives: (1) explore the genetic diversity of tobacco germplasm; (2) identify outstanding variant loci associated with TBWDI; (3) screen materials with stable resistance and multiple outstanding variant loci for potential hybrid combinations.

**Table 1**
Origins and types of 78 tobacco accessions.

| No. | Accession | Origin | Type | No. | Accession | Origin | Type |
|---|---|---|---|---|---|---|---|
| 1 | D101 | USA | Flue-cured | 40 | Xiaohuangjin | Guangdong China | Flue-cured |
| 2 | RG17 | USA | Flue-cured | 41 | Xujin No.1 | Henan China | Flue-cured |
| 3 | Changbohuang | Henan China | Flue-cured | 42 | Yanyan97 | Fujian China | Flue-cured |
| 4 | C151 | Guangdong China | Flue-cured | 43 | Yinnizhong | Indonesia | Flue-cured |
| 5 | C212 | Guangdong China | Flue-cured | 44 | Yunyan87 | Yunnan China | Flue-cured |
| 6 | CO139 | USA | Flue-cured | 45 | Zhongyan14 | Shandong China | Flue-cured |
| 7 | CO176 | USA | Flue-cured | 46 | 123-13-2 | China (breeding) | Sun-cured |
| 8 | CV087 | Shandong China | Flue-cured | 47 | 5669-1-qing-1 | China (breeding) | Sun-cured |
| 9 | G28 | USA | Flue-cured | 48 | Baiguniuli | Guangdong China | Sun-cured |
| 10 | K149 | USA | Flue-cured | 49 | Daqiugen-2 | Guangdong China | Sun-cured |
| 11 | K326 | USA | Flue-cured | 50 | Dashangou | Yunnan China | Sun-cured |
| 12 | K346 | USA | Flue-cured | 51 | Datong | Guangdong China | Sun-cured |
| 13 | K358 | USA | Flue-cured | 52 | Dayemihe | Guangdong China | Sun-cured |
| 14 | K399 | USA | Flue-cured | 53 | Dazhongbaimao | Guangdong China | Sun-cured |
| 15 | K730 | USA | Flue-cured | 54 | Fengkai-10 | Guangdong China | Sun-cured |
| 16 | NC60 | USA | Flue-cured | 55 | Guju | Guangdong China | Sun-cured |
| 17 | NC95 | USA | Flue-cured | 56 | Guangyeshugeng | Guangdong China | Sun-cured |
| 18 | OX2028 | USA | Flue-cured | 57 | Guanghuang No.5 | China (breeding) | Sun-cured |
| 19 | R-158 | USA | Flue-cured | 58 | Hainanyan | Hainan China | Sun-cured |
| 20 | RG11 | USA | Flue-cured | 59 | Jiangyouyan | Sichuan China | Sun-cured |
| 21 | RG-22 | USA | Flue-cured | 60 | Jincaiding-2 | Guangdong China | Sun-cured |
| 22 | RG-8 | USA | Flue-cured | 61 | Kuanjian | Guangdong China | Sun-cured |
| 23 | ATNARELLO | Mauritius | Burley | 62 | Lingshuiligong-3 | Hainan China | Sun-cured |
| 24 | KY26 | USA | Burley | 63 | Maoyantieganzi | Sichuan China | Sun-cured |
| 25 | Baxishailiangyan | Brazil | Burley | 64 | Mihezai | Guangdong China | Sun-cured |
| 26 | Gezaji | jiangxi China | Flue-cured | 65 | Miyayan | Guangdong China | Sun-cured |
| 27 | Cuibi NO.1 | Fujian China | Flue-cured | 66 | Mianchengyan | Guangdong China | Sun-cured |
| 28 | Dabaijin599 | Shandong China | Flue-cured | 67 | Nahuo | Guangdong China | Sun-cured |
| 29 | Fengzi No.1 | Guangdong China | Flue-cured | 68 | Niuliyan | Guangdong China | Sun-cured |
| 30 | Gexin No.6 | Shandong China | Flue-cured | 69 | Qinggeng | Guangdong China | Sun-cured |
| 31 | Honghuadajinyuan | Yunnan China | Flue-cured | 70 | Qiongzhongwuzhishan | Hainan China | Sun-cured |
| 32 | Jinxing | Shandong China | Flue-cured | 71 | Renheyan | Guangdong China | Sun-cured |
| 33 | Anbalima No.2 | China (breeding) | Sun-cured | 72 | Yunluo01 | Guangdong China | Sun-cured |
| 34 | Lingnong No.1 | Guangxi China | Flue-cured | 73 | Xingzizhouye-1 | Guangdong China | Sun-cured |
| 35 | Lvyin No.1 | China (unknown) | Flue-cured | 74 | Yulaoyan | Guangdong China | Sun-cured |
| 36 | Luodingjinxing | Guangdong China | Flue-cured | 75 | Zhongshan | Guangdong China | Sun-cured |
| 37 | Meifu No.4 | Guizhou China | Flue-cured | 76 | Ziyan | Guangdong China | Sun-cured |
| 38 | Qingshengyan | Henan China | Flue-cured | 77 | Yanshanyan | Guangdong China | Sun-cured |
| 39 | xiangyan | Hunan China | Sun-cured | 78 | 2040 | Cuba | Cigar |

## 2. Materials and methods

### 2.1. Plant materials and phenotyping

A natural population of 78 tobacco accessions, provided by the Nangxiong Tobacco Research Institute in Guangdong Province, China, was used as the mapping population. These materials encompassed four tobacco types—flue-cured, cigar, sun-cured, and burley—originating from the United States of America (USA), Cuba, Brazil, Indonesia, Mauritius, and China (see Table 1 for details).

The 78 tobacco accessions were cultivated in an experimental field at the Nangxiong Tobacco Research Institute in Guangdong Province, China, in 2011 (denoted E1) and 2013 (denoted E2). Plants were spaced 0.5 m apart in rows with a row spacing of 1.2 m. Each accession, comprising 20 plants, was arranged in a randomized complete block design, and field management practices followed local guidelines. Bacterial wilt disease, caused by the biochemical type III race-1 strain of *Ralstonia solanacearum*, was introduced via stem puncture inoculation during the vigorous growth phase using a cell suspension adjusted to approximately $1 \times 10^8$ CFU $mL^{-1}$ (OD600 nm = 0.1) [12,18]. Disease assessment followed established protocols [12], and the TBWDI was calculated following the method described by Lan et al. [1] and categorized as follows: 0 < TBWDI ≤25 (highly resistant), 25 < TBWDI ≤50 (moderately resistant), 50 < TBWDI ≤75 (moderately susceptible), and 75 < TBWDI ≤100 (highly susceptible).

### 2.2. Phenotypic data analysis

The skewness and kurtosis of TBWDI for the 78 tobacco accessions were calculated using IBM SPSS 19.0 software [19,20]. Analysis of variance (ANOVA) and deviation analysis were performed to assess normal distribution. A normal distribution plot of TBWDI for the tobacco accessions was generated using the 'NORMDIST' function [21], and the range, mean, variance, and coefficient of variation were computed. Broad-sense heritability ($H^2$) of TBWDI was calculated as follows:

$$H^2 = \sigma g^2 \ / \ (\sigma g^2 + \sigma e^2 / n)$$

where $\sigma g^2$ is the genetic variance and $\sigma e^2$ is the environmental variance, and n is the number of environmental factors.

### 2.3. DNA extraction and genotyping

Fresh young leaves were collected from each accession for DNA extraction using a modified CTAB method [22]. The quality and concentration of the extracted DNA were assessed via agarose gel electrophoresis and a NanoDrop UV–Vis spectrophotometer, respectively. The extracted DNA was diluted to 30 ng $L^{-1}$ for use as a PCR amplification template.

A combination of previously reported tobacco SSR markers [14–16] and SSR/InDel markers developed through RAD-seq (restriction-site associated DNA sequencing) using '118–3' and 'Yueyan 98' in our laboratory [17] were used. These markers, which displayed clear bands, good specificity, and polymorphism in 'YanYan 97', 'Honghuadajinyuan', '118–3' and 'Yueyan 98' resulted in 512 SSR/InDel markers for genotyping the population.

