

Article

Trait Mapping of Phenolic Acids in an Interspecific (*Vaccinium corymbosum* var. *caesariense* × *V. darrowii*) Diploid Blueberry Population

Ira A. Herniter ^{1,*} , Yurah Kim ¹, Yifei Wang ¹, Joshua S. Havill ² , Jennifer Johnson-Cicalese ^{1,3}, Gary J. Muehlbauer ², Massimo Iorizzo ^{4,5} and Nicholi Vorsa ^{1,3}

¹ Department of Plant Biology, Rutgers University, New Brunswick, NJ 08901, USA

² Department of Agronomy and Plant Genetics, University of Minnesota, St. Paul, MN 55108, USA

³ Phillip E. Marucci Center for Blueberry and Cranberry Research and Extension, 125a Lake Oswego Road, Chatsworth, NJ 08019, USA

⁴ Department of Horticultural Science, North Carolina State University, Raleigh, NC 27695, USA

⁵ Plants for Human Health Institute, North Carolina State University, Kannapolis, NC 28081, USA

* Correspondence: ira.herniter@rutgers.edu

Abstract: Blueberries (*Vaccinium* sect. *Cyanococcus*) are a dietary source of phenolic acids, including chlorogenic acid (CGA) and related compounds such as acetylated caffeoylquinic acid (ACQA) and caffeoylarbutin (CA). These compounds are known to be potent antioxidants with potential health benefits. While the chemistry of these compounds has been extensively studied, the genetic analysis has lagged behind. Understanding the genetic basis for traits with potential health implications may be of great use in plant breeding. By characterizing genetic variation related to fruit chemistry, breeders can make more efficient use of plant diversity to develop new cultivars with higher concentrations of these potentially beneficial compounds. Using a large interspecific F1 population, developed from a cross between the temperate *V. corymbosum* var. *caesariense* and the subtropical *V. darrowii*, with 1025 individuals genotyped using genotype-by-sequencing methods, of which 289 were phenotyped for phenolic acid content, with data collected across 2019 and 2020, we have identified loci associated with phenolic acid content. Loci for the compounds clustered on the proximal arm of Vc02, suggesting that a single gene or several closely associated genes are responsible for the biosynthesis of all four tested compounds. Within this region are multiple gene models similar to hydroxycinnamoyl CoA shikimate/quinic acid hydroxycinnamoyltransferase (HCT) and UDP glucose:cinnamate glucosyl transferase (UGCT), genes known to be involved in the CGA biosynthesis pathway. Additional loci on Vc07 and Vc12 were associated with caffeoylarbutin content, suggesting a more complicated biosynthesis of that compound.

Keywords: blueberry; trait mapping; chlorogenic acids; candidate genes; fruit quality; nutrition



Citation: Herniter, I.A.; Kim, Y.; Wang, Y.; Havill, J.S.; Johnson-Cicalese, J.; Muehlbauer, G.J.; Iorizzo, M.; Vorsa, N. Trait Mapping of Phenolic Acids in an Interspecific (*Vaccinium corymbosum* var. *caesariense* × *V. darrowii*) Diploid Blueberry Population. *Plants* **2023**, *12*, 1346. <https://doi.org/10.3390/plants12061346>

Academic Editors: Guiliang Tang, Shaomin Bian and Ting Lan

Received: 9 January 2023

Revised: 6 March 2023

Accepted: 6 March 2023

Published: 16 March 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The species commonly referred to as blueberry (*Vaccinium* sect. *Cyanococcus*) are perennial flowering plants native to North America. In the last thirty years, blueberry production has increased by over 600%, to 858,886 tonnes of blueberry harvested worldwide in 2020, with over 70% of that production occurring in the United States of America, Peru, and Canada [1]. This increase in production has occurred in the wake of the publication of a number of studies, beginning in the mid-1990s, examining the potential health benefits associated with blueberry consumption, especially in regard to antioxidant activity, for which blueberry has among the highest levels compared to other fruits and vegetables [2–4]. In addition to the use of the berries as fresh or processed foodstuff, the plants have also been used in traditional medicines [5]. Even today in the realm of natural medicine, blueberries are recommended for use in treating Type II diabetes [6].

Compounds with antioxidant activity, such as phenolic acids, anthocyanins, and anthocyanidins, have an important function in the plant. Plants regularly experience environmental stresses resulting in the production of reactive oxygen species (ROS), excessive levels of which disrupt normal metabolic function by damaging lipids, proteins, and nucleic acids, all of which negatively impact plant growth and development [7–9]. Compounds with antioxidant activity scavenge ROS, protecting normal metabolic function [10]. While the evidence for effect of antioxidant compounds obtained through food on human health is slim, the possibility of efficaciousness has encouraged substantial research on the topic.

