

UC Davis

UC Davis Previously Published Works

Title

Uncovering genetic and metabolite markers associated with resistance against anthracnose fruit rot in northern highbush blueberry.

Permalink

<https://escholarship.org/uc/item/8m66p4mg>

Journal

Horticulture Research, 10(10)

ISSN

2662-6810

Authors

Song, Guo-Qing

Miles, Timothy

Edger, Patrick

et al.

Publication Date

2023-10-01

DOI

10.1093/hr/uhad169

Peer reviewed

## Article

# Uncovering genetic and metabolite markers associated with resistance against anthracnose fruit rot in northern highbush blueberry

MacKenzie Jacobs<sup>1,2,3</sup>, Samantha Thompson<sup>3,4</sup>, Adrian E. Platts<sup>2</sup>, Melanie J.A. Body<sup>2</sup>, Alexys Kelsey<sup>2</sup>, Amanda Saad<sup>2</sup>, Patrick Abeli<sup>2,5</sup>, Scott J. Teresi<sup>2,6</sup>, Anthony Schillmiller<sup>7</sup>, Randolph Beaudry<sup>2</sup>, Mitchell J. Feldmann<sup>8</sup>, Steven J. Knapp<sup>8</sup>, Guo-qing Song<sup>2</sup>, Timothy Miles<sup>3,4,6,\*</sup> and Patrick P. Edger<sup>2,3,6,\*</sup>

<sup>1</sup>Department of Biochemistry and Molecular Biology, Michigan State University, East Lansing, MI 48824, USA

<sup>2</sup>Department of Horticulture, Michigan State University, East Lansing, MI 48824, USA

<sup>3</sup>Molecular Plant Science Program, Michigan State University, East Lansing, MI 48824, USA

<sup>4</sup>Department of Plant, Soil and Microbial Sciences, Michigan State University, East Lansing, MI 48824, USA

<sup>5</sup>Department of Horticulture and Natural Resources, Kansas State University, Olathe, KS 66061, USA

<sup>6</sup>Genetics and Genome Sciences Program, Michigan State University, East Lansing, MI 48824, USA

<sup>7</sup>Mass Spectrometry & Metabolomics Core, Michigan State University, East Lansing, MI 48824, USA

<sup>8</sup>Department of Plant Sciences, University of California, Davis, CA 95616, USA

\*Corresponding authors. E-mail: [milesti2@msu.edu](mailto:milesti2@msu.edu); [edgerpat@msu.edu](mailto:edgerpat@msu.edu)

## Abstract

Anthracnose fruit rot (AFR), caused by the fungal pathogen *Colletotrichum fioriniae*, is among the most destructive and widespread fruit disease of blueberry, impacting both yield and overall fruit quality. Blueberry cultivars have highly variable resistance against AFR. To date, this pathogen is largely controlled by applying various fungicides; thus, a more cost-effective and environmentally conscious solution for AFR is needed. Here we report three quantitative trait loci associated with AFR resistance in northern highbush blueberry (*Vaccinium corymbosum*). Candidate genes within these genomic regions are associated with the biosynthesis of flavonoids (e.g. anthocyanins) and resistance against pathogens. Furthermore, we examined gene expression changes in fruits following inoculation with *Colletotrichum* in a resistant cultivar, which revealed an enrichment of significantly differentially expressed genes associated with certain specialized metabolic pathways (e.g. flavonol biosynthesis) and pathogen resistance. Using non-targeted metabolite profiling, we identified a flavonol glycoside with properties consistent with a quercetin rhamnoside as a compound exhibiting significant abundance differences among the most resistant and susceptible individuals from the genetic mapping population. Further analysis revealed that this compound exhibits significant abundance differences among the most resistant and susceptible individuals when analyzed as two groups. However, individuals within each group displayed considerable overlapping variation in this compound, suggesting that its abundance may only be partially associated with resistance against *C. fioriniae*. These findings should serve as a powerful resource that will enable breeding programs to more easily develop new cultivars with superior resistance to AFR and as the basis of future research studies.

## Introduction

Anthracnose fruit rot (AFR), caused by the fungal pathogen *Colletotrichum fioriniae* Marcelino & Gouli (=syn. *Glomerella acutata* var. *fioriniae*, formerly *Colletotrichum acutatum* Simmonds), is among the most destructive and widespread fruit disease of blueberries (*Vaccinium* sp.) [1]. The infection of *C. fioriniae* impacts fruit quality and can result in a complete loss of post-harvest yield [2]. *Colletotrichum* species have been reported to infect numerous other high-valued fruit crops, including apple, citrus, and strawberry [3]. Infections occur as early as fruit set but remain latent until the fruit ripens, complicating the disease's detection and protection. Initially, sunken areas develop on the fruit surface, followed by the exudation of salmon-colored spore masses (Fig. 1).

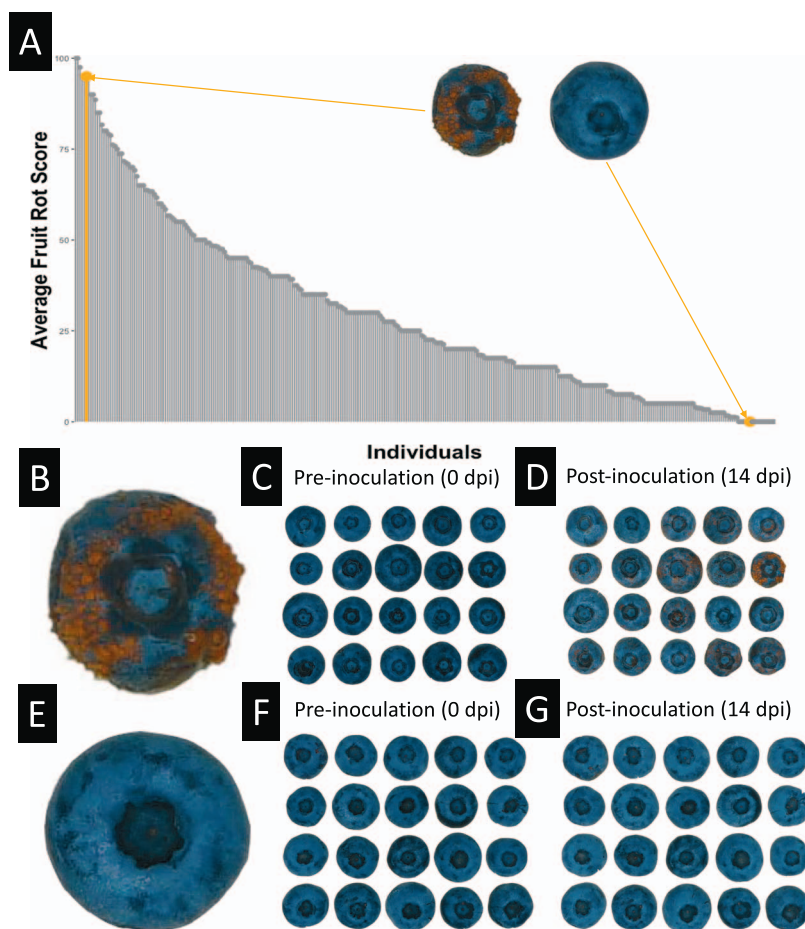
Most blueberry cultivars are highly to moderately susceptible to AFR [4]. Fungicides remain the primary method to mitigate

AFR infection in cultivated blueberry [5]. However, they are often expensive and not a favorable option for growers. Moreover, some of these fungicides are suspected carcinogens, whereas others are prone to fungicide resistance development [6]. Often, fungicide sprays are more frequently used than necessary because of the difficulty in optimizing spray timing due to the long latency period and variable weather conditions influencing the pathogen life cycle [5]. Therefore, the development of AFR-resistant cultivars is highly desired by the blueberry industry [7].

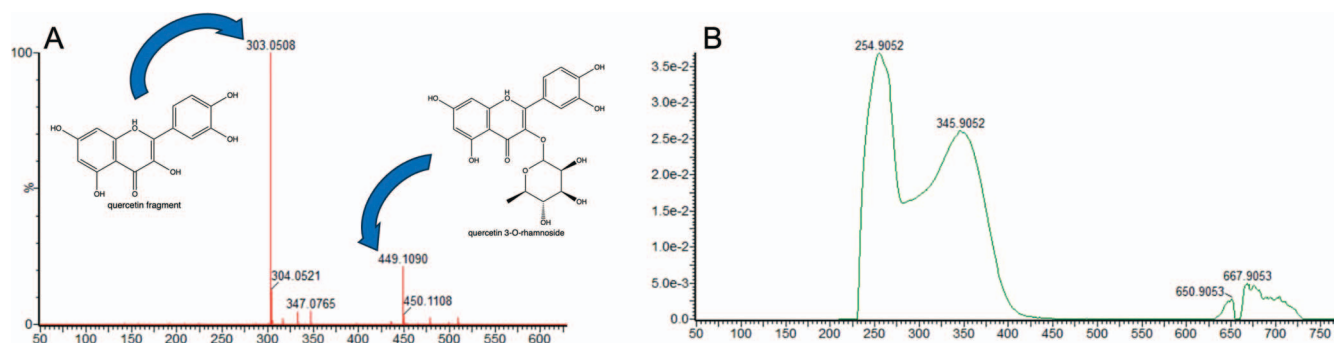
Several highly resistant cultivars have been identified, including northern highbush *Vaccinium corymbosum* L. 'Draper', which display strong resistance in the field and in laboratory inoculation studies [8, 9]. The genome of 'Draper' [10] was previously sequenced for three primary reasons: (1) it is a commonly utilized parent in breeding programs, (2) it is widely cultivated

Received: 4 January 2023; Accepted: 20 August 2023; Published: 29 August 2023; Corrected and Typeset: 1 October 2023

© The Author(s) 2023. Published by Oxford University Press on behalf of Nanjing Agricultural University. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.