The PCR reaction mixture for SSR and InDel marker amplification (10 μL total) included 2 μL DNA template, 0.3 μL forward primer (1 μmol $L^{-1}$), 0.3 μL reverse primer (1 μmol $L^{-1}$), 0.8 μL $MgCl_2$ (25 mmol $L^{-1}$), 0.1 μL Taq DNA polymerase (5 U·μ$L^{-1}$), 1 μL 10 × PCR buffer, 0.3 μL dNTPs (10 mmol $L^{-1}$), and 5.2 μL double-distilled water. The PCR amplification program included an initial denaturation step at 94 °C for 5 min, followed by 38 cycles of denaturation at 94 °C for 45 s, annealing at the corresponding primer annealing temperature for 45 s, and extension at 72 °C for 1 min. The final extension step was performed at 72 °C for 10 min, with the resulting PCR products analyzed using 6 % non-denaturing polyacrylamide gel electrophoresis and a silver staining method [23]. Gel images were scanned with a BenQ M800 scanner and analyzed using GelBuddy software [24]. Allelic variations were determined by band presence or absence, and a binary data matrix (1, 0) was constructed for population genotyping. All reagents were purchased from Sangon Biotech (Shanghai, China) Co., Ltd.

### 2.4. Population structure, kinship, linkage disequilibrium, and association analysis

Principal component analysis (PCA) was conducted using the 'Dcenter' and 'Eigen' subroutines in NYSYS-PC 2.1 software [25]. The unweighted pair-group method with arithmetic means (UPGMA) clustering analysis was performed in the same software using the 'Clusting' and 'SAHN' subroutines. Population structure was analyzed using Structure 2.3 software [25], estimating the optimal number of population clusters (K), with the K value range set to 1–11, with the software run to ascertain the inflection point of the posterior probability value [LnP(K)] to determine K. The maximum ΔK value method [26] was applied to determine K and obtain the corresponding population correction coefficient Q-value.

Genotyping data were imported into TASSEL 3.0 software [27]. The 'Kinship' and 'Linkage Disequilibrium' subroutines were used to calculate the kinship coefficient between accessions and the LD $R^2$ value between loci, respectively. Haploview software [28] generated LD plots for key loci.

Association loci related to traits were identified using a general linear model with Q (GLM_Q) and a mixed linear model with Q and kinship coefficient (MLM_Q + Kinship) in TASSEL3.0 software [27]. The models were validated through Q-Q plot analysis and lambda value (λ-value) = median ($chi^2$)/0.456 [29] analysis using R_4.0.4 software. Peak loci were determined with a significance level of $P < 0.05$ [30], and false discovery rate (FDR) was controlled using the method of Benjamini and Hochberg [31]. Associated markers with

FDR (q)-values <0.05 ($P < 0.05$) were retained.

### 2.5. Sequence alignment and three-dimensional structure prediction

Homologous proteins were identified using the NCBI (https://blast.ncbi.nlm.nih.gov) and TAIR (https://www.arabidopsis.org/) databases. Sequence alignment was conducted using DNAMAN7.0 software, and three-dimensional structure prediction was performed using SWISS-MODEL software (https://swissmodel.expasy.org/) [32].

### 2.6. Phenotypic effect and evaluation of tobacco bacterial wilt disease index

The phenotypic effect value (Ai) was calculated using the methods of Zhang et al. [33] and Zhu et al. [34], as follows:

$$Ai = \sum Xij \Big/ ni - \sum Nk \Big/ nk$$

where Ai is the phenotypic effect value of the i-th variant locus, Xij is the measured value of the j-th material phenotype carrying the i-th variant locus, ni is the number of materials with the i-th variant locus, and ∑Nk/nk represents the average value of all material phenotype measurements. A negative or positive Ai value indicates a negative (denoted by '–') or positive (denoted by '+') phenotypic effect, respectively.

Based on the average disease index, tobacco materials were evaluated for disease resistance over multiple years. A higher disease index indicates greater disease severity and lower disease resistance. Therefore, TBWDI was used as a correlated trait, and the phenotypic effect of the associated loci was calculated to assess tobacco material resistance. Negative phenotypic effects indicate a reduction in TBWDI, while positive effects indicate an increase in TWBDI.

## 3. Results

### 3.1. Statistical analysis of tobacco bacterial wilt disease index

The TBWDI exhibited substantial variation (>89), with a high coefficient of variation (>51 %) across both environments (E1 and E2). This variability suggests the presence of highly resistant varieties, which could serve as foundational materials for breeding superior varieties (Table 2; Fig. 1A and B). The skewness and kurtosis values for TBWDI in both environments had absolute values < 1 (Table 2), and the normality deviations were evenly distributed around the zero axis (Fig. 1C and D), indicating a normal distribution. Furthermore, the broad-sense heritability of TBWDI exceeded 70 % in both environments, highlighting a substantial genetic influence on TBWDI. ANOVA analysis revealed significant differences in TBWDI between E1 and E2 (Fig. 1E), indicating that environmental factors also affected TBWDI. These findings confirm significant variability in TBW disease resistance among the selected germplasm population, consistent with the quantitative trait characteristics and complex genetic patterns.

### 3.2. Genetic diversity, population structure, and linkage disequilibrium analysis

Using 1377 loci from SSR and InDel marker genotyping, we calculated 3003 pairwise kinship coefficients for the 78 tobacco accessions. The results revealed some kinship relationships among the accessions (Fig. 2A; Table S1). However, 2714 combinations had kinship coefficients below 0.5, accounting for 90.38 % of the total combinations (Table S1), suggesting that while most accessions are related, a significant amount of allelic variation exists, which is beneficial for identifying relevant gene loci through phenotype mining. These kinship coefficients can also be used as correction factors to reduce false associations [35].

Population structure was assessed using UPGMA clustering (Fig. S1; Table S2), PCA (Fig. 2B; Table S2), and population structure analysis (Fig. 2C and D; Table S2), which grouped the population into two subgroups (subgroup-1/2). The classifications from UPGMA clustering (subgroup-u1/u2) and PCA (subgroup-p1/p2) were consistent. Subgroup 1 comprised 19 flue-cured tobacco materials, while subgroup 2 included 59 materials: 36 air-cured tobacco, 19 flue-cured tobacco, three burley tobacco, and one cigar tobacco. The population structure analysis (subgroup-s1/s2) differed only for the flue-cured tobacco variety 'Zhongyan14' from China, but the classifications for the other materials were consistent across methods, with subgroup-s1 comprising 18 materials and subgroup-s2 containing 60 materials. These consistent results indicate that population structure analysis, corrected Q-values, can be used for subsequent association analysis.

Linkage disequilibrium, which influences association analysis, was measured by the $R^2$ value, representing the correlation between two loci [6,36]. The $R^2$ values for the 1377 loci (947,376 combinations) ranged from 0 to 1 (Fig. 3A; Table S3). Strong LD combinations

**Table 2**
Statistical analysis of tobacco bacterial wilt disease index.

| Year | Environment | Range | Mean | Skewness | Kurtosis | CV (%) | $H^2$ (%) |
|---|---|---|---|---|---|---|---|
| 2011 | E1 | 4.44–100 | 47.66 | 0.221 | −0.742 | 51.49 | 72.29 |
| 2013 | E2 | 1.67–91.11 | 34.31 | 0.555 | −0.527 | 68.99 | 73.72 |

Note: CV is the coefficient of variation; $H^2$ is broad-sense heritability.