One class of naturally occurring compounds known to have significant antioxidant capacity are phenolic acids, compounds containing a phenol moiety highly suited for trapping free radicals [10]. Chlorogenic acids are a family of polyphenol esters formed between *trans*-cinnamic acids and quinic acids [11]. Chlorogenic acids are one of the most well-studied families of polyphenols, due to their abundance in plant-based food and drinks [12]. They are widespread in plants and can be found in nearly all plant species [13–16].

One important subgroup of chlorogenic acids is the caffeoylquinic acids, which consist of esterifications of caffeic acid (Figure 1A), of which there are several isomeric forms. The most abundant of these isomers present in plants is 5-*O*-caffeoylquinic acid (5-CQA) [12], and it is the most widely studied due to its commercial availability [14]. Generally, the term “chlorogenic acid” (CGA) refers to 5-CQA [12]. It should be noted, however, that the nomenclature of the isomers 5-CQA (Figure 1B) and 3-CQA (3-*O*-caffeoylquinic acid) (Figure 1C) can create some confusion. In 1976, the International Union of Pure and Applied Chemistry (IUPAC) reversed the order of the numbering of atoms on the quinic acid ring [17]. Consequently, the previously identified 3-CQA was renamed as 5-CQA [17], and the current 3-CQA refers to neochlorogenic acid in accordance with the new numbering system [12,16]. However, due to the name change, there is still considerable confusion in the literature, and many papers do not specify, when they use the term “chlorogenic acid”, whether they mean to refer to 5-CQA or 3-CQA [16]. This paper follows Clifford et al. [16] and uses the current IUPAC numbering, with CGA as 5-CQA.

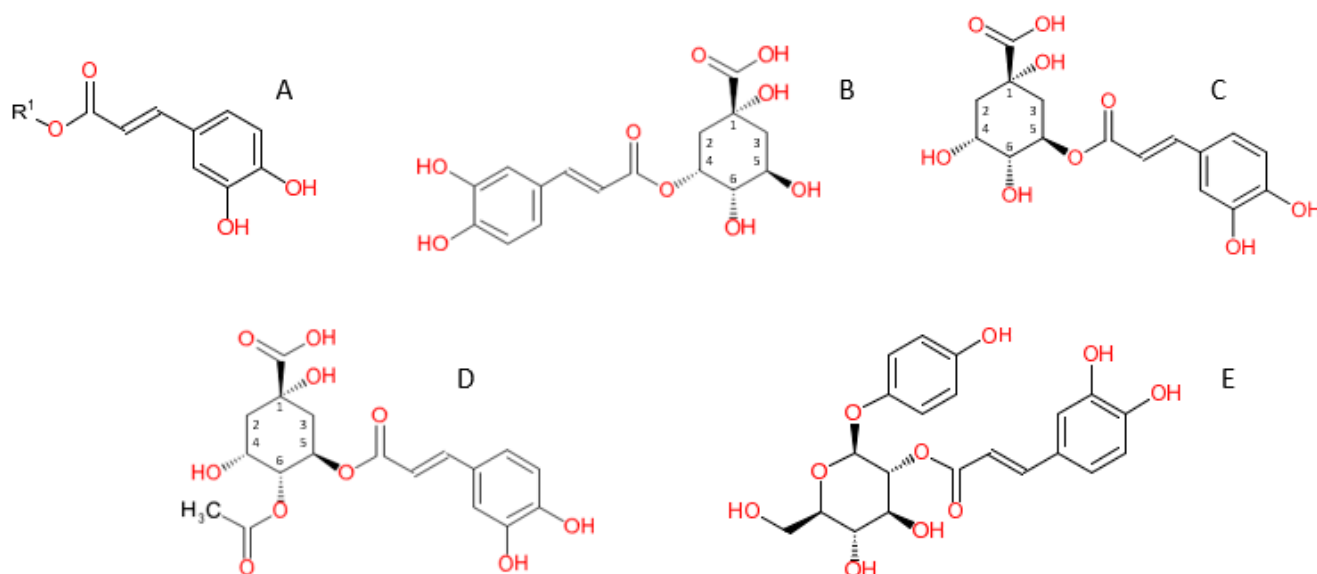


Figure 1. Chemical structures of esterification products of caffeic acid. (A) caffeic acid $R = H$, (B) 3-*O*-caffeoylquinic acid (neochlorogenic acid), (C) 5-*O*-caffeoylquinic acid (chlorogenic acid), (D) acetyl-caffeoylquinic acid (tentative), (E) 2-*O*-caffeoylarbutin.

