Received: 13 August 2019 | Accepted: 24 September 2019

DOI: 10.1111/pbr.12773

ORIGINAL ARTICLE

Plant Breeding WILEY

Single-marker and haplotype-based association analysis of anthracnose (*Colletotrichum dematium*) resistance in spinach (*Spinacia oleracea*)

Henry O. Awika[1] | Kimberly Cochran[2,3] | Vijay Joshi[2,4] | Renesh Bedre[1] | Kranthi K. Mandadi[1,3] | Carlos A. Avila[1,4]

[1]Texas A&M AgriLife Research, Weslaco, Texas

[2]Texas A&M AgriLife Extension, Uvalde, Texas

[3]Department of Plant Pathology and Microbiology, Texas A&M University, College Station, Texas

[4]Department of Horticultural Sciences, Texas A&M University, College Station, Texas

Correspondence
Carlos A. Avila, Texas A&M AgriLife Research and Extension Center, 2415 E. Hwy 83, Weslaco, TX 78596.
Email: Carlos.Avila@ag.tamu.edu

Funding information
Texas A&M AgriLife Vegetable Seed Grant, Grant/Award Number: FY17- and FY18-124353-96181

Communicated by: Henryk Flachowsky

Abstract

Anthracnose (*Colletotrichum dematium*) is an important disease in spinach (*Spinacia oleracea*). Sources of resistance must be identified, and molecular tools must be developed to expedite cultivar development. In this study, a diverse collection of 276 spinach accessions was scored for anthracnose disease severity. We then evaluated marker identification approaches by testing how well haplotype-based trait modelling compares to single markers in identifying strong association signals. Alleles in linkage disequilibrium were tagged in haplotype blocks, and anthracnose-associated molecular markers were identified using single-SNP (sSNP), pairwise haplotype (htP) and multi-marker haplotype (htM) SNP tagging approaches. We identified 49 significantly associated markers distributed on several spinach chromosomes using all methods. The sSNP approach identified 13 markers, while htP identified 24 (~63% more) and htM 34 (~162% more). Of these markers, four were uniquely identified by the sSNP approach, nine by htP and nineteen by htM. The results indicate that resistance to anthracnose is polygenic and that haplotype-based analysis may have more power than sSNP. Using a combination of these methods can improve the identification of molecular markers for spinach breeding.

KEYWORDS
anthracnose, *Colletotrichum dematium*, GWAS, haplotype, SNP, spinach

1 | INTRODUCTION

Spinach anthracnose (*Colletotrichum dematium*) is an economically important disease widespread across spinach (*Spinacia oleracea*)-growing regions of the USA and around the world (Correll, Black, Koike, Brandenberger, & Dalnello, 1994). Several anthracnose-causing *Colletotrichum* species have demonstrated considerable versatility in plant host preference and virulence in different environments. For example, the anthracnose pathogen species *C. fruticola* and *C. siamense* can both infect white jute, a fibre crop in China (Niu, Gao, Chen, & Qi, 2016; Sharma, Kumar, Weir, Hyde, & Shenoy, 2013). *C. brevisporum*, *C. gloeosporiodes* and *C. truncatum* cause pepper anthracnose disease in India and China (Liu, Tang, et al., 2016a). *Colletotrichum siamense* has been found on partridge tea, citrus and mango in China, India, South Africa and Brazil (Cheng et al., 2013; Liu et al., 2017; Liu, Chen, Liu, & Hou, 2018). Similarly, *C. dematium* has been isolated from a range of crops including common knotgrass in China, cowpea in South Africa, tomato fruits and mulberry in Argentina, the ornamental plant German Statice in Bulgaria, and

Abbreviations: GO, gene ontology; GWAS, genome-wide association study; h^2, heritability; htM, multi-marker haplotype tagging; htP, pairwise haplotype tagging; HW, Hardy–Weinberg; IBD, identity by descent; LD, linkage disequilibrium; MAF, minor allele frequency; MLM, mixed linear model; QTL, quantitative trait locus; SNP, single nucleotide polymorphism; sSNP, single SNP.

spinach in Australia and the USA (Correll et al., 1994; Smith & Aveling, 1997; Yoshida & Shirata, 1999; Dal Bello, 2000; Washington et al., 2006; Bobev, Jelev, Zveibil, Maymon, & Freeman, 2009; Liu, Jin, et al., 2016b). These studies illustrate that different species and strains of *Colletotrichum* may thrive in the same environment on the same host plant, on different host plant species, or in the case of *C. dematium*, on different host plants in a wide range of environments. No resistance sources have been characterized until now; therefore, sources of genetic resistance need to be identified and molecular tools need to be developed to expedite cultivar development.

Spinach is a diploid with 2n = 2X = 12 chromosome configuration (Ellis & Janick, 1960) and possesses a medium size genome (989 Mb; Arumuganathan & Earle, 1991), and can serve as a model to other Amaranthaceae species. Enhanced genomic mapping can be used to unlock our understanding of pathogenicity and host resistance for anthracnose and many other non-characterized diseases in spinach. However, limited genomic-disease trait characterization studies have been reported in spinach in the past. Pioneer genetic mapping techniques did not target spinach diseases but were important precursors for the current improved genomic capabilities for spinach. For instance, an earlier QTL mapping used 101 AFLP (amplified fragment length polymorphism) and nine microsatellite markers for sex linkage/QTL mapping (Khattak, Torp, & Andersen, 2006), SLAF (specific-locus amplified fragment-seq). This study preceded the use of single nucleotide polymorphic (SNP) variants for mapping genomic features associated with spinach diseases such as Stemphylium leaf spot (Shi, Mou, Correll, Koike, et al., 2016a) and verticillium wilt (Shi, Mou, Correll, Motes, et al., 2016b), albeit to a limited extent. Several of these SNP-based genomic mapping efforts have concentrated on single-marker regressions to identify association signals for quantitative trait locus (QTL).

Spinach is dioecious, with separate male and female plants, although monoecious plants can occasionally be found that contain both male and female flowers (Morelock & Correll, 2008). As the female and male parents used in spinach crosses are normally family pools of genetic material, obtaining a highly homozygous parent or inbred lines for linkage QTL studies is difficult. In this regard, association studies offer a good alternative since no biparental populations need to be developed.

There has been significant progress towards the sequencing of the first chromosome-anchored draft genome for spinach, and related genomic features have been organized into databases (Xu et al., 2017; Yang, Tan, & Zhu, 2016). Thus, as in many relatively well-studied crops, molecular procedures previously unavailable for scientists can now be fully integrated into spinach genomic studies. One such implementation is the use of physical map distance-based linkage disequilibrium (LD) and haplotype blocking to dissect the architecture of important traits. SNP tagging to haplotype regions of the genome can be extremely useful for testing associations with qualitative and quantitative traits. For example, haplotype tagging has shown superior power compared to single SNPs in QTL detection and mapping accuracy in multiple species (Calus et al., 2009). However, variable results have been reported from haplotypes versus single SNPs. Earlier simulation studies produced conflicting results, with some reporting that single-marker regression and identity by descent (IBD) showed greater power than haplotype-based mapping (Long & Langley, 1999) and others reporting no difference between single-marker and haplotype LD mapping (Zhao, Fernando, & Dekkers, 2007).

Although single-SNP-based regressions have been used successfully to identify markers associated with important traits in spinach, whether or not haplotyping can offer an added advantage over single-marker regression in association studies has not been explored. Several studies using empirical data have shown advantages of using haplotypes in LD mapping. For example, most haplotypes fall into a few classes with little evidence of recombination and thus can dramatically reduce the number of tests and hence the type I error rate (Zhao, Aranzana, et al., 2007). A number of studies have rationalized grouping SNPs into haplotype blocks because of the ability to improve power, robustness and accuracy of association mapping in humans (Gabriel et al., 2002), animals (pigs and cattle) (Karimi, Sargolzaei, JaB, & Schenkel, 2018; Meuwissen, Odegard, Andersen-Ranberg, & Grindflek, 2014) and plants (barley) (Lorenz, Hamblin, & Jannink, 2010). Using haplotypes may also capture more LD between haplotypes and causal variants (Braz et al., 2019; Wu et al., 2014); provide more power than single SNPs when an allelic series exists at a locus (Morris Rw, 2002); potentially capture epistatic interactions between variants within a haplotyped locus (Clark, 2004; Hamon et al., 2006); and even allow informed testing between two alleles showing IBD and clades of haplotype alleles by capturing information from evolutionary history (Calus et al., 2009; Zhao, 2003).

Since every association mapping population has a unique population history (both recent and ancient) that shapes its patterns of genetic variation, it is necessary to determine case by case which mapping method works best for each crop. Here, we have attempted to expand molecular breeding efforts in spinach by exploring how well haplotype-based SNP–trait modelling compares to single-SNP marker in identifying polymorphic signals associated with *C. dematium* disease resistance. To our knowledge, no previous study has explored the utility of haplotypes in genetic studies in spinach.

2 | MATERIALS AND METHODS

2.1 | Plant materials

Plant material was previously described by Shi, Mou, Correll, Koike, et al. (2016a); Shi et al. (2017); Awika et al. (2019). Briefly, the 276 spinach (*Spinacia oleracea* L.) accessions used consisted of a diverse panel originating from 33 countries that is part of a diversity collection maintained and provided by the USDA-National Plant Germplasm System (NPGS) at Ames, Iowa, USA. A detailed description of each accession can be found at https://npgsweb.arsgrin.gov/gringlobal/view2.aspx?dv=web_site_taxon_accessionlist&params=:taxonomyid=35256;:siteid=16. For the purpose of this study, we grouped the accessions into seven continental subregions based on countries of origin as follows: Region 1 (39 accessions from Afghanistan, Egypt,

Iran, Iraq and Syria); Region 2 (13 accessions from India, Nepal and Pakistan); Region 3 (58 accessions from the USA); Region 4 (1 accession from Ethiopia); Region 5 (29 accessions from Belgium, Denmark, France, Hungary, Italy, the Netherlands, Spain, Sweden and the United Kingdom); Region 6 (30 accessions from China, Japan, Korea, Mongolia, Taiwan and Thailand); and Region 7 (106 accessions from Georgia, Greece, Macedonia, Montenegro, Russia, Serbia and Turkey).

2.2 | Field design and growth environment

The 276 spinach accessions were grown at the Texas A&M AgriLife Research and Extension Center located in Uvalde, Texas, at 29°12′56″ latitude and 99°45°21″ longitude. Accessions were randomized in the field intercalating rows of the susceptible control cultivar 'Viroflay'. Each plot consisted of a single row of 30 plants at a distance of 4 cm between plants and 50 cm between plots. Seeds were sown manually, and plots were fertilized with a generalized N-P_2O_5-K_2O rate of 135–84–90 kg/ha. The spinach field was irrigated using an overhead irrigation system as needed.

2.3 | Inoculum preparation, inoculation and phenotyping

Plants were inoculated with a single spore isolate of *C. dematium* obtained from spinach plants in a commercial field near Uvalde, Texas, and maintained on potato dextrose agar media (Difco Laboratories Inc). To prepare inoculum, oat (Great Value Old Fashioned Oats, Walmart) was placed in 0.02-μm vented mycology mushroom spawn bags (53.34 cm × 20.96 cm × 12.07 cm, MycoSupply Inc) and autoclaved twice, 24 hr apart, at 121°C for 45 min and cooled to room temperature. After autoclaving, 200 ml of sterile water per kg of sterile oats was added to each inoculum bag. Each bag was then inoculated with approximately twelve 5 × 5 mm pieces of 10-day-old *C. dematium* colonies grown on potato dextrose agar (PDA) media. The inoculum bags were incubated at room temperature and shaken well, breaking up any clumps every day to distribute fungal growth evenly. The readiness of inoculum was determined when an even growth throughout the oats could be seen (~10 days).

At 21 days after spinach seedling emergence, each plant in a plot was inoculated uniformly with *C. dematium* by evenly sprinkling a level cup scoop of ~ 90 g oat inoculum and using gloved fingers on the seedling canopy. Prior to inoculation, an amount sufficient to inoculate all plots was bulked in a clean new five-gallon bucket and uniformly mixed, being careful to break up any clumps to ensure even size inoculum for dispersal. Field inoculation was done in the morning, while the morning dew was still present to enhance the initial inoculum–plant surface contact. Moisture was supplemented by light sprinkle irrigation every day (afternoon) for five days to ensure favourable conditions for disease development.

Anthracnose symptoms were evaluated using a disease severity rating (DSR) based on the proportion of surface area of fully formed leaves covered by anthracnose lesions per plant. The plants were examined perpendicular from top and at three angles from the sides. The surface area covered was averaged per plant and then totalled as a per cent on a plot basis. The per-plot percentages were then converted to nominal scores of 1 to 10, where a score of 1 denotes 0 to 10% infection, 2 denotes 11 to 20% infection, and so on up to a score of 10 denoting 91 to 100% infection. Genotypes were evaluated twice, at 14 days after inoculation (DAI) and at 28 DAI. A mean of the two measurements was reported for each plot.