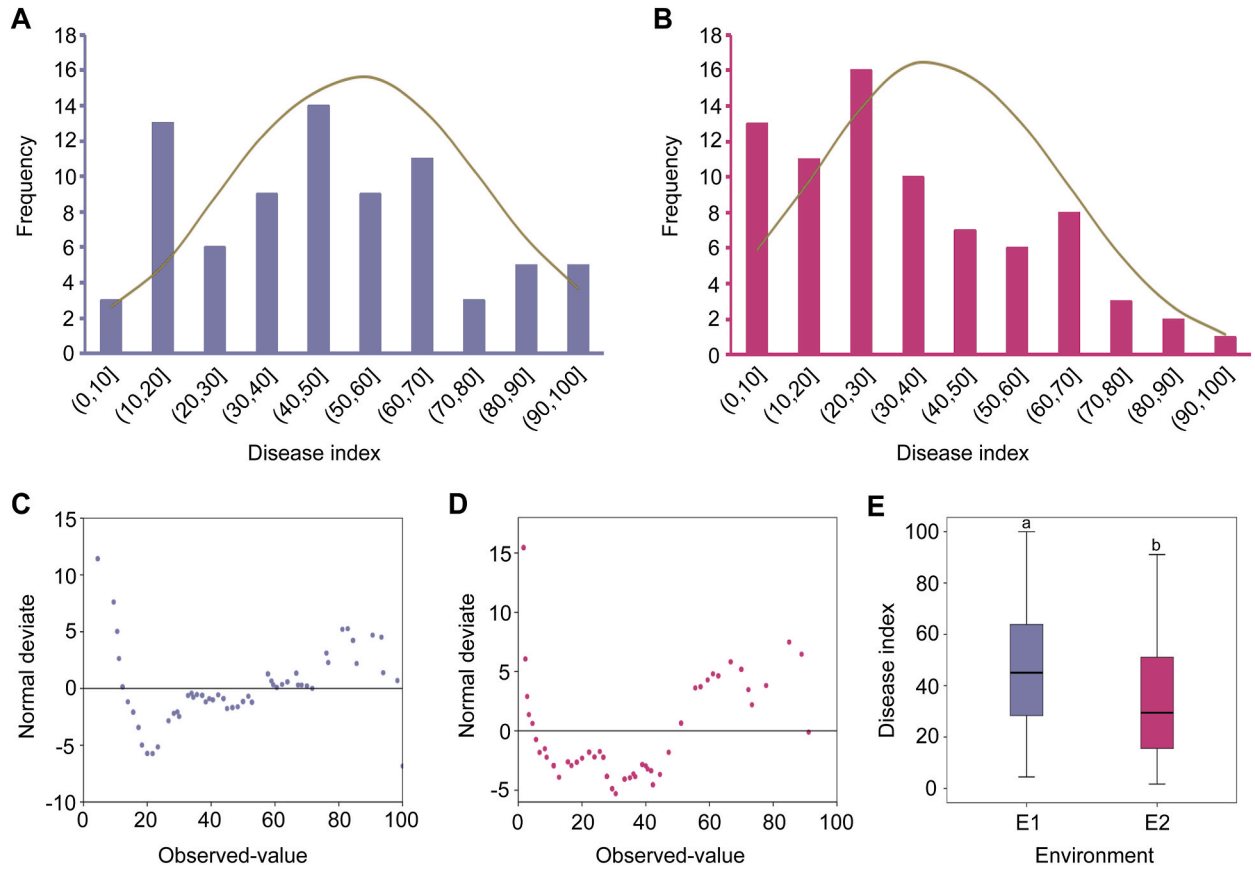


**Fig. 1.** Statistical analysis of the tobacco bacterial wilt disease index (TBWDI). Distribution of TBWDI in two experiments: (A) E1 (2011) and (B) E2 (2013). Normal distribution deviations for TBWDI in (C) E1 and (D) E2 were detected using the NORMDIST function, where the x-axis represents observed values of the disease index. (E) ANOVA results comparing TBWDI between E1 and E2. Different lowercase letters on the box plot indicate significant differences ($P < 0.01$).

($R^2 \geq 0.2$) accounted for 9.42 % of the total combinations, with very strong LD combinations ($R^2 > 0.8$) [37] comprising only 0.54 %. Most polymorphic loci exhibited weak LD ($0 < R^2 < 0.2$) or no LD ($R^2 = 0$), with combinations showing $R^2 < 0.1$ accounting for 82.22 % of the total (Fig. 3B; Table S3). These results suggest that the selected polymorphic molecular markers are highly specific and effective for capturing genomic information.

### 3.3. Association mapping to identify loci

Q-values and kinship coefficients were used as correction factors in association analysis to reduce false associations. The GLM_Q and MLM_Q + Kinship models were used to associate genetic loci with phenotype traits, verified through Q-Q plot analysis and λ-value calculations, respectively. The GLM_Q model exhibited weak alignment with expected values for the phenotype data from E1 (λ-value = 2.5434) (Fig. 4A), but better performance from E2 (λ-value = 1.5331) (Fig. 4B). In contrast, the MLM_Q + Kinship model aligned more reliably with expected values for both environments (λ-values of 0.6949 for E1 and 0.6565 for E2) (Fig. 4C and D). Therefore, we focused on marker loci identified by these methods.

Based on the marker loci identified using these methods (Fig. 5A; Table S4), the GLM_Q model detected 91 allelic variation loci associated with TBWDI in E2 (P-values: 4.88E-02 to 7.07E-05), explaining 5.08–19.10 % of the phenotypic variation. The MLM_Q + Kinship model detected 46 allelic variation loci in E1 (P-values: 4.94E-02 to 2.35E-03), explaining 5.25–14.22 % of the phenotypic variation and 34 loci in E2 (P-values: 4.98E-02 to 9.31E-04), explaining 5.36–16.37 % of the phenotypic variation. Notably, the GLM_Q and MLM_Q + Kinship models each detected 22 allelic variation loci in E2 (Fig. 5A; Table S5), the MLM_Q + Kinship model detected one allelic variation locus in both environments (Fig. 5A; Table S4), and the MLM_Q + Kinship (E1) and GLM_Q (E2) models detected eight allelic variation loci in different environments (Fig. 5A; Table S4). The 29 loci (Table S4) identified by multiple methods or in multiple environments were considered candidate loci for further analysis.