Chlorogenic acids are the major hydroxycinnamic acids present in blueberries [18], likely accounting for a large proportion of their antioxidant activity [19,20], with CGA being the major component [21]. CGA constitutes 10–16% of total acids in the blueberry

fruit [22–24]. It is present in concentrations of 98–208 mg per 100 g FW (fresh weight) in *V. corymbosum* cultivars [21,24,25]. Other phenolic acids such as caffeic acid, p-coumaric acid, and ferulic acid are present in concentrations under 1% [21,22].

CGA is one of the most abundant beneficial polyphenols in the human diet and is well known as a nutritional antioxidant in plant-based foods [26,27]. Dietary consumption of CGA is associated with the prevention of certain oxidative and degenerative, age-related diseases [28–30]. Compelling evidence indicates that dietary CGA can promote a wide range of pharmacological effects and biological activities in various tissues and organs [12]. Numerous studies have demonstrated the antioxidant activities of CGA, which include inhibiting the formation or scavenging of ROS [31]. CGA is also negatively correlated with the risk of various harmful conditions, such as oxidative and inflammatory stresses [32], type 2 diabetes mellitus [33,34], cardiovascular disease [35], neurodegenerative disease [36], and cancer [37].

Compounds closely related to CGA are the acetylated caffeoylquinic acids (ACQA, $C_{18}H_{20}O_{10}$), which have previously been identified in blueberry [38]. These compounds have not been well characterized and the configurations of the compounds are not known, but the acetylation is likely to be on the quinic acid moiety [38]. Figure 1D shows a potential chemical structure for ACQA proposed by Jaiswal et al. [38]. Note that the regiochemistry of the acetyl group in Figure 1D is an arbitrarily selected example; the compound identified in the present analysis may not be the 4-acetyl caffeoylquinic acid isomer.

While CGA has been studied for its potential for improving human health, similar efforts have yet to be made to understand the potential health impacts of consumption of ACQA1 and ACQA2. Further research is required to understand the potential benefits of the compounds, as well as their bioavailability when consumed in food or drink.

Another related compound to CGA is caffeoylarbutin (CA). As with CGA, CA is an ester of caffeic acid, though with an arbutin group instead of quinic acid (Figure 1E). CA has been previously identified in the leaves of blueberry [39], as well as in the leaves of other *Vaccinium* species, such as lingonberry (*V. vitis-idaea*) [40,41], *V. dunalianum* [42,43], and bilberry (*V. myrtillus*) [39,41].

Other minor and specialty crops have seen great advances in the availability of genetic resources over the past decade. However, blueberry has lagged behind, with only a few published genome sequences [44–46] and a limited number of genotyped mapping populations [46–49], all of which were constructed through crosses between tetraploid cultivars of *V. corymbosum*. The development of mapping populations in blueberry is complicated by long generation time and partial self-sterility but is of exceeding importance to crop improvement efforts, as the identification of genetic markers and trait loci can increase the speed at which new varieties can be developed, and aid in understanding the mechanisms of compound biosynthesis.

The blueberry (*Cyanococcus*) section of *Vaccinium* is highly diverse, including species of differing ploidy levels and adapted to different environments. The highly interfertile nature of the section offers great opportunities for geneticists and plant breeders alike to identify potentially valuable variations which could be used to develop new blueberry cultivars with improved nutritional value and increased resilience in the face of climate change. The most commercially important blueberries are tetraploid ($4n = 48$) highbush plants (*V. corymbosum*). Accordingly, most of the available evidence on the beneficial health effects of blueberry deals with compounds identified in tetraploids. However, many wild species exist throughout North America [50], naturally occurring as diploids, tetraploids, and hexaploids [20]. Among these are numerous diploid blueberry species which offer diverse germplasm [51], though most have limited analyses of their genetics. Two species likely to be of use in this effort are *V. darrowii* and *V. corymbosum*. *V. darrowii* is a subtropical lowbush diploid blueberry native to the American southeast which has been used as a source of variation in breeding programs [52,53], while *V. corymbosum* is a temperate highbush blueberry native to the mid-Atlantic region of the United States. The varying climates to which these two species are adapted, as well as genetic drift over evolutionary

time, have resulted in wide differentiation on many traits including its fruit chemistry which is quite distinct from that of the highbush species [32,36]. This differentiation is exploited in this study by using hybrid populations of these two species for trait mapping. By identifying quantitative trait loci (QTL) in a population containing *V. darrowii* ancestry, greater insight can be obtained into the genetic architecture of blueberry.

