



OPEN

## Extreme-phenotype genome-wide association study (XP-GWAS) of powdery mildew race 2 W tolerance in the USDA *Citrullus* germplasm collection

Sandra E. Branham<sup>1</sup>  , Venkata Rao Ganaparthi<sup>1</sup>, W. Patrick Wechter<sup>1</sup>, Younghoon Park<sup>3</sup>, Todd Wehner<sup>4</sup>, Angela Davis<sup>5</sup>, Antonia Tetteh<sup>6</sup>, Laura Massey<sup>2</sup>, Sue Hammar<sup>7</sup>, Rebecca Grumet<sup>7</sup>, Shaker Kousik<sup>2</sup> & Amnon Levi<sup>1</sup> 

Powdery mildew outbreaks, caused by *Podosphaera xanthii*, cause reduced watermelon yields as the plants produce fewer and smaller fruits due to premature leaf senescence. The reduced leaf canopy can decrease fruit quality due to sun scalding. Sources of powdery mildew tolerance were previously identified by screening the USDA *Citrullus* germplasm collection with *P. xanthii* races 1 W and 2 W. However, not all gene loci associated with tolerance to race 2 W have been identified and markers tightly linked to such loci have not been developed. We employed a bulked segregant analysis approach using historical data from the USDA Germplasm Resource Information Network for an extreme-phenotype genome-wide association study (XP-GWAS) of tolerance to *P. xanthii* race 2 W in *Citrullus* accessions ( $N=1,147$ ). XP-GWAS identifies variants that segregate between pools of individuals chosen from the extremes of a phenotypic distribution from a diversity panel. Whole-genome resequencing of 45 individuals bulked from tolerant and susceptible extremes resulted in 301,059 high-quality biallelic SNPs. Two adjacent SNPs on chromosome 7 were significantly associated with *P. xanthii* race 2 W tolerance in the bulks and two additional SNPs had a strong signal in the XP-GWAS analysis. Kompetitive Allele-Specific PCR (KASP) markers were designed for sixteen SNPs across the three genomic regions. The KASP markers were validated by genotyping 186 accessions from the extremes of the disease response distribution of the *Citrullus* collection. Analysis of variance determined that thirteen of the markers were significantly associated, with the best marker in each region explaining 21–31% of the variation in powdery mildew tolerance.

**Keywords** Disease resistance, XP-GWAS, Powdery mildew, Watermelon, Resequencing, KASP

Watermelon is an important specialty crop in United States, primarily grown in the Southeastern region. Production has decreased over the last five decades, resulting in increased imports necessary to meet consumer demand<sup>1</sup>. Imports peaked in 2019, representing one-third of U.S. consumption. Powdery mildew was rated as the second most prevalent and consistent foliar disease in watermelon<sup>2</sup>. Powdery mildew outbreaks cause yield losses through decreased fruit weight and can also reduce fruit quality through sun-scalding from reduced leaf canopy<sup>3</sup>.

Powdery mildew is a fungal disease caused by several species in genus *Podosphaera*, affecting a wide range of crops across the globe. Different races of *Podosphaera xanthii* and *Golovinomyces cichoracearum* are responsible for causing powdery mildew on Cucurbits, especially melons, and races from one cucurbit can readily infect other cucurbit species. Lack of variation in host tolerance and high pathogen variability complicate race classification in watermelon<sup>4</sup>. Races of *P. xanthii* have primarily been classified using muskmelon (*Cucumis melo*) differentials.

<sup>1</sup>Coastal Research and Education Center, Clemson University, 2700 Savannah Hwy, Charleston, SC, USA. <sup>2</sup>Agricultural Research Service - Southeast Area, New Orleans, USA. <sup>3</sup>Pusan National University, Miryang, Republic of Korea.

<sup>4</sup>North Carolina State University, Raleigh, NC, USA. <sup>5</sup>United States Department of Agriculture, Washington, USA.

<sup>6</sup>Kwame Nkrumah University of Science and Technology, Kumasi, Ghana. <sup>7</sup>Michigan State University, East Lansing, MI, USA.  email: sebranh@clemson.edu; amnon.levi@usda.gov

Twenty-eight races of *P. xanthii*, along with eight variants of race 1 and six variants of race 2, were identified based on 22 melon cultigens<sup>5</sup>. Two races of *P. xanthii* classified based on melon (*Cucumis melo*) differentials, that can also infect watermelon, have been reported in the U.S and designated as races 1 W and 2 W<sup>6,7</sup>. However, detailed classification of powdery mildew races based on watermelon differentials is not available. Evidence for the presence of two races based on watermelon differentials has been reported in U.S<sup>8</sup>, and similarly the potential for presence of numerous races based on watermelon differentials have been reported in Europe<sup>9</sup>. However, further studies are needed to define these races across US and Europe. Numerous fungicides are available for control of powdery mildew, but generally are more preventative than curative of an active infection<sup>10</sup>. Risk of evolution of fungicide tolerance is high and has already been reported for some fungicides in South Carolina and other states<sup>11,12</sup>, thus making host tolerance an essential component of disease management. Field evaluations of commercial watermelon cultivars found a single edible cultivar with tolerance to locally prevailing strains of powdery mildew that are predominantly race 1 (1 W) based on melon differentials<sup>13</sup>, highlighting the need for new cultivars with improved resistance. More recent field evaluations have identified a few seedless (triploid) cultivars with varying levels of tolerance to powdery mildew (race not reported)<sup>14</sup>. However, because of the presence of multiple races, new information with presence of new genetic loci potentially controlling resistance will be useful.