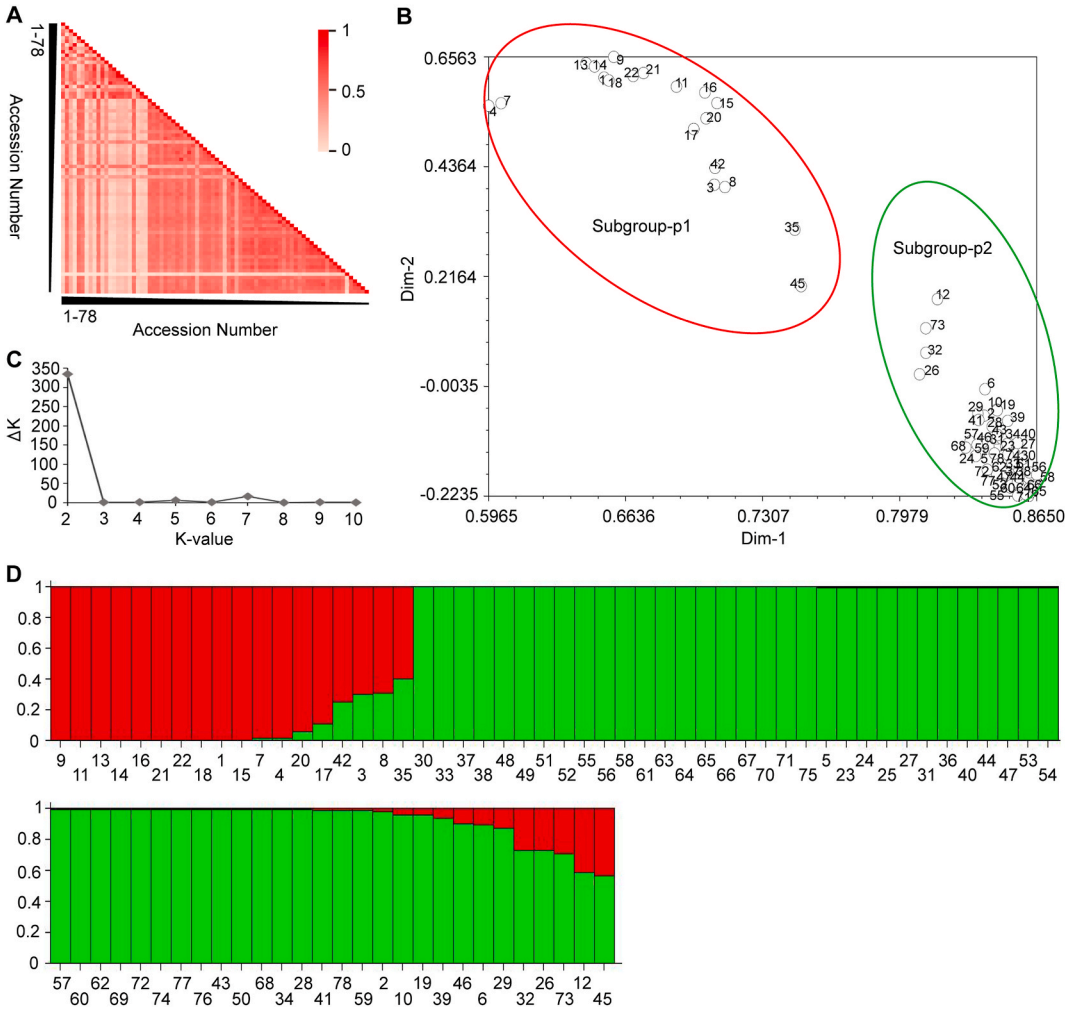


**Fig. 2.** Phylogenetic relationship and population classification analysis. (A) Phylogenetic coefficients between pairwise combinations of 78 tobacco materials. The triangular slant represents the material numbers in ascending order, with each rectangle representing the phylogenetic coefficient between two materials. Redder colors indicate stronger phylogenetic relationships. (B) Principal component analysis. Each circle represents a material, with numbers representing material codes. Dur to high affinity, some materials overlap and are not fully visible. Materials within the red circle are classified as Subgroup-s1, and those within the green circle as Subgroup-s2. The first and second principal components are labeled Dim-1 and Dim-2, respectively. Table S2 provides detailed classification information. (C) Population structure analysis. The maximum posterior probability value of the population structure is used, with clustering based on the maximum ΔK calculated as mean (|LnP(K–1)+|LnP(K+1)–2LnP(K)|)/sd (LnP (K). (D) Material classification results based on the population structure analysis. Red and green rectangles represent Subgroup-s1 and Subgroup-s2, respectively, with material numbers indicated below the rectangles.

### *3.4. Selection of reliable and stable loci*

Materials containing or lacking a candidate allelic variation locus were compared to identify reliable loci. The locus was deemed reliable if the average disease index of materials consistently increased or decreased compared to the overall average. Nineteen of the 29 loci (Fig. 5B; Table S5) met this criterion. None of these loci reached strong linkage levels.

Materials containing or lacking specific reliable loci were further analyzed using ANOVA. Significant differences ($P < 0.05$) occurred between materials containing or lacking the variation locus PT30311b (or InDel534 b or T1Y-93 b) in E1, with no significant differences observed in E2 (Fig. S2A; Table S5). Conversely, materials containing or lacking the variation locus InDel251 b (or T1Y-94a) significantly differed in E2 but not E1 (Fig. S2B; Table S5), indicating the influence of environmental factors. Notably, locus InDel261 b significantly differed in both environments (Fig. 5C; Table S5) and was also identified through QTL mapping (Fig. 5D) with LOD $\geq$2 [38,39] using the established linkage map (unpublished), accounting for more than 8 % of the phenotypic variation (Fig. 5D; Table S5). Moreover, in the 17.58 Kbp downstream of InDel261 b, a candidate gene (LOC107795335) encoding a *Nicotiana tabacum* sodium-coupled neutral amino acid transporter 5-like protein (NtAT5-like) on the scaffold358925 of tobacco whole genome shotgun sequence (available at GenBank under accession number AWOJ01S358925) [40] was discovered, and this gene was reported to be

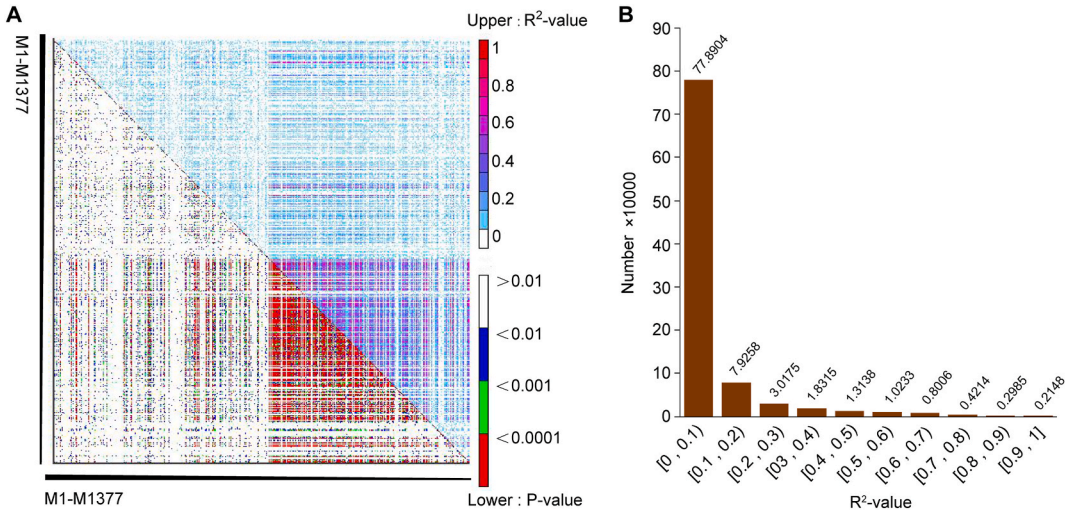


**Fig. 3.** (A) Linkage disequilibrium analysis of 1377 marker loci. The triangular slant represents the numbering of 1377 marker loci, sorted from small to large. The upper right corner corresponds to the $R^2$ value between pairwise marker loci, indicating the strength of linkage disequilibrium. The lower left corner represents the P-values of the $R^2$-value analysis, showing the confidence level of the detected linkage distribution. (B) Histogram of the distribution (X-axis) and frequency (Y-axis) of $R^2$ values.

associated with bacterial wilt resistance.

Using the NCBI and TAIR databases, we identified homologous proteins with high similarity to NtAT5-like: amino acid transporter NtAVT6-like (LOC104092802) in *Nicotiana tomentosiformis* (99.78 % similarity) and AtAVT6 (AT3G30390) in *Arabidopsis* (72 % similarity) (Fig. 6A). The predicted three-dimensional structures of NtAT5-like (Fig. 6B), NtAVT6-like (Fig. 6C) and AtAVT6 (Fig. 6D) further suggest that these proteins share similar functions.

Although other individual marker loci did not significantly differ across environments (Table S5), when evaluated together, as seen in locus InDel251 b and T1Y-94a, significant differences were observed in both environments (Fig. S2B; Table S5), indicating the cumulative effect of multiple genes and leading to the selection of 19 marker loci for subsequent material screening.

### *3.5. Highly resistant material screening and hybrid combination prediction*

Screening using the 19 marker loci identified 11 materials highly resistant to bacterial wilt (Table S6), all containing at least one excellent locus. Of the top materials, 'K326' and 'NC60' each contained 17 excellent loci but lacked the M383(Y) and M422(Y) excellent loci (Table S6). Their average TWBDI values across both environments were 23.61 and 14.17, respectively. Other notable materials included 'D101' and 'RG-22', each containing 15 excellent loci, with 'RG-22' containing the M422(Y) excellent locus. Their average TWBDI values across both environments were 3.06 ('D101') and 16.11 ('RG-22'), respectively. 'Yanyan97' contained 14 excellent loci, with an average TWBDI value across both environments of 10.

Interesting, none of the 11 materials (Table S6) possessed all the excellent loci, including M383(Y) and M422(Y). 'Dayemihe' with seven excellent loci, and 'Fengkai-10' with five excellent loci, had average TWBDI values of 12.78 and 8.33, respectively. In addition to 'RG-22', 'OX2028' also contained the M422(Y) locus (with 13 excellent loci), with an average TWBDI of 9.45.

Therefore, to harness more excellent loci and breed more stable resistant varieties, several hybrid combinations were proposed (Table 3), including 'K326' × 'RG-22', 'NC60' × 'RG-22', 'K326' × 'Dayemihe', 'NC60' × 'Dayemihe' and 'RG-22' × 'Dayemihe', for subsequent screening of superior resistant offspring.

## 4. Discussion

Traditional breeding processes often rely on breeders' experiences and preferences for selecting desired phenotypes. However, as most plant traits are quantitative [13], this approach can significantly extend the breeding cycle. In this study, a two-year analysis of TBWDI revealed that genetic resistance to tobacoo bacterial wilt disease is complex, with no materials exhibiting complete immunity. Kinship analysis further demonstrated that more than 90 % of the materials had a kinship coefficient below 0.5, indicating significant genetic variation among the loci of these materials, which likely influences TBWDI resistance. This genetic variation contributes to the wide distribution and variability observed in the disease index. Consequently, highly resistant candidate materials were identified among the 78 tobacco accessions, serving as valuable resources for identifying key genetic loci and expediting the breeding process for disease-resistant tobacco varieties.