2.4 | Genotyping by sequencing and SNP calling

Genotyping and SNP calling were performed as described by Awika et al. (2019). Briefly, the 276 spinach accessions were genotyped by sequencing using ddRADseq methodology (Peterson, Weber, Kay, Fisher, & Hoekstra, 2012). Illumina short-read sequencing (HiSeq 2500) and demultiplexing using individual indexes were performed by the Texas A&M AgriLife Genomics and Bioinformatics services. Filtered and adapter-trimmed sequence reads containing >5% uncalled bases and with average quality score ≤ 20 were discarded using a pipeline developed in-house using Python programming (https://github.com/reneshbedre/RseqFilt) (Bedre et al., 2015). High-quality cleaned sequence data were aligned to the draft spinach reference genome (v1) (Xu et al., 2017) using the bowtie2 alignment tool (Langmead & Salzberg, 2012). Sequences from the 276 spinach accessions were then run on Stacks (v1.48) (Catchen, Hohenlohe, Bassham, Amores, & Cresko, 2013; Rochette & Catchen, 2017) using pstacks, cstacks, sstacks and rxstacks modules to identify and filter the genotypes (Catchen et al., 2013). The Ada cluster from the TAMU High Performance Research Computing Center (http://hprc.tamu.edu/) was used to perform bioinformatics analysis. The SNP pipeline-end cleanup criteria also included removal of SNPs not anchoring to the six published draft chromosome sequences (Xu et al., 2017). A minimum minor allele frequency (MAF) of > 0.05 was used. The 6,167 resulting biallelic SNPs in VCF v4.2 (Danecek et al., 2011) were used for downstream analyses. SNP by spinach accession data can be downloaded from the Table S1 reported in Awika et al., 2019.

2.5 | Linkage disequilibrium (LD) and LD blocks

We used pairwise comparison of each marker pair by computing the squared correlation coefficient (r^2) of their allele population frequencies. As evidence of no historical recombination, the Hardy–Weinberg (HW) equilibrium cut-off *p*-value for each marker pair was set at .001, which is the probability that their frequency deviation from HW equilibrium could be explained by chance (Barrett, Fry, Maller, & Daly, 2005; De Bakker et al., 2005). Genotypes (individual or singletons in our case) not meeting a minimum 75% threshold of non-missing genotypes for each test marker were excluded. The minimum MAF was set at 0.001. To reduce duplicity of marker positions, where two or more markers in an input file had the same chromosomal position, the marker with the least-complete genotype was highlighted in the markers panel and ignored by default. Pairwise comparison between any two markers was ignored at >500 kb apart.

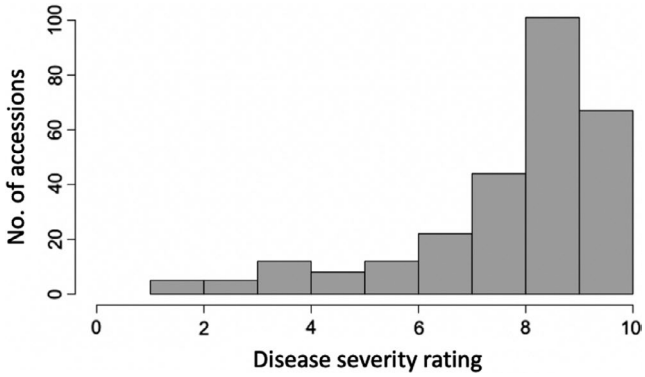

FIGURE 1 Anthracnose disease symptoms are distributed across all index values on evaluated spinach accessions. Rating scale of symptomatic leaf tissue was as follows: 1 = 1%–10%; 2 = 11%–20%; 3 = 21%–30%; 4 = 31%–40%; 5 = 41%–50%; 6 = 51%–60%; 7 = 61%–70%; 8 = 71%–80%; 9 = 81%–90%; 10 = 91=100%. Accessions with disease index of 1–2 were considered as resistant; those with disease index of 3–10 were considered as susceptible

TABLE 1 Summary data for the three tag–trait association tests

	sSNP	htP	htM
Number of singletons	276	274	276
Number of tagged alleles	6,167	861	2057
Number of blocks	–	532	971
Average number of alleles tagged per block	–	1.62	2.12
Total SNPs for GWA (alleles captured by tests)	6,167	4,231	3,848
Fraction of alleles captured by tests	1	1	0.91
Mean maximum r^2 for tests	–	0.80	0.99
Trait h^2	0.82	0.96	0.99
Power of H_a hypothesis test	0.81	0.84	0.87
Number of significant markers	19	26	35
Number of filtered significant markers[a]	13	24	34

[a]After co-anchored markers were removed.

The genotype set consisted of 276 singletons (treated as independent families) and zero trios. LD was declared when the r^2 between two alleles was at least 0.2. LD values were plotted and overlaid with other analysis tracks of genomic metadata in Haploview v4.2 (Barrett et al., 2005).

We adopted Haploview's internally developed "Solid Spine of LD" (Barrett et al., 2005) to search for a "spine" of strong LD running from one marker to another along the legs of the triangle in the LD chart. This meant that the first and last markers in a block were in strong LD with all intermediate markers, but that the intermediate markers may not necessarily be in LD with each other. Haplotypes for selected blocks were estimated using the accelerated expectation-maximum (EM) algorithm (Excoffier, 1995) similar to the partition/ligation method described by Qin, Niu, and Liu (2002). This creates highly accurate population frequency estimates of the phased haplotypes based on the maximum likelihood as determined from the unphased input.

2.6 | Tagging haplotypes

All 6,167 SNP markers from the single-marker SNPs were available for tag selection in haplotype blocks. To tag markers in haplotype blocks, two tagging strategies were implemented: pairwise marker (pairwise mode), to develop marker-pair-based haplotype tag SNPs (htP), or multiple marker (aggressive mode), to develop multi-marker-based haplotype tag SNPs (htM). Both the pairwise and aggressive methods are well described by De Bakker et al. (2005). Briefly, in both pairwise and multiple-marker strategies, a minimal set of markers were selected such that all alleles to be captured were correlated with a marker in that set at an r^2 greater than our user-defined 0.8 threshold. Where two or more SNPs had identical positions, the less-completely genotyped duplicate was automatically deselected. To avoid overfitting and unbounded haplotype tests in the downstream association phase, only those multiallelic combinations in which the alleles were themselves in strong LD as measured by a pairwise LOD score set at 3.0 were recorded, assigned genotypes and used as pairwise haplotype tags (htP). The multi-marker haplotype tagging (htM) strategy went two steps further by first using multi-marker tests constructed from the set of markers chosen as pairwise tags to try to capture SNPs that could not be captured in the pairwise step. It then replaced certain tags with multi-marker tests to map new multi-marker tags.

Parameters for marker checks were HW *p*-value cut-off .001, minimum genotype for each allele at 75% and minimum MAF of 0.001. In picking tags to declare a haplotype, possible SNP–SNP interactions that may occur within a gene or gene cluster were also recognized (Su et al., 2013). Since there is little evidence to suggest that epistasis occurs frequently between randomly chosen SNPs hundreds of kilobases apart (Lorenz et al., 2010), the maximum distance within which to declare a haplotype was set at 200 kb. A minimum distance between picked tags within a haplotype block was set at 60 bp to ensure only informative SNPs were included and to avoid overfitting the allele test models (Lorenz et al., 2010). All haplotype alleles within a block were displayed using AGCT notation, and only the polymorphic variants tagged within a block were applied in association tests.

2.7 | Population structure

The allelic ancestry-based admixture model (Alexander, Novembre, & Lange, 2009; Falush, Stephens, & Pritchard, 2007) was applied to account for population stratification bias for all the strategies (a total of 6,167 single markers, sSNP; 4,231 pairwise marker haplotype tags, htP; and 3,848 multi-marker haplotype tags, htM). Identity-by-state (IBS) similarity was also called, which assumed that two

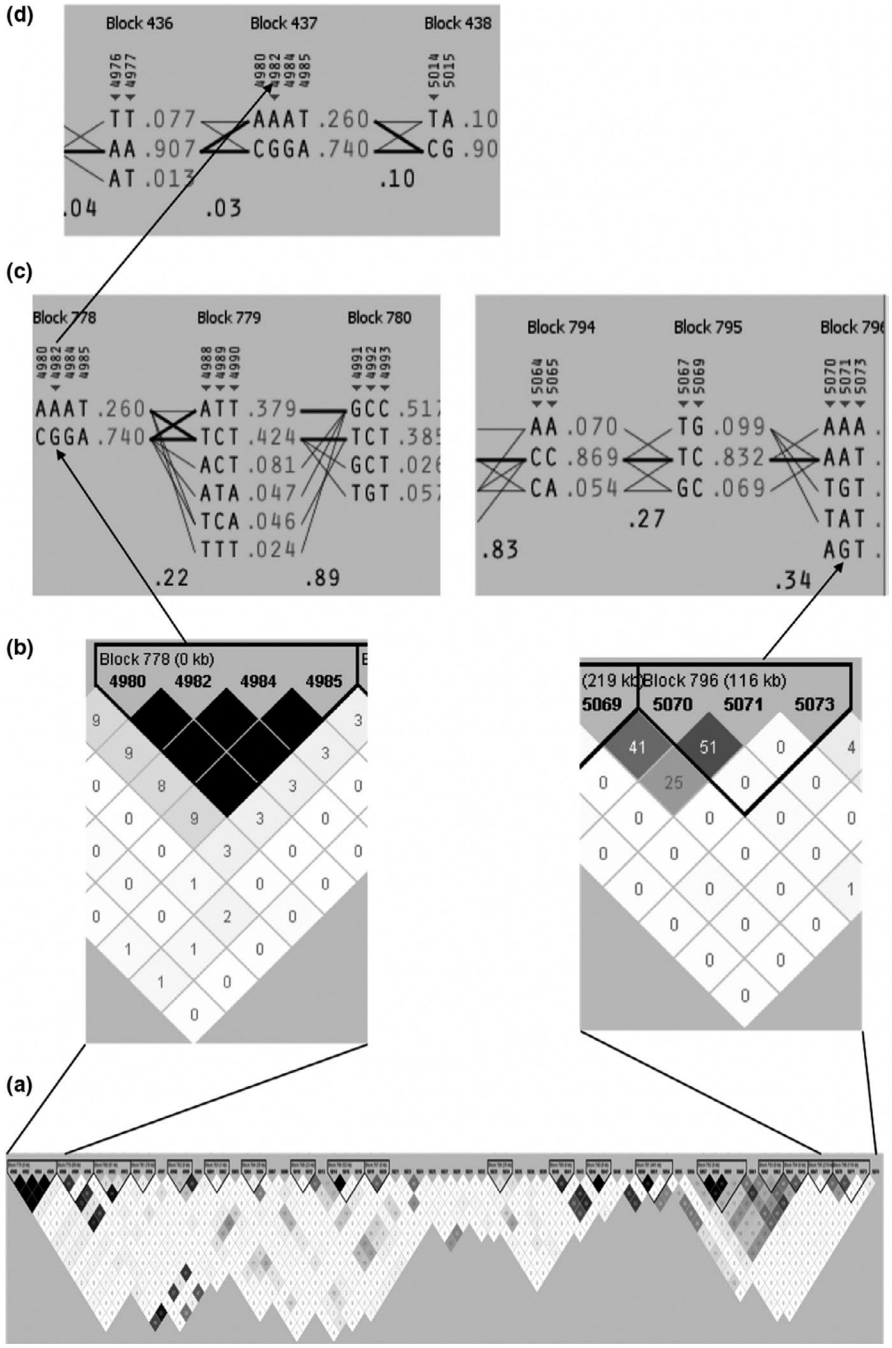

FIGURE 2 Graphical representation of the relationship between linkage disequilibrium (LD) blocks, the two haplotype tagging approaches, and SNP markers used for association studies. LD blocks are shown in solid spines of LD. (a) LD scheme generated by the multi-marker method spanning a ~ 460 kb section on chromosome 5; declared blocks are highlighted with a thick, black outline. (b) Enlargement of two htM blocks (block number and block size are shown; 0 kb means < 1 kb) harbouring alleles associated with a tagged marker. LD is represented by small squares with values ranging from 0 to 1 (decimals points are omitted to avoid clutter, e.g.: if 1 is shown in a square, it should be read as r^2 = .01, 8 is r^2 = .08, 25 is r^2 = .25, and so on); r^2 = 0 shown in white and complete LD, r^2 = 1 in black. (c) Display of 23 haplotypes in six blocks from the htM method, and (d) seven haplotypes in three blocks from the htP approach. Each haplotype in a block is shown with its population frequency and connections from one block to the next haplotype. Haplotypes within a block are read as rows, while alleles of a marker are read as columns. Black arrows connect a marker with its haplotype and block

random alleles drawn from the same locus are the same. The distance of an individual from itself was set to 0. The admixture model was computed using STRUCTURE (v2.3.4) (Pritchard, Stephens, & Donnelly, 2000). STRUCTURE was run with 10,000 burn-in periods and 16 replications on all SNPs in the three strategies (sSNP, htP and htM). The optimal (ΔK) and *K* population structure (*Q*) of each strategy were estimated using the Evanno method (Evanno, Regnaut, & Goudet, 2005). Briefly, to identify the most likely value of *K*, the rate of change in likelihoods between successive values of *K* was examined. To do this, the proportion of each subpopulation assigned to each cluster was first determined, and the run with the highest log likelihood among the runs for the best *K* value was identified. The optimal Δ*K* was used in determining the subgroup membership of each accession by 10,000 iterations for each *K* from 2 to 7 (7 being the number of arbitrary regions into which the countries of origin were grouped for the purpose of this study). The graphics were visualized in STRUCTURE HARVESTER (Earl & Vonholdt, 2012), which also applies the CLUMPP algorithm (Jakobsson & Rosenberg, 2007).