Disease screenings of the entire USDA *Citrullus* germplasm collection for response to artificial inoculation with *P. xanthii* races 1 W<sup>6</sup> and 2 W<sup>7</sup> have been reported, resulting in the identification of several accessions with strong resistance. Inheritance studies have found a complex genetic basis of resistance (e.g. partial dominance, incomplete dominance, multigenic and epistasis) that varies between populations and tissue types for both races<sup>15,16</sup>. Genome-based analyses of powdery mildew resistance in watermelon have been limited. QTL mapping of a biparental watermelon population segregating for resistance to *P. xanthii* race 1 W identified a single major QTL on chromosome 2, *pmr2.1*, that explained 80% of the variation in disease response<sup>17</sup>. They developed CAPS markers associated with resistance. Mandal et al. (2020) used comparative whole genome resequencing analysis of resistant inbred lines and their parents, using a different resistance source, and found a single major QTL for *P. xanthii* race 1 W resistance, collocated with *pmr2.1*<sup>18</sup>. A genome-wide association study of *P. xanthii* race 2 W resistance using historical data for the USDA *Citrullus* collection identified 43 SNPs across nine chromosomes, including a strong signal of several significant SNPs collocated with *pmr2.1*<sup>19</sup>. The GWAS employed low-density genotyping-by-sequencing (384-plex) which can result in false negatives in genomic regions of low coverage where there are no markers in linkage disequilibrium with the causal variant<sup>20</sup>. However, whole-genome resequencing of thousands of accessions is prohibitively expensive. A cost-effective alternative to traditional GWAS is bulked segregant analysis of extreme phenotypic pools from a diversity panel, termed an extreme-phenotype genome-wide association study (XP-GWAS)<sup>21</sup>. Along with reducing the number of accessions to be genotyped, XP-GWAS can enrich the frequency of rare alleles. XP-GWAS has been effective in identifying QTL in multiple crops, including quality characteristics in coffee<sup>22</sup> and apples<sup>23</sup>, and alkyl cannabinoid in cannabis<sup>24</sup>. We performed an XP-GWAS of disease response of the USDA *Citrullus* collection to *P. xanthii* race 2 W using historical data. Our objectives were to compare the results of traditional GWAS<sup>19</sup> to XP-GWAS using the same historical dataset and design KASP markers for marker-assisted breeding of powdery mildew resistance in watermelon.

## Materials and methods

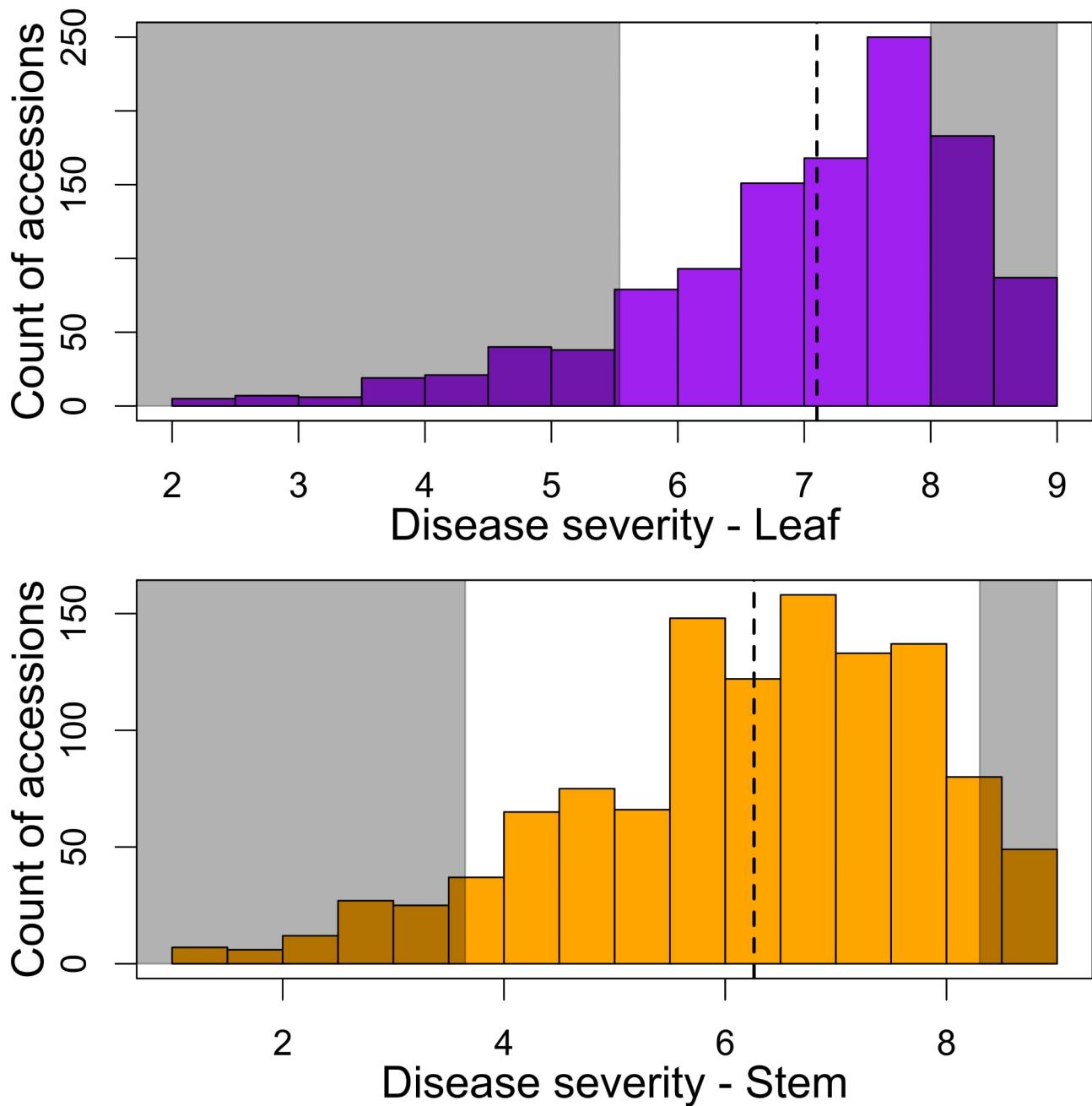
### Bulked segregant analysis

Historical data for disease response to artificial inoculation with *P. xanthii* race 2 W<sup>7</sup> for the cultivated species, *C. lanatus* ( $N=1,095$ ), and its sister species, *Citrullus mucospermus* ( $N=52$ ) were obtained from the USDA Germplasm Resource Information Network at <https://www.ars-grin.gov/>. Disease severity (DS) ratings for both stem and leaf tissues were used to select 45–46 accessions for each of three bulks. In addition to a bulk from each extreme of the phenotypic distribution, the XP-GWAS method requires a random bulk with individuals chosen at random with respect to phenotype<sup>21</sup>. The random bulk is used as a control to account for background population allele frequencies. Accessions with a disease severity rating of less than 4 for stems and 5 for leaves did not have visible mycelium growth on these tissues<sup>7</sup>. We chose individuals for the resistant bulk that had an average disease severity score of less than 5 across tissues (Fig. 1; Supplementary table S1). Accessions with a mean disease severity score of less than 8.3 across tissues were chosen for the susceptible bulk ( $N=46$ ) to obtain an approximately equal number of accessions as the resistant bulk ( $N=45$ ) (Fig. 1; Supplementary table S1). A disease severity rating of 8 was given to plants with 50–70% mycelium growth on that tissue (leaf or stem) with large necrotic areas and a 9 for tissues fully covered with mycelium or for a dead plant<sup>7</sup>.