Differences in geographical environments and natural selection pressures can cause subgroups to form within populations. Studies have shown that approximately 9.3 % of phenotypic variation can be attributed to population structure, leading to erroneous associations with non-functional loci [10]. In this study, UPGMA clustering, PCA analysis, and population structure analysis consistently

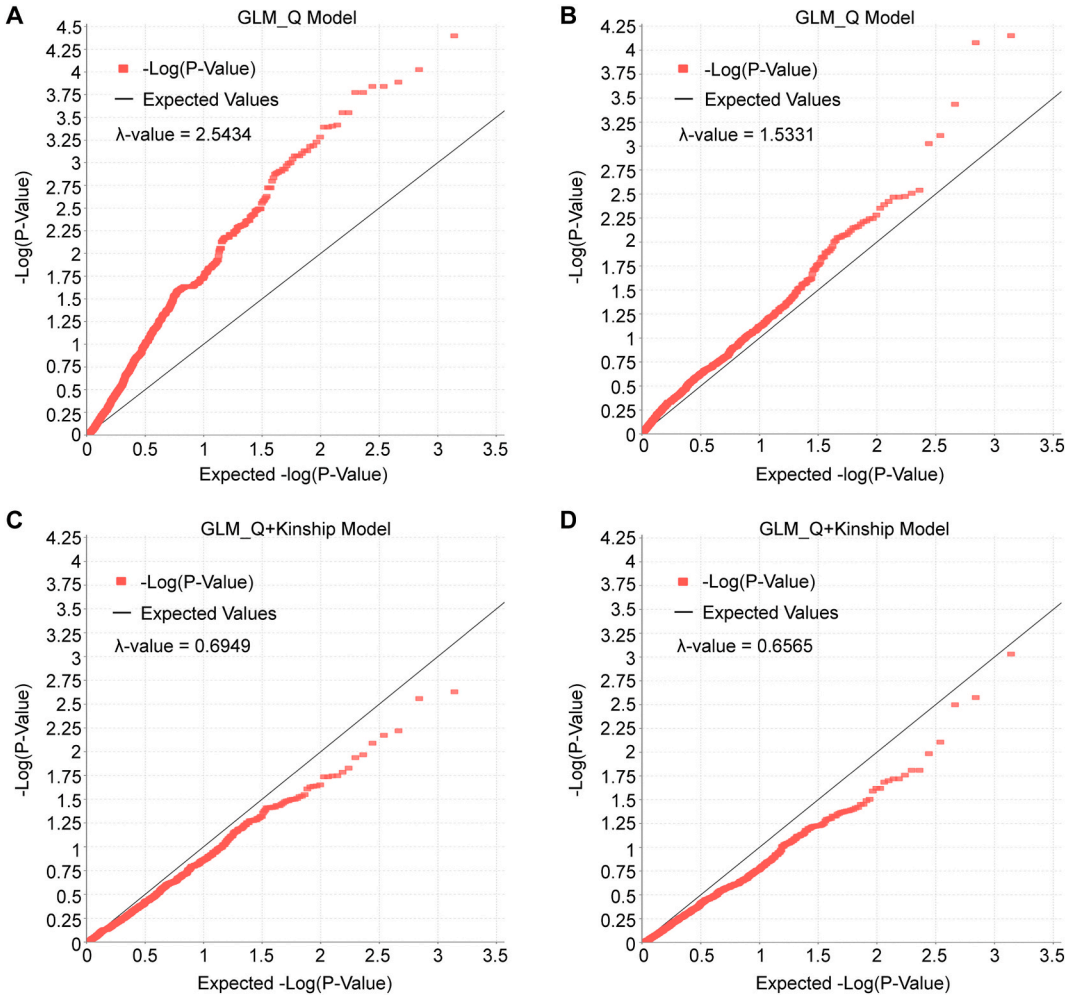


**Fig. 4.** Q-Q plot analysis of association models using the GLM_Q and MLM_Q + Kinship models based on phenotype values in two environments: (A, C) E1 (2011) and (B, D) E2 (2013), with the expected –log (P-value) on the X-axis and the observed –log (P-value) on the Y-axis. The closer the red squares are to the Y = X line, the higher the model's accuracy in predicting associations. The lambda (λ) value indicates the degree of inflation in the test statistics, with a value closer to '1', representing higher model reliability and less inflation.

divided the population into two subgroups, with only 'Zhongyan14' exhibiting classification discrepancies. This consistency suggests a relatively simple population structure, reducing the risk of false positives in association analyses [41]. Thus, the Q-value correction coefficient derived from population structure analysis can be reliably to improve the accuracy of subsequent association analyses.

The coefficient of relatedness is an important factor affecting the accuracy of association analyses [35], and different association methods can detect reliable marker loci in multiple environments [42,43]. Considering the correction abilities of both Q-values and kinship values, this study used two association models, GLM_Q and MLM_Q + Kinship, to analyze the correlation between TBWDI and marker loci in two environments. This approach accounted for environmental factors, selection thresholds, and phenotypic values. We identified 19 reliable variant loci associated with bacterial wilt resistance, providing a foundation for molecular marker-assisted selection of resistant tobacco materials. Unlike previous studies that relied on linkage analysis and only used two parental materials [1,7,8], this study's use of diverse materials allowed for higher mapping accuracy for precisely positioning individual marker loci [44]. Interestingly, one QTL identified through QTL mapping (Fig. 5D) was associated with the marker InDel261, which was consistent with the findings of this study, further confirming the reliability of this associated marker.

Notably, a candidate gene (LOC107795335) located 17.58 Kbp downstream of InDel261 was identified. This gene encodes a protein homologous to NtAT5-like [40], and shares significant sequence and structural similarity with NtAVT6A-like and AtAVT6, a protein associated with disease resistance in *Arabidopsis* [45,46]. Therefore, NtAT5-like is highly likely to play an important role in disease resistance, particularly against TBW.

Finally, based on the 19 excellent loci, the best hybrid combinations with low disease indexes were predicted for both environments. The results indicated that one parent participated in multiple hybridizations, consistent with previous studies [34,43,47]. These newly identified excellent loci can be used in marker-assisted breeding to enhance tobacco resistance to bacterial wilt.