**Figure 1.** Susceptibility assessment of 'Draper' x 'Liberty' Hybrids to AFR. A. Susceptibility of hybrids to AFR. Bars highlighted in orange represent the average fruit rot susceptibility score of the individuals shown in panels (B) and (E). The average fruit rot score represents the average number of berries that developed fruit rot infection per each replicate of 5 berries. So, an average score of 100% demonstrates all berries developing infection. B. Infected fruit from a susceptible individual. C. Susceptible fruits prior to infection via *C. fioriniae*. D. Susceptible fruits 14 days after infection via *C. fioriniae*. E. Infected fruit from a susceptible individual. F. Resistant fruits prior to infection via *C. fioriniae*. G. Resistant fruits 14 days after infection via *C. fioriniae*.



**Figure 2.** MS and UV/Vis Spectra of Quercetin Rhamnoside. (A) Ions detected consistent with expected  $m/z$  scores of a quercetin rhamnoside and its deglycosylated daughter ion. Here, the intact parent ion (right) is shown as quercetin 3-O-rhamnoside, however, the glycosylation may be in a different position. (B) UV/Vis spectra of the detected compound, consistent with the anticipated absorbance spectra of quercetin 3-O-rhamnoside.

worldwide as an early mid-season ripening variety, and (3) it is highly resistant to AFR. However, to our knowledge, no cultivars exhibit complete resistance [11, 12]. In these studies [11, 12], *C. fioriniae* had differential infection strategies and infection rates in resistant versus susceptible cultivars. Furthermore, Miles and Hancock [4] recently reported that resistance to AFR infection is highly heritable and argue that there are likely only a few loci involved in resistance. However, the underlying genetic mechanism(s) of resistance to AFR remains poorly understood in blueberry and other fruit crops.

Blueberry fruits contain a high concentration of many phytochemicals, including compounds with known antifungal properties [13, 14]. One potential component of resistance to AFR could involve specialized metabolites. For example, quercetin 3-O-rhamnoside is a flavonol glycoside synthesized from the amino acid precursor L-phenylalanine via the phenylpropanoid pathway [15] whose antimicrobial activity has been demonstrated against *C. fioriniae* [13], *Pseudomonas maltophilia*, and *Enterobacter cloacae* [16]. In fact, treating susceptible blueberry fruits with a 4% solution of extract from resistant fruit containing quercetin

3-O-rhamnoside, among other anthocyanins and non-anthocyanin flavonoids, decreased *C. fioriniae* infection by 88% [13]. Quercetin and its glycosides have been studied in other systems, but the dynamics of these compounds remain poorly understood in blueberry.

Quercetin glycosides may be deglycosylated, leaving the bioactive core, quercetin [17]. Structural analysis of plant-derived flavonoids revealed that quercetin contains numerous structural components important in bioactivity against certain pathogens, including methicillin-resistant *Staphylococcus aureus*, vancomycin-resistant enterococci, and *Burkholderia cepacia* [18]. Furthermore, quercetin may be oxidized to form quinones, antifungal compounds previously shown to be effective against certain *Colletotrichum* species [19]. However, previous studies have also proposed that AFR resistance in ripe blueberries may be due to an interaction between simple phenolic compounds and organic acids and not necessarily individual fungitoxic compounds [20–22].

Here, we used a genetic mapping approach to identify genomic loci associated with resistance to AFR infection in northern highbush blueberry. We generated an RNAseq dataset to identify which genes are differentially expressed during infection in ‘Draper’ mature fruits. Finally, we performed metabolite profiling in mature fruits and identified a metabolite with properties consistent with a quercetin rhamnoside whose abundance is positively correlated with AFR resistance.

## Results

### Identification of genetic markers associated with AFR resistance

Individuals within the ‘Draper’ × ‘Liberty’ genetic mapping population exhibited variation in susceptibility to AFR infection (Fig. 1). High resistance without showing any obvious AFR symptoms was observed in ~5% of the individuals surveyed. In comparison, ~4.7% exhibited high susceptibility to AFR—individuals averaging 4 or more out of 5 berries per replicate developing signs of rot (Fig. 1; Supplemental Figure S1; Supplemental Data S1, S2, and S3). Association mapping used genotyping data and fruit rot susceptibility scores as inputs into the GAPIT tool to determine the association of each SNP with the resistance phenotype under three models (GLM, BLINK, FarmCPU; Supplemental Figure S2). Overall, BLINK and FarmCPU models each suggested six QTLs after the Benjamini-Hochberg FDR adjustment, while the GLM model (Fig. 3; Supplemental Figure S3) suggested four distinct loci (12 distinct QTL loci over the three models). All approaches determined a significant QTL around a marker at Chromosome 17 position 22 625 275, while BLINK and GLM alone found a significant QTL at Chromosome 23 position 3 482 889, and an SNP on Chromosome 28 at position 31 421 447 was significant only in the BLINK and FarmCPU models (Fig. 3). The same regions were identified using R/qtl, with both a standard single QTL (qtl::scanone) and composite interval mapping (qtl::cim) and but with relatively lower LOD scores (Supplemental Figure S3). It’s important to note that Chromosomes 17, 23 and 28 are not homoeologous chromosomes, which are derived from the polyploidization event (Supplemental Figures S4 & S5).

The SNP on Chromosome 17 at position 22 625 275 lies in a candidate regulatory locus 1.8 Kilo bases (Kb) upstream of a poorly characterized protein with a moderate 34% similarity to the Arabidopsis cytokine signaling protein AT4G33800, a putative member of the SOB-FIVE- LIKE (SOFL) family of cytokinin genes [55]. SOFLs are a plant-specific gene family whose functions

remain poorly understood even in model species such as Arabidopsis [55]. Cytokinins are plant hormones that have diverse functions, including various aspects influencing plant growth and development, but they are also known to impact immunity and resistance against pathogen infections [56].

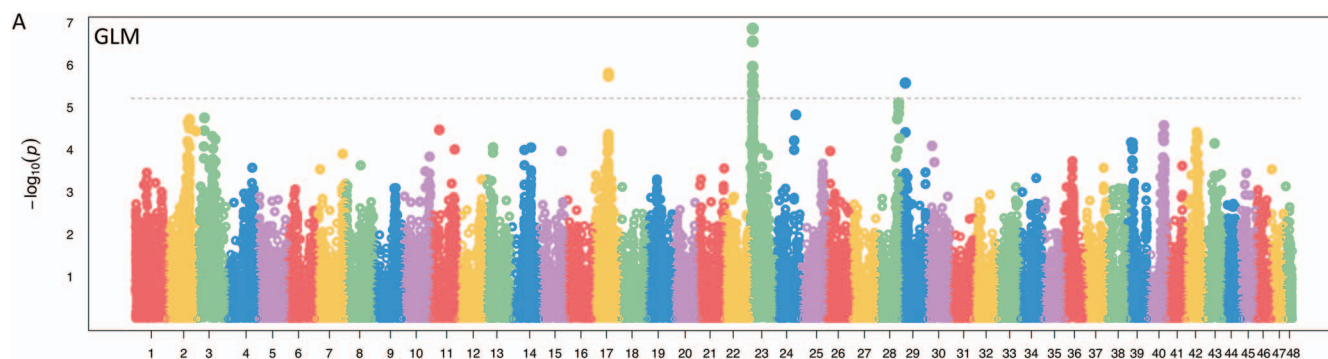
On chromosome 23 at position 3 482 889, there is an SNP that lies somewhat upstream (6 Kb) of a gene model orthologous to the Arabidopsis YABBY family transcription factor (TF) gene AT2G26580 (Yab5) as determined by both gene synteny and protein similarity (BLASTp 75%). While the most significant marker in this locus is somewhat distal to the gene’s transcription start site (TSS), the association peak in the GLM model is broadly characteristic of a sweep and includes linked SNPs both upstream and downstream of the gene. The YABBY family of TFs, notably Yab5, have been reported to interact physically with elements in the jasmonate pathway and may be involved indirectly in pathogen defense [57, 58] and potentially acting as upstream regulators of phenylalanine ammonia-lyase (PAL) [59]. PAL is an early component of the quercetin 3-O-rhamnoside biosynthesis pathway, where it catalyzes the conversion of L-phenylalanine to cinnamate and ammonium [60], a step common to the biosynthesis of anthocyanins and proanthocyanidins as well [60].

The SNP located on Chromosome 28 at position 31 421 447 was flagged for its highly significant association with fruit rot resistance by both BLINK and FarmCPU approaches. This SNP is genic and one of three synonymous SNPs located at consecutive coding 4D sites in a gene model predicted by Augustus [61] starting at the highly conserved residue 186 and positioned in exons on either side of the splice junction for the terminal coding exon. The gene model is syntenic with and has 73% BLASTp identity to the Arabidopsis gene GGL17 (AT3G11210), an SGNH hydrolase-type esterase superfamily protein that is predicted to function in anthocyanin metabolic processes [62, 63].

### Assessment of differential gene expression in response to anthracnose infection

Gene expression changes were evaluated in the fruit of ‘Draper’ following inoculation with *C. fioriniae*. Among the 128 559 protein-coding genes annotated in the haplotype-phased tetraploid genome of ‘Draper’ [10], a total of 2948 DEGs were identified during any one of the five time points during infection between the treated fruit and the corresponding controls (SRA SUB12410201). The most DEGs were identified after 2 DPI (48 hrs) and the lowest after 1 hr following inoculation. The number of identified DEGs per stage are as follows: Day 0 (450 upregulated, 57 downregulated), Day 1 (793 upregulated, 189 downregulated), Day 2 (1423 upregulated, 284 downregulated), Day 3 (600 upregulated, 13 downregulated), and Day 4 (1016 upregulated, 95 downregulated). Most DEGs, 2559 out of 2948 (~87%), exhibit single rhythmic patterns.

First, we evaluated the functional enrichment for DEGs across all time points (134 genes total). Five KEGG pathways were identified as being enriched with significant DEGs; 1. ath01100 Metabolic pathways (FDR P-value=0.0020), 2. ath00520 Amino sugar and nucleotide sugar metabolism (FDR P-value=0.0087), 3. ath00944 Flavone and flavonol biosynthesis (FDR P-value=0.0087), 4. ath00900 Terpenoid backbone biosynthesis (FDR P-value=0.0471), and 5. ath01110 Biosynthesis of secondary metabolites (FDR P-value=0.0471) (Table 1). No Gene Ontology (GO) Biological Processes terms were significantly enriched among these genes.