A cost-effective method to produce genetic information for newly developed populations is genotyping-by-sequencing (GBS) as it simultaneously performs single nucleotide polymorphism (SNP) discovery and genotyping, eliminating a step required by other methods [54]. Using the newly available genome sequences, we can create highly accurate maps based on the physical ordering of the observed SNP markers. To the authors' knowledge, phenolic acid content in blueberry has only been genetically mapped once previously, by Mengist et al. [48], who identified a QTL on Vc02 for CGA in an F₁ tetraploid *V. corymbosum* population. While CA has been identified in a number of different species, as yet no QTL for CA has been identified. The ACQA compounds have not been fully characterized, and as yet have not been identified outside of blueberry, let alone been mapped.

In this paper, we present the first genotyped diploid interspecific mapping population derived from crosses between *V. darrowii* and *V. corymbosum* var. *caesariense* (a diploid variety of *V. corymbosum*), two divergent species. This large population, recently developed at the Marucci Blueberry and Cranberry Research Center in Chatsworth, NJ, segregates for many traits of interest to breeders, including fruit chemistry. We present here mapping of the genetic control of phenolic acid content in blueberry fruit.

2. Results

2.1. Qualitative and Quantitative Analysis of Phenolics

Figure 2 illustrates a selected HPLC chromatogram of blueberry phenolic compounds. Table 1 summarizes the retention times and MS spectra of the studied blueberry phenolic compounds. The exact isomer of caffeoylquinic acid could not be determined using the mass data. However, based on the literature, the peak identified was assumed to be CGA (5-CQA). Similarly, two of the peaks were identified as isomers of ACQA. To clarify which one is being referred to, these compounds have been labeled as ACQA1 and ACQA2 based on the order in which they were eluted.

Table 1. Phenolic acids identified in blueberries. ESI-MS-MS: electrospray ionization-tandem mass spectrometry.

Peak Number	Retention Time (min)	[M-H]	Main Fragment Ions in ESI-MS-MS (<i>m/z</i>)	Formula	ppm	Tentative Peak Identity
1	6.657	353.0873	191.0556 (100)	C ₁₆ H ₁₈ O ₉	−1.3	Chlorogenic acid (5- <i>O</i> -caffeoylquinic acid, 5-CQA, CGA)
2	10.010	395.098	191.0556 (100), 233.0662 (75)	C ₁₈ H ₂₀ O ₁₀	−1.0	Acetyl-caffeoylquinic acid isomer 1 (ACQA1)
3	11.437	395.0979	233.0662 (100)	C ₁₈ H ₂₀ O ₁₀	−1.2	Acetyl-caffeoylquinic acid isomer 2 (ACQA2)
4	15.005	433.1139	161.0236 (100), 323.0763 (9)	C ₂₁ H ₂₂ O ₁₀	−0.3	Caffeoylarbutin (CA)

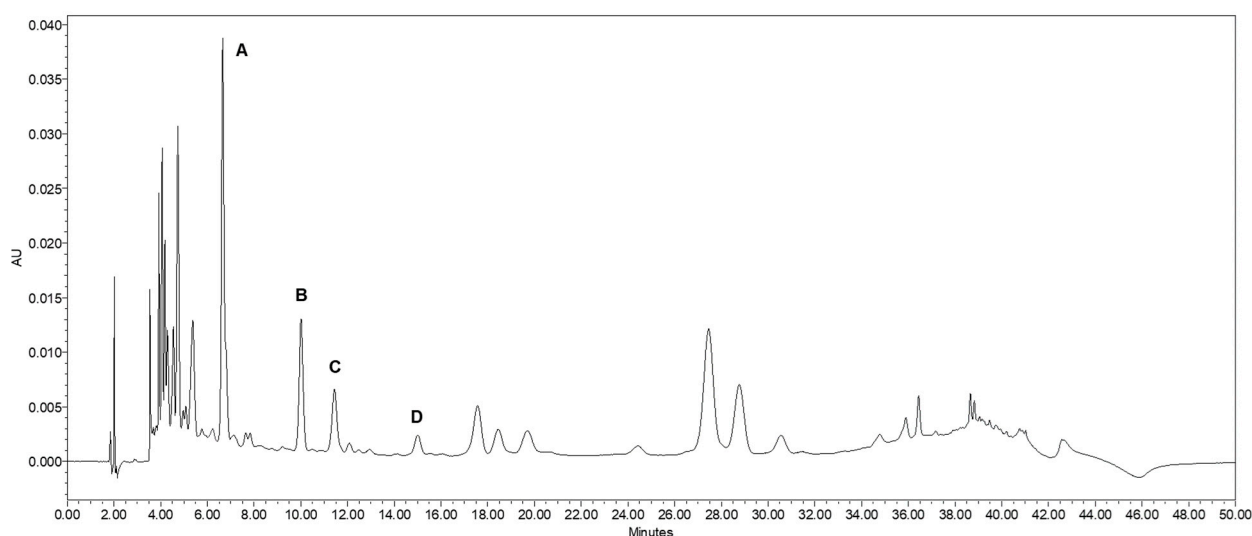


Figure 2. Representative PDA chromatogram of blueberry accession BNJ05-218-9 acquired at 366 nm. The characteristic peaks are labelled 1 through 4. (A) chlorogenic acid, (B) acetyl-caffeoylquinic isomer 1, (C) acetyl-caffeoylquinic isomer 2, (D) caffeoylarbutin.