To account for possible hidden allele sharing (Blouin, 2003) that may bias association among the study population clusters, a kinship matrix (*K*) was implemented under the non-shrunk (Bradbury et al., 2007) context of realized relational matrix (Endelman & Jannink, 2012), because the number of markers was greater than the number of individuals genotyped. Monomorphic sites were removed before calculating kinship using the numerical genotype method (Endelman & Jannink, 2012).

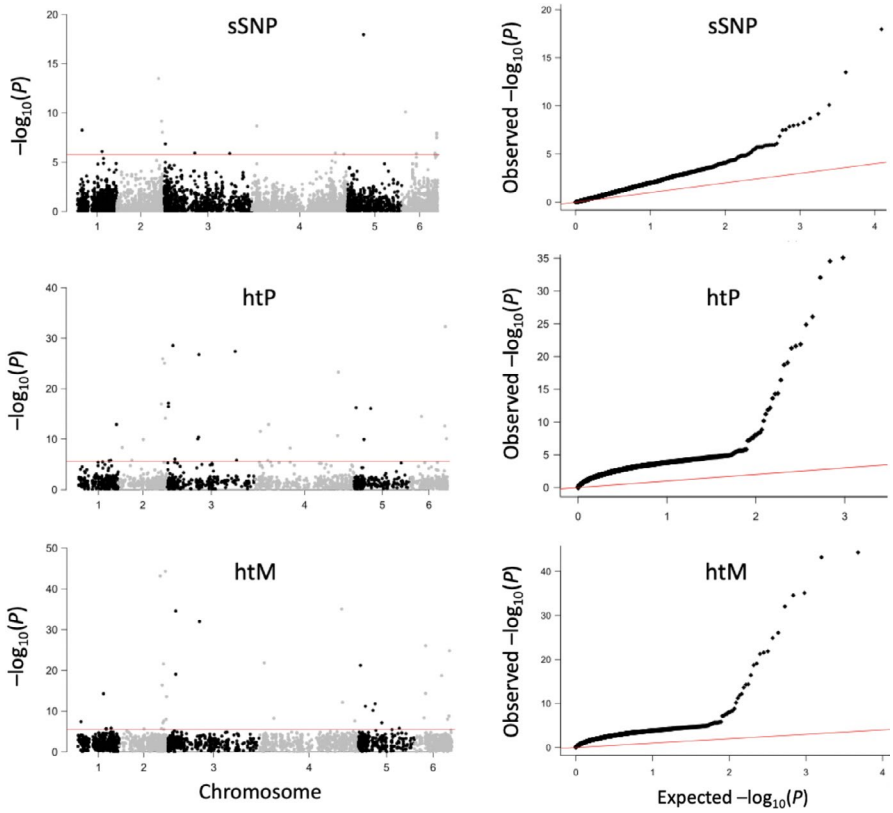

FIGURE 3 Total number of significant markers identified by the different association tagging methods. Manhattan plots (left) showing chromosome designations on the horizontal axis. Significance thresholds calculated using the Bonferroni multiple comparison method at 0.01 are shown by red horizontal lines. Markers above the threshold lines are considered significant association signals for the purpose of this study. QQ plots (right) showing P-value deviation of markers from expected values [Colour figure can be viewed at wileyonlinelibrary.com]

2.8 | Genome-wide association (GWA) tests

Single-locus, pairwise marker haplotype and multi-marker haplotype association tests were computed. The haplotype association tests were performed on the set of tagged SNPs in the haplotype blocks. Each model incorporated kinship into the *Anthracnose resistance = Population structure + Marker effect + Individuals +Residual* model. TASSEL 5.2.5 (Bradbury et al., 2007) (version released July 1, 2017; accessed October 27, 2018–December 25, 2018) was used to implement compressed mixed linear model (cMLM) and its complementary approach of "population parameters previously determined (P3D)" (Zhang et al., 2010). The cMLM decreases the effective sample size of such data sets by clustering individuals into groups, while P3D eliminates the need to recompute variance components for each marker. The P3D is also expected to reduce the dimensionality of the kinship matrix (K) and computational time, and to improve model fitting (Bradbury et al., 2007). Compression on *Q* + *K* models and P3D was applied for the sSNP, htP and htM SNPs with *Q* as the covariate and *K* matrix as the random coefficient. This also enabled estimation of genetic and residual variances for each combination at the trait level.

To estimate variance components for each marker, cMLM without the P3D was also implemented on *Q* + *K* for sSNP, htP and htM, allowing for each taxon to belong to its own group and to test each marker independently. Effects were determined for each marker irrespective of whether P3D was used or not. For each compression level, likelihood, genetic variance and error variance were determined, and the compression level with the lowest value of −2LnLk for the trait–compression combination was used for testing the markers. A genome-wide significance threshold was determined using the Bonferroni multiple correction method (Bonferroni, 1935) at 0.01 significance for sSNP, htP and htM. Markers not meeting the cut-offs were considered not significantly associated with anthracnose resistance in spinach in this study. Manhattan and QQ plots were made using the R statistical package, qqman (Turner, 2018). The genotypic data for the haplotype tags can be found in Table S2 for htP and Table S3 for htM; the genotypic information for the single SNPs (sSNPs) can be downloaded from a previous report (Awika et al., 2019).

2.9 | Heritability and variant (SNP) effects

Narrow-sense heritability (h^2) was calculated for each marker using the outputs from the compressed MLM without P3D model parameters. Per trait, h^2 was obtained from the mean genetic variance and the mean residual variance from the compression models with P3D, as additive variance (σ^2_a), divided by additive plus residual variances ($\sigma^2_a + \sigma^2_e$). For each compression level, likelihood, genetic variance and error variance were determined.

2.10 | Genomic context of polymorphic variants

The tagged SNP positions on chromosomes were used to query the JBrowse 1.15.3 (Buels et al., 2016; Skinner, Uzilov, Stein, Mungall, & Holmes, 2009) plugin in SpinachBase (http://www.spinachbase.org), and GO annotations were obtained in GOlr 2 frontend (Ashburner et al., 2000; Carbon et al., 2009; The_Gene_Ontology_Consortium, 2019). In silico genomic addresses of the tagged SNPs

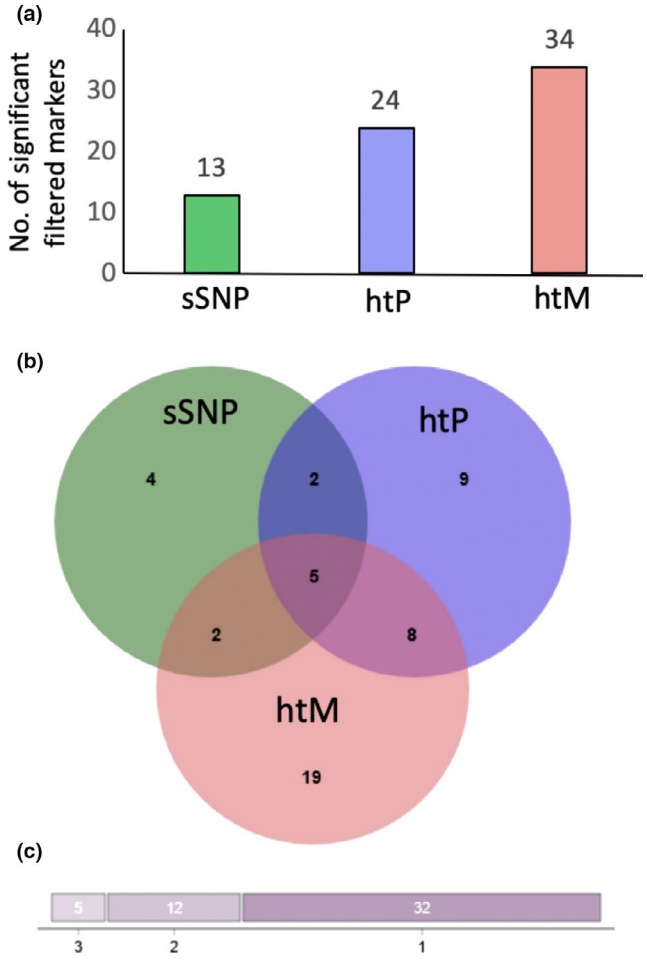

FIGURE 4 Association tagging methods identified unique and overlapped significant markers associated with anthracnose resistance. (a) Total number of significant filtered markers by haplotype strategy; (b) Venn representation of the intersection of the marker signals; **c** number of specific or shared markers identified

were used to locate genomic features anchored in the spinach draft reference genome. Only markers showing non-significant LD decay and meeting the significance thresholds in the GWA study were used. The gene at the associated polymorphic site itself is reported. In cases when no anchor gene was found at the associated site, genes within an arbitrary 200 kb upstream or downstream were reported as the representative gene. Putative functional annotations were determined through UniProtKB (The_Uniprot_Consortium, 2019).

2.11 | Ontology classification and pathway enrichment

Genes were classified into biological process (BP) and molecular function (MF) using the tool GOSlim terms in QuickGO (Binns et al., 2009) accessed on February 2, 2019. Pathway enrichment analysis for gene products was implemented at a 0.1 significance level on a Benjamini–Hochberg (B-H) (Benjamini & Hochberg, 1995) false discovery rate (FDR) correction threshold in SpinachBase. Metabolic pathway nodes were predicted and drawn in MetaCyc (Caspi et al., 2018). MetaCyc was selected because it is curated only with evidence-based data.

3 | RESULTS

3.1 | Anthracnose disease index shows a continuous distribution of infection severity

We classified the germplasm from the most resistant to the most susceptible based on visual disease severity rating (DSR) scheme of leaf surface area with anthracnose lesions. The DSR was determined on a mean basis of 30 plants in each spinach plot/accession. A plot with an average DSR of 1 or 2 was considered highly resistant (R) and a plot with DSR of 9 or 10, very susceptible (VS). Of the 276 accessions in this study, only five (2% of total) accessions (NSL 32629, PI 648954, NSL 28216, NSL 42771 and PI 648946), all originally sourced or developed in the USA (Table S1), were classified as highly resistant. By contrast, most of the spinach accessions were classified as susceptible, of which 168 (61%) among which included PI 604783, PI 648939, PI 648953, PI 648959, PI 648960, NSL 6092 from the USA, and PI 648962, PI 648964, PI 647852 PI 604784 from Afghanistan, China, Georgia and Mongolia, respectively, were VS (DSR = 9 or 10) (Table S1). The rest 103 (37%) accessions showed intermediate disease symptoms between the two extremes, R and VS (Figure 1 and Table S1). The presence of intermediate levels of disease symptoms suggests that the resistance is quantitative.

3.2 | Linkage disequilibrium, haplotypes and number of tests

Linkage disequilibrium (LD) indicates how correlated polymorphisms are due to their shared history of mutation and recombination (Flint-Garcia, Thornsberry, & Buckler Iv, 2003). Thus, positive selection of SNP in LD surrounding the causative locus of interest can lead to the maintenance of longer haplotypes at high frequencies within the population. The LD between marker pairs ranged from $r^2 = 0$ to $r^2 = 1$ ($SD = 0.147$), and LD block sizes ranged from 4 bp to 500 kb ($SD = 150{,}133$ bp). We declared 532 haplotype blocks in pairwise marker tests and 971 haplotype blocks for multi-marker tests (Table 1). To declare these haplotype blocks in the pairwise tests, we used 3,312 SNPs in the same number of tests. For example, 3,312 unique SNPs were chosen to create 3,312 tests, which could either be one of the sets of 3,312 SNPs or some combination of those SNPs. We tested 7,545 allele pair combinations and captured 4,231 of 4,231 (100%) alleles at a tagging r^2 threshold ≥ 0.8 (mean $r^2 = .993$). "Alleles captured" simply shows how many of the SNPs in the data set have been successfully tagged by the set of chosen tests, and the mean r^2 represents the mean for only those SNPs successfully captured.