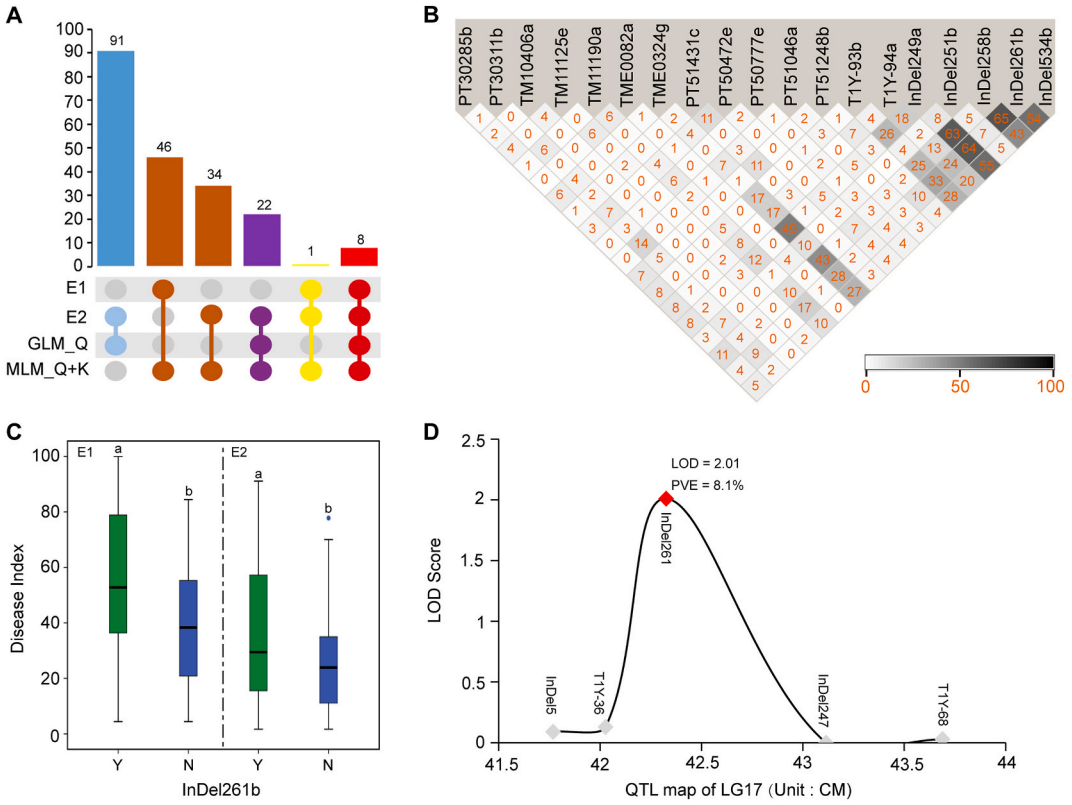


**Fig. 5.** Important marker loci related to tobacco bacterial wilt resistance. (A) Bar chart showing the number of marker loci associated with different environments and methods. Colored circles and connecting bars below each bar indicate the association model type used for the corresponding environment. Red circles represent the GLM_Q model for the disease index based on Experiment 1 (E1) in 2011 and the MLM_Q + Kinship model for the disease index based on Experiment 2 (E2) in 2013. (B) Pairwise linkage disequilibrium analysis between 19 stable marker loci. The numbers inside each rectangular box represent the $R^2$ value × 100, reflecting the strength of the linkage between pairs of loci. (C) Box plot comparing the disease index of materials with (denoted Y) or without (denoted N) the marker locus InDel261 b. ANOVA was performed, with different lowercase letters indicating significant differences ($P < 0.05$). (D) Partial LOD scan results of the previous linkage group 17 (LG17). QTL mapping based on the disease index of the RIL population in 2018 identified one QTL identified at LOD = 2.01 (red diamond), corresponding to the marker InDel261, with a phenotypic variation explained (PVE) of 8.1 %.

## 5. Conclusion

This study conducted an association analysis using SSR/InDel markers across 78 tobacco accessions to investigate TBW disease resistance. Two association models identified 19 reliable marker loci with 19 excellent loci that explained 5.37–19.10 % of their phenotypic variation for TBW disease resistance. Based on these loci, five hybrid combinations—'K326' × 'RG-22', 'NC60' × 'RG-22', 'K326' × 'Dayemihe', 'NC60' × 'Dayemihe' and 'RG-22' × 'Dayemihe'— were proposed for subsequent screening of superior resistant offspring. In addition, InDel261 b, one of the identified loci, was also detected as a QTL for TBW resistance through linkage analysis, and a candidate gene *NtAT5-like* (LOC107795335), was located nearby (downstream at 17.58 Kbp). The findings from this study provide valuable insights into the genetic architecture of TBW resistance and offer a foundation for marker-assisted selection to improve tobacco resistance in breeding programs.

## Funding

This work was funded by the Scientific and Technological Projects of Guangdong Tobacco Corporation (201702). The funder was not involved in the study design, data collection, data analysis, data interpretation, article writing, or the publication submission decision.

## Ethics approval and consent to participate

All methods were performed in accordance with the relevant guidelines and regulations. No specific permissions were required for the collection of plant material that was conducted in this study. Tobacco accessions collected from the United States of America (USA) and China were kindly provided by Mr. Weicai Zhao (Guangdong Research Institute of Tobacco Science, Shaoguan 512029, China).

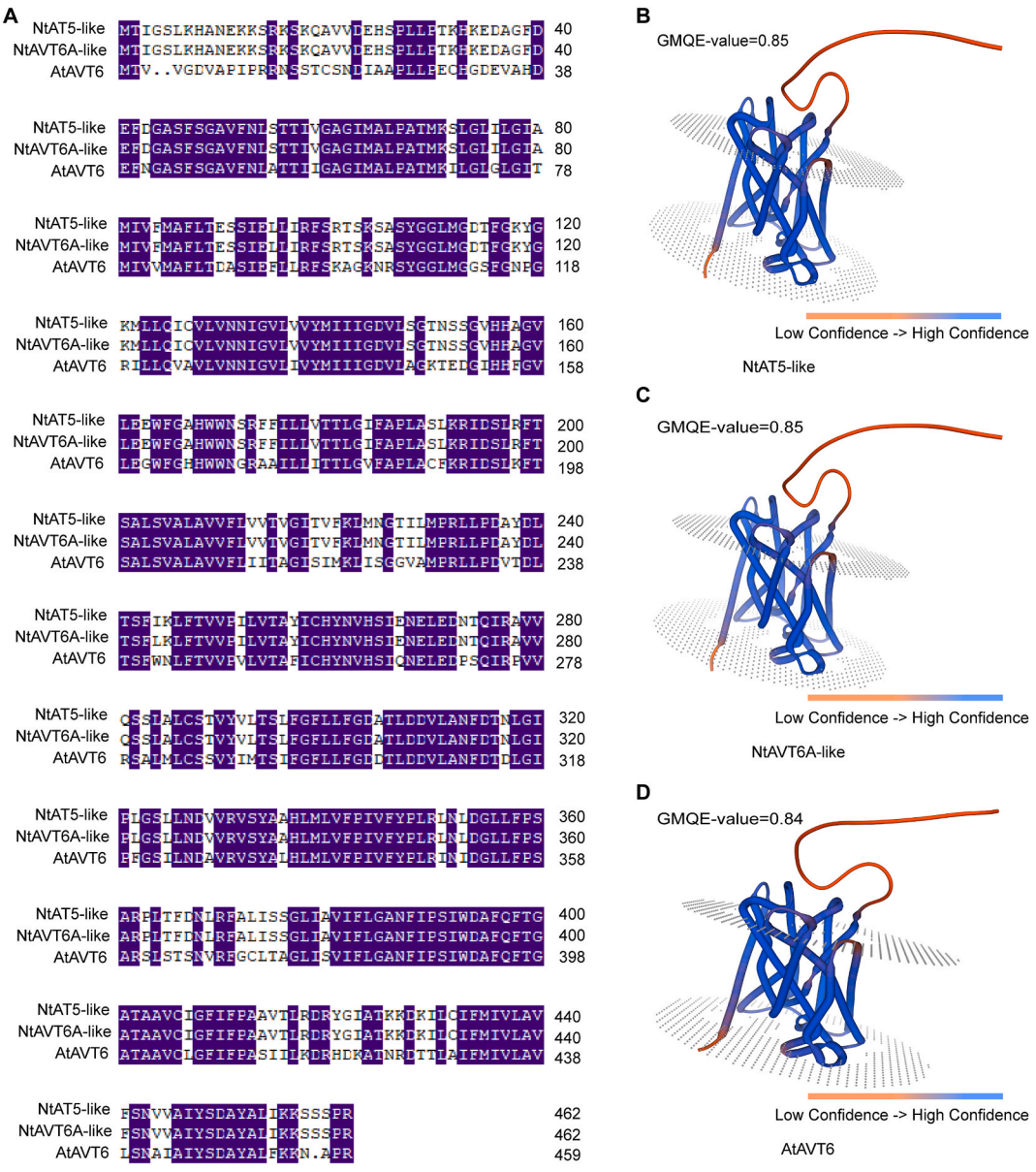


**Fig. 6.** Amino acid sequence alignment and protein three-dimensional structure prediction. (A) Amino acid sequence alignment of NtAT5-like, NtAVT6-like, and AtAVT6. Positions where all three proteins share the same amino acid are highlighted with a dark blue background. Three-dimensional structure prediction of (B) NtAT5-like and (C) NtAVT6-like and AtAVT6 using the Global Model Quality Estimation (GMQE) method. The GMQE value indicates the accuracy of the prediction, with values closer to 1 representing higher accuracy.