**Figure 3.** Genome-wide association study of AFR resistance in blueberry. Manhattan plot of SNP association with fruit rot resistance using a GLM model (Benjamini-Hochberg corrected significance threshold:  $5.9e-6$ , dotted line). Four loci on chromosomes 17, 23 (2 loci; Supplemental Figure S6), and 29 show a significance beyond the FDR threshold.

**Table 1.** KEGG pathway enrichment.

| KEGG ID  | Pathway description                                   | Timepoint 1 | Timepoint 2 | Timepoint 3 | Timepoint 4 | Timepoint 5 |
|----------|---|-------------|-------------|-------------|-------------|-------------|
| ath01100 | Metabolic pathways                                    | 1.50E-05    | 0.00018     | 5.27E-16    | 6.01E-09    | 3.21E-14    |
| ath01110 | Biosynthesis of secondary* metabolites                | 0.0015      | 4.98E-05    | 5.39E-17    | 6.01E-09    | 2.00E-14    |
| ath00908 | Zeatin biosynthesis                                   | NS          | 0.028       | 0.008       | 0.0417      | 0.00095     |
| ath00940 | Phenylpropanoid biosynthesis                          | NS          | 0.0083      | 4.89E-05    | NS          | 1.23E-05    |
| ath00941 | Flavonoid biosynthesis                                | NS          | 0.0019      | 0.0098      | NS          | 0.0015      |
| ath00945 | Stilbenoid, diarylheptanoid and gingerol biosynthesis | NS          | 0.0048      | 0.0141      | NS          | 0.0314      |
| ath00480 | Glutathione metabolism                                | NS          | NS          | 0.0237      | 0.00026     | 0.00061     |
| ath00900 | Terpenoid backbone biosynthesis                       | NS          | NS          | 4.89E-05    | 0.0152      | 0.0182      |
| ath00500 | Starch and sucrose metabolism                         | 0.0477      | NS          | 0.0461      | NS          | NS          |
| ath00010 | Glycolysis / Gluconeogenesis                          | NS          | NS          | NS          | 0.0152      | 0.0182      |
| ath00592 | alpha-Linolenic acid metabolism                       | NS          | NS          | 0.0486      | NS          | 0.0182      |
| ath01200 | Carbon metabolism                                     | NS          | NS          | 0.0276      | NS          | 0.0332      |
| ath01230 | Biosynthesis of amino acids                           | NS          | NS          | 0.0237      | NS          | 0.0332      |
| ath00270 | Cysteine and methionine metabolism                    | NS          | NS          | 0.0463      | NS          | NS          |
| ath00460 | Cyanoamino acid metabolism                            | NS          | NS          | 0.0463      | NS          | NS          |
| ath00330 | Arginine and proline metabolism                       | NS          | NS          | 0.0408      | NS          | NS          |
| ath00520 | Amino sugar and nucleotide sugar metabolism           | NS          | NS          | NS          | NS          | 0.0297      |
| ath04016 | MAPK signaling pathway                                | NS          | NS          | NS          | NS          | 0.0314      |
| ath00130 | Ubiquinone and other terpenoid-quinone biosynthesis   | NS          | NS          | NS          | NS          | 0.0332      |
| ath00904 | Diterpenoid biosynthesis                              | NS          | NS          | NS          | NS          | 0.0332      |

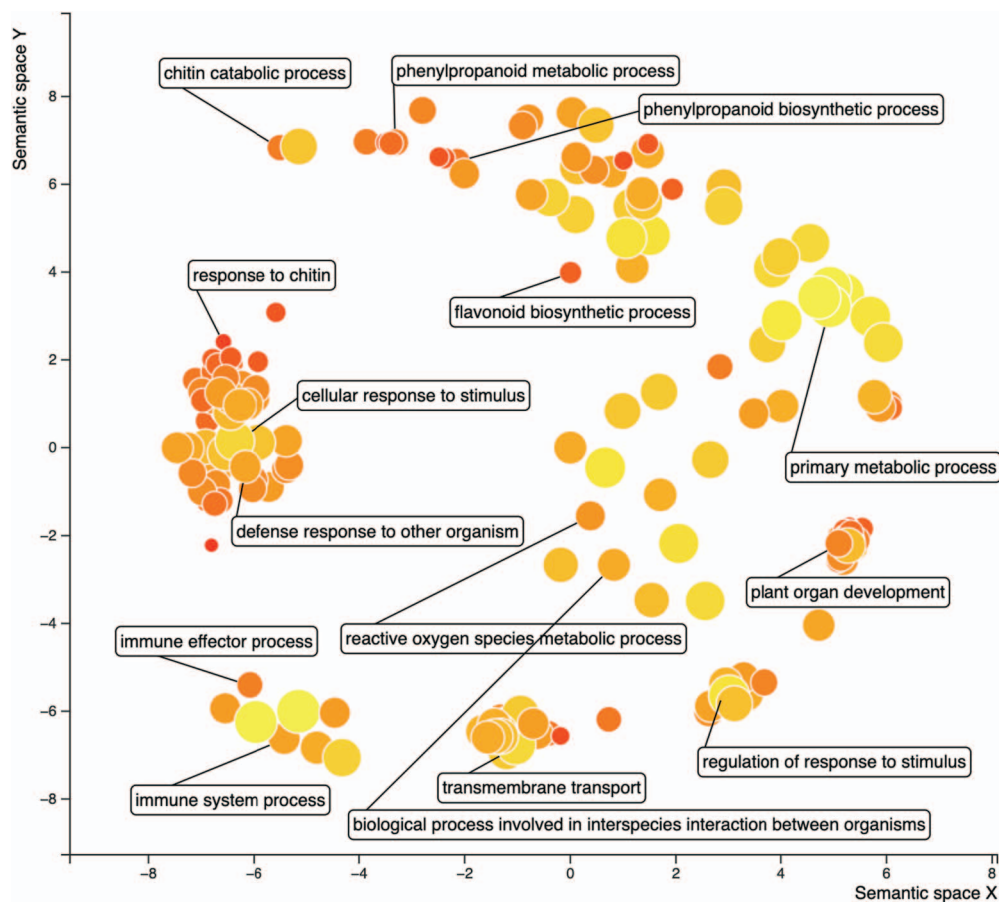
\*Referred to as 'specialized metabolites' throughout.

Next, we evaluated gene enrichment for differentially expressed genes across any time point. A total of 16 KEGG pathways were identified as significantly enriched with differentially expressed genes, including four of the five KEGG pathways identified when analyzing genes significantly differentially expressed during all time points (Table 1). The only KEGG pathway that was not enriched here was 'ath00944 Flavone and flavonol biosynthesis' but instead 'ath00941 Flavonoid biosynthesis (FDR  $P$ -value = 0.0012)', as well as the metabolite precursor biosynthetic pathways, ath00940 'Phenylpropanoid biosynthesis (FDR  $P$ -value = 0.0009)' and 'ath00360 Phenylalanine metabolism (FDR  $P$ -value = 0.0356), for quercetin 3-O-rhamnoside, was identified as being significantly enriched. A total of 183 GO terms for biological processes were identified as being enriched with significantly differentially expressed genes, including the majority being either related to stress (e.g. GO:0006950 Response to stress, FDR  $P$ -value =  $9.59e-20$ ), defense (e.g. GO:0006952 Defense response, FDR  $P$ -value =  $2.41e-07$ ) and/or metabolism (e.g. GO:0009698 Phenylpropanoid metabolic process, FDR  $P$ -value =  $2.16e-08$ ) (Fig. 4). The list of significantly differentially expressed genes identified here also significantly overlaps with genes identified by Toth et al. (2016) that investigated powdery mildew colonization

in grape (FDR  $P$ -value =  $9.40e-10$ ) [65]. Similarly, there was a significant overlap of differentially expressed genes discovered here with those identified from previous studies that focused on reactive oxygen species (ROS) signaling (e.g. [66], FDR  $P$ -value =  $3.84e-07$ ), fungal pathogen resistance (e.g. [67], FDR  $P$ -value = 0.00031), biosynthesis of flavonol glycosides (e.g. [68], FDR  $P$ -value = 0.0012) and phenylpropanoid metabolism in response to fungal interactions (e.g. [69], FDR  $P$ -value = 0.0018) (Supplemental Data S6).

Lastly, we evaluated each of the individual time points. During the first time point (1 hr post-inoculation), three KEGG pathways and 56 GO biological process terms were identified as enriched with significantly differentially expressed genes. The three KEGG pathways are 1. Ath01100 Metabolic pathways (FDR  $P$ -value =  $1.50e-05$ ), 2. Ath01110 'Biosynthesis of secondary metabolites' (FDR  $P$ -value = 0.0015) and 3. Ath00500 'Starch and sucrose metabolism' (FDR  $P$ -value = 0.0477) (Table 1). Highly enriched GO terms include GO:0006950 'Response to stress' (FDR  $P$ -value =  $5.04e-05$ ) and GO:0050896 'Response to stimulus' (FDR  $P$ -value =  $1.23e-05$ ). In addition, these genes had an enrichment of PFAM protein domains associated with 'innate immunity' (FDR  $P$ -value = 0.0052) (Supplemental Data S6).





**Figure 4.** Scatterplot of significantly enriched gene ontology (GO) terms of differentially expressed genes in blueberry fruit following treatment with the fungal pathogen *C. fioriniae*. The scatterplot generated using Revigo [64] displays the relationship among GO terms identified among significantly differentially expressed genes following inoculation of blueberries with *C. fioriniae*. The size of the bubble representing each GO term indicates the frequency in the underlying GOA database, while hue indicates P-value significance (dark being most significant). Most highly similar GO terms form clusters. The identity of select GO terms are shown with arrows and boxes.