For multi-marker haplotyping, 2,924 SNPs were used in the same number of tests, and 3,848 of 4,228 (91%) markers were tagged

TABLE 2 Markers mapped by marker tagging method

Tagged marker	Chr with htTAG	Position	Alleles		Single SNP (sSNP)						Pairwise haplotype (htP)						Multi-marker haplotype (htM)					
					p-value	Effect allele	Effect	MAF	PVE	h^2	*p*-value	Effect allele	Effect	MAF	PVE	h^2	*p*-value	Effect allele	Effect	MAF	PVE	h^2
44430_95	6	4,847,811	T	A	8.01E−11	A	−3.35	0.471	0.216	0.949	–	–	–	–	–	–	–	–	–	–	–	–
44220_132	6	44,775,432	C	A	1.07E−08	A	−3.36	0.469	0.181	0.833	–	–	–	–	–	–	–	–	–	–	–	–
32006_119	4	107,172,537	T	A	1.18E−06	A	−2.77	0.482	0.144	0.978	–	–	–	–	–	–	–	–	–	–	–	–
41566_67	6	18,775,069	C	T	1.32E−06	C	−2.06	0.475	0.144	0.629	–	–	–	–	–	–	–	–	–	–	–	–
26785_118	3	1,532,278	A	T	–	–	–	–	–	–	7.59E−18	A	−13.95	0.321	0.106	0.908	–	–	–	–	–	–
20941_21	1	48,227,888	C	G	–	–	–	–	–	–	1.27E−13	C	8.26	0.067	0.002	0	–	–	–	–	–	–
44219_78	6	44,775,486	C	T	–	–	–	–	–	–	7.95E−11	T	8.26	0.471	0.294	0.875	–	–	–	–	–	–
28140_95	3	37,770,803	G	A	–	–	–	–	–	–	8.94E−11	A	−5.91	0.395	0.368	0.943	–	–	–	–	–	–
37730_147	5	12,033,640	G	A	–	–	–	–	–	–	1.12E−10	A	16.12	0.142	0.002	0.922	–	–	–	–	–	–
34572_112	4	43,343,096	C	T	–	–	–	–	–	–	6.06E−09	T	11.73	0.089	0.003	0	–	–	–	–	–	–
21915_12	2	16,845,611	G	T	–	–	–	–	–	–	1.54E−06	G	6.18	0.172	0.001	0.833	–	–	–	–	–	–
20014_86	1	41,230,635	G	A	–	–	–	–	–	–	1.62E−06	A	16.99	0.104	0.062	0	–	–	–	–	–	–
19722_61	1	39,055,145	C	A	–	–	–	–	–	–	2.01E−06	C	7.26	0.202	0.001	0	–	–	–	–	–	–
19263_44	1	34,763,714	C	A	–	–	–	–	–	–	–	–	–	–	–	–	2.25E−06	A	44.96	0.236	0.167	0.656
19563_17	1	3,782,036	G	A	–	–	–	–	–	–	–	–	–	–	–	––	3.78E−08	A	−21.26	0.191	0.082	0.947
19962_15	1	40,835,316	C	A	–	–	–	–	–	–	–	–	–	–	–	–	1.51E−06	A	223.38	0.088	0.013	0.843
24234_63	2	50,772,717	G	T	–	–	–	–	–	–	–	–	–	–	–	–	6.21E−44	G	−7.12	0.464	0.148	0.976
24516_60	2	52,366,936	A	G	–	–	–	–	–	–	–	–	–	–	–	–	1.81E−06	G	93.84	0.175	0.039	0.975
24705_120	2	54,129,583	A	G	–	–	–	–	–	–	–	–	–	–	–	–	6.95E−08	G	2.78	0.408	0.271	0.933
24844_46	2	55,114,067	C	T	–	–	–	–	–	–	–	–	–	–	–	–	2.62E−08	C	18.36	0.165	0.082	0.496
25384_66	2	58,651,886	C	T	–	–	–	–	–	–	–	–	–	–	–	–	2.42E−14	T	−28.42	0.41	0.23	0
31191_147	3	9,539,285	T	A	–	–	–	–	–	–	–	–	–	–	–	–	8.65E−20	A	7.52	0.225	0.167	0.626
31597_16	4	101,831,638	T	A	–	–	–	–	–	–	–	–	–	–	–	–	8.16E−36	T	−4.87	0.16	0.037	0.861
33698_12	4	18,346,130	G	A	–	–	–	–	–	–	–	–	–	–	–	–	5.37E−09	G	−15.27	0.086	0.038	0.995
38278_107	5	17,702,567	C	T	–	–	–	–	–	–	–	–	–	–	–	–	6.56E−11	C	−630.59	0.12	0.132	0
39005_53	5	28,531,658	A	T	–	–	–	–	–	–	–	–	–	–	–	–	6.61E−08	T	−9.18	0.206	0.092	0.83
40099_56	5	50,109,435	C	T	–	–	–	–	–	–	–	–	–	–	–	–	1.59E−06	C	24.02	0.077	0.039	0
40948_89	5	8,362,430	A	G	–	–	–	–	–	–	–	–	–	–	–	–	6.26E−12	A	6.96	0.14	0.137	0.858
41186_110	6	12,079,239	C	T	–	–	–	–	–	–	–	–	–	–	–	–	4.26E−09	C	−25.61	0.184	0.151	0
41286_72	6	13,952,296	C	A	–	–	–	–	–	–	–	–	–	–	–	–	8.79E−27	C	−10.8	0.162	0.258	0
42874_64	6	33,655,811	T	C	–	–	–	–	–	–	–	–	–	–	–	–	1.96E−19	T	17.63	0.342	0.171	0
43777_80	6	41,015,140	T	C	–	–	–	–	–	–	–	–	–	–	–	–	8.74E−09	C	−20.6	0.289	0.35	0.933
28237_70	3	39,199,668	A	G	1.18E−06	A	1.3	0.094	0.117	0.959	1.74E−27	G	−7.95	0.094	0.3	0.999	–	–	–	–	–	––
30552_38	3	83,995,831	C	A	1.30E−06	C	−1.44	0.247	0.118	0.968	4.01E−28	C	6.21	0.247	0.497	0	–	–	–	–	–	–
18919_39	1	31,435,812	C	A	7.98E−07	A	−2.28	0.463	0.118	0.835	–	–	–	–	–	–	5.07E−15	C	3.17	0.463	0.239	0
32755_39	4	1.18E + 08	T	C	1.47E−06	T	1.44	0.189	0.13	0.844	–	–	–	–	–	–	2.50E−08	C	−5.03	0.189	0.104	0

(Continues)

TABLE 2 (Continued)

Tagged marker	Chr with htTAG	Position	Alleles		Single SNP (sSNP)						Pairwise haplotype (htP)						Multi-marker haplotype (htM)					
					p-value	Effect allele	Effect	MAF	PVE	h^2	*p*-value	Effect allele	Effect	MAF	PVE	h^2	*p*-value	Effect allele	Effect	MAF	PVE	h^2
24806_50	2	54,880,205	C	A	-	-	-	-	-	-	1.17E−26	A	16.53	0.433	0.309	0.972	2.73E−22	A	3.98	0.433	0.082	0.706
28190_146	3	38,721,303	A	T	-	-	-	-	-	-	2.95E−29	A	18.27	0.241	0.371	0.996	9.12E−33	T	−8.1	0.24	0.138	0.998
31191_149	3	9,539,283	C	A	-	-	-	-	-	-	1.96E−11	A	10.41	0.232	0	0.921	2.78E−35	C	−2.89	0.232	0.137	0.938
31716_82	4	1.03E + 08	A	G	-	-	-	-	-	-	1.25E−13	A	8.43	0.246	0.001	0.089	6.94E−13	A	−3.71	0.246	0.102	0
38635_43	5	2,323,118	C	T	-	-	-	-	-	-	5.95E−17	C	−21.72	0.097	0.134	0.98	5.45E−22	T	31.91	0.097	0.104	0.985
41284_58	6	13,942,768	A	G	-	-	-	-	-	-	3.02E−15	G	−14.02	0.17	0.068	0.904	4.22E−15	G	−24.92	0.17	0.034	0
44036_48	6	43,347,911	C	T	-	-	-	-	-	-	4.73E−33	T	−16.29	0.105	0.173	0.948	1.39E−25	T	−42.54	0.104	0.157	0.998
22624_143	2	30,891,579	C	T	-	-	-	-	-	-	1.14E−10	C	−2.82	0.48	0.023	0.68	2.15E−06	T	17.59	0.48	0.167	0.813
38421_39	5	20,353,750	C	T	1.10E−18	T	−8.47	0.448	0.26	0.958	7.95E−17	C	−3.62	0.448	0.179	0	1.41E−12	T	−7.1	0.448	0.138	0.925
24623_126	2	53,247,408	G	A	3.17E−14	G	0.49	0.33	0.155	0.888	1.09E−17	A	−9.15	0.331	0.517	0.929	3.91E−17	G	−25.34	0.33	0.099	0.917
35570_138	4	6,527,209	C	T	2.06E−09	T	−3.02	0.435	0.226	0.832	2.79E−12	T	−3.64	0.435	0.137	0.945	1.33E−22	C	−4.32	0.435	0.075	0.831
25305_48	2	58,152,897	T	A	8.86E−09	T	5.65	0.056	0.18	0.98	8.43E−26	T	4.44	0.056	0.29	0	4.68E−45	A	−21.31	0.056	0.179	0.068
43942_57	6	42,706,126	C	T	1.20E−06	C	−1.26	0.228	0.117	0	2.59E−13	C	5.63	0.229	0.304	0.75	1.54E−09	T	2.63	0.228	0.091	0

in marker–anthracnose resistance association tests (Table 1). The slightly smaller number of tagged markers in multi-marker tests (3,848 compared to 4,231 in pairwise marker tests) is likely due to the replacement of some markers chosen in the pairwise tests with multi-marker tests and higher failure rate of those tests to meet the set LOD cut-off. This was so, even though the number of blocks harbouring the tagged SNPs was larger for multi-marker tests than for pairwise marker tests. This observation points to the conclusion that multi-marker-based haplotypes reduce the number of tests required compared to both pairwise marker haplotypes and single-SNP testing. Figure 2 shows the relationship between LD blocks, haplotype tagging strategies htP and htM, and SNP markers on a ~ 460 kb section of chromosome 5.

3.3 | Population structure identified two main clusters

We used the SNP allele distribution to determine the population structure of the spinach panel. The admixture model grouped the 276 accessions into two population clusters (*Q*), as observed based on the highest ΔK (K = 2, Figure S1). The K stands for the number of clusters, while each cluster is denoted *Q*, hence *Q1* and *Q2* in the K = 2 clusters. However, there was a level of genetic admixture (Figure S1c) suggesting a level of historical interbreeding or from the shared ancestral genetic stock. We used the mixed ancestry method to account for this admixed structure.

3.4 | Haplotype SNP tagging approaches increase signal robustness

The three methods showed different levels of sensitivity and power in identifying significant association signals as reported previously (Lorenz et al., 2010) at a Bonferroni correction factor of 0.01 (Figure 3). After co-anchored markers were filtered, the htM approach identified the largest number of markers (34 markers) with strong association, followed by the htP strategy with 24 markers and the sSNP strategy with 13 markers (Figure 4a). From the three strategies, a total of 49 different SNP variants showed strong association with anthracnose resistance. Some of these signals showed up exclusively using only one method, while others were identified by two or all three methods (Figure 4b-c). For instance, of the 13 strong signals identified in the association study using sSNP mapping, only four were unique to sSNP: two were shared with the htP method and another two were shared with the htM methods; five markers were commonly identified by the two haplotype-based strategies (Figure 4b; Table 2). The haplotype htP versus htM methods shared eight strongly associated markers not identified by the sSNP approach. However, htM tagged 19 unique associated signals, while pairwise tagging (htP) only tagged nine association signals not identified by the other methods (Figure 4b). These observations show that haplotype block tagging of polymorphic markers improved the robustness of marker–phenotype signal detection compared to the single-SNP association study and suggest that even with a reduced

number of tests, the multi-marker haplotype method was more robust at detecting associations than the pairwise marker method in this study. The htP and htM spinach genotypic data per accession are shown in Tables S2 and S3.