**Table 3**
Parental cross combinations for tobacco bacterial wilt resistance from superior alleles.

| P1 | P2 | P1-Phen. | P2-Phen. | P1-Exce. | P2-Exce. | Offspring loci |
|---|---|---|---|---|---|---|
| K326 | RG-22 | 23.61 | 16.11 | 17 | 15 | 18 |
| NC60 | RG-22 | 14.17 | 16.11 | 17 | 15 | 18 |
| K326 | Dayemihe | 23.61 | 12.78 | 17 | 7 | 18 |
| NC60 | Dayemihe | 14.17 | 12.78 | 17 | 7 | 18 |
| RG-22 | Dayemihe | 16.11 | 12.78 | 15 | 7 | 16 |

Note: Phen. is the average tobacco bacterial wilt disease index in two environments; Exce. is excellent loci.

## Consent for publication

Not applicable.

## Data availability

All data used in this study are included in the article and its supplementary materials.

## CRediT authorship contribution statement

**Ruiqiang Lai:** Writing – original draft, Software, Formal analysis. **Yanshi Xia:** Writing – review & editing, Validation, Software, Project administration, Methodology, Investigation, Formal analysis, Data curation. **Ronghua Li:** Writing – review & editing, Validation, Supervision, Software, Project administration, Methodology, Investigation, Formal analysis, Data curation. **Qinghua Yuan:** Writing – review & editing, Validation, Resources, Project administration, Funding acquisition, Data curation. **Weicai Zhao:** Writing – review & editing, Validation, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Data curation. **Kadambot H.M. Siddique:** Writing – review & editing, Validation, Software, Methodology. **Peiguo Guo:** Writing – review & editing, Writing – original draft, Supervision, Software, Resources, Methodology, Funding acquisition, Data curation.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgments

All authors would like to thank the Guangdong Research Institute of Tobacco Science for providing tobacco accessions.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heliyon.2024.e38939.

## References

[1] T. Lan, S. Zheng, L. Yang, S. Wu, B. Wang, S. Zhang, et al., Mapping of quantitative trait loci conferring resistance to bacterial wilt in tobacco (*Nicotiana tabacum* L.), Plant Breed. 133 (5) (2014) 672–677, https://doi.org/10.1111/pbr.12202.
[2] B. Lemaga, D. Siriri, P. Ebanyat, Effect of soil amendments on bacterial wilt incidence and yield of potatoes in southwestern Uganda, Afr. Crop Sci. J. 9 (1) (2001) 257–266, https://doi.org/10.4314/acsj.v9i1.27648.
[3] K. He, S.Y. Yang, H. Li, H. Wang, Z.L. Li, Effects of calcium carbonate on the survival of *Ralstonia solanacearum* in soil and control of tobacco bacterial wilt, Eur. J. Plant Pathol. 140 (4) (2014) 665–675, https://doi.org/10.1007/s10658-014-0496-4.
[4] K. Wu, S. Yuan, G. Xun, W. Shi, B. Pan, H. Guan, Root exudates from two tobacco cultivars affect colonization of *Ralstonia solanacearum* and the disease index, Eur. J. Plant Pathol. 141 (4) (2015) 667–677, https://doi.org/10.1007/s10658-014-0569-4.
[5] H. Zhang, M. Ikram, R. Li, Y. Xia, W. Zhao, Q. Yuan, et al., Uncovering the transcriptional responses of tobacco (*Nicotiana tabacum* L.) roots to *Ralstonia solanacearum* infection: a comparative study of resistant and susceptible cultivars, BMC Plant Biol. 23 (1) (2023) 620, https://doi.org/10.1186/s12870-023-04633-w.
[6] I. Mackay, W. Powell, Methods for linkage disequilibrium mapping in crops, Trends Plant Sci. 12 (2) (2007) 57–63, https://doi.org/10.1016/j.tplants.2006.12.001.
[7] T. Nishi, T. Tajima, S. Noguchi, H. Ajisaka, H. Negishi, Identification of DNA markers of tobacco linked to bacterial wilt resistance, Theor. Appl. Genet. 106 (4) (2003) 765–770, https://doi.org/10.1007/s00122-002-1096-9.
[8] Y. Qian, X. Wang, D. Wang, L. Zhang, C. Zu, Z. Gao, et al., The detection of QTLs controlling bacterial wilt resistance in tobacco (*N. tabacum* L.), Euphytica 192 (2) (2013) 259–266, https://doi.org/10.1007/s10681-012-0846-2.
[9] K. Drake-Stowe, N. Bakaher, S. Goepfert, B. Philippon, R. Mark, P. Peterson, et al., Multiple disease resistance loci affect soilborne disease resistance in tobacco (*Nicotiana tabacum*), Phytopathology 107 (9) (2017) 1055–1061, https://doi.org/10.1094/PHYTO-03-17-0118-R.
[10] S.A. Flint-Garcia, A.C. Thuillet, J. Yu, G. Pressoir, S.M. Romero, S.E. Mitchell, et al., Maize association population: a high-resolution platform for quantitative trait locus dissection, Plant J. 44 (6) (2005) 1054–1064, https://doi.org/10.1111/j.1365-313X.2005.02591.x.
[11] E.S. Ersoz, J. Yu, E.S. Buckler, Applications of linkage disequilibrium and association mapping in crop plants, in: R.K. Varshney, R. Tuberosa (Eds.), Genomics-Assisted Crop Improvement, Springer, Dordrecht., 2007, pp. 97–119, https://doi.org/10.1007/978-1-4020-6295-7_5.
[12] R. Lai, M. Ikram, R. Li, Y. Xia, Q. Yuan, W. Zhao, et al., Identification of novel quantitative trait nucleotides and candidate genes for bacterial wilt resistance in tobacco (*Nicotiana tabacum* L.) using genotyping-by-sequencing and multi-locus genome-wide association studies, Front. Plant Sci. 12 (2021) 744175, https://doi.org/10.3389/fpls.2021.744175.
[13] G. Yang, Y. Li, Q. Wang, Y. Zhou, Q. Zhou, B. Shen, et al., Detection and integration of quantitative trait loci for grain yield components and oil content in two connected recombinant inbred line populations of high-oil maize, Mol. Breed. 29 (2) (2012) 313–333, https://doi.org/10.1007/s11032-011-9548-z.
[14] G. Bindler, R. Van Der Hoeven, I. Gunduz, J. Plieske, M. Ganal, L. Rossi, et al., A microsatellite marker based linkage map of tobacco, Theor. Appl. Genet. 114 (2) (2007) 341–349, https://doi.org/10.1007/s00122-006-0437-5.
[15] G. Bindler, J. Plieske, N. Bakaher, I. Gunduz, N. Ivanov, R. van der Hoeven, et al., A high density genetic map of tobacco (*Nicotiana tabacum* L.) obtained from large scale microsatellite marker development, Theor. Appl. Genet. 123 (2) (2011) 219–230, https://doi.org/10.1007/s00122-011-1578-8.