Accessions for the random bulk were chosen using the sample function in R<sup>25</sup>. The tolerant and random bulks were comprised of 22 and 40 *C. lanatus* accessions, and 23 and 6 *C. mucospermus* accessions, respectively (See Supplementary table S1 online). The susceptible bulk was composed entirely of *C. lanatus* accessions. All the tolerant accessions originated in Africa and most of the susceptible accessions were from Europe or Asia. Seeds for the bulked accessions were obtained from the USDA National Plant Germplasm System.

### Whole-genome resequencing

DNA was extracted from young leaf tissue of a single plant per accession with a ChargeSwitch DNA Kit (Invitrogen, Waltham, MA). DNA was quantified with a Qubit Fluorometer (Invitrogen, Waltham, MA) and 10 ng of genomic DNA per accession was combined to form the three bulks. Bulked DNA was sent to the W.M. Kick Center at the University of Illinois (Urbana, IL) for library preparation and whole genome resequencing. Genomic libraries were prepared with a Hyper Library Construction Kit (Kapa Biosystems, Roche, Basel, Switzerland), quantified with qPCR and sequenced on a single lane of an Illumina Novaseq 6000 (Illumina, San Diego, CA) with a NovaSeq S2 Reagent Kit (Illumina, San Diego, CA). Paired-end reads (150 bp) were



**Fig. 1.** Histogram of disease severity on stems and leaves after inoculation of the USDA *Citrullus* collection with *Podosphaera xanthii* race 2 W<sup>7</sup>. Shaded boxes indicate the bulked accessions. The accession mean for disease severity across the collection is indicated by a dashed vertical line. Figure generated in R<sup>25</sup>.

generated with 151 cycles from each end of the fragment. Reads were demultiplexed and adaptors trimmed with the bcl2fastq v2.20 Software (Illumina, San Diego, CA). Raw sequence files have been deposited in the NCBI Sequence Read Archive and are available under BioProject PRJNA1197509.

#### Variant calling and XP-GWAS

Duplicate reads were removed with a custom Perl script obtained from [https://github.com/Sunhh/NGS\\_data\\_processing/blob/master/drop\\_dup\\_both\\_end.pl](https://github.com/Sunhh/NGS_data_processing/blob/master/drop_dup_both_end.pl). Trimmomatic v0.38<sup>26</sup> was used to remove low-quality reads with the following filtering criteria: SLIDINGWINDOW:4:20 LEADING:3 TRAILING:3 HEADCROP:10 MINLEN:40. The remaining reads were aligned to the 'Charleston Gray' watermelon reference genome v2<sup>19</sup> obtained from the CuGenDBv2<sup>27,28</sup> using BWA v0.5.9<sup>29</sup>. Reads originating from a single fragment were tagged and assigned to read groups using Picard Toolkit v2.18.7 (<http://broadinstitute.github.io/picard/>). A sequence dictionary and index for the reference genome was created with Picard and Samtools v0.1.8<sup>30</sup>, respectively. Variant calling followed the GATK Best Practices Workflow<sup>31</sup> in GATK v3.6<sup>32</sup>. Vcftools v0.1.15<sup>33</sup> was used to

retain bi-allelic SNPs with no missing data, a GQ score greater than 30 and the maximum read depth within one standard deviation of the mean read depth. Variant counts of the reference and alternative alleles for each pool were calculated with the CollectAllelicCounts function of GATK and used as input for XP-GWAS.

The R package, XP-GWAS<sup>21</sup>, was used to identify SNPs associated with tolerance to *P. xanthii* race 2 W with the depth filter set to 50. The package computes a likelihood ratio test statistic on the output of a generalized linear model for each SNP. The statistic was divided by the inflation factor  $\lambda^{34}$  to control for population structure. A false discovery rate (FDR) of 5% was applied for multiple testing correction<sup>35</sup>.

### KASP marker development

Significant SNPs from XP-GWAS were submitted to LGC Genomics (Teddington, Middlesex, UK) for KASP by Design Services. Alleles and flanking regions were used to design an assay mix able to distinguish alleles with allele-specific, fluorophore-labelled primers (See Supplementary table S2 online). PCR cocktails consisted of 15 ng of sample DNA, 2.5  $\mu$ L of 2x Master Mix (LGC Genomics, Teddington, Middlesex, UK), 0.07  $\mu$ L of Primer Mix (LGC Genomics, Teddington, Middlesex, UK) and molecular grade water to bring the total reaction volume up to 5  $\mu$ L. A touchdown PCR protocol was followed with an initial 15 min denaturing step at 94 °C, followed by ten touchdown cycles with a 20 s denaturing step at 94 °C and annealing temperatures starting at 61 °C and dropping by 0.6 °C each cycle. The protocol finished with twenty-six additional cycles of 94 °C for 20 s and 55 °C for 60 s. Allele discrimination was determined through quantification of fluorescence using a Stratagene Mx3005P qPCR Machine (Agilent Technologies, Santa Clara, CA) with MxPro v4.10 Software (Agilent Technologies, Santa Clara, CA). Association of the KASP markers with *P. xanthii* race 2 W tolerance was verified with an expanded set of accessions ( $N=186$ ; See Supplementary table S3 online) from the extremes of the distribution of the previously phenotyped USDA *Citrullus* collection<sup>7</sup>. Seeds of the accessions were provided by the USDA National Plant Germplasm System. Association between the markers and disease response was tested using analysis of variance (ANOVA) with the aov function<sup>36</sup> in R.

### Candidate genes

QTL intervals around the significant QTL extended to the base of the QTL peak in each direction. Functional annotation of the SNPs and genes in these regions were used to determine the most promising candidate genes for powdery mildew tolerance. SNPs were prioritized by annotation to include those located in promoter regions as they may affect transcription of a causal gene and missense/nonsense mutations as they would change protein structure. Genes encoding proteins previously implicated in disease resistance are the most promising candidate genes for this study. Genome annotation for the 'Charleston Gray' v2 genome<sup>19</sup> was obtained from the CuGenDB<sup>27</sup>. ANNOVAR was used for functional annotation of the SNPs<sup>37</sup>.

## Results

### Extreme-phenotype genome-wide association study

Sequencing of the genomic libraries generated 92.7 to 99.1 million reads per bulk for a minimum of 33.2x genomic coverage. The variant calling pipeline identified 1,157,440 SNPs between the bulks. Filtering of the raw SNP calls resulted in 301,059 high-quality biallelic SNPs for analysis. Marker density varied across the genome, with a range of 19,548 to 59,943 SNPs per chromosome. Two SNPs were significantly (FDR < 0.05, with genomic control) associated with disease response to artificial inoculation with *P. xanthii* race 2 W (Fig. 2). The SNPs were adjacent (4,865,001 bp and 4,865,003 bp) on chromosome 7. There were two additional SNPs on chromosomes 2 and 4 that approached significance (FDR = 0.11 and 0.14, respectively) and were therefore included for marker design (Fig. 2).