Table 2 summarizes the mean and standard deviation values of all phenolic compounds among the genotypes. There were significant observed differences between average concentrations for all phenolic compounds in fruit among two genotypes of *V. corymbosum* var. *caesariense* (OPB-8 and OPB-15), two genotypes of *V. darrowii* (NJ88-12-41 and NJ88-14-03), and two interspecific hybrids (BNJ05-218-9 and BNJ05-237-8) (Figure 3). ACQA1 and ACQA2 concentrations were higher in the *V. corymbosum* var. *caesariense* blueberries, while CGA and CA concentrations were higher in *V. darrowii* blueberries, with intermediate levels in the hybrid blueberries (Figure S1). In the BNJ16-4 population, the mean CGA concentration was 0.25 ± 0.15 mg/g FW in 2019 and 0.23 ± 0.17 mg/g FW in 2020. The mean ACQA1 concentration was 0.14 ± 0.07 mg/g FW in 2019 and 0.13 ± 0.07 mg/g FW in 2020. The mean ACQA2 concentration was 0.11 ± 0.04 mg/g FW in 2019 and 0.09 ± 0.04 mg/g FW in 2020. The mean CA concentration was 0.05 ± 0.02 mg/g FW in 2019 and 0.05 ± 0.02 mg/g FW in 2020. Results for parents can also be found in Table 2. Graphs showing the range of concentrations found in the population and the parents can be found in Figure 3.

2.2. Trait Variation and Correlation

In the parents and grandparent plants there was no observed harvest year effect. The Kruskal–Wallis rank sum test did not identify significant differences between 2019 and 2020 in compound concentrations in the parents and grandparents considered together (Table 3). Despite this OPB-15 did show great variation, with observed phenolic compound concentrations reduced by about half. In the BNJ16-4 population, harvest year effects were observed for all the tested compounds except ACQA1.

Table 2. Mean, median, standard deviation, and range of phenolic acid concentrations (mg/g fresh weight) in blueberry fruit harvested in 2019 and 2020.

2019		<i>V. corymbosum</i> var. <i>caesariense</i>		Hybrids.		<i>V. darrowii</i>		F1
	Genotype	OPB-8	OPB-15	BNJ05-237-8	BNJ05-218-9	NJ88-12-41	NJ88-14-03	BNJ16-4
CGA	Range	0.04–0.21	0.23–0.32	0.21–0.28	0.14–0.42	0.47–1.76	0.18–1.27	0.06–0.90
	Mean	0.09	0.27	0.24	0.24	1.03	0.72	0.25
	Median	0.07	0.27	0.24	0.21	1.01	0.57	0.22
	SD	0.05	0.04	0.03	0.1	0.29	0.36	0.15
ACQA1	Range	0.04–0.23	0.24–0.33	0.18–0.23	0.13–0.28	0.03–0.15	0.02–0.12	0.02–0.40
	Mean	0.1	0.29	0.2	0.19	0.07	0.07	0.14
	Median	0.08	0.28	0.19	0.15	0.05	0.04	0.13
	SD	0.06	0.04	0.03	0.07	0.03	0.03	0.07
ACQA2	Range	0.04–0.18	0.17–0.23	0.11–0.18	0.09–0.21	0.03–0.11	0.02–0.13	0.04–0.25
	Mean	0.08	0.2	0.15	0.14	0.06	0.07	0.11
	Median	0.06	0.18	0.15	0.12	0.05	0.06	0.11
	SD	0.04	0.03	0.03	0.05	0.03	0.03	0.05
CA	Range	0–0.10	0.05–0.11	0.10–0.17	0.05–0.13	0.17–0.60	0.03–0.22	0–0.15
	Mean	0.03	0.09	0.14	0.09	0.39	0.15	0.05
	Median	0.03	0.1	0.14	0.1	0.39	0.15	0.04
	SD	0.03	0.02	0.03	0.03	0.11	0.06	0.02
2020	Genotype	OPB-8	OPB-15	BNJ05-237-8	BNJ05-218-9	NJ88-12-41	NJ88-14-03	BNJ16-4
CGA	Range	0.06–0.16	0.06–0.24	0.13–0.41	0.08–0.20	0.91–1.60	0.51–1.49	0.05–1.26
	Mean	0.1	0.13	0.26	0.15	1.22	1.08	0.23
	Median	0.1	0.11	0.29	0.17	1.19	1.13	0.18
	SD	0.04	0.06	0.08	0.04	0.23	0.28	0.17
ACQA1	Range	0.08–0.19	0.08–0.30	0.15–0.29	0.11–0.18	0.04–0.05	0.04–0.05	0.04–0.37
	Mean	0.12	0.17	0.21	0.14	0.05	0.04	0.13
	Median	0.11	0.17	0.23	0.13	0.05	0.04	0.12
	SD	0.04	0.08	0.04	0.03	0	0	0.07
ACQA2	Range	0.05–0.11	0.06–0.18	0.09–0.17	0.07–0.13	0.04–0.05	0.04–0.05	0.04–0.23
	Mean	0.07	0.11	0.12	0.09	0.05	0.04	0.09
	Median	0.07	0.1	0.12	0.09	0.05	0.04	0.08
	SD	0.02	0.04	0.03	0.02	0	0	0.04
CA	Range	0–0.04	0–0.04	0.07–0.11	0.05–0.06	0.21–0.42	0.11–0.18	0–0.17
	Mean	0.01	0.02	0.09	0.06	0.33	0.15	0.05
	Median	0	0	0.09	0.06	0.34	0.15	0.04
	SD	0.02	0.02	0.01	0	0.07	0.02	0.02