During the second time point (24 hrs post-inoculation), six KEGG pathways and 56 GO biological process terms were identified that were enriched significantly differentially expressed genes. The six KEGG pathways are 1. Ath01110 'Biosynthesis of secondary metabolites' (FDR  $P$ -value =  $4.98 \times 10^{-5}$ ), 2. Ath01100 'Metabolic pathways' (FDR  $P$ -value = 0.00018), 3. Ath00940 'Phenylpropanoid biosynthesis' (FDR  $P$ -value = 0.0083), 4. Ath00941 'Flavonoid biosynthesis' (FDR  $P$ -value = 0.0019), 5. Ath00945 'Stilbenoid, diarylheptanoid and gingerol biosynthesis' (FDR  $P$ -value = 0.0048), and 6. Ath00908 'Zeatin biosynthesis' (FDR  $P$ -value = 0.0280) (Table 1). Highly enriched GO terms include GO:0010200 'Response to chitin' (FDR  $P$ -value =  $1.18 \times 10^{-7}$ ), GO:0006952 'Defense response' (FDR  $P$ -value =  $3.22 \times 10^{-6}$ ), and GO:0009698 'Phenylpropanoid metabolic process' (FDR  $P$ -value =  $5.98 \times 10^{-5}$ ). In addition, these genes had an enrichment of PFAM protein domains associated with 'stress response' (FDR  $P$ -value =  $6.07 \times 10^{-5}$ ), 'chitin degradation' (FDR  $P$ -value = 0.0337), and chitin-binding (FDR  $P$ -value = 0.0103). A significant overlap in shared differentially expressed genes was observed at this time point with Toth et al. (2016), which investigated powdery mildew colonization in grape [65] (FDR  $P$ -value =  $2.5 \times 10^{-5}$ ) (Supplemental Data S6).

During the third timepoint (48 hrs post-inoculation), fifteen KEGG pathways and 158 GO biological process terms were identified that were enriched with significantly differentially expressed genes. The top three most significant KEGG pathways

are 1. Ath01110 'biosynthesis of secondary metabolites' (FDR  $P$ -value =  $5.39 \times 10^{-17}$ ), 2. Ath01100 'metabolic pathways' (FDR  $P$ -value =  $5.27 \times 10^{-16}$ ) and 3. Ath00940 'Phenylpropanoid biosynthesis' (FDR  $P$ -value =  $4.89 \times 10^{-5}$ ) (Table 1). Flavonoid biosynthesis (Ath00941) is also enriched with significantly differentially expressed genes (FDR  $P$ -value = 0.0098). Highly enriched GO terms include 1. GO:0050896 'Response to stimulus' (FDR  $P$ -value =  $4.68 \times 10^{-19}$ ), 2. GO:0009698 'Phenylpropanoid metabolic process' (FDR  $P$ -value =  $4.66 \times 10^{-9}$ ), 3. GO:0051707 'Response to other organism' (FDR  $P$ -value =  $2.84 \times 10^{-7}$ ), 4. GO:0009699 'Phenylpropanoid biosynthetic process' (FDR  $P$ -value =  $4.18 \times 10^{-6}$ ), and 5. GO:0045087 'Innate immune response' (FDR  $P$ -value =  $7.11 \times 10^{-6}$ ). In addition, these genes had an enrichment of PFAM protein domains associated with 'chitin degradation' (FDR  $P$ -value = 0.009) and 'plant defense' (FDR  $P$ -value = 0.0267). Greater overlap in shared differentially expressed genes was observed in this timepoint with Toth et al. (2016) (FDR  $P$ -value =  $7.86 \times 10^{-11}$ ) [65] (Supplemental Data S6).

During the fourth time point (72 hrs post-inoculation), six KEGG pathways and 61 GO biological process terms were identified that were enriched significantly differentially expressed genes. The six KEGG pathways are 1. Ath01110 'Biosynthesis of secondary metabolites' (FDR  $P$ -value =  $6.01 \times 10^{-9}$ ), 2. Ath01100 'Metabolic pathways' (FDR  $P$ -value =  $6.01 \times 10^{-9}$ ), 3. Ath00480 'Glutathione metabolism' (FDR  $P$ -value = 0.00026), 4. Ath00900 'Terpenoid backbone biosynthesis' (FDR  $P$ -value = 0.0152), 5. Ath00010 'Glycolysis /

Gluconeogenesis' (FDR P-value=0.0152) and 6. Ath00908 'Zeatin biosynthesis' (FDR P-value=0.0417) (Table 1). Highly enriched GO terms include 1. GO:0050896 'Response to stimulus' (FDR P-value = 4.63e-09), 2. GO:0051707 'Response to other organism' (FDR P-value = 0.00016), 3. GO:0045087 'Innate immune response' (FDR P-value=0.00096), 4. GO:0009699 'Phenylpropanoid biosynthetic process' (FDR P-value=0.0024), and 5. GO:0006952 'Defense response' (FDR P-value=0.0197). In addition, these genes had an enrichment of PFAM protein domains associated with 'chitin-binding' (FDR P-value = 0.0048) and 'chitin degradation' (FDR P-value=0.0153). A significant overlap, but lower than compared to the previous time-point, in shared differentially expressed genes was observed at this timepoint with Toth et al. (2016) (FDR P-value = 1.28e-08) [65] (Supplemental Data S6).

During the final time point (96 hrs post-inoculation), 16 KEGG pathways and 112 GO biological process terms were identified that were enriched significantly differentially expressed genes. The top three most significant KEGG pathways are 1. Ath01110 'biosynthesis of secondary metabolites (FDR P-value = 2e-14), 2. Ath01100 'metabolic pathways' (FDR P-value = 3.21e-14) and 3. Ath00940 'Phenylpropanoid biosynthesis (FDR P-value = 1.23e-05) (Table 1). Flavonoid biosynthesis (Ath00941) is also enriched with significantly differentially expressed genes (FDR P-value = 0.0015). The overall patterns observed for this time point are similar to the third time point. Highly enriched GO terms include 1. GO:0050896 'Response to stimulus (FDR P-value = 5.15e-16), 2. GO:0009698 'Phenylpropanoid metabolic process' (FDR P-value = 1.15e-09), 3. GO:0009699 'Phenylpropanoid biosynthetic process' (FDR P-value = 3.13e-07), 4. GO:0009813 'Flavonoids biosynthesis process' (FDR P-value = 0.00043) and 5. GO:0045087 'Innate immune response' (FDR P-value = 0.0014) (Supplemental Data S6).

### Variation in metabolite abundance between resistant and susceptible individuals

In the initial screening, one replicate of fruit extracts from six resistant (average fruit rot score = 0%) and six susceptible (average fruit rot score  $\geq$  80%) individuals from the genetic mapping population were analyzed in an untargeted manner using liquid chromatography coupled with mass spectrometry (LC-MS). We compared the chromatograms to identify peaks with the largest fold changes between the resistant and susceptible individuals. A peak annotated as the quercetin daughter fragment of quercetin rhamnoside showed the largest fold change at  $\sim$ 2.5-fold higher abundance in resistant lines. This finding is supported by work published by Miles et al. (2013), stating that quercetin 3-O-rhamnoside extracted from blueberry fruit exhibits antifungal properties and is effective against the causative agent of fruit rot [13].

To further explore the potential role of the suspected quercetin rhamnoside in AFR resistance, targeted metabolite analysis was performed with additional replicates for each individual to quantify the abundance of the suspected quercetin rhamnoside in these extracts. Once again, we observed a significantly higher concentration of suspected quercetin rhamnoside in the extracts of resistant lines compared to susceptible lines (Fig. 5). However, some susceptible individuals did contain levels of these compounds more similar to resistant individuals (Supplemental Data S4, S5).

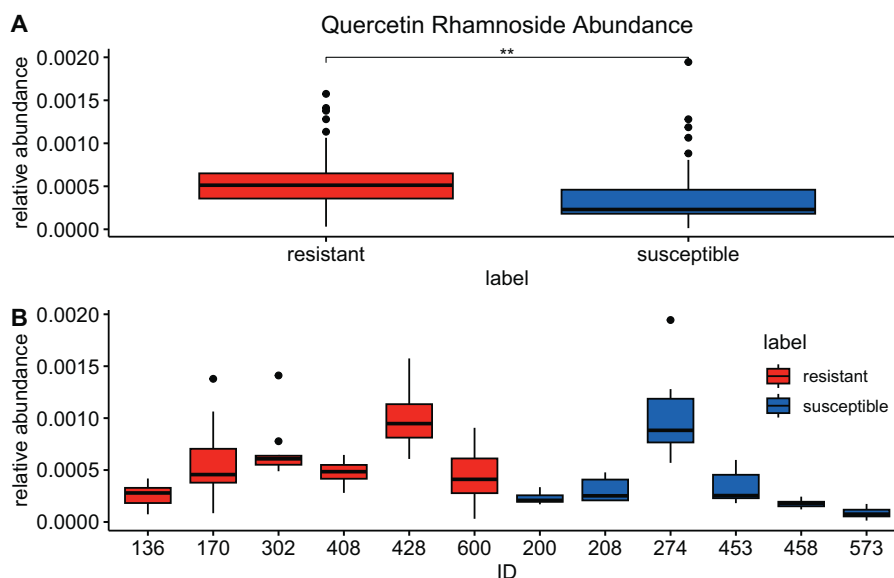
### Discussion

AFR is a top disease priority for the blueberry industry, as it can result in up to 100% post-harvest yield loss [2, 70]. Thus, growers have largely relied on fungicides to maximize yields. Both the

infection and resistance mechanisms of AFR are highly variable among and within crops (see review [71]). Resistance may arise from passive mechanisms such as physiological fruit characteristics and pre-existing compounds with antifungal properties. Immature fruits often exhibit many features that lend themselves to resistance to anthracnose, such as firmness [72, 73], pH [74], and antimicrobial compounds [75]. However, these resistance factors tend to dampen as fruit matures [72, 73]. Further, the accumulation of soluble sugars in conjunction with ascorbic acid was previously associated with anthracnose resistance in guava [76]. Work in blueberry indicates a connection between sugar content and anthracnose resistance, but some moderately susceptible cultivars have high sugar concentrations [71]. This suggests that sugar content may be only one piece of a multi-factor resistance mechanism. Additionally, the abundance of certain fruit volatiles, including (E)-Hex-2-enal, has been linked to fruit rot resistance in strawberry [77]. While some of these volatiles are also found in blueberry, their presence and quantity are not correlated with resistance [8, 78]. Various antimicrobial compounds are correlated with AFR resistance, including 1-acetoxy-2-hydroxy-4-oxoheneicosa-12,15-diene [75] and avocadynone acetate [79] in avocado, and dopamine and dopamine-derivatives in banana [80]. In blueberry, resistance was previously hypothesized to come from the interaction of several compounds, including phenolics and organic acids [20]. Supportive of this hypothesis, two non-anthocyanin flavonoids from blueberry fruit, quercetin 3-O-rhamnoside and putative syringetin-rhamnoside, were found to have antifungal activity against the causative agent of AFR [13].