### Marker validation

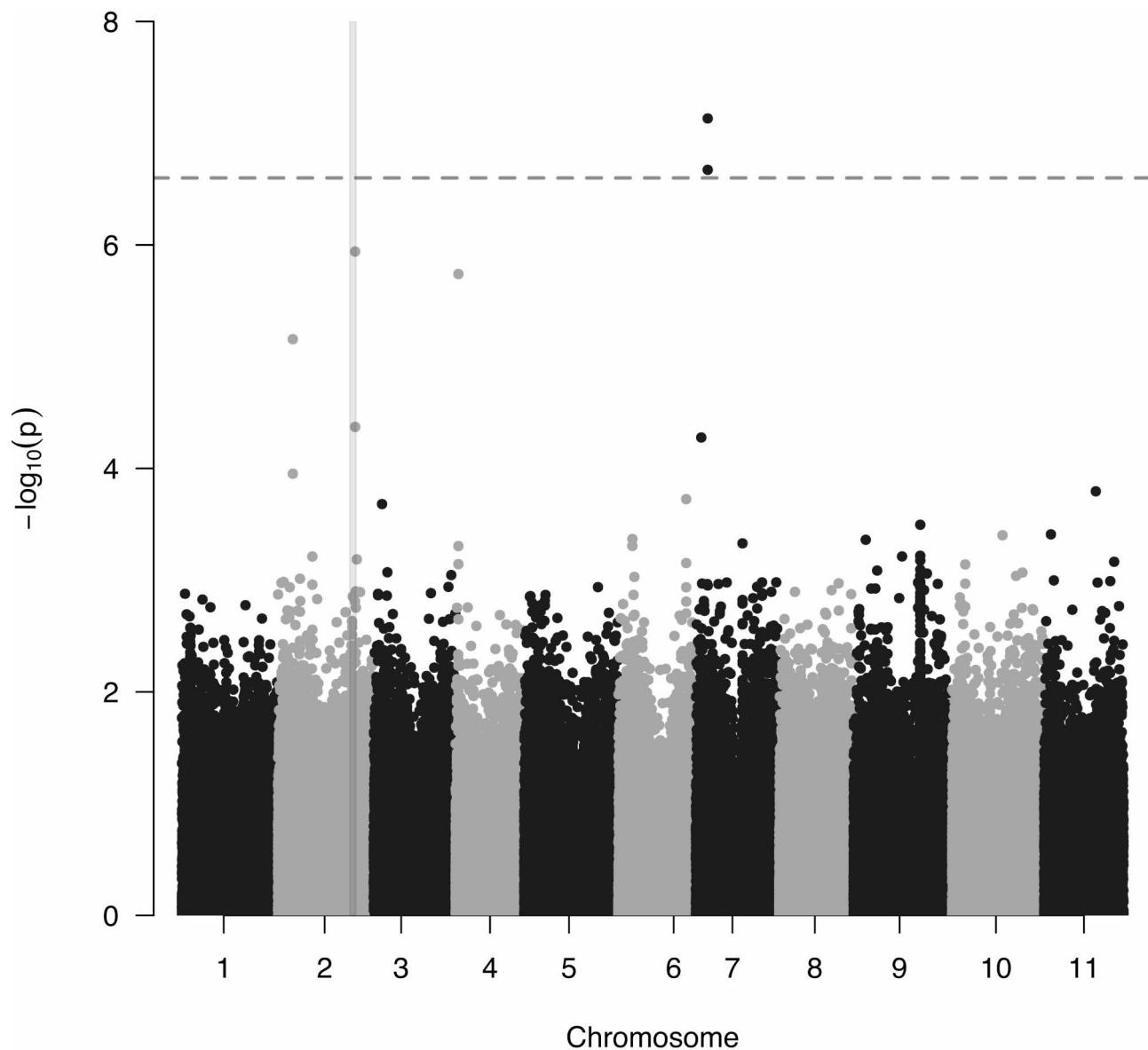
KASP markers were designed for the four SNPs (S2\_31304458, S4\_1119489, S7\_4865001 and S7\_4865003) with the strongest signal from XP-GWAS and two adjacent SNPs up and downstream of each for a total of 16 markers (See Supplementary table S2 online). An expanded set of accessions ( $N=186$ ) from the same phenotyped collection were genotyped with the KASP markers to validate their association with powdery mildew tolerance (See Supplementary table S3 online). Two markers were monomorphic and removed from downstream analyses. Thirteen of the fourteen remaining markers were significantly ( $P < 0.05$ ) associated with tolerance (Fig. 3; See Supplementary table S3 online).

The SNPs with the strongest association on each chromosome were CS2\_31304458 ( $\rho = 0.26$ ), CS4\_1119489 ( $\rho = 0.21$ ), and CS7\_4855905 ( $\rho = 0.31$ ) (Fig. 3). The pattern of marker association with disease severity was the same for both tissue types so allele effect plots are only shown for leaf ratings (Figs. 3 and 4). Two accessions of *C. mucosospermus* represent the most promising sources of donor material to use with these three markers for introgression of *P. xanthii* race 2 W resistance into elite *C. lanatus* lines. Accessions PI560020 and PI560005 had homozygous resistant genotypes at all three markers and disease severity scores indicating less than 20% chlorotic lesions on leaves (DS of 2.1 and 2.4, respectively) and less than three necrotic spots on their stems (DS of 1 and 1.4) (Supplementary table S3).

The additive effect of the tolerance allele for these markers ranged from 1.6 to 1.9 for the stem ratings and 1.4 to 1.6 for the leaf ratings. There was no evidence of enhanced tolerance in accessions with tolerance alleles at multiple loci (Fig. 4).