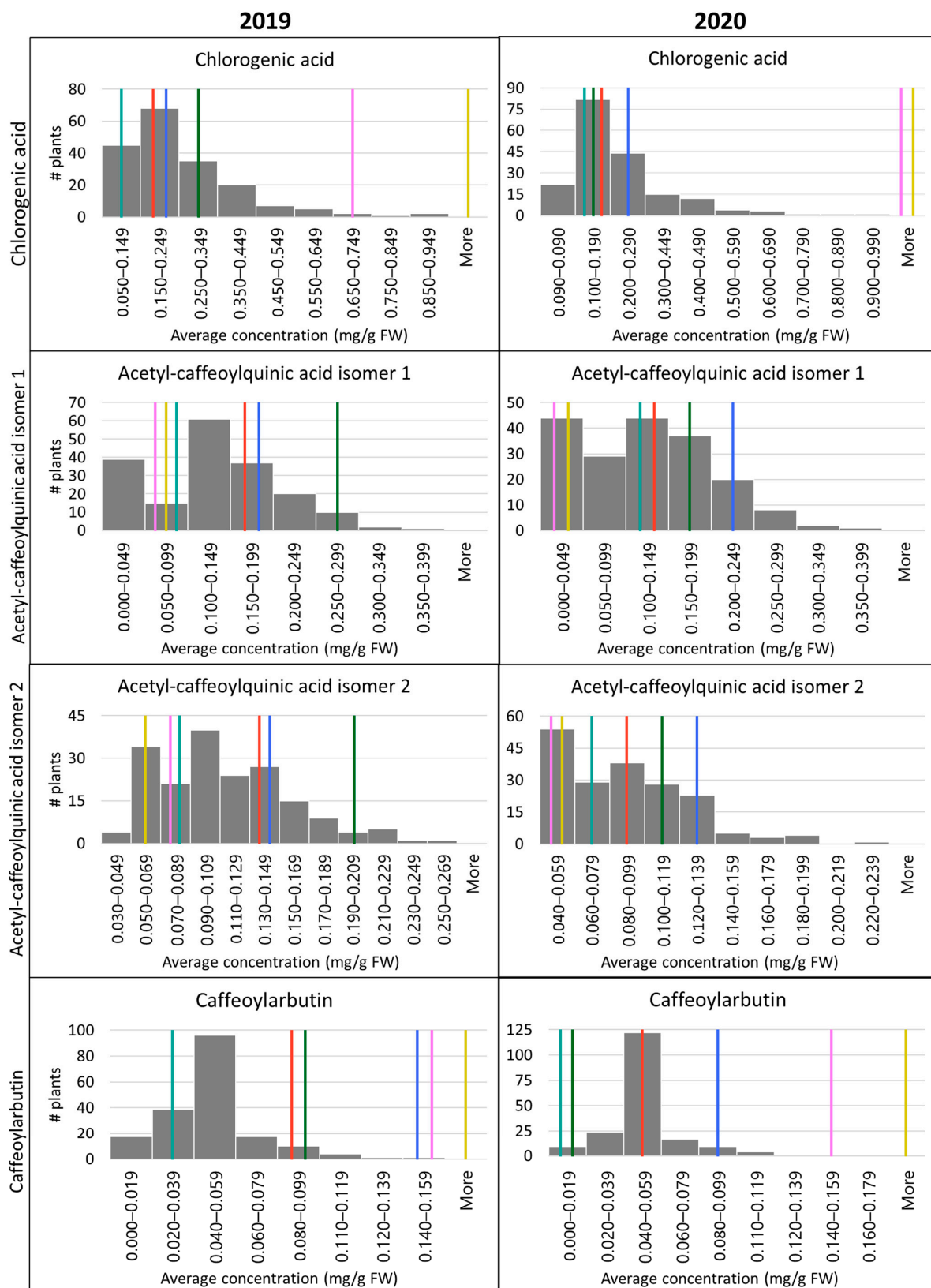


Figure 3. Range and frequency of concentrations of tested compounds in the BNJ16-4 population (mg/g fresh weight). Values from the grandparents and F₁s are noted with vertical lines in the appropriate bins: OPB-8 (teal), OPB-15 (green), NJ88-14-03 (pink), NJ88-12-41 (yellow), BNJ05-218-9 (red) and BNJ05-237-8 (blue).

Table 3. Kruskal–Wallis rank sum test with phenolic acid concentrations (mg/g fresh weight) in blueberry fruit by harvest year (alpha = 0.05) in the parents, grandparents, and BNJ16-4 population and genotype-by-environment (G×E) variation in the BNJ16-4 population.

Parents and Grandparents				BNJ16-4 F ₁ s			
Compound	χ ²	df	p-Value	χ ²	df	p-Value	G×E Variation
CGA	0.00	1	1	21.09	1	4.38 × 10 ^{−6}	3.66 × 10 ^{−3}
ACQA1	0.41	1	0.52	2.20	1	0.14	4.71 × 10 ^{−4}
ACQA2	1.64	1	0.2	50.95	1	9.48 × 10 ^{−13}	1.28 × 10 ^{−4}
CA	0.92	1	0.34	8.82	1	0.00298	8.17 × 10 ^{−5}

In the parents and grandparents, each tested compound showed statistically significant correlations between itself and the other tested compounds (Table 4). ACQA1 and ACQA2 showed a strong correlation in both 2019 (0.97, $p < 0.001$) and 2020 (0.98, $p < 0.001$). CGA and CA also showed a strong correlation in 2019 (0.87, $p < 0.001$) and 2020 (0.85, $p < 0.001$). In addition, CGA and CA both showed moderate negative correlation with ACQA1 and ACQA2. In the BNJ16-4 population, ACQA1 and ACQA2 showed a strong correlation both in 2019 (0.89, $p < 0.001$) and 2020 (0.98, $p < 0.001$). CGA and CA showed a weak correlation in 2019 (0.26, $p < 0.001$) and a moderate correlation in 2020 (0.37, $p < 0.001$).

Table 4. Correlational data for the Parents and Grandparents and the BNJ16-4 population between tested compounds. Italicized numbers indicate the p -values of the observed correlations. p -value < 0.05 was considered significant. Correlations were ranked as weak ($|<0.3|$), moderate ($|0.3–0.7|$), or strong ($|>0.7|$).