[16] Z. Tong, Z. Yang, X. Chen, F. Jiao, X. Li, X. Wu, et al., Large-scale development of microsatellite markers in *Nicotiana tabacum* and construction of a genetic map of flue-cured tobacco, Plant Breed. 131 (5) (2012) 674–680, https://doi.org/10.1111/j.1439-0523.2012.01984.x.
[17] H. Li, M. Ikram, Y. Xia, R. Li, Q. Yuan, W. Zhao, et al., Genome-wide identification and development of InDel markers in tobacco (*Nicotiana tabacum* L.) using RAD-seq, Physiol. Mol. Biol. Plants 28 (5) (2022) 1077–1089, https://doi.org/10.1007/s12298-022-01187-3.
[18] H. Zhang, M. Ikram, R. Li, Y. Xia, W. Zhao, Q. Yuan, et al., Uncovering the transcriptional responses of tobacco (*Nicotiana tabacum* L.) roots to *Ralstonia solanacearum* infection: a comparative study of resistant and susceptible cultivars, BMC Plant Biol. 23 (1) (2023) 620, https://doi.org/10.1186/s12870-023-04633-w.
[19] T.S. Kumar, P. Rani, V.R. Hemanth Kumar, S. Samal, S. Parthasarathy, M. Ravishankar, Quality of labor epidural analgesia and maternal outcome with levobupivacaine and ropivacaine: a double-blinded randomized trial, Anesth. Essays Res. 11 (1) (2017) 28–33, https://doi.org/10.4103/0259-1162.194573.
[20] W. Chen, X. Li, X. Zhang, Z. Chachar, C. Lu, Y. Qi, et al., Genome-wide association study of trace elements in maize kernels, BMC Plant Biol. 24 (1) (2024) 724, https://doi.org/10.1186/s12870-024-05419-4.
[21] D. Stöckl, L.M. Thienpont, About the z-multiplier in total error calculations, Clin. Chem. Lab. Med. 46 (11) (2008) 1648–1649, https://doi.org/10.1515/cclm.2008.309.
[22] S. Porebski, L.G. Bailey, B.R. Baum, Modification of a CTAB DNA extraction protocol for plants containing high polysaccharide and polyphenol components, Plant Mol. Biol. Rep. 15 (1) (1997) 8–15, https://doi.org/10.1007/BF02772108.
[23] L. Huang, X. Deng, R. Li, Y. Xia, G. Bai, K.H.M. Siddique, et al., A fast silver staining protocol enabling simple and efficient detection of ssr markers using a non-denaturing polyacrylamide gel, J. Vis. Exp. 2018 (134) (2018) 57192, https://doi.org/10.3791/57192.
[24] T. Zerr, S. Henikoff, Automated band mapping in electrophoretic gel images using background information, Nucleic Acids Res. 33 (9) (2005) 2806–2812, https://doi.org/10.1093/nar/gki580.
[25] S.J. Jang, K.J. Sa, Z.Y. Fu, J.K. Lee, Association mapping analysis for cultivated and weedy types of Perilla crop collected from South Korea using morphological characteristics and SSR markers, Heliyon 10 (5) (2024) e26720, https://doi.org/10.1016/j.heliyon.2024.e26720.
[26] G. Evanno, S. Regnaut, J. Goudet, Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study, Mol. Ecol. 14 (8) (2005) 2611–2620, https://doi.org/10.1111/j.1365-294X.2005.02553.x.
[27] P.J. Bradbury, Z. Zhang, D.E. Kroon, T.M. Casstevens, Y. Ramdoss, E.S. Buckler, TASSEL: software for association mapping of complex traits in diverse samples, Bioinformatics 23 (2007) 2633–2635, https://doi.org/10.1093/bioinformatics/btm308.
[28] J.C. Barrett, B. Fry, J. Maller, M.J. Daly, Haploview: analysis and visualization of LD and haplotype maps, Bioinformatics 21 (2) (2005) 263–265, https://doi.org/10.1093/bioinformatics/bth457.
[29] L. Paternoster, D.M. Evans, E.A. Nohr, C. Holst, V. Gaborieau, P. Brennan, et al., Genome-wide population-based association study of extremely overweight young adults - the GOYA Study, PLoS One 6 (9) (2011) e24303, https://doi.org/10.1371/journal.pone.0024303.
[30] W.M. Otte, C.H. Vinkers, P.C. Habets, D.G.P. van Ijzendoorn, J.K. Tijdink, Analysis of 567,758 randomized controlled trials published over 30 years reveals trends in phrases used to discuss results that do not reach statistical significance, PLoS Biol. 20 (2) (2022) e3001562, https://doi.org/10.1371/journal.pbio.3001562.
[31] Y.H. Benjamini, Y. Hochberg, Controlling the false discovery rate - a practical and powerful approach to multiple testing, J. Roy. Stat. Soc. B: Methodological 57 (1995) 289–300, https://doi.org/10.1111/j.2517-6161.1995.tb02031.x.
[32] R. Lai, J. Jiang, J. Wang, J. Du, J. Lai, C. Yang, Functional characterization of three maize SIZPIAS-type SUMO E3 ligases, J. Plant Physiol. 268 (2022) 153588, https://doi.org/10.1016/j.jplph.2021.153588.
[33] T. Zhang, N. Qian, X. Zhu, H. Chen, S. Wang, H. Mei, et al., Variations and transmission of QTL alleles for yield and fiber qualities in upland cotton cultivars developed in China, PLoS One 8 (2) (2013) e57220, https://doi.org/10.1371/journal.pone.0057220.
[34] H. Zhu, R. Lai, W. Chen, C. Lu, C. Zaid, S. Lu, et al., Genetic dissection of maize (*Zea mays* L.) trace element traits using genome-wide association studies, BMC Plant Biol. 23 (2023) 631, https://doi.org/10.1186/s12870-023-04643-8.
[35] J. Yu, G. Pressoir, W.H. Briggs, I.V. Bi, M. Yamasaki, J.F. Doebley, et al., A unified mixed-model method for association mapping that accounts for multiple levels of relatedness, Nat. Genet. 38 (2) (2006) 203–208, https://doi.org/10.1038/ng1702.
[36] A.Y. Al-Maskri, M. Sajjad, S.H. Khan, Association mapping: a step forward to discovering new alleles for crop improvement, Int. J. Agric. Biol. 14 (1) (2012) 153–160, https://doi.org/10.13140/2.1.1925.9524.
[37] M.S. Phillips, R. Lawrence, R. Sachidanandam, A.P. Morris, D.J. Balding, M.A. Donaldson, et al., Chromosome-wide distribution of haplotype blocks and the role of recombination hot spots, Nat. Genet. 33 (3) (2003) 382–387, https://doi.org/10.1038/ng1100.
[38] J. Zhuang, H. Lin, J. Lu, H. Qian, S. Hittalmani, N. Huang, et al., Analysis of QTL x environment interaction for yield components and plant height in rice, Theor. Appl. Genet. 95 (1997) 799–808, https://doi.org/10.1007/s001220050628.
[39] R. Raman, S. Diffey, J. Carling, R.B. Cowley, A. Kilian, D.J. Luckett, et al., Quantitative genetic analysis of grain yield in an Australian *Brassica napus* doubled-haploid population, Crop Pasture Sci. 67 (4) (2016) 298–307, https://doi.org/10.1071/CP15283.
[40] N. Sierro, J.N.D. Battey, S. Ouadi, N. Bakaher, L. Bovet, A. Willig, et al., The tobacco genome sequence and its comparison with those of tomato and potato, Nat. Commun. 5 (2014) 3833, https://doi.org/10.1038/ncomms4833.
[41] W.C. Knowler, R.C. Williams, D.J. Pettitt, A.G. Steinberg, Gm3;5,13,14 and type 2 diabetes mellitus: an association in American Indians with genetic admixture, Am. J. Hum. Genet. 43 (4) (1988) 520–526. https://europepmc.org/article/PMC/1715499.
[42] Y.M. Zhang, Z. Jia, J.M. Dunwell, Editorial: the applications of new multi-locus GWAS methodologies in the genetic dissection of complex traits, Front. Plant Sci. 10 (2019) 00100, https://doi.org/10.3389/fpls.2019.00100.
[43] M. Ikram, R. Lai, Y. Xia, R. Li, W. Zhao, K. Siddique, et al., Genetic dissection of tobacco (*Nicotiana tabacum* L.) plant height using single-locus and multi-locus genome-wide association studies, Agronomy 12 (2022) 1047, https://doi.org/10.3390/agronomy12051047.
[44] S.A. Flint-Garcia, J.M. Thornsberry, E.S. Buckler, Structure of linkage disequilibrium in plants, Annu. Rev. Plant Biol. 54 (4) (2003) 357–374, https://doi.org/10.1146/annurev.arplant.54.031902.134907.
[45] Z.N. Yang, X.R. Ye, J. Molina, M.L. Roose, T.E. Mirkov, Sequence analysis of a 282-kilobase region surrounding the citrus Tristeza virus resistance gene (Ctv) locus in *Poncirus trifoliata* L. Raf, Plant Physiol. 131 (2) (2003) 482–492, https://doi.org/10.1104/pp.011262.
[46] P. Dhatterwal, S. Mehrotra, A.J. Miller, R. Aduri, R. Mehrotra, Effect of ACGT motif in spatiotemporal regulation of AtAVT6D, which improves tolerance to osmotic stress and nitrogen-starvation, Plant Mol. Biol. 109 (1–2) (2022) 67–82, https://doi.org/10.1007/s11103-022-01256-x.
[47] X. Han, Z.R. Xu, L. Zhou, C.Y. Han, Y.M. Zhang, Identification of QTNs and their candidate genes for flowering time and plant height in soybean using multi-locus genome-wide association studies, Mol. Breed. 41 (6) (2021) 39, https://doi.org/10.1007/s11032-021-01230-3.