Alternatively, resistant fruits may prompt active resistance mechanisms such as inducible defense-related proteins, including chitinases and  $\beta$ -1-3-glucanases that degrade the cell wall [81–84]. Expression of defense-related proteins in response to anthracnose fungi has been previously reported in pepper [85–88], apple [89, 90], raspberry [91], tomato [92], and blueberry [11]. Putative defense genes in blueberry include those that encode cell wall degrading protein, pathogenesis-related protein 10 (PR10), metallothionein-like protein, and monodehydroascorbate reductase [11]. Moreover, resistant fruits may actively combat anthracnose infection through the formation of reactive oxygen species (ROS). Production of ROS in response to anthracnose has been documented in tomato [13, 93], strawberry [94], and blueberry [12, 13, 93]. In all three crops, ROS occurs at or near the time of attempted penetration of the pathogen. Concurrently, infected fruits upregulate expression of oxidative stress response genes to minimize potential harmful effects of the ROS on the host's own tissues [11].

In this study, we sought to identify genomic loci associated with resistance to AFR by examining a genetic mapping population with tremendous variation in susceptibility. In order to capture allele and subgenome-specific variation that might be associated with resistance, we mapped to a fully haplotype resolved genome assembly with no explicit association made between the locations of the equivalent bases in the different subgenomes/haplotypes. Cultivated highbush blueberry (*V. corymbosum*) has a rich history of introgression from multiple other wild blueberry species from breeding efforts, plus natural gene flow that has previously shown to occur among sympatric blueberry species [95]. We have an unpublished analysis, as part of another study, that suggests that the genome of cultivar 'Draper', one of the parents analyzed here, has minimally 23% introgressed from at least three other *Vaccinium* species. This may, in part, be the reason why we identified significant candidate genomic regions in only one of the homoeologous chromosomes. Homoeologous chromosome



**Figure 5.** Abundance of metabolic markers in resistant and susceptible fruits. Normalized abundances of the suspected quercetin rhamnoside compound by fruit rot susceptibility (A) and by individual (B). Throughout the figure, boxes represent the interquartile ranges, bold horizontal lines depict the median of the observed abundances, vertical lines represent upper and lower extremes, and individual points represent outliers.

sets are shown in [Supplemental Figure S5](#). The different homoeologous chromosomes (i.e. haplotypes) of ‘Draper’ exhibit high sequence divergence. In effect, the tetraploid was treated in alignment terms as a single fully expanded haploid.

QTL mapping tools are available that function for polyploids, but they frequently depend on other tools such as *polymap* [96] that require dosage information that we are not exploring in the current work or require rather specific types of population that were not generated as a part of this study. Meanwhile, the GWAS approaches are necessarily somewhat robust to variable levels of population relatedness, explicitly incorporating this in approach-specific ways [26, 27], while more recent GWAS approaches adopt elements of QTL mapping to improve their sensitivity. Both the *FARMCPU* approach and *BLINK* approach for example exploit the generation of pseudo-markers and marker aggregation, *FARMCPU* using block-interval marker aggregation with block-size determined by MLE and *BLINK* marker aggregation in discovered bayesian linkage groups to recover essentially much of the power lost in approaches such as GWAS GLM/MLM where markers are considered essentially individually. Nonetheless, the different GWAS approaches selected very similar sites with only the association strength being notably different. *BLINK* and *FARMCPU* in this respect achieved higher levels of significance at these consistent sites likely due to their mapping-like marker aggregation. We have added to the supplementals a QTL mapping study results that includes standard interval mapping, composite interval mapping, and multiple QTL mapping to test whether the QTL and GWAS approaches are essentially in agreement. The QTL approaches largely support the GWAS sites albeit with the limitations on marker selection discussed above not always leading to greatly enhanced levels of significance relative to *FARMCPU/BLINK* ([Supplemental Figures S2 and S3](#)). In addition, we reran GWAS and QTL analysis to examine results with a binary categorization of susceptibility ( $\leq 1:0$ ,  $> 1:1$ ). There was no evident increase in association power when classifying the genotypes in a binary form ([Supplemental Figure S7](#)).

We identified three loci located on pseudomolecules 17, 23, and 28 in the *V. corymbosum* ‘Draper’ genome that are significantly

associated with the resistance phenotype ([Fig. 3](#)). Having several candidate causal SNPs is consistent with a polygenic resistance trait that may partly contribute to the observed continuous fruit-rot susceptibility distribution. Variants in several linked biosynthesis pathways, including cytokines, anthocyanins, and other flavonoids such as quercetin, may have become established in the population due to a protective role gained over time against multiple fungal pathogens. The identification of these loci will be useful for developing a molecular marker-assisted selection protocol for blueberry, which will greatly facilitate future screening for AFR resistance early in the breeding and selection process. Several candidate genes within these QTLs were previously associated with resistance against pathogens and/or flavonoid biosynthesis, including anthocyanins.

In comparing gene expression of blueberry fruits infected with *C. fioriniae* to uninfected fruits, numerous biological processes showed differential expression across the infection time course. Within 1 hour of infection, infected samples began upregulating genes related to metabolic pathways, including starch and sucrose metabolism and biosynthesis of secondary metabolites. In the following days, genes involved in metabolic pathways and biosynthesis of secondary metabolites continued to be upregulated. Additionally, the biosynthesis and metabolism of other primary and specialized metabolites and plant hormones were upregulated. Notably, many of the differential KEGG pathways show associations with the metabolite marker quercetin 3-O-rhamnoside. Most clearly, phenylpropanoid and flavonoid biosynthesis can be tied directly to quercetin 3-O-rhamnoside content as this metabolite falls within the phenylpropanoid and flavonoid families of specialized metabolites. Furthermore, the biosynthesis of amino acids may also be connected to quercetin 3-O-rhamnoside, as L-phenylalanine, an amino acid, is a precursor of quercetin 3-O-rhamnoside.

Our transcriptome analyses identified an enrichment of significantly differentially expressed genes associated with certain specialized metabolic pathways (e.g. flavonol biosynthesis) and pathogen resistance. Plus, several candidate genes within these QTLs were previously associated with resistance against



pathogens and/or flavonoid biosynthesis, including anthocyanins. However, it's important to note that none of these genes were identified as significantly differentially expressed. Low expressed genes, including certain regulators (e.g. transcription factors), may not be identified as significantly differently expressed but can cause a cascade of transcriptional changes, including inducing certain genetic pathways, and impact certain phenotypic traits. Thus, it's possible that our transcriptome analyses were unable to identify them based on the current data and significance thresholds. We hope that the community will engage in combining these results with their own to further explore the data and identify additional potential candidate genes.

Lastly, we identified a flavonol glycoside with accurate mass, fragmentation, and absorbance characteristics consistent with a quercetin rhamnoside, whose abundance is significantly greater in berries from resistant individuals. These findings confirm data previously reported by Miles et al. (2013), which identified quercetin 3-O-rhamnoside as an antifungal component of blueberry fruit [13]. Several efforts were taken to normalize any variation in metabolite content due to fruit location on the bush, collection time, and fruit ripeness. Fruits were collected from the entire bush and randomly separated into three subsets. Fruits within each subset were ground together and homogenized, resulting in three separate homogeneous mixtures of powdered fruit from each individual. Variability of the quercetin rhamnoside abundances was observed within resistant and susceptible individuals (Fig. 5). Furthermore, levels of quercetin were previously shown to vary greatly between different harvests - indicating that the biosynthesis of these compounds may be influenced by the environment [97].

In general, flavonol glycosides are known to have antioxidant activity (even higher than anthocyanin antioxidant properties [98]), and blueberry fruit flavonol extracts demonstrate antioxidative activity against peroxy and superoxide anion radicals [99]. Taken with the hypotheses of Cipollini and Stiles (1992a, 1992b, 1993), these findings suggest that quercetin 3-O-rhamnoside, a phenolic compound, is likely a major component of resistance to AFR in blueberry, but not the only component. Perhaps this molecule acts with other resistance mechanisms to protect the fruits from AFR.

Traditional breeding efforts in blueberry have contributed to major improvements of various target traits, but it is a lengthy and expensive process for perennial crops [31]. The use of molecular markers to guide breeding efforts has long been shown to greatly accelerate the development of superior cultivars, including selecting disease-resistant individuals [100]. There has been a strong community effort to develop and implement cost-effective methods for blueberry breeding programs [70]. Collectively, we hope that these findings reported in this study will allow breeders to develop new resistant cultivars to AFR more efficiently and rapidly by leveraging these new genetic regions to identify and select resistant individuals in their breeding programs. The phenotype predictive power from the combination of the three markers on chromosomes 17, 23 and 28 was assessed using a multiple regression approach. Individually the Pearson product moment ( $r$ ) between phenotype on the 0–5 scale and genotype ranged from a low of 0.25 (chromosome 17 SNP) to a high of 0.30 (chromosome 23 SNP), jointly a predictive model generated from all 3 sites had an  $r$  of 0.49 ( $p \sim 0$ ,  $dF = 3$ ,  $F = 30.38$ ). In summary, to the best of our knowledge, this is the first study to identify potential markers associated with resistance to AFR in blueberry. We anticipate that additional markers and candidate genes will likely be identified as part of future studies of this important

target trait in blueberry. We see the research presented in this manuscript as a stepping stone toward uncovering the underlying mechanism(s) that contributes to anthracnose resistance in fruit of northern highbush blueberry.