3.5 | Models differ in minor allele frequency, allelic effect on the phenotypic variance explained

The higher-order allele clustering (htM) tagged the most unique minor alleles with effects (19 unique markers) with a 32.48 average per allele effect, which was above the mean for all significant minor allele effects (38 alleles, average effect 20.79). The average per allele effect of htM-tagged alleles was nearly 2.5 times as large as the average per allele effect for htP (13 alleles, effect 10.84) and nearly eight times that of sSNP (6 alleles, average effect 3.88) (Table 3). Even when the effect (223.38, in Table 2) of the minor allele *A*, marker *19962_15*, chr 1, located at 40,835,316 (generated by htM), was assumed to be an outlier and the value normalized by winsorization (Dixon, 1960; Hastings, Mosteller, Tukey, & Winsor, 1947), the average effects of the minor alleles became 27.79 for htM and 17.30 for the total, which were not substantial drops from the original means. This trend was reversed with regard to the MAF, r^2 and per marker h^2. Generally, the sSNP method produced the largest average of MAF, minor allele phenotypic variation explained (PVE, r^2) and h^2, followed by the htP method. Both sSNP and htP had above-average MAF, PVE and h^2, while htM had below-average MAF, PVE and h^2. This lower average MAF, PVE and h^2 may have been due to the larger variability observed of these association components in the haplotype-based approaches compared to the single-SNP approach. However, this increased variability did not alter the integrity of the markers detected by the haplotype approaches (hence the higher marker effects). Such variability is associated with rare alleles, and it is expected not to lead to false positives (Tabangin, Woo, & Martin, 2009).

3.6 | Anchored genes show a wide range of functions with putative roles in anthracnose resistance

The 49 polymorphic variants were anchored in or close to 49 genes (35 in genes and 14 within 30 kb of a gene). Of the 49 genes, 45 (~92%) code for known proteins and four (~8%) code for unknown proteins (Table 4). Considering each method (sSNP, htP and htM) separately, the markers identified were generally proportionately distributed across all six chromosomes (Table 2). Functional annotation reveals several biological processes and molecular functions putatively involved in host resistance such as those involved in plant stress responses and metabolic processes (Figure 5).

Each method separately or jointly identified genes implicated in plant–pathogen interactions and other important plant metabolic functions. For instance, the marker *31716_82* (on chr4) identified by htP and htM is a polymorphism of the gene *Spo07294*, an ADP-ribosylation factor-like protein implicated in effective plant host defences and in pathogenic bacteria–plant host interactions (Feng, Liu, Shan, & He, 2016). The marker *25201_21* (chr2), identified by sSNP only, is in the gene *Spo21787* that codes for FAD (flavin adenine dinucleotide) synthase and is important in plant defence signalling, while the marker *43942_57*, identified by all the three methods (sSNP, htP and htM), is in the putative susceptibility factor gene *Spo09720* encoding beta-amylase, an enzyme present in fungi, bacteria and plants that may be used by microbes to hydrolyse insoluble starch in plant hosts (Rodríguez-Sanoja, Oviedo, & Sánchez, 2005). These results suggest that the use of both single markers and haplotypes may be combined to increase the robustness of association studies.

3.7 | The relationship between resistant and susceptible accessions tested using the 49 significant polymorphic variants

In order to determine the propensity of the identified signals to associate with either resistance or susceptibility, we tested ancestral relatedness among 17 accessions consisting of the five most resistant (R, DSR 1 or 2) and 12 of the most susceptible (S, DSR 9 or 10) accessions (Figure 6). We aligned and determined the phylogeny of the nucleotide sequences of each of the 17 accessions at the 49 significant SNP loci using MEGA version X (Kumar, Stecher, Li, Knyaz, & Tamura, 2018) based on the neighbour-joining method (Saitou & Nei, 1987). Phylogeny was tested using 500 bootstrap replications (Felsenstein, 1985) for the combined 17 (R and S) accessions. The substitution model based on maximum composite likelihood method (Tamura, Nei, & Kumar, 2004) was implemented for the nucleotides at the 49 significant polymorphic sites (see below) in each of the accession. We used the default settings of uniform rates and assumed homogenous pattern among lineages. Pairwise deletion was selected for sites with gaps or missing data. The results show there are two main divergent hereditary branches for the markers among the most R and S accessions (Figure 6). One branch consisted of six accessions, which were all classified as susceptible (S). The second branch showed a measure of shared ancestry between the five R and the rest six S lines, with all the five R lines (NSL 28216, NSL 32629, NSL 42771, PI 648946 and PI 648954). This suggests that some of the marker signals were shared between the R and the S accessions, and thus that such polymorphism signals may not be important resistance factors. However, the R accession nodes were disproportionately enriched in the second arm of the significant marker hereditary tree (Figure 6), suggesting that some common SNPs shared among the R accessions are associated with resistance and may be important for marker-assisted selection. Particularly, two of the R accessions (PI 648946 and PI 648954) specifically shared the most distant node from the admixed R and S lines, suggesting that these spinach accessions may be harbouring SNPs more closely associated with resistance compared to the rest of the accessions. These two R lines and the six S lines in the all-susceptible branch could be candidate parents in developing biparental populations segregating for the anthracnose resistance and for the resistance

marker validation. Since none of the five R lines have been previously reported as possible resistance sources for anthracnose, they may form part of potential new sources for anthracnose-resistant breeding in spinach.

4 | DISCUSSION

The availability of genomic resources for spinach has grown appreciatively over the last decade. Here, we leveraged tagging of alleles in LD in haplotypes as a method implemented successfully to improve detection of marker–trait association signals in other crops and animal systems. We also performed single-marker SNP tests to offer a comparison of their sensitivity to that of haplotype-based markers in the same population. Differences in power and sensitivity of single SNPs compared to haplotypes and how these are calculated have been shown by other studies using simulated and empirical data in pigs and cattle, in humans and in barley (Braz et al., 2019; Clark, 2004; Gabriel et al., 2002; Hamon et al., 2006; Karimi et al., 2018; Lorenz et al., 2010; Meuwissen et al., 2014; Morris Rw, 2002; Wu et al., 2014). In the present study, we therefore concentrated on applying those principles with the aim of exploiting the advantages of single-SNP and haplotype methods to enhance the detection of polymorphisms with potential roles in spinach–anthracnose interactions. Such advantages include a dramatically reduced number of tests and hence the type I error rate since most haplotypes fall into a few classes with little evidence of recombination (Zhao, Fernando, et al., 2007). This also corresponds to improved power and robustness to accurately detect true association (Gabriel et al., 2002; Morris Rw, 2002). However, they were also cognizant of the hypothetical problems associated with pooling across those heterogeneous samples and hence some phasing ambiguity (Andrés et al., 2007). However, the robust application of IBD algorithms should have reduced this ambiguity to a bare minimum (Browning & Browning, 2011). We also highlighted where empirical differences between results from the three approaches were apparent.

4.1 | Blocking-and-tagging results in enhanced detection of true associations

We used the conservative Bonferroni correction (BC) at 0.01 significance level to adjust *p*-value for the number of tests conducted. BC helps to reveal the strength of the test for method comparison by reducing the family-wise error rate or the number of false positives, although we acknowledge this approach may have resulted in an increased number of false-negative markers. In order to estimate the power of the additive effect model to detect true associations in each approach, we considered the number of tagged markers in each approach as a "sample size", with a mean and a variance of the marker effects, and a BC threshold. Because of the reduced testing burden (in numbers) (Zhao, Fernando, et al., 2007), the BC-associated threshold P-value was slightly larger (2.60E-06) in htM than in htP (2.36E-06) and sSNP (1.62E-06) (Figure 2). The haplotype approaches therefore had slightly greater but not significantly different power from that of the sSNP approach. The values were 0.81, 0.84 and 0.87 from sSNP, htP and htM, respectively (Table 1). As expected, the two main determinants of power (population size and alpha) (Lorenz et al., 2010) appeared to compensate for power such that the larger population (test SNPs) in sSNP compared to htP and htM compensated for the smaller alpha (BC threshold) in sSNP compared to the larger alpha in htP and htM (Figure 4; Tables 1 and 2). In a simulation of QTL effect by combining SNPs, a previous report found that blocking methods that combined a SNP pair had the greatest power to detect significant markers in most cases (Schaid, 2004). In our case, the differences in power were very slight but might explain the ~ 63% and ~ 162% more marker-anchoring genes identified by the haplotype tagged SNPs (htP and htM, respectively), than by sSNP. This does not preclude the fact that the single-marker procedure could still uniquely identify four marker-anchoring genes, even though this number was lower than the number of unique marker-anchoring genes in the pairwise and multi-marker haplotype procedures. Therefore, the use of haplotype tags may be appropriate for detecting additional true association signals missed by the single-SNP tagging alone.

4.2 | Haplotype SNP tagging results on higher allelic effects

The allelic effects were generally higher for the same marker when tagged in a haplotype block than as a single marker (Table 2). On average, there was a significantly larger effect (Student's *t*-test, $p \leq .05$) with increase in marker clustering (htM > htP>sSNP). For example, for the first five markers identified by all three strategies, the alleles of sSNP had an average effect (absolute) of 3.73 (R^2 = 18.8), while those of htP and htM had average effects of 5.29 (R^2 = 28.8) and 12.6 (R^2 = 10.836), respectively. A similar trend was shown for the eight markers common between htP and htM, with alleles of htP having an average effect of 13.6 (R^2 = 13.5) and those of htM having an average effect of 17.0 (R^2 = 11.5). However, on average, the minor alleles in sSNP explained the least variance (3.88%), followed by the

TABLE 3 Mean minor allele statistics of significant markers by tagging method. Standard deviations are shown in parentheses

Method	Number of minor alleles	Mean effects	Mean MAF	Mean PVE	Mean h^2
sSNP	6	3.88 (2.28)	0.46 (0.02)	0.19 (0.05)	0.89 (0.07)
htP	13	10.84 (4.38)	0.25 (0.15)	0.18 (0.17)	0.72 (0.41)
htM	19	32.48 (51.04)	0.25 (0.13)	0.14 (0.08)	0.64 (0.40)
All	38	20.79 (37.70)	0.28 (0.13)	0.16 (0.11)	0.71 (0.38)

TABLE 4 Functional annotation of sample markers with significant association signals

Method	Marker	Chr	Position	Gene version	Functional annotation
sSNP	32006_119	4	1.07E + 08	Spo15466.1	BHLH transcription factor, Basic helix–loop–helix-like protein
	41566_67	6	18,775,069	Spo06040.1	TransSNP, htMembrane nine 1
	44220_132	6	44,775,432	Spo23358.1	Carotenoid cleavage dioxygenase 4
	44430_95	6	4,847,811	Spo17060.1	Unknown protein
htP	26787_118	3	1,532,247	Spo12930.1	Protein LONGIFOLIA 2 (Protein TON1 RECRUITING MOTIF 1)
	20941_21	1	48,227,888	Spo04891.1	Phosphoinositide phospholipase C (3.1.4.11)
	20014_86	1	41,230,635	Spo10761.1	Pectin lyase-like protein
	19722_61	1	39,055,145	Spo25784.1	5'-AMP-activated protein kinase-related
	21915_12	2	16,845,611	Spo02089.1	Potassium channel
	28140_95	3	37,770,803	Spo08254.1	Zinc finger A20 and AN1 domain-containing stress-associated protein 8
	34572_112	4	43,343,096	Spo16791.1	BZIP transcription factor family protein 3
	37730_147	5	12,033,640	Spo05241.1	Pectin lyase-like protein
	44219_78	6	44,775,486	Spo23358.1	Carotenoid cleavage dioxygenase 4
htM	24234_63	2	50,772,717	Spo01606.1	(Glycosyltransferase family 1 protein) (glycosyltransferase)
	31597_16	4	1.02E + 08	Spo11081.1	Mov34/MPN/PAD−1 family protein
	41286_72	6	13,952,296	Spo08623.1	Rubisco methyltransferase family protein
	42874_64	6	33,655,811	Spo12155.1	Nodulin MtN21/EamA-like transporter family protein
	19263_44	1	34,763,714	Spo06085.1	Probable magnesium transporter NIPA9
	19563_17	1	3,782,036	Spo25392.1	MYB transcription factor
	19962_15	1	40,835,316	Spo10734.1	Unknown protein
	24516_60	2	52,366,936	Spo23665.1	(DNA-binding protein) (MYB transcription factor)
	24705_120	2	54,129,583	Spo23855.1	F-box family protein
	24844_46	2	55,114,067	Spo21552.1	Heat shock 70 kDa protein 4
	25384_66	2	58,651,886	Spo01185.1	Zinc finger, C3HC4 type (RING finger) protein
	31191_147	3	9,539,285	Spo26660.1	Gamma-tocopherol methyltransferase
	33698_12	4	18,346,130	Spo06452.1	Unknown protein
	38278_107	5	17,702,567	Spo19761.1	Nucleolar protein NOP56
	39005_53	5	28,531,658	Spo01004.1	GHMP kinase ATP-binding protein, putative
	40099_56	5	50,109,435	Spo16236.1	Plant protein of unknown function (DUF641)
	40948_89	5	8,362,430	Spo27269.1	Arginine/serine-rich coiled coil protein 1
	41186_110	6	12,079,239	Spo04996.1	(Oligopeptidase A, putative) (3.4.24.70)
	43777_80	6	41,015,140	Spo25995.1	Unknown protein
sSNP, htP	28237_70	3	39,199,668	Spo20779.1	S-Adenosyl-L-methionine-dependent methyl-transferases superfamily protein
	30552_38	3	83,995,831	Spo08703.1	Unknown protein
sSNP, htM	18919_39	1	31,435,812	Spo00536.1	TransSNP, htMembrane protein, putative
	32755_39	4	1.18E + 08	Spo11036.1	DUF688 family protein