### Functional annotation and candidate genes

The QTL intervals on chromosomes 2 (31,287,589 to 31,388,599 bp), 4 (1,014,968 to 1,123,501 bp), and 7 (4,799,532 to 4,880,782 bp) encompassed 9, 4, and 4 genes, respectively (See Supplementary table S4 online). Functional annotation of the 246 SNPs within the QTL intervals was: 1 exonic, 7 upstream, 5 downstream, 15

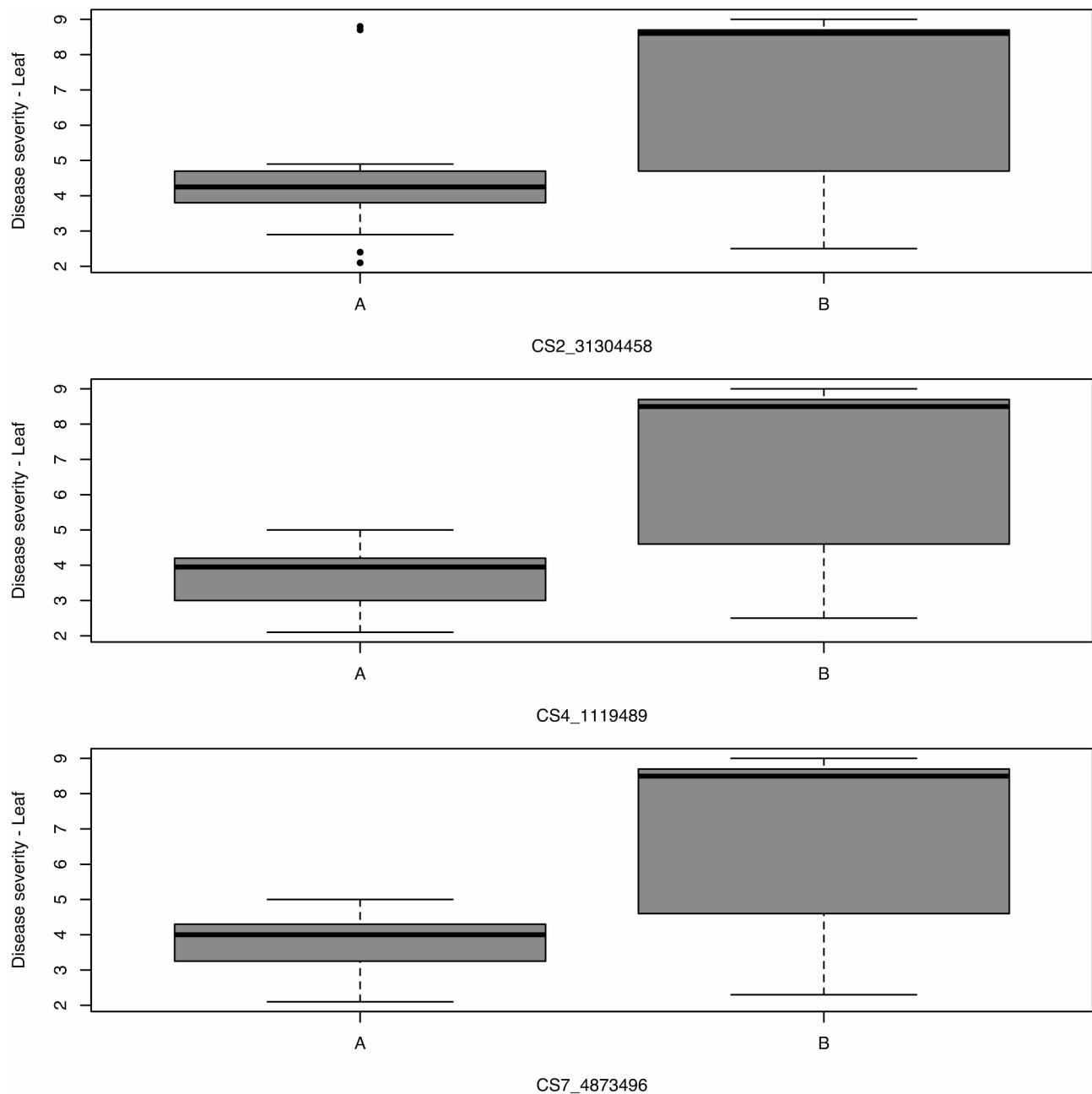


**Fig. 2.** Manhattan plot of the XP-GWAS results for disease response in the USDA *Citrullus* collection after inoculation with *Podosphaera xanthii* race 2 W<sup>7</sup>. The shaded box indicates the previously identified QTL associated with resistance to *P. xanthii* races 1 W<sup>17,19</sup> and 2 W<sup>19</sup>. The horizontal dashed line indicates an FDR significance threshold of 0.05. Figure generated with the qqman package<sup>38</sup> in R<sup>25</sup>.

intronic, and 218 intergenic (See Supplementary table S5 online). The exonic SNP was a nonsynonymous change from an A to a G at 1,075,967 bp on chromosome 4 causing an amino acid change from phenylalanine to serine in gene *ClCG04G000310*.

## Discussion

Genetic mapping studies of resistance to powdery mildew in watermelon have been limited to three previous studies, including a traditional GWAS using historical data for *P. xanthii* race 2 W<sup>19</sup>, a bi-parental QTL mapping study for race 1 W<sup>17</sup>, and a comparative genomic analysis for race 1 W<sup>18</sup>. Traditional GWAS of the same historical powdery mildew screening data<sup>7</sup> utilized here identified 43 significant SNPs across 9 of the 11 chromosomes (not chromosomes 5 or 11) of watermelon<sup>19</sup>. Two QTL identified here, on chromosomes 2 and 4, collocate with those identified through traditional GWAS. The QTL on chromosome 7 was not significantly associated with powdery mildew resistance in the GWAS study. Insufficient marker density may have contributed to this false negative as the accessions in the GWAS were genotyped with low-coverage GBS (384-plex)<sup>19</sup>, while in the present study the bulks were genotyped with whole-genome resequencing. The flanking SNPs from the GBS dataset were 48 kb and 264 kb away from the significant SNPs (S7\_4865001 and S7\_4865003) identified in the present study. The additional GWAS associations that were not significant with XP-GWAS may be due to the reduced number of