## Materials and methods

### Plant materials

A genetic mapping population (F1) of northern highbush blueberry cultivars was derived from a cross between cultivars 'Draper' and 'Liberty'. A total of 323 individual bushes growing at the Michigan State University Horticulture Teaching and Research Center were used during the 2020 and 2021 growing seasons in this study.

### Phenotyping - infection assay

To determine susceptibility to AFR, 5 to 20 full ripe blueberry fruits from the 323 F1 bushes were collected and assayed during July and August of 2020 (Supplemental Data S1) and 2021 (Supplemental Data S2). Fruits were sprayed evenly with *C. fioriniae* at a spore suspension of  $1 \times 10^6$ /mL, placed in humidifier trays under ambient conditions (21°C 12 h:12 h light: dark), and inoculated fruits were monitored for signs of anthracnose infection 12 days post-inoculation (DPI). Fruits were assigned, prior to the inoculation, to replicates containing 5 berries each, and the number of berries that developed symptoms of infection per replicate was assessed (Fig. 1). A replicate score of 100% indicates that all berries developed visible infection symptoms. All replicate 'scores' from 320 individuals were averaged to produce the 'average fruit rot scores' shown in Fig. 1 (Supplemental Data S3).

### Genotyping

Genotyping for single nucleotide polymorphisms (SNPs) was performed by RAPiD Genomics (Gainesville, FL, USA) using the Capture-Seq targeted system. A total of 31 063 biotinylated probes of 120-mer were originally designed based on the 'W8520' draft genome v. 2013 [23]. The probe sequences were then remapped against the 'Draper' genome [10] to retrieve probes that strictly align to only a single homoeologous chromosome group [24]. From the filtered set, 10 000 probes were selected throughout the 12 largest haploid chromosomes set and small contigs to be used for targeted sequencing. The presence of SNPs in the selected probes was tested on the parental 'Draper' and 'Liberty' before sequencing the entire population. Sequencing was then carried out on the Illumina HiSeq platform using a 150 cycle paired-end layout, generating just over 200 GBase of trimmed read data around the 10 000 probe sites. After read-trimming and variant calling 27 SNPs were filtered with VCFTools [25] with parameters —min-meanDP 5 —minQ 30 —maf 0.05 —remove-indels —max-missing 0.25 —min-alleles 2 —max-alleles 2 —recode. Downstream analyses only analyzed samples with at least 10× sequence depth and this comprised 323 samples with phenotype data and a mean sequencing depth of  $79.14 \times$  (395.7 Mb of data).

### QTL identification

A genome-wide association study was performed using the genotyping data and fruit rot susceptibility scores, with potential quantitative trait loci (QTL) associations assessed using the GAPIT tool (version 3) [26]. Several models were used to account for kinship and population structure (see the GAPIT manual for full model details: [https://zzlab.net/GAPIT/gapit\\_help\\_document.pdf](https://zzlab.net/GAPIT/gapit_help_document.pdf)). Overall Q-Q plots suggested that GLM weakly compensated for demographic artifacts, while the FarmCPU [27] and BLINK [28]

models both appeared to work well. Concordance between the most significant markers after false discovery rate (FDR: Benjamini and Hochberg [29];  $< 0.05$ ) correction in the association models was used to prioritize QTL locations for further investigation [30].

Genome wide association study was chosen over QTL mapping for several reasons. Northern highbush cultivars, including Draper and Liberty, are highly heterozygous and large amounts of introgression from multiple other wild blueberry species [31]. Thus, as part of this study, our goal was to analyze each haplotype separately, as a single fully expanded haploid. QTL mapping tools, which analyze allele dosage, are available for polyploid organisms, but this is outside of the scope of this project. Further, recent GWAS analysis methods utilize elements of QTL mapping to improve sensitivity. Here, we performed GWAS analysis using GLM, MLM, BLINK, and FarmCPU models as well as a QTL mapping study (R/qtl) that includes standard interval mapping, composite interval mapping, and multiple QTL mapping to test whether the QTL and GWAS approaches are in agreement (Supplementary Methods QTL mapping).

## RNA sequencing

Ripe berries were collected from 'Draper' and inoculated with either *C. fioriniae*,  $1 \times 10^6$  spores per milliliter with sterile distilled water, or sterile distilled water to serve as the 'mock-inoculated' control. Five time points were collected: one hour after inoculation (0 days) and 1, 2, 3, and 4 DPI for both inoculated and mock-inoculated with three biological replicates each (30 libraries total). Fruit tissue was sampled using a single-edged razor blade to excise all tissue within 2–3 mm of the calyx. For each of the 30 samples, RNA was extracted using the MagMax™ plant RNA Isolation Kit (Applied Biosystems/ThermoFisher Scientific) from ~50 mg of frozen ground tissue according to the manufacturer's recommended protocol for use with the KingFisher™ Flex Magnetic Particle Processor 96DW (ThermoFisher Scientific). RNase-free PVP40 (2% w/v) was added to the lysis buffer according to the manufacturer's recommendation to deal with polyphenol- and polysaccharide-rich samples. On-Column DNase digestion with the RNase-free DNase Set was used to remove DNA in the RNA samples (Qiagen, Valencia, CA, USA). RNAseq libraries were prepared using the mRNA HyperPrep Kit for Illumina® Platforms and the Dual-Indexed Adapter Kit for Illumina platforms and then sequenced using 150-bp paired-end reads on an Illumina HiSeq4000 system in MSU Research Technology Support Facility Genomics Core.

A total of 188.77 billion bases was sequenced, with an average of 20.97 million 150 bp paired end read per library (30 libraries total). Read quality was assessed using FastQC [32]. All reads were processed to remove adapters, low-quality leading or trailing bases, and minimum length using Trimmomatic v0.38 [33] with the following options: ILLUMINACLIP:TruSeq3-PE-2.fa:2:30:10:8:TRUE LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:36. Paired reads were aligned using Bowtie2 [34] against the 'Draper' reference genome and abundances estimated by the Expectation Maximum (RSEM) method [35]. The R package edgeR [36–46] (R version 3.4.1, edgeR version 3.18.1) was used to identify significantly differentially expressed genes (DEGs) following FDR correction ( $P < 0.05$ ) to determine differential expression between the control and inoculated berries. Gene enrichment analyses were formed using ortholog predictions to *Arabidopsis thaliana* [47], based on a combination of synteny- [48, 49] and BLASTp [50, 51] based approaches, and the STING database [52] containing KEGG

pathway [53] and Gene Ontology [54] data for Arabidopsis [33] (Supplemental Data S6).

## Metabolite analysis

Extraction buffer was prepared by combining 800 mL of HPLC grade methanol (Sigma-Aldrich,  $\geq 99.9\%$ ), 200 mL of HPLC grade H<sub>2</sub>O, 1 mL of formic acid, and 1 mL of 100  $\mu$ M telmisartan (internal standard). Ripe blueberry fruits from the 12 individuals exhibiting the strongest resistance or susceptibility to AFR in the 2020 fruit rot scoring were collected in 2021 and frozen (Supplemental Table S1). Frozen berries were ground to powder with a small amount of dry ice using a mortar and pestle in liquid nitrogen to prevent melting. Portions of ground tissue were transferred to 2 mL snap tubes, and 1 mL of extraction buffer was added to each tube. Tubes were vortexed until the tissue was thawed and then centrifuged at  $13 \times g$  for 10 minutes to pellet insoluble material. Supernatants were decanted and diluted 1:1 in HPLC grade H<sub>2</sub>O+0.1% formic acid. Snap tubes containing the pelleted insoluble material were dried at 60°C for one week (complete dryness), and the dry mass of extracted tissue was determined. Diluted extracts were analyzed on a Waters Acquity UPLC System coupled to a Waters Xevo G2-XS quadrupole time-of-flight mass spectrometer. Separations were performed using an Acquity UPLC HSS-T3 (1.8  $\mu$ m;  $2.1 \times 100$  mm) at 40°C. Solvent A consisted of H<sub>2</sub>O+0.1% formic acid, and solvent B was acetonitrile. A 10-minute elution gradient was used to separate 10  $\mu$ L of injected sample as follows (%A/%B): 0.00 min (100/0), 0.50 min (100/0), 6.00 min (50/50), 6.50 min (1/99), 8.50 min (1/99), 8.51(100/0), 10.00 (100/0) [10]. Positive ionization mode was used with a capillary voltage of 3.0 kV. Analysis of metabolites was completed in two phases. First, chromatograms and spectra were imported into Progenesis QI (Nonlinear Dynamics) which identifies peaks and quantifies them through the summation of intensities of compound adducts. The reported abundances for compounds in each sample were normalized to the dry mass of the extracted blueberry fruit tissue. The fold change of normalized metabolite intensities between resistant and susceptible lines was compared. The compound with the largest fold change, whose identity could be speculated, was consistent with the quercetin daughter fragment of a parent ion with qualities consistent with a quercetin rhamnoside (Fig. 2). This daughter ion has the same retention time as the quercetin rhamnoside parent, indicating that fragmentation likely occurred during ionization. The suspected quercetin rhamnoside was chosen for further pursuit. Next, the second round of LCMS analysis was performed on additional biological and technical replicates from each individual. A targeted QuanLynx processing method was designed to determine the intensity of the quercetin rhamnoside parent ion and the internal standard telmisartan (Supplemental Data S4). Parent ion intensities were normalized to telmisartan intensities and the dry mass of the tissue used to prepare the extract.

## Acknowledgments

This work was supported by United States Department of Agriculture - Agriculture and Food Research Initiative (USDA-AFRI) #2018-67013-27592 to G.S., T.M. and P.P.E.; United States Department of Agriculture - Specialty Crop Research Initiative (USDA-SCRI) #2019-51181-30015 to P.P.E.; and National Institute of Health (NIH) Training Grant #T32-GM110523 and National Science Foundation (NSF) Training Grant DGE-1828149 to M.J. and P.P.E. This work was also supported by Michigan State University AgBioRe-

search and United States Department of Agriculture – HATCH MICL02742 to P.P.E.