(Continues)

TABLE 4 (Continued)

Method	Marker	Chr	Position	Gene version	Functional annotation
htP, htM	24806_50	2	54,880,205	Spo17851.1	Single-stranded DNA-binding protein WHY1, chloroplastic (Protein WHIRLY 1) (ZmWHY1) (Precursor)
	28190_146	3	38,721,303	Spo21815.1	Proteasome subunit alpha type (3.4.25.1)
	31191_149	3	9,539,283	Spo26660.1	Gamma-tocopherol methyltransferase
	31716_82	4	1.03E + 08	Spo07294.1	ADP-ribosylation factor-like protein 5
	41284_58	6	13,942,768	Spo08608.1	O-Fucosyltransferase family protein
	44036_48	6	43,347,911	Spo23418.1	Peroxidase (1.11.1.7)
	22624_143	2	30,891,579	Spo07988.1	Lipid transfer protein
	38635_43	5	2,323,118	Spo02299.1	(Cytochrome P450, putative) (1.14.14.1)
sSNP, htP, htM	24623_126	2	53,247,408	Spo23748.1	WD and tetratricopeptide repeat protein, putative
	25305_48	2	58,152,897	Spo01146.1	Cullin-like protein 1
	35570_138	4	6,527,209	Spo15721.1	Heavy metal ATPase 5
	38421_39	5	20,353,750	Spo05063.1	Pentatricopeptide repeat superfamily protein
	43942_57	6	42,706,126	Spo09720.1	Beta-amylase

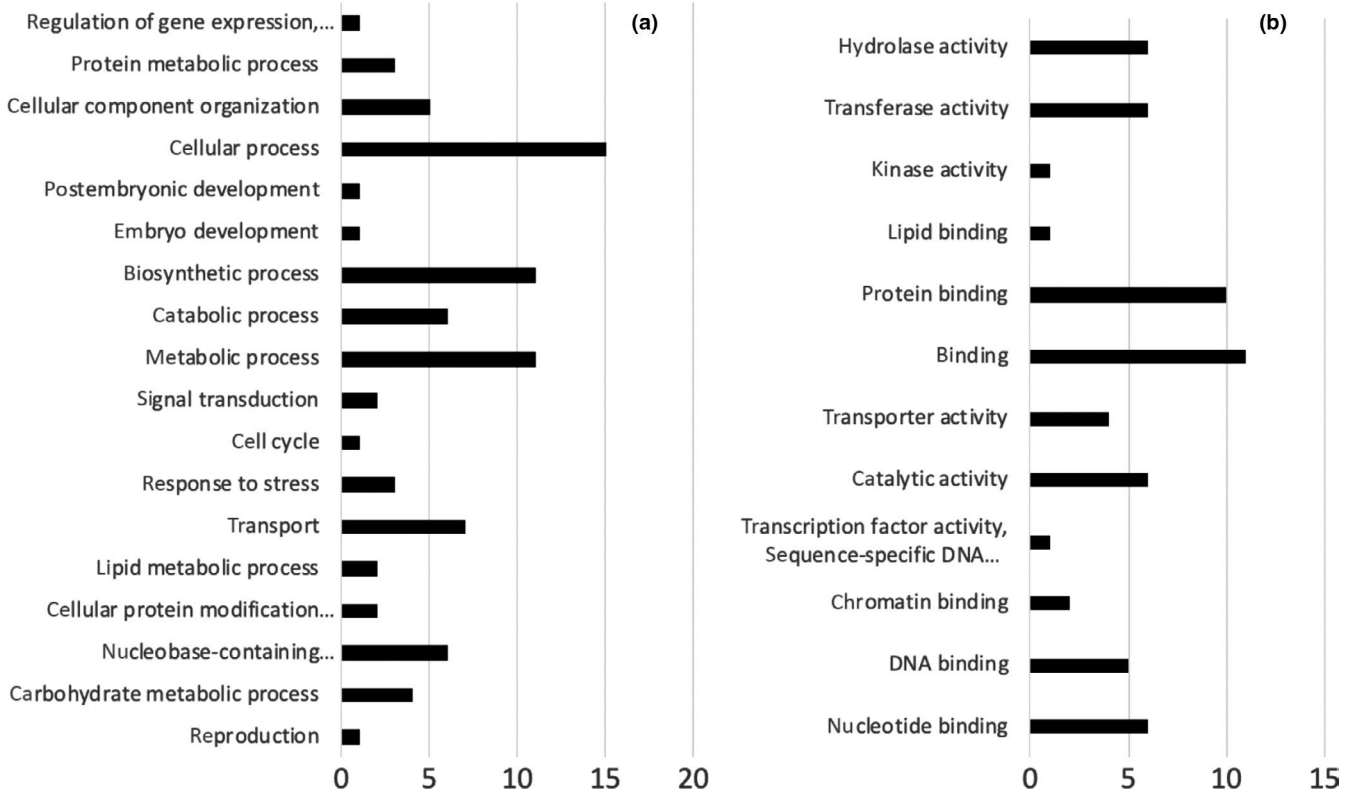

FIGURE 5 Gene ontogeny classification of genes anchoring polymorphic variants strongly associated with anthracnose resistance ratings by (a) biological process and (b) molecular function

minor alleles in htP (10.84%) and those in htM (32.48%; Table 3). The number of minor alleles as a percentage of the total was larger in haplotype tags (52% for htP and 59% for htM) than for sSNP (40%). For the anthracnose resistance trait, the estimated effect-size distribution among minor alleles may suggest the existence of increasingly large numbers of susceptibility SNPs with decreasingly small effects.

4.3 | Genes associated with anthracnose resistance are ubiquitously dispersed throughout the genome, and some participate in many biological processes and functions

The number, location and wide functional diversity of genes identified suggests that anthracnose in spinach is a multigenic trait

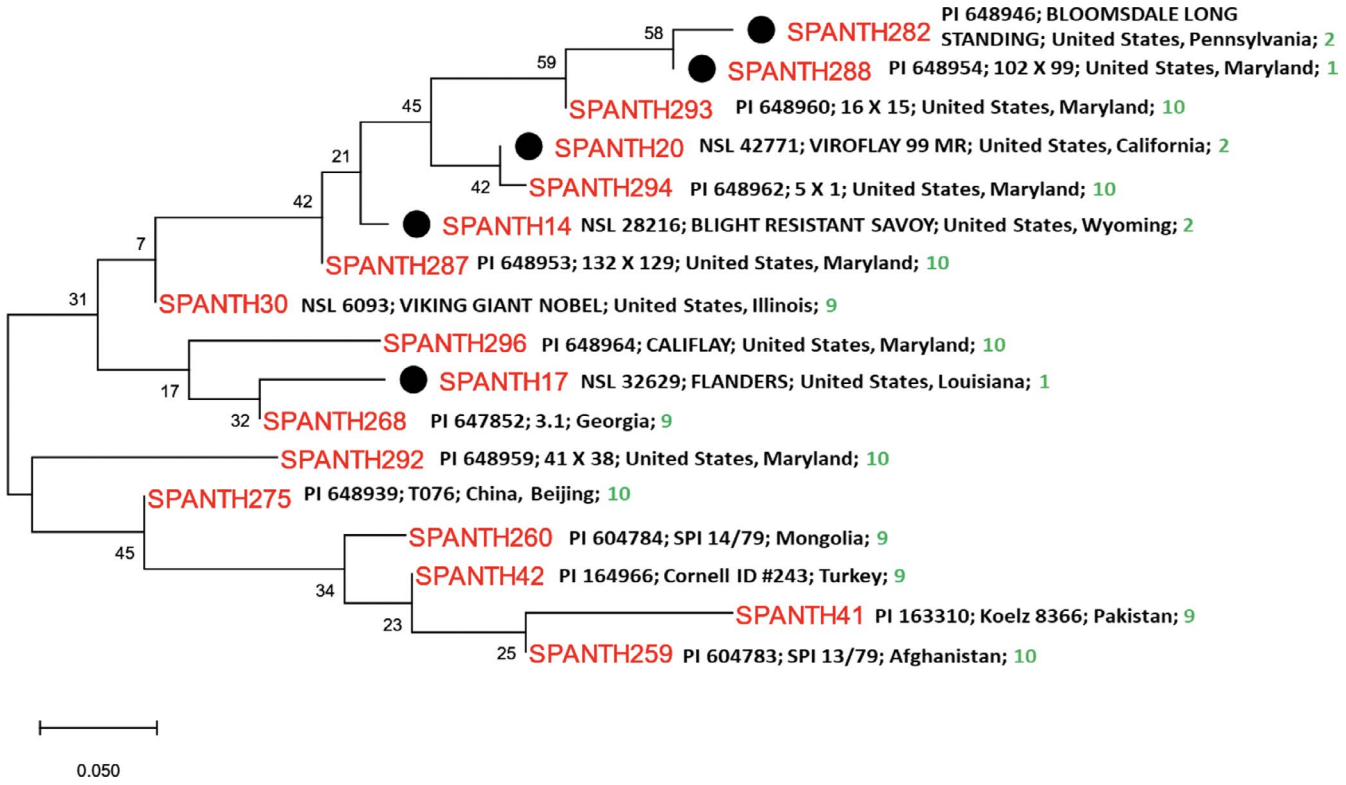

FIGURE 6 Ancestry relationships of spinach accessions with respect to 49 significant polymorphic sites (see in GWAS results). The ancestry history was inferred using the neighbour-joining method. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (500 replicates) is shown next to the branches. The evolutionary distances were computed using the maximum composite likelihood method and are in the units of the number of base substitutions per site. Field entries as used with the genotypic data are shown in red, followed by the corresponding accession identifier, the name, origin and disease severity rating—DSR. All ambiguous positions were removed for each sequence pair (pairwise deletion option). Analyses were conducted in MEGA X (Kumar et al., 2018). Analysis involved 17 nucleotide sequences (five resistant lines, in filled circles, tested together with 12 lines classified as susceptible [Colour figure can be viewed at wileyonlinelibrary.com]

(Figures 3 and 5). In this study, for instance, the gene ontology (GO) term (GO:0031625) "ubiquitin protein ligase binding" was particularly enriched for the genes *Spo21552* (identified by htM only) and *Spo01146* (identified by sSPN, htP and htM). Ubiquitination is widespread in the genome of many organisms and is a well-known protein-modification system that may be deployed to regulate plant defences against pathogens and as essential components of (R)-gene-mediated resistance (Devoto, Muskett, & Shirasu, 2003). Some of the genes we identified are known to mediate diverse biological processes (Table 4) and may be, intuitively, involved in polygenic resistance in many pathosystems including *C. dematium–S. oleracea*.

At the marker resolution and population sizes used, both the pairwise (htP) and multiple-marker (htM) methods enabled the identification of additional markers hosted by genes associated with biosynthetic processes involving important defence molecules. These pathways were not represented by markers in the sSNP strategy. For instance, by htP and htM haplotype tagging strategies, we identified the pathways PWY-7436 (tocotrienol) and PWY-1422 (tocopherol), which are vitamin E pathways in which the gene *Spo26660* codes for the enzyme gamma-tocopherol methyltransferase, which mediates the conversion of gamma-tocopherol to alpha-tocopherol in vitamin E biosynthesis — vitamin E is an antioxidant suggested to mediate in abiotic and biotic stresses (Abbasi, Hajirezaei, Hofius, Sonnewald, & Voll, 2007). The *Spo26660* gene annexes (186 bp upstream) the marker *31191_149*, which is a C-A inversion located at 9539283 on chr3.