Parents									
2019	CGA	ACQA1	ACQA2	CA	2020	CGA	ACQA1	ACQA2	CA
CGA					CGA				
ACQA1	−0.36				ACQA1	−0.62			
	(<i><0.001</i>)					(<i><0.001</i>)			
ACQA2	−0.30	0.97			ACQA2	−0.55	0.98		
	(<i>0.003</i>)	(<i><0.001</i>)				(<i><0.001</i>)	(<i><0.001</i>)		
CA	0.87	−0.35	−0.34		CA	0.85	−0.48	−0.42	
	(<i><0.001</i>)	(<i><0.001</i>)	−0.001			(<i><0.001</i>)	(<i><0.001</i>)	(<i><0.001</i>)	
BNJ16-4									
2019	CGA	ACQA1	ACQA2	CA	2020	CGA	ACQA1	ACQA2	CA
CGA					CGA				
ACQA1	−0.11				ACQA1	−0.02			
	0.027					0.672			
ACQA2	−0.09	0.89			ACQA2	0.03	0.98		
	0.064	(<i><0.001</i>)				0.491	(<i><0.001</i>)		
CA	0.26	0.12	0.07		CA	0.37	0.03	0.03	
	(<i><0.001</i>)	0.014	0.12			(<i><0.001</i>)	0.61	0.595	

2.3. Map Information

The output VCF from the “population” function on STACKS produced a VCF with 47,594 markers, of which 47,199 were located on chromosome assemblies. After filtering to remove monomorphic markers in the parents (BNJ05-237-8 and BNJ05-218-9), or where one

parent had a missing call, or where the minor allele frequency (MAF) > 0.05 , 34,445 markers remained. Following processing using a custom R script to turn the data into the JoinMap format, 24,134 markers remained. Within JoinMap, the non-segregating markers were removed, leaving 17,633 markers. We further separated markers physical chromosomes, where markers on each chromosome were filtered to remove markers with ≥ 0.95 similarity and significantly distorted markers ($< 1 \times 10^{-4}$), leaving a total of 9952 markers for formation of linkage groups. The physical marker order was used to set the marker order and the Kosambi algorithm was used for map construction. During map construction, 7862 markers were removed for lack of information provided or were otherwise not mapped, leaving 2090 markers in the final map.