## Data Availability

The genomic and transcriptomic sequence reported in this paper has been deposited in the National Center for Biotechnology Information BioProject database (#PRJNA1005331). The phenotyping data, metabolite measurement data, and Gene Ontology enrichment results are available in the supplementary materials associated with this manuscript.

## Conflict of Interests

The authors have no conflicts of interest to declare.

## Supplementary Data

Supplementary data is available at Horticulture Research online.

## References

- Milholland RD. Anthracnose fruit rot (ripe rot). In: Caruso FL, Ramsdell DC, eds. *Compendium of Blueberry and Cranberry Diseases*. St. Paul, MN, USA: APS Press, 1995, 17
- Caruso FL. *Compendium of Blueberry and Cranberry Diseases*. Amer Phytopathological Society; 1995:
- Freeman S, Katan T, Shabi E. Characterization of *Colletotrichum* species responsible for anthracnose diseases of various fruits. *Plant Dis*. 1998;**82**:596–605
- Miles TD, Hancock JF. Inheritance of resistance to anthracnose fruit rot caused by *Colletotrichum fioriniae* in highbush blueberry. *Int J Fruit Sci*. 2022;**22**:160–9
- Wise JC, Miles LA, Acimovic D et al. Sprayer type and water volume influence spatial patterns of pesticide deposition and control of diseases and insect pests of highbush blueberries. *Int J Fruit Sci*. 2020;**20**:S1805–18
- Forcelini BB, Peres NA. Widespread resistance to QoI fungicides of *Colletotrichum acutatum* from strawberry nurseries and production fields. *Plant Health Prog*. 2018;**19**:338–41
- Gallardo RK, Klingthong P, Zhang Q et al. Breeding trait priorities of the cranberry industry in the United States and Canada. *HortScience*. 2018;**53**:1467–74
- Polashock JJ, Ehlenfeldt MK, Stretch AW et al. Anthracnose fruit rot resistance in blueberry cultivars. *Plant Dis*. 2005;**89**:33–8
- Miles TD, Hancock JF, Callow P et al. Evaluation of screening methods and fruit composition in relation to anthracnose fruit rot resistance in blueberries. *Plant Pathol*. 2012;**61**:555–66
- Colle M, Leisner CP, Wai CM et al. Haplotype-phased genome and evolution of phytonutrient pathways of tetraploid blueberry. *Gigascience*. 2019;**8**:giz012
- Miles TD, Day B, Schilder AC. Identification of differentially expressed genes in a resistant versus a susceptible blueberry cultivar after infection by *Colletotrichum acutatum*. *Mol Plant Pathol*. 2011;**12**:463–77
- Wharton PS, Schilder AC. Novel infection strategies of *Colletotrichum acutatum* on ripe blueberry fruit. *Plant Pathol*. 2007;**0**:070928213247003–???
- Miles TD, Vandervoort C, Nair MG et al. Characterization and biological activity of flavonoids from ripe fruit of an anthracnose-resistant blueberry cultivar. *Physiol Mol Plant Pathol*. 2013;**83**:8–16
- Wang H, Guo X, Hu X et al. Comparison of phytochemical profiles, antioxidant and cellular antioxidant activities of different varieties of blueberry (*Vaccinium* spp.). *Food Chem*. 2017;**217**:773–81
- Quattrocchio F, Baudry A, Lepiniec L et al. The Regulation of Flavonoid Biosynthesis. In: *The Science of Flavonoids*. New York, USA: Springer Science+Business Media, Inc, 2006, 97–122
- Waage SK, Hedin PA. Quercetin 3-O-galactosyl-(1 → 6)-glucoside, a compound from narrowleaf vetch with antibacterial activity. *Phytochemistry*. 1985;**24**:243–5
- Takahama U, Hirota S. Deglucosidation of quercetin glucosides to the aglycone and formation of antifungal agents by peroxidase-dependent oxidation of quercetin on browning of onion scales. *Plant Cell Physiol*. 2000;**41**:1021–9
- Xu H-X, Lee SF. Activity of plant flavonoids against antibiotic-resistant bacteria. *Phytother Res*. 2001;**15**:39–43
- Meazza G, Dayan FE, Wedge DE. Activity of quinones on *Colletotrichum* species. *J Agric Food Chem*. 2003;**51**:3824–8
- Cipollini ML, Stiles EW. Antifungal activity of ripe ericaceous fruits: Phenoliz-acid interactions and palatability for dispersers. *Biochem Syst Ecol*. 1992;**20**:501–14
- Cipollini ML, Stiles EW. Relative risks of microbial rot for fleshy fruits: significance with respect to dispersal and selection for secondary defense. In: Begon M, Fitter AH, eds. *Advances in Ecological Research*. San Diego, CA: Academic Press, 1992, 35–91
- Cipollini ML, Stiles EW. Fungi as biotic defense agents of fleshy fruits: alternative hypotheses, predictions, and evidence. *Am Nat*. 1993;**141**:663–73
- Bian Y, Ballington J, Raja A et al. Patterns of simple sequence repeats in cultivated blueberries (*Vaccinium* section *Cyanococcus* spp.) and their use in revealing genetic diversity and population structure. *Mol Breed*. 2014;**34**:675–89
- Benevenuto J, Ferrão LFV, Amadeu RR et al. How can a high-quality genome assembly help plant breeders? *Gigascience*. 2019;**8**:8
- Danecek P, Auton A, Abecasis G et al. The variant call format and VCFtools. *Bioinformatics*. 2011;**27**:2156–8
- GAPIT. <https://zzlab.net/GAPIT/> (accessed 7 November 2022).
- Liu X, Huang M, Fan B et al. Iterative usage of fixed and random effect models for powerful and efficient genome-wide association studies. *PLoS Genet*. 2016;**12**:1–24
- Huang M, Liu X, Zhou Y et al. BLINK: a package for the next level of genome-wide association studies with both individuals and markers in the millions. *Gigascience*. 2019;**8**:8
- Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Series B Stat Methodol*. 1995;**57**:289–300
- Kaler AS, Gillman JD, Beissinger T et al. Comparing different statistical models and multiple testing corrections for association mapping in soybean and maize. *Front Plant Sci*. 2019;**10**:1794
- Edger PP, Iorizzo M, Bassil NV et al. There and back again; historical perspective and future directions for *Vaccinium* breeding and research studies. *Hortic Res*. 2022;**9**:uhac083
- Staley B, Freeman L, Tipungwuti B et al. Barriers to inclusion: aboriginal pre-service teachers' perspectives on inclusive education in their remote Northern Territory schools. *Learn Community Int J Learn Soc Contexts*. 2019;**25**:78–89
- Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014;**30**:2114–20
- Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods*. 2012;**9**:357–9