Additional genes implicated in plant immunity signalling pathways were identified by haplotype tagging. One of these genes, *Spo04891*, was annotated to three related but distinct pathways including D-*myo*-inositol (1,4,5)-trisphosphate biosynthesis, suggested to be an important precursor in plant signalling for energy homeostasis and a signalling molecule in immune responses (Sengupta, Mukherjee, Basak, & Majumder, 2015). Others included *Spo21787*, coding for a central precursor molecule of FAD synthase, which has antioxidant activity (Sandoval, Zhang, & Roje, 2008). FAD synthase 2 is a chloroplastic protein known for its crucial role in defence signalling against oxidative stress (Sandoval et al., 2008). Similarly, antioxidant vitamin tocopherols, including α-tocopherol, β-tocopherol, γ-tocopherol and δ-tocopherol (Cheng et al., 2003; Porfirova, Bergmuller, Tropf, Lemke, & Dormann, 2002), have been ubiquitously identified in co-expression analysis during fungal and plant–host interactions (Samsatly, Copley, & Jabaji, 2018). The strong signal associated with these antioxidant genes may suggest similar defence signalling in *C. dematium* and *S. oleracea* L. interactions. Functional analyses of the identified candidate genes can be performed in future work to explore their involvement in resistance mechanisms.

5 | CONCLUSIONS

Based on our current data in spinach, haplotype tagging methods htP and htM may be more powerful in detecting marker–phenotype signals than the sSNP method. However, these methods failed to detect some signals found only to be significant when using the sSNP method. Therefore, the combined use of sSNP, htP and htM strategies should be employed to enhance the detection of polymorphisms with a potential role in spinach resistance against anthracnose since the highest number of detected markers resulted from the combination of all three approaches. After detection, these potential markers need to be validated in segregating populations to determine their value in marker-assisted selection programmes.

ACKNOWLEDGMENTS

This study was supported in part by funds from the Texas A&M AgriLife Research Vegetable Seed Grant FY17 and FY18 124353-96181 to C.A.A., V.J. and K.C.

CONFLICT OF INTERESTS

The authors declare that they have no competing interest.

AUTHOR CONTRIBUTIONS

H.O.A. performed population and genomic analyses and drafted the manuscript; K.C. and V.J. conducted field experiments and anthracnose resistance phenotyping; R.B. and K.K.M. performed SNP variant call analysis; C.A.A. designed and supervised the experiments. All the authors contributed to writing and reviewing the manuscript.

ORCID

Carlos A. Avila https://orcid.org/0000-0002-1969-4706

REFERENCES

Abbasi, A.-R., Hajirezaei, M., Hofius, D., Sonnewald, U., & Voll, L. M. (2007). Specific roles of alpha- and gamma-tocopherol in abiotic stress responses of transgenic tobacco. *Plant Physiology*, *143*, 1720–1738. https://doi.org/10.1104/pp.106.094771

Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, *19*, 1655–1664. https://doi.org/10.1101/gr.094052.109

Andrés, A., Clark, A., Shimmin, L., Boerwinkle, E., Sing, C., & Hixson, J. (2007). Understanding the accuracy of statistical haplotype inference with sequence data of known phase. *Genetic Epidemiology*, *31*, 659–671. https://doi.org/10.1002/gepi.20185

Arumuganathan, K., & Earle, E. (1991). Nuclear DNA content of some important plant species. *Plant Molecular Biology Reporter*, *9*, 208–218. https://doi.org/10.1007/BF02672069

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., … Sherlock, G. (2000). Gene ontology: Tool for the unification of biology. *Nature Genetics*, *25*, 25–29. https://doi.org/10.1038/75556

Awika, H. O., Bedre, R., Yeom, J., Marconi, T. G., Enciso, J., Mandadi, K., … Avila, C. A. (2019). Developing Growth-associated molecular markers via high-throughput phenotyping in Spinach. *The Plant Genome Journal*, *12*(3), 1–19. https://doi.org/10.3835/plantgenome2019.03.0027

Barrett, J. C., Fry, B., Maller, J., & Daly, M. J. (2005). Haploview: Analysis and visualization of LD and haplotype maps. *Bioinformatics*, *21*, 263–265. https://doi.org/10.1093/bioinformatics/bth457

Bedre, R., Rajasekaran, K., Mangu, V., Sanchez Timm, L., Bhatnagar, D., & Baisakh, N. (2015) Genome-wide transcriptome analysis of cotton (*Gossypium hirsutum* L.) identifies candidate gene signatures in response to aflatoxin producing fungus *Aspergillus flavus*. *PLoS ONE*, *10*, e0138025.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society*, *1*, 289–300.

Binns, D., Dimmer, E., Huntley, R., Barrell, D., O'Donovan, C., & Apweiler, R. (2009). QuickGO: A web-based tool for Gene Ontology searching. *Bioinformatics*, *25*, 3045–3046. https://doi.org/10.1093/bioinformatics/btp536

Blouin, M. S. (2003). DNA-based methods for pedigree reconstruction and kinship analysis in natural populations. *Trends in Ecology & Evolution*, *18*, 503–511. https://doi.org/10.1016/S0169-5347(03)00225-8

Bobev, S. G., Jelev, Z. J., Zveibil, A., Maymon, M., & Freeman, S. (2009). First report of anthracnose caused by *Colletotrichum dematium* on Statice (*Goniolimon tataricum*, Synonym *Limonium tataricum*) in Bulgaria. *Plant Disease*, *93*, 552–552. https://doi.org/10.1094/PDIS-93-5-0552C

Bonferroni, C. (1935) *Il calcolo delle assicurazioni su gruppi di teste*. Studi in onore del professore salvatore ortu carboni:13–60.

Bradbury, P., Zhang, Z., Kroon, D., Casstevens, T., Ramdoss, Y., & Buckler, E. (2007). TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinformatics*, *23*, 2633–2635. https://doi.org/10.1093/bioinformatics/btm308

Braz, C. U., Taylor, J. F., Bresolin, T., Espigolan, R., Feitosa, F. L. B., Carvalheiro, R., … de Oliveira, H. N. (2019). Sliding window haplotype approaches overcome single SNP analysis limitations in identifying genes for meat tenderness in Nelore cattle. *BMC Genetics*, *20*, 8. https://doi.org/10.1186/s12863-019-0713-4

Browning, S., & Browning, B. (2011). Haplotype phasing: Existing methods and new developments. *Nature Reviews Genetics*, *12*, 703–714. https://doi.org/10.1038/nrg3054

Buels, R., Yao, E., Diesh, C. M., Hayes, R. D., Munoz-Torres, M., Helt, G., … Holmes, I. H. (2016). JBrowse: A dynamic web platform for genome visualization and analysis. *Genome Biology*, *17*, 66. https://doi.org/10.1186/s13059-016-0924-1

Calus, M. P. L., Meuwissen, T. H. E., Windig, J. J., Knol, E. F., Schrooten, C., Vereijken, A. L. J., & Veerkamp, R. F. (2009). Effects of the number of markers per haplotype and clustering of haplotypes on the accuracy of QTL mapping and prediction of genomic breeding values. *Genetics Selection Evolution*, *41*, 11. https://doi.org/10.1186/1297-9686-41-11

Carbon, S., Ireland, A., Mungall, C. J., Shu, S. Q., Marshall, B., & Lewis, S. (2009). AmiGO: Online access to ontology and annotation data. *Bioinformatics*, *25*, 288–289. https://doi.org/10.1093/bioinformatics/btn615

Caspi, R., Billington, R., Fulcher, C. A., Keseler, I. M., Kothari, A., Krummenacker, M., … Karp, P. D. (2018). The MetaCyc database of metabolic pathways and enzymes. *Nucleic Acids Research*, *46*, D633–D639. https://doi.org/10.1093/nar/gkx935

Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., & Cresko, W. A. (2013). Stacks: An analysis tool set for population genomics. *Molecular Ecology*, *22*, 3124–3140. https://doi.org/10.1111/mec.12354

Cheng, B. P., Huang, Y. H., Song, X. B., Peng, A. T., Ling, J. F., & Chen, X. (2013). First Report of *Colletotrichum siamense* Causing leaf

drop and fruit spot of *Citrus reticulata* Blanco cv. Shiyue Ju in China. *Plant Disease*, *97*, 1508–1508. https://doi.org/10.1094/PDIS-04-13-0352-PDN

Cheng, Z., Sattler, S., Maeda, H., Sakuragi, Y., Bryant, D. A., & Dellapenna, D. (2003). Highly divergent methyltransferases catalyze a conserved reaction in tocopherol and plastoquinone synthesis in cyanobacteria and photosynthetic eukaryotes. *The Plant Cell*, *15*, 2343–2356. https://doi.org/10.1105/tpc.013656

Clark, A. G. (2004). The role of haplotypes in candidate gene studies. *Genetic Epidemiology*, *27*, 321–333. https://doi.org/10.1002/gepi.20025

Correll, J. C. M. E. T., Black, M. C., Koike, S. T., Brandenberger, L. P., & Dalnello, F. J. (1994). Economically important diseases of Spinach. *Plant Disease*, *78*, 653–660. https://doi.org/10.1094/PD-78-0653

Dal Bello, G. M. (2000). First Report of *Colletotrichum dematium* on tomato in Argentina. *Plant Disease*, *84*, 198–198. https://doi.org/10.1094/PDIS.2000.84.2.198A

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., … Durbin, R. (2011). The variant call format and VCFtools. *Bioinformatics*, *27*, 2156–2158. https://doi.org/10.1093/bioinformatics/btr330

De Bakker, P. I., Yelensky, R., Pe'er, I., Gabriel, S. B., Daly, M. J., & Altshuler, D. (2005). Efficiency and power in genetic association studies. *Nature Genetics*, *37*, 1217–1223. https://doi.org/10.1038/ng1669

Devoto, A., Muskett, P. R., & Shirasu, K. (2003). Role of ubiquitination in the regulation of plant defence against pathogens. *Current Opinion in Plant Biology*, *6*, 307–311. https://doi.org/10.1016/S1369-5266(03)00060-8

Dixon, W. J. (1960). Simplified estimation from censored normal samples. *The Annals of Mathematical Statistics*, *31*, 385–391. https://doi.org/10.1214/aoms/1177705900

Earl, D. A., & Vonholdt, B. M. (2012). STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, *4*, 359–361. https://doi.org/10.1007/s12686-011-9548-7

Ellis, J. R., & Janick, J. (1960). The chromosomes of *Spinacia oleracea*. *American Journal of Botany*, *47*, 210–214. https://doi.org/10.1002/j.1537-2197.1960.tb07115.x

Endelman, J. B., & Jannink, J.-L. (2012). Shrinkage estimation of the realized relationship matrix. *G3: Genes. Genomes*, *Genetics*, *2*, 1405–1413. https://doi.org/10.1534/g3.112.004259

Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software structure: A simulation study. *Molecular Ecology*, *14*, 2611–2620. https://doi.org/10.1111/j.1365-294X.2005.02553.x

Excoffier, L. S. M. (1995). Maximum-likelihood estimation of molecular haplotype frequencies in a diploid population. *Molecular Biology and Evolution*, *12*, 912–917.