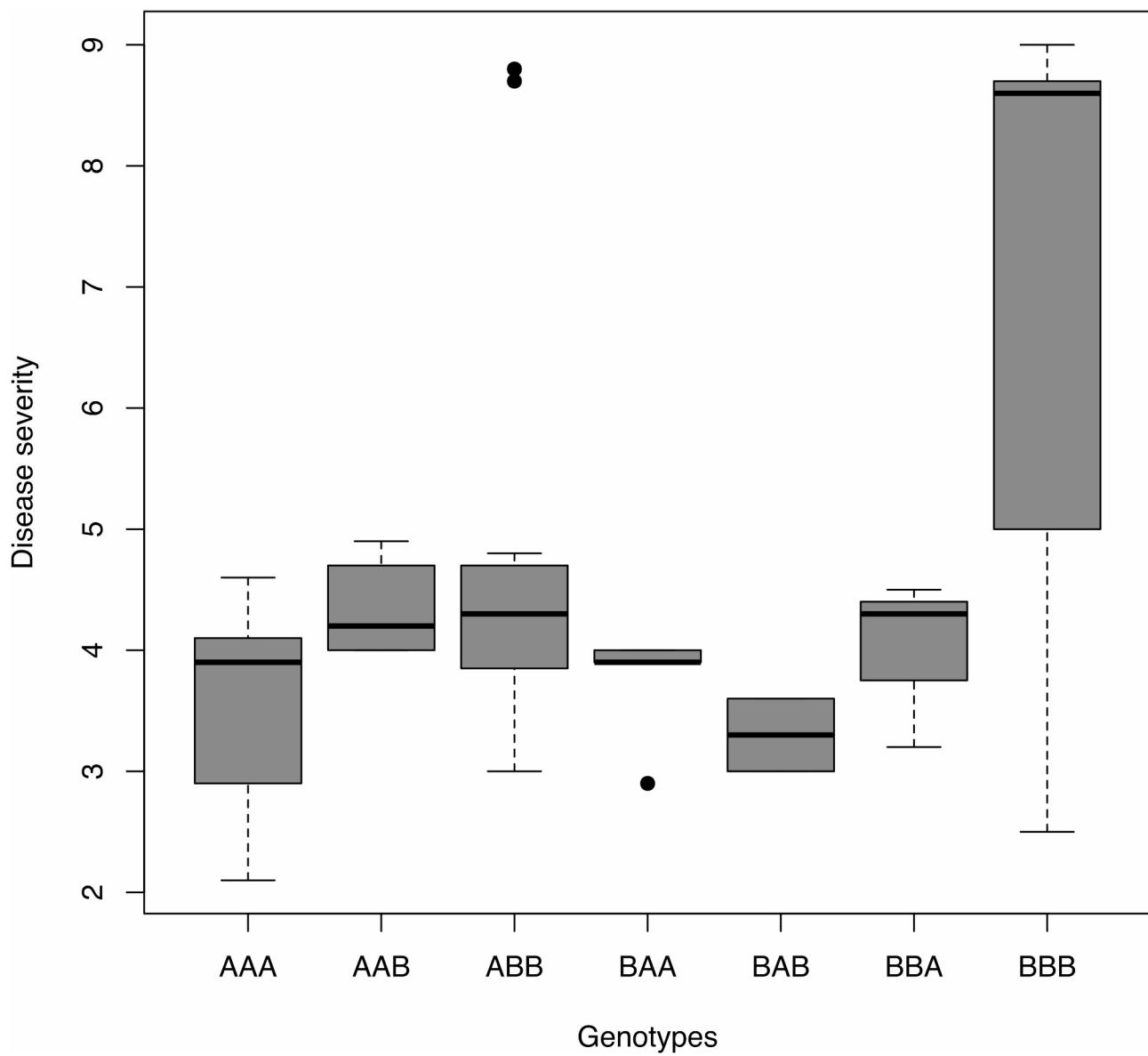


**Fig. 3.** Boxplots of genotypic effects of the most significant KASP markers developed for each QTL identified through XP-GWAS of disease response of the USDA *Citrullus* collection after inoculation with *Podosphaera xanthii* race 2 W<sup>7</sup>. Association between the markers and disease response was tested using analysis of variance (ANOVA) with the `aov` function<sup>30</sup> in R<sup>25</sup>. Homozygous resistant is labelled 'A' and homozygous susceptible is 'B'. The number of accessions with each allele at the SNPs were: SNP CS2\_31304458 A = 42 and B = 130; SNP CS4\_1119489 A = 26 and B = 158; and SNP CS7\_4873496 A = 23 and B = 162. Figure generated in R<sup>25</sup>.

accessions, from 1,147 to 91. Fewer accessions can decrease the power to detect associations, particularly for rare alleles and/or the resistance alleles may not have been present in the bulked accessions<sup>21</sup>.

QTL mapping of *P. xanthii* race 1 W resistance in a bi-parental mapping population<sup>17</sup> and comparative genomic analysis of a different resistance source for race 1 W<sup>18</sup> identified a single major QTL on chromosome 2. The XP-GWAS signal on chromosome 2 for *P. xanthii* race 2 W resistance collocated with this QTL, indicating that this region may provide broad-spectrum tolerance to *P. xanthii* in watermelon. The source of resistance derived from PI 494,531 used in studies by Mandal et al. (2020)<sup>18</sup> was also shown to be resistant to 11 isolates of powdery mildew collected from across the United States<sup>8</sup>.

There were no obvious candidate genes based upon SNP annotation or predicted gene function in the QTL interval on chromosome 2. The most significant SNPs were in an intergenic region between *ClCG02G016820* and *ClCG02G016830*, which are predicted to encode a flowering locus K homology domain and an unknown



**Fig. 4.** Boxplots of haplotype effects of the most significant KASP markers (as determined by ANOVA) developed for each QTL identified through XP-GWAS of disease response of the USDA *Citrullus* collection after inoculation with *Podosphaera xanthii* race 2 W<sup>7</sup>. Marker order in the genotype labels is S2\_31304458, S4\_1119489, and S7\_4873496. Homozygous resistant is labelled 'A' and homozygous susceptible is 'B'. The number accessions for each of the genotypes is: AAA = 9, AAB = 5, ABB = 27, BAA = 6, BAB = 2, BBA = 4, and BBB = 116. Figure generated in R<sup>25</sup>.

protein, respectively. Overexpression of *RLK-V* in a susceptible wheat line improved resistance to powdery mildew (caused by *Blumeria graminis* f. sp. *tritici*) and silencing of this gene in a resistant line reduced the response to infection<sup>39</sup>. An RLK was the closest R gene (41 kb) to the resistance QTL on chromosome 2, however, none of the SNPs within this gene were significantly associated with resistance with XP-GWAS. There were two promising candidate genes for the QTL on chromosome 4 based upon gene functional annotation, including one that encodes a P-loop containing nucleoside triphosphate hydrolase superfamily protein (*ClCG04G000310*) and a polygalacturonase inhibitor-like protein (PGIP; *ClCG04G000300*). Mutational analysis of several plant R-genes found that a functional P-loop domain is necessary for plant defense signaling<sup>40</sup>. The non-synonymous mutation in this gene may cause reduced ability to initiate the plant defense signaling cascade causing higher disease severity. PGIPs are cell wall proteins that recognize the cell-wall degrading enzymes, polygalacturonases, critical for fungal invasion into host cells<sup>41</sup>. Exogenous application of methyl jasmonate in a grape vineyard caused a dramatic increase in the transcription of pathogenesis-related proteins, including a PGIP, leading to a 73% reduction in powdery mildew-infected leaf surface area in treated plants<sup>42</sup>. Four SNPs were located in the promoter region of *ClCG04G000300* which could decrease expression of this gene, allowing more effective pathogen invasion. One promising candidate gene was found in this QTL interval. Peptidyl-prolyl cis-trans

isomerases (PPIs) function in protein folding allowing them to regulate protein structure, activity and stability<sup>43</sup>. Some PPIs activate a structural change in pathogen effector proteins triggering the plant immune response<sup>44</sup>. One SNP was located in the promotor region of *CICG07G004160*, which encodes a PPI. Decreased expression of this gene could cause a decreased immune response in powdery mildew susceptible accessions. These candidate genes represent the most promising targets for future gene editing studies to test their effects on powdery mildew disease response.