35. Liu P. RNA-Seq gene expression estimation with read mapping uncertainty. <http://deweylab.github.io/RSEM/> (accessed 7 November 2022).
36. Robinson MD, Smyth GK. Small-sample estimation of negative binomial dispersion, with applications to SAGE data. *Biostatistics*. 2008;**9**:321–32
37. Robinson MD, Smyth GK. Moderated statistical tests for assessing differences in tag abundance. *Bioinformatics*. 2007;**23**:2881–7
38. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*. 2010;**26**:139–40
39. Robinson MD, Oshlack A. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol*. 2010;**11**:R25
40. McCarthy DJ, Chen Y, Smyth GK. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Res*. 2012;**40**:4288–97
41. Lun ATL, Chen Y, Smyth GK. It's DE-licious: a recipe for differential expression analyses of RNA-seq experiments using quasi-likelihood methods in edgeR. *Methods Mol Biol*. 2016;**1418**: 391–416
42. Chen Y, Lun ATL, Smyth GK. Differential expression analysis of complex RNA-seq experiments using edgeR. In: Datta S, Nettleton D, eds. *Statistical Analysis of Next Generation Sequencing Data*. Springer: Cham, 2014,51–74
43. Zhou X, Lindsay H, Robinson MD. Robustly detecting differential expression in RNA sequencing data using observation weights. *Nucleic Acids Res*. 2014;**42**:e91–1
44. Dai Z, Sheridan JM, Gearing LJ et al. edgeR: a versatile tool for the analysis of shRNA-seq and CRISPR-Cas9 genetic screens. *F1000Res*. 2014;**3**:95
45. Chen Y, Lun ATL, Smyth GK. From reads to genes to pathways: differential expression analysis of RNA-Seq experiments using Rsubread and the edgeR quasi-likelihood pipeline. *F1000Res*. 2016;**5**:1438
46. Chen Y, Pal B, Visvader JE et al. Differential methylation analysis of reduced representation bisulfite sequencing experiments using edgeR. *F1000Res*. 2017;**6**:2055
47. Lamesch P, Berardini TZ, Li D et al. The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. *Nucleic Acids Res*. 2012;**40**:D1202–10
48. Lyons E, Pedersen B, Kane J et al. The value of nonmodel genomes and an example using SynMap within CoGe to dissect the hexaploidy that predates the rosids. *Trop Plant Biol*. 2008;**1**: 181–90
49. Tang H, Bomhoff MD, Briones E et al. SynFind: compiling syntenic regions across any set of genomes on demand. *Genome Biol Evol*. 2015;**7**:3286–98
50. Altschul SF, Gish W, Miller W et al. Basic local alignment search tool. *J Mol Biol*. 1990;**215**:403–10
51. Gish W, States DJ. Identification of protein coding regions by database similarity search. *Nat Genet*. 1993;**3**:266–72
52. Szklarczyk D, Gable AL, Nastou KC et al. The STRING database in 2021: customizable protein–protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Res*. 2020;**49**:D605–12
53. Kanehisa M, Goto S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res*. 2000;**28**:27–30
54. Mi H, Muruganujan A, Ebert D et al. PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. *Nucleic Acids Res*. 2019;**47**:D419–26
55. Tayengwa R, Zhao J, Pierce CF et al. Synopsis of the SOFL plant-specific gene family. *G3 (Bethesda)*. 2018;**8**:1281–90
56. Akhtar SS, Mekureyaw MF, Pandey C et al. Role of cytokinins for interactions of plants with microbial pathogens and pest insects. *Front Plant Sci*. 2019;**10**:1777
57. Zhang L, Zhang F, Melotto M et al. Jasmonate signaling and manipulation by pathogens and insects. *J Exp Bot*. 2017;**68**: 1371–85
58. Boter M, Golz JF, Giménez-Ibañez S et al. Filamentous flower is a direct target of JAZ3 and modulates responses to jasmonate. *Plant Cell*. 2015;**27**:3160–74
59. Kayani S-I, Shen Q, Rahman S-U et al. Transcriptional regulation of flavonoid biosynthesis in *Artemisia annua* by AaYABBY5. *Hortic Res*. 2021;**8**:257
60. Jaakola L, Määttä K, Pirttilä AM et al. Expression of genes involved in anthocyanin biosynthesis in relation to anthocyanin, proanthocyanidin, and flavonol levels during bilberry fruit development. *Plant Physiol*. 2002;**130**:729–39
61. Stanke M, Waack S. Gene prediction with a hidden Markov model and a new intron submodel. *Bioinformatics*. 2003;**19 Suppl 2**:ii215–25
62. Ashburner M, Ball CA, Blake JA et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet*. 2000;**25**:25–9
63. Gene Ontology Consortium. The gene ontology resource: enriching a GOLD mine. *Nucleic Acids Res*. 2021;**49**:D325–34
64. Supek F, Bošnjak M, Škunca N et al. REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS One*. 2011;**6**:1–9
65. Toth Z, Winterhagen P, Kalapos B et al. Expression of a grapevine NAC transcription factor gene is induced in response to powdery mildew colonization in salicylic acid-independent manner. *Sci Rep*. 2016;**6**:30825
66. Mata-Pérez C, Sánchez-Calvo B, Begara-Morales JC et al. Transcriptomic profiling of linolenic acid-responsive genes in ROS signaling from RNA-seq data in *Arabidopsis*. *Front Plant Sci*. 2015;**6**:122
67. Galindo-González L, Deyholos MK. RNA-seq transcriptome response of flax (*Linum usitatissimum* L.) to the pathogenic fungus *Fusarium oxysporum* f. sp. lini. *Front Plant Sci*. 2016;**7**:1766
68. Neugart S, Krumbein A, Zrenner R. Influence of light and temperature on gene expression leading to accumulation of specific flavonol glycosides and hydroxycinnamic acid derivatives in kale (*Brassica oleracea* var. sabellica). *Front Plant Sci*. 2016;**7**:326
69. Bajaj R, Huang Y, Gebrechristos S et al. Transcriptional responses of soybean roots to colonization with the root endophytic fungus *Piriformospora indica* reveals altered phenylpropanoid and secondary metabolism. *Sci Rep*. 2018;**8**: 10227
70. Gallardo RK, Karina Gallardo R, Zhang Q et al. Breeding trait priorities of the blueberry industry in the United States and Canada. *HortScience*. 2018;**53**:1021–8
71. Miles TD, Schilder AC. Host defenses associated with fruit infection by *Colletotrichum* species with an emphasis on anthracnose of blueberries. *Plant Health Prog*. 2013;**14**:14
72. Sacher JA. Senescence and postharvest physiology. *Annu Rev Plant Physiol*. 1973;**24**:197–224
73. Brady CJ. Fruit ripening. *Annu Rev Plant Physiol*. 1987;**38**:155–78
74. Prusky D, Koblier I, Ardi R et al. Resistance mechanisms of subtropical fruits to *Colletotrichum gloeosporioides*. In: Prusky D, Freeman S, Dickman MB, eds. *Colletotrichum: Host Specificity, Pathology, and Host-Pathogen Interaction*. American Phytopathological Society: St. Paul, MN, 2000,232–44



75. Prusky D, Keen NT, Eaks I. Further evidence for the involvement of a preformed antifungal compound in the latency of *Colletotrichum Gloeosporioides* on unripe avocado fruits. *Physiol Plant Pathol.* 1983;**22**:189–98
76. Singh J, Sharma S. Screening and chemical basis of resistance in guava varieties to anthracnose (*Glomerella cingulata*). *Haryana J Hortic Sci.* 1981;**10**:155–7
77. Arroyo FT, Moreno J, Daza P et al. Antifungal activity of strawberry fruit volatile compounds against *Colletotrichum acutatum*. *J Agric Food Chem.* 2007;**55**:5701–7
78. Polashock JJ, Saftner RA, Kramer M. Postharvest highbush blueberry fruit antimicrobial volatile profiles in relation to anthracnose fruit rot resistance. *J Am Soc Hortic Sci.* 2007;**132**: 859–68
79. Prusky D, Kobiler I, Fishman Y et al. Identification of an antifungal compound in unripe avocado fruits and its possible involvement in the quiescent infections of *Colletotrichum gloeosporioides*. *J Phytopathol.* 1991;**132**:319–27
80. Muirhead IF, Deverall BJ. Evaluation of 3,4-dihydroxybenzaldehyde, dopamine and its oxidation products as inhibitors of *Colletotrichum musae* (Berk. & Curt.) Arx in green Banana fruits. *Aust J Bot.* 1984;**32**:575
81. Casado-Díaz A, Encinas-Villarejo S, de los Santos B et al. Analysis of strawberry genes differentially expressed in response to *Colletotrichum* infection. *Physiol Plant.* 2006;**128**:633–50
82. Goodwin P. Comparative analysis of expressed sequence tags from *Malva pusilla*, *Sorghum bicolor*, and *Medicago truncatula* infected with *Colletotrichum* species. *Plant Sci.* 2004;**167**:481–9
83. Lafitte C, Barthe JP, Montillet JL et al. Glycoprotein inhibitors of *Colletotrichum lindemuthianum* endopolygalacturonase in near isogenic lines of *Phaseolus vulgaris* resistant and susceptible to anthracnose. *Physiol Plant Pathol.* 1984;**25**:39–53
84. Wijesundera RLC, Bailey JA, Byrde RJW et al. Cell wall degrading enzymes of *Colletotrichum lindemuthianum*: their role in the development of bean anthracnose. *Physiol Mol Plant Pathol.* 1989;**34**:403–13
85. Oh BJ, Ko MK, Kostenyuk I et al. Coexpression of a defensin gene and a thionin-like via different signal transduction pathways in pepper and *Colletotrichum gloeosporioides* interactions. *Plant Mol Biol.* 1999;**41**:313–9
86. Oh BJ, Ko MK, Kim YS et al. A cytochrome P450 gene is differentially expressed in compatible and incompatible interactions between pepper (*Capsicum annum*) and the anthracnose fungus, *Colletotrichum gloeosporioides*. *Mol Plant-Microbe Interact.* 1999;**12**:1044–52
87. Kim YS, Park JY, Kim KS et al. A thaumatin-like gene in non-climacteric pepper fruits used as molecular marker in probing disease resistance, ripening, and sugar accumulation. *Plant Mol Biol.* 2002;**49**:125–35
88. Ko MK, Jeon WB, Kim KS et al. A *Colletotrichum gloeosporioides*-induced esterase gene of nonclimacteric pepper (*Capsicum annum*) fruit during ripening plays a role in resistance against fungal infection. *Plant Mol Biol.* 2005;**58**:529–41
89. Gregori R, Mari M, Bertolini P et al. Reduction of *Colletotrichum acutatum* infection by a polygalacturonase inhibitor protein extracted from apple. *Postharvest Biol Technol.* 2008;**48**:309–13
90. Yao C, Conway WS, Ren R et al. Gene encoding polygalacturonase inhibitor in apple fruit is developmentally regulated and activated by wounding and fungal infection. *Plant Mol Biol.* 1999;**39**:1231–41
91. Johnston DJ, Williamson B, McMillan GP. The interaction in plants of polygalacturonases from *Botrytis cinerea* with a cell wall-bound polygalacturonase-inhibiting protein (PGIP) in raspberry fruits. *J Exp Bot.* 1994;**45**:1837–43
92. Stotz HU, Contos JJ, Powell AL et al. Structure and expression of an inhibitor of fungal polygalacturonases from tomato. *Plant Mol Biol.* 1994;**25**:607–17
93. Mellersh DG, Foulds IV, Higgins VJ et al. H<sub>2</sub>O<sub>2</sub> plays different roles in determining penetration failure in three diverse plant-fungal interactions. *Plant J.* 2002;**29**:257–68
94. Brown SH, Yarden O, Gollop N et al. Differential protein expression in *Colletotrichum acutatum*: changes associated with reactive oxygen species and nitrogen starvation implicated in pathogenicity on strawberry. *Mol Plant Pathol.* 2008;**9**:171–90
95. Manzanero BR, Kulkarni KP, Vorsa N et al. Genomic and evolutionary relationships among wild and cultivated blueberry species. *BMC Plant Biol.* 2023;**23**:126
96. Bourke PM, van Geest G, Voorrips RE et al. polymapR—linkage analysis and genetic map construction from F1 populations of outcrossing polyploids. *Bioinformatics.* 2018;**34**:3496–502
97. Cvetković M, Kočić M, Dabić Zagorac D et al. When is the right moment to pick blueberries? Variation in agronomic and chemical properties of blueberry (*Vaccinium corymbosum*) cultivars at different harvest times. *Meta.* 2022;**12**:12
98. Mannino S, Brenna O, Buratti S et al. A new method for the evaluation of the 'antioxidant power' of wines. *Electroanalysis.* 1998;**10**:908–12
99. Cho MJ, Howard LR, Prior RL et al. Flavonol glycosides and antioxidant capacity of various blackberry and blueberry genotypes determined by high-performance liquid chromatography/mass spectrometry. *J Sci Food Agric.* 2005;**85**:2149–58
100. Moose SP, Mumm RH. Molecular plant breeding as the foundation for 21st century crop improvement. *Plant Physiol.* 2008;**147**: 969–77