Falush, D., Stephens, M., & Pritchard, J. K. (2007). Inference of population structure using multilocus genotype data: Dominant markers and null alleles. *Molecular Ecology Notes*, *7*, 574–578. https://doi.org/10.1111/j.1471-8286.2007.01758.x

Felsenstein, J. (1985). Confidence limits on phylogenies: An approach using the bootstrap. *Evolution*, *39*, 783–791. https://doi.org/10.1111/j.1558-5646.1985.tb00420.x

Feng, B., Liu, C., Shan, L., & He, P. (2016). Protein ADP-Ribosylation takes control in plant-bacterium interactions. *PLoS Path*, *12*, e1005941–e1005941. https://doi.org/10.1371/journal.ppat.1005941

Flint-Garcia, S. A., Thornsberry, J. M., & Buckler Iv, E. S. (2003). Structure of linkage disequilibrium in plants. *Annual Review of Plant Biology*, *54*, 357–374. https://doi.org/10.1146/annurev.arplant.54.031902.134907

Gabriel, S. B., Schaffner, S. F., Nguyen, H., Moore, J. M., Roy, J., Blumenstiel, B.,…Altshuler, D. (2002). The structure of haplotype blocks in the human genome. *Science*, *296*, 2225–2229. https://doi.org/10.1126/science.1069424

Hamon, S. C., Kardia, S. L. R., Boerwinkle, E., Liu, K., Klos, K. L. E., Clark, A. G., & Sing, C. F. (2006). Evidence for consistent intragenic and intergenic interactions between SNP effects in the APOA1/C3/A4/A5 gene cluster. *Human Heredity*, *61*, 87–96. https://doi.org/10.1159/000093384

Hastings, C., Mosteller, F., Tukey, J. W., & Winsor, C. P. (1947). Low moments for small samples: A comparative study of order statistics. *The Annals of Mathematical Statistics*, *18*, 413–426. https://doi.org/10.1214/aoms/1177730388

Jakobsson, M., & Rosenberg, N. A. (2007). CLUMPP: A cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics*, *23*, 1801–1806. https://doi.org/10.1093/bioinformatics/btm233

Karimi, Z., Sargolzaei, M., JaB, R., & Schenkel, F. S. (2018). Assessing haplotype-based models for genomic evaluation in Holstein cattle. *Canadian Journal of Animal Science*, *98*, 750–759. https://doi.org/10.1139/cjas-2018-0009

Khattak, J. Z. K., Torp, A. M., & Andersen, S. B. (2006). A genetic linkage map of *Spinacia oleracea* and localization of a sex determination locus. *Euphytica*, *148*, 311–318. https://doi.org/10.1007/s10681-005-9031-1

Kumar, S., Stecher, G., Li, M., Knyaz, C., & Tamura, K. (2018). MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Molecular Biology and Evolution*, *35*, 1547–1549. https://doi.org/10.1093/molbev/msy096

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, *9*, 357–U354. https://doi.org/10.1038/Nmeth.1923

Liu, F., Tang, G., Zheng, X., Li, Y., Sun, X., Qi, X., … Gong, G. (2016a). Molecular and phenotypic characterization of Colletotrichum species associated with anthracnose disease in peppers from Sichuan Province. China. *Scientific Reports*, *6*, 32761. https://doi.org/10.1038/srep32761

Liu, L. P., Jin, X. S., Yu, L., Lu, B. H., Liu, Y. N., Yang, L. Y., … Hsiang, T. (2016b). *Colletotrichum dematium:* Causing anthracnose on common knotgrass (*Polygonum aviculare*) in China. *Plant Disease*, *100*, 1240–1240. https://doi.org/10.1094/PDIS-10-15-1161-PDN

Liu, L. P., Shu, J., Zhang, L., Hu, R., Chen, C. Q., Yang, L. N., …Hsiang, T. (2017). First Report of post-harvest anthracnose on mango (*Mangifera indica*) caused by *Colletotrichum siamense* in China. *Plant Disease*, *101*, 833–833. https://doi.org/10.1094/PDIS-08-16-1130-PDN

Liu, T., Chen, D., Liu, Z., & Hou, J. M. (2018). First report of *Colletotrichum siamense* causing anthracnose on partridge tea (*Mallotus oblongifolius*) in China. *Plant Disease*, *102*, 1669–1669. https://doi.org/10.1094/PDIS-12-17-1957-PDN

Long, A. D., & Langley, C. H. (1999). The power of association studies to detect the contribution of candidate genetic loci to variation in complex traits. *Genome Research*, *9*, 720–731. https://doi.org/10.1101/gr.9.8.720

Lorenz, A. J., Hamblin, M. T., & Jannink, J.-L. (2010). Performance of single nucleotide polymorphisms versus haplotypes for Genome-Wide Association Analysis in Barley. *PLoS ONE*, *5*, e14079. https://doi.org/10.1371/journal.pone.0014079

Meuwissen, T. H. E., Odegard, J., Andersen-Ranberg, I., & Grindflek, E. (2014). On the distance of genetic relationships and the accuracy of genomic prediction in pig breeding. *Genetics*, *Selection*, *Evolution*, *46*, 49–49. https://doi.org/10.1186/1297-9686-46-49

Morelock, T. E., & Correll, J. C. (2008) Spinach. In J. Prohens & F. Nuez (Eds), *Vegetables I: Asteraceae*, *brassicaceae*, *chenopodicaceae*, *and cucurbitaceae* (pp. 189–218). New York, NY: Springer, New York. https://doi.org/10.1007/978-0-387-30443-4-6.

Morris, R. W., & Kaplan, N. L. (2002). On the advantage of haplotype analysis in the presence of multiple disease susceptibility alleles. *Genetic Epidemiology*, *23*, 221–233. https://doi.org/10.1002/gepi.10200

Niu, X. P., Gao, H., Chen, Y., & Qi, J. M. (2016). First Report of Anthracnose on White Jute (*Corchorus capsularis*) caused by *Colletotrichum fructicola* and *C. siamense* in China. *Plant Disease*, *100*, 1243–1243. https://doi.org/10.1094/PDIS-12-15-1418-PDN

Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., & Hoekstra, H. E. (2012). Double digest RADseq: An inexpensive method for *de novo* SNP discovery and genotyping in model and non-model species. *PLoS ONE*, *7*, e37135. https://doi.org/10.1371/journal.pone.0037135

Porfirova, S., Bergmuller, E., Tropf, S., Lemke, R., & Dormann, P. (2002). Isolation of an Arabidopsis mutant lacking vitamin E and identification of a cyclase essential for all tocopherol biosynthesis. *Proceedings of the National Academy of Sciences of the United States of America*, *99*, 12495–12500. https://doi.org/10.1073/pnas.182330899

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, *155*, 945.

Qin, Z. S., Niu, T., & Liu, J. S. (2002). Partition-ligation-expectation-maximization algorithm for haplotype inference with single-nucleotide polymorphisms. *American Journal of Human Genetics*, *71*, 1242–1247. https://doi.org/10.1086/344207

Rochette, N. C., & Catchen, J. M. (2017). Deriving genotypes from RAD-seq short-read data using Stacks. *Nature Protocols*, *12*, 2640–2659. https://doi.org/10.1038/nprot.2017.123

Rodríguez-Sanoja, R., Oviedo, N., & Sánchez, S. (2005). Microbial starch-binding domain. *Current Opinion in Microbiology*, *8*, 260–267. https://doi.org/10.1016/j.mib.2005.04.013

Saitou, N., & Nei, M. (1987). The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution*, *4*, 406–425. https://doi.org/10.1093/oxfordjournals.molbev.a040454

Samsatly, J., Copley, T. R., & Jabaji, S. H. (2018). Antioxidant genes of plants and fungal pathogens are distinctly regulated during disease development in different *Rhizoctonia solani* pathosystems. *PLoS ONE*, *13*, e0192682. https://doi.org/10.1371/journal.pone.0192682

Sandoval, F. J., Zhang, Y., & Roje, S. (2008). Flavin nucleotide metabolism in plants: Monofunctional enzymes synthesize fad in plastids. *Journal of Biological Chemistry*, *283*, 30890–30900. https://doi.org/10.1074/jbc.M803416200

Schaid, D. J. (2004). Evaluating associations of haplotypes with traits. *Genetic Epidemiology*, *27*, 348–364. https://doi.org/10.1002/gepi.20037

Sengupta, S., Mukherjee, S., Basak, P., & Majumder, A. L. (2015). Significance of galactinol and raffinose family oligosaccharide synthesis in plants. *Frontiers in Plant Science*, *6*, https://doi.org/10.3389/fpls.2015.00656

Sharma, G., Kumar, N., Weir, B. S., Hyde, K. D., & Shenoy, B. D. (2013). The ApMat marker can resolve *Colletotrichum* species: A case study with *Mangifera indica*. *Fungal Diversity*, *61*, 117–138. https://doi.org/10.1007/s13225-013-0247-4

Shi, A., Mou, B., Correll, J., Koike, S. T., Motes, D., & Qin, J., …Yang, W. (2016a). Association analysis and identification of SNP markers for Stemphylium Leaf Spot (*Stemphylium botryosum* f. sp. *spinacia*) resistance in Spinach (*Spinacia oleracea*). *American Journal of Plant Sciences*, *7*, 1600–1611. https://doi.org/10.4236/ajps.2016.712151

Shi, A., Mou, B., Correll, J., Motes, D., Weng, Y., Qin, J., & Yang, W. (2016b). SNP association analysis of resistance to Verticillium wilt (*Verticillium dahliae* Kleb.) in spinach. *Australian Journal of Crop Science*, *10*, 1188–1196. https://doi.org/10.21475/ajcs.2016.10.08.p7893

Shi, A., Qin, J., Mou, B., Correll, J., Weng, Y., Brenner, D., … Ravelombola, W. (2017). Genetic diversity and population structure analysis of spinach by single-nucleotide polymorphisms identified through genotyping-by-sequencing. *PLoS ONE*, *12*, e0188745. https://doi.org/10.1371/journal.pone.0188745

Skinner, M. E., Uzilov, A. V., Stein, L. D., Mungall, C. J., & Holmes, I. H. (2009). JBrowse: A next-generation genome browser. *Genome Research*, *19*, 1630–1638. https://doi.org/10.1101/gr.094607.109

Smith, J. E., & Aveling, S. (1997). *Colletotrichum dematium*: Causal agent of a new cowpea stem disease in South Africa. *Plant Disease*, *81*, 832–832. https://doi.org/10.1094/PDIS.1997.81.7.832D

Su, W.-H., Yao Shugart, Y., Chang, K.-P., Tsang, N.-M., Tse, K.-P., & Chang, Y.-S. (2013). How Genome-wide SNP-SNP interactions relate to nasopharyngeal carcinoma susceptibility. *PLoS ONE*, *8*, e83034. https://doi.org/10.1371/journal.pone.0083034

Tabangin, M. E., Woo, J. G., & Martin, L. J. (2009). The effect of minor allele frequency on the likelihood of obtaining false positives. *BMC Proceedings*, *3*, S41. https://doi.org/10.1186/1753-6561-3-S7-S41

Tamura, K., Nei, M., & Kumar, S. (2004). Prospects for inferring very large phylogenies by using the neighbor-joining method. *Proceedings of the National Academy of Sciences (USA)*, *101*, 11030–11035. https://doi.org/10.1073/pnas.0404206101

The_Gene_Ontology_Consortium (2019). The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic Acid Research*, *47*, D330–D338.

The_Uniprot_Consortium (2019). UniProt: A worldwide hub of protein knowledge. *Nucleic Acids Research*, *47*, D506–D515.

Turner, S. D. (2018). qqman: An R package for visualizing GWAS results using Q-Q and manhattan plots. *Journal of Open Source Software*, *3*(25), 731. https://doi.org/10.21105/joss.00731

Washington, W. S., Irvine, G., Aldaoud, R., Dealwis, S., Edwards, J., & Pascoe, I. G. (2006). First record of anthracnose of spinach caused by *Colletotrichum dematium* in Australia. *Australasian Plant Pathology*, *35*, 89–91. https://doi.org/10.1071/AP05095

Wu, Y., Fan, H., Wang, Y., Zhang, L., Gao, X., Chen, Y., … Gao, H. (2014). Genome-wide association studies using haplotypes and individual SNPs in Simmental cattle. *PLoS ONE*, *9*, e109330–e109330. https://doi.org/10.1371/journal.pone.0109330

Xu, C., Jiao, C., Sun, H., Cai, X., Wang, X., Ge, C., … Wang, Q. (2017). Draft genome of spinach and transcriptome diversity of 120 Spinacia accessions. *Nature Communications*, *8*, 15275. https://doi.org/10.1038/ncomms15275 https://www.nature.com/articles/ncomms15275#supplementary-information

Yang, X.-D., Tan, H.-W., & Zhu, W.-M. (2016). SpinachDB: A Well-characterized genomic database for gene family classification and SNP information of spinach. *PLoS ONE*, *11*, e0152706. https://doi.org/10.1371/journal.pone.0152706

Yoshida, S., & Shirata, A. (1999). Survival of *Colletotrichum dematium* in soil and infected mulberry leaves. *Plant Disease*, *83*, 465–468. https://doi.org/10.1094/PDIS.1999.83.5.465

Zhang, Z., Ersoz, E., Lai, C.-Q., Todhunter, R. J., Tiwari, H. K., Gore, M. A., … Buckler, E. S. (2010). Mixed linear model approach adapted for genome-wide association studies. *Nature Genetics*, *42*, 355–360. https://doi.org/10.1038/ng.546

Zhao, H. H., Fernando, R. L., & Dekkers, J. C. M. (2007). Power and precision of alternate methods for linkage disequilibrium mapping of quantitative trait loci. *Genetics*, *175*, https://doi.org/10.1534/genetics.106.066480

Zhao, H., Pfeiffer, R., & Gail, M. H. (2003). Haplotype analysis in population genetics and association studies. *Pharmacogenomics*, *4*, 171–178. https://doi.org/10.1517/phgs.4.2.171.22636

Zhao, K., Aranzana, M. J., Kim, S., Lister, C., Shindo, C., Tang, C., … Nordborg, M. (2007). An Arabidopsis example of association mapping in structured samples. *PLOS Genetics*, *3*, e4. https://doi.org/10.1371/journal.pgen.0030004

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Awika HO, Cochran K, Joshi V, Bedre R, Mandadi KK, Avila CA. Single-marker and haplotype-based association analysis of anthracnose (*Colletotrichum dematium*) resistance in spinach (*Spinacia oleracea*). *Plant Breed*. 2020;139:402–418. https://doi.org/10.1111/pbr.12773