Public germplasm repositories have a wealth of historical phenotypic data associated with freely available accessions, much of which was collected prior to the genomics era. Whole-genome resequencing of bulked extremes of a diversity panel (XP-GWAS) offers a fast, effective, and relatively inexpensive mapping method. The ability to use historical data with this method will allow researchers with limited space and resources to pursue a variety of projects previously unattainable. We used XP-GWAS to successfully develop markers associated with tolerance to powdery mildew (*P. xanthii* race 2 W) using publicly available historical data for the USDA *Citrullus* collection. The KASP markers released here may be used, after validation, to incorporate powdery mildew tolerance into elite watermelon cultivars through marker-assisted selection.

## Data availability

The raw sequencing data generated during the current study are available in the NCBI SRA repository, <https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA1197509>. The remaining datasets generated during and/or analysed during the current study are included in the supplementary tables.

Received: 17 May 2024; Accepted: 5 February 2025

Published online: 08 February 2025

## References

- United States Department of Agriculture, Economic Research Service. Fruit and Tree Nuts Yearbook. <https://www.ers.usda.gov/ata-products/fruit-and-tree-nuts-data/fruit-and-tree-nuts-yearbook-tables/> (2019).
- Rennberger, G., Gerard, P. & Keinath, A. P. Occurrence of foliar pathogens of watermelon on commercial farms in South Carolina estimated with stratified cluster sampling. *Plant. Dis.* **102**, 2285–2295. <https://doi.org/10.1094/pdis-03-18-0468-re> (2018).
- Keinath, A. P. & Hassel, R. L. Evaluation of diploid and triploid watermelon cultivars for field resistance to anthracnose and gummy stem blight. *Biol. Cult. Tests.* **14**, 188 (2000).
- Cohen, R., Burger, Y. & Katzir, N. Monitoring physiological races of *Podosphaera Xanthii* (syn. *Sphaerotheca Fuliginea*), the causal agent of powdery mildew in cucurbits: factors affecting race identification and the importance for research and commerce. *Phytoparasitica* **32**, 174–183. <https://doi.org/10.1007/BF02979784> (2004).
- McCreight, J. D. Melon-powdery mildew interactions reveal variation in Melon cultigena and *Podosphaera Xanthii* races 1 and 2. *J. Am. Soc. Hortic. Sci.* **131**, 59–65. <https://doi.org/10.21273> (2006).
- Davis, A. R. et al. Evaluation of watermelon and related species for resistance to race 1 W Powdery Mildew. *J. Am. Soc. Hortic. Sci. J. Amer Soc. Hort Sci.* **132**, 790–795. <https://doi.org/10.21273/JASHS.132.6.790> (2007).
- Tetteh, A. Y., Wehner, T. C. & Davis, A. R. Identifying resistance to powdery mildew race 2 W in the USDA-ARS watermelon germplasm collection. *Crop Sci.* **50**, 933–939. <https://doi.org/10.2135/cropsci2009.03.0135> (2010).
- Kousik, C. S. & Ikerd, J. L. Evidence for cucurbit powdery mildew pathogen races based on watermelon differentials. In *Proceedings of Cucurbitaceae 2014*. (eds Havey, M. et al.) ISBN 978-0-9830932-2-0, 32–34 (American Society of Horticultural Science, 2014).
- Tores, J. A., Fernández-Ortuño, D., Daniel Jiménez, D., Guiderdone, S. & Bellón-Doña, D. Evidence of physiological races of *Podosphaera xanthii* in watermelon in Southern Europe. EUCARPA, Cucurbitaceae, 2021, P6-5, Page 81. <https://cucurbit.info/wp-content/uploads/2021/09/cuc2021proceedings.pdf> (2021).
- Keinath, A. P. & Dubose, V. B. Evaluation of fungicides for prevention and management of powdery mildew on watermelon. *Crop Prot.* **23**, 35–42. [https://doi.org/10.1016/S0261-2194\(03\)00165-0](https://doi.org/10.1016/S0261-2194(03)00165-0) (2004).
- McGrath, M. T. & Wyenandt, C. A. First detection of bosalid resistance in *Podosphaera Xanthii* in the United States associated with failure to control cucurbit powdery mildew in New York and New Jersey in 2009. *Plant. Health Prog.* **18**, 93–93. <https://doi.org/10.1094/PHP-03-17-0014-BR> (2017).
- Keinath, A. P., Rennberger, G. & Kousik, C. S. First report of resistance to bosalid in *Podosphaera Xanthii*, Cucurbit Powdery Mildew, in South Carolina. *Plant. Health Prog.* **19**, 220–221. <https://doi.org/10.1094/PHP-03-18-0009-BR> (2018).
- Kousik, C. S., Ikerd, J. L. & Mandal, M. Relative susceptibility of commercial watermelon varieties to powdery mildew. *Crop Prot.* **125**, 104910. <https://doi.org/10.1016/j.cropro.2019.104910> (2019).
- Rosado-Rivera, Y. & Quesada-Ocampo, L. Evaluation of watermelon cultivars for powdery mildew management, Goldsboro 2022. *Plant. Disease Manage. Rep.* **17**, V099 (2023).
- Tetteh, A. Y., Wehner, T. C. & Davis, A. R. Inheritance of resistance to the new race of powdery mildew in watermelon. *Crop Sci.* **53**, 880–887. <https://doi.org/10.2135/cropsci2012.07.0453> (2013).
- Ben-Naim, Y. & Cohen, Y. Inheritance of resistance to powdery mildew race 1 W in watermelon. *Phytopathology* **105**, 1446–1457. <https://doi.org/10.1094/PHYTO-02-15-0048-R> (2015).
- Kim, K. H. et al. Major quantitative trait loci and putative candidate genes for powdery mildew resistance and fruit-related traits revealed by an intraspecific genetic map for watermelon (*Citrullus lanatus* var. *lanatus*). *PLoS One.* **10** <https://doi.org/10.1371/journal.pone.0145665> (2015).
- Mandal, M. K., Suren, H. & Kousik, C. Elucidation of resistance signaling and identification of powdery mildew resistant mapping loci (ClaPMR2) during watermelon-podosphaera xanthii interaction using RNA-Seq and whole-genome resequencing approach. *Sci. Rep.* **10**, 14038. <https://doi.org/10.1038/s41598-020-70932-z> (2020).
- Wu, S. et al. Genome of 'Charleston Gray', the principal American watermelon cultivar, and genetic characterization of 1,365 accessions in the U.S. National Plant Germplasm System watermelon collection. *Plant. Biotechnol. J.* **17**, 2246–2258. <https://doi.org/10.1111/pbi.13136> (2019).
- Bian, Y. et al. Limits on the reproducibility of marker associations with southern leaf blight resistance in the maize nested association mapping population. *BMC Genom.* **15**, 1068. <https://doi.org/10.1186/1471-2164-15-1068> (2014).
- Yang, J. et al. Extreme-phenotype genome-wide association study (XP-GWAS): a method for identifying trait-associated variants by sequencing pools of individuals selected from a diversity panel. *Plant J.* **84**, 587–596. <https://doi.org/10.1111/tpj.13029> (2015).
- Tran, H. T. M. et al. SNP in the *Coffea arabica* genome associated with coffee quality. *Tree Genet. Genomes.* **14**, 72. <https://doi.org/10.1007/s11295-018-1282-9> (2018).
- Kumar, S. et al. Extreme-phenotype GWAS unravels a complex nexus between apple (*Malus domestica*) red-flesh colour and internal flesh browning. *Fruit Res.* **2**, 1–14. <https://doi.org/10.48130/frures-2022-0012> (2022).

24. Welling, M. T. et al. An extreme-phenotype genome-wide association study identifies candidate cannabinoid pathway genes in Cannabis. *Sci. Rep.* **10**, 18643. <https://doi.org/10.1038/s41598-020-75271-7> (2020).
25. R Core Team. *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, 2021).
26. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120. <https://doi.org/10.1093/bioinformatics/btu170> (2014).
27. Zheng, Y. et al. Cucurbit Genomics Database (CuGenDB): a central portal for comparative and functional genomics of cucurbit crops. *Nucleic Acids Res.* **47**, D1128–D1136. <https://doi.org/10.1093/nar/gky944> (2019).
28. Yu, J. et al. CuGenDBv2: an updated database for cucurbit genomics. *Nucleic Acids Res.* **51**, D1457–D1464. <https://doi.org/10.1093/nar/gkac921> (2023).
29. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760. <https://doi.org/10.1093/bioinformatics/btp324> (2009).
30. Li, H. et al. The sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079. <https://doi.org/10.1093/bioinformatics/btp352> (2009).
31. Van der Auwera, G. A. et al. From fastQ data to high-confidence variant calls: the genome analysis toolkit best practices pipeline. *Curr. Protoc. Bioinf.* <https://doi.org/10.1002/0471250953.bi1110s43> (2013).
32. McKenna, A. et al. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303. <https://doi.org/10.1101/gr.107524.110> (2010).
33. Danecek, P. et al. The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158. <https://doi.org/10.1093/bioinformatics/btr330> (2011).
34. Devlin, B. & Roeder, K. Genomic control for association studies. *Biometrics* **55**, 997–1004. <https://doi.org/10.1111/j.0006-341X.1999.00997.x> (1999).
35. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. Roy. Stat. Soc.: Ser. B (Methodol.)* **57** (1), 289–300 (1995).
36. Chambers, J., Hastie, T., Pregibon, D. & Dubrovnik Statistical models in S. In *Compstat: proceedings in computational statistics*. In 9th Symposium held at Yugoslavia, 1990, 317–321 (Physica-Verlag HD, 1992).
37. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **38** <https://doi.org/10.1093/nar/gkq603> (2010).
38. Turner, S. Qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. *J. Open. Source Softw.* <https://doi.org/10.21105/joss.00731> (2018). :.
39. Hu, P. et al. A malectin-like/leucine-rich repeat receptor protein kinase gene, RLK-V, regulates powdery mildew resistance in wheat. *Mol. Plant. Pathol.* **19**, 2561–2574. <https://doi.org/10.1111/mpp.12729> (2018).
40. Tameling, W. I. L. et al. The tomato R gene products i-2 and Mi-1 are functional ATP binding proteins with ATPase activity. *Plant. Cell.* **14**, 2929–2939. <https://doi.org/10.1105/tpc.005793> (2002).
41. De Lorenzo, G. & Ferrari, S. Polygalacturonase-inhibiting proteins in defense against phytopathogenic fungi. *Curr. Opin. Plant. Biol.* **5**, 295–299 (2002).
42. Belhadj, A. et al. Methyl jasmonate induces defense responses in grapevine and triggers protection against Erysiphe necator. *J. Agric. Food Chem.* **54**, 9119–9125. <https://doi.org/10.1021/jf0618022> (2006).
43. Wang, P. & Heitman, J. The cyclophilins. *Genome Biol.* **6** (2005).
44. Coaker, G. et al. Eukaryotic cyclophilin as a molecular switch for effector activation. *Mol. Microbiol.* **61**, 1485–1496. <https://doi.org/10.1111/j.1365-2958.2006.05335.x> (2006).

## Acknowledgements

Clemson University is acknowledged for their generous allotment of compute time on the Palmetto Cluster. This research was made possible, in part, with support from the Clemson University Genomics and Bioinformatics Facility, which receives support from the College of Science and two Institutional Development Awards (IDeA) from the National Institute of General Medical Sciences of the National Institutes of Health under grant numbers P20GM146584 and P20GM139769. This research was supported by grants from United States Department of Agriculture, National Institute of Food and Agriculture, Specialty Crop Research Initiative (grant numbers 2015-51181-24285 and 2020-51181-32139).

## Author contributions

SEB, WPW, and AL contributed to the study conception and design. Material preparation was performed by SEB, VRG, SK, LM, SH, and RG. Data collection was performed by SEB, VRG, YP, TW, AD, and AT. Analyses were performed by SEB and VRG. Figures were generated by SEB. The first draft of the manuscript was written by SEB and VRG, and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

## Funding

This research was supported by grants from United States Department of Agriculture, National Institute of Food and Agriculture, Specialty Crop Research Initiative (grant numbers 2015-51181-24285 and 2020-51181-32139).

## Declarations

### Competing interests

The authors declare no competing interests.

### Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-025-89445-8>.

**Correspondence** and requests for materials should be addressed to S.E.B. or A.L.

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

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

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

© The Author(s) 2025