Of the 2090 markers in the final map, 771 markers were heterozygous in both parents, while 640 markers were heterozygous only in BNJ05-237-8, and 679 markers were heterozygous in only BNJ05-218-9. The total length of the map was 1591.6 cM, with the chromosomes ranging in length from 113.9 cM (Vc07) to 152.6 cM (Vc02), with an average length of 132.6 cM. Markers were generally well distributed across the genome, with an average of 0.8 cM between markers on both male and female parent-derived markers, with the exceptions of a 14.7 cM gap on Vc10, and gaps greater than 8 cM on Vc01, Vc06, Vc08, Vc11, and Vc12 (Table S1). The distribution of markers and marker types across the genome is shown in Figure 4.

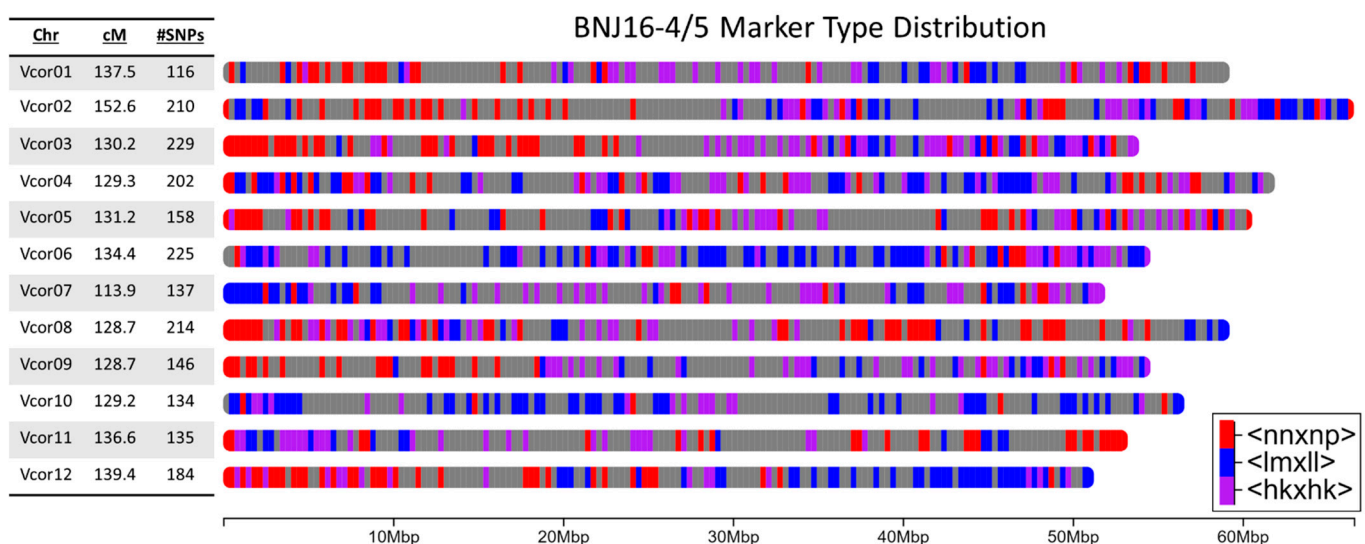


Figure 4. Distribution of markers in the genetic map across the physical chromosomes. <nnxnp> markers are those which are homozygous in BNJ05-237-8 and heterozygous in BNJ05-218-9, <lmxll> are the opposite, and <hkxhk> are heterozygous in both parents.

Graphs showing the relationship between physical and genetic maps of each chromosome can be found in Figure S2. Collinearity between the physical and genetic maps was very strong, with the Spearman' correlation rho value being > 0.999 for each chromosome. Polynomial formulae describing the instantaneous recombination rate across each chromosome were calculated and can be found in Table S2. Most chromosomes displayed the characteristic "S" shape, with the ends of the chromosome demonstrating higher recombination rates than the centromeric regions. In most of the chromosomes, regions of lower recombination were located in or near the physical center of the chromosome, with the exception of Vc06, in which the region of reduced recombination was shifted towards the proximal end of the chromosome.

2.4. Trait Mapping and Heritability

Following phenotypic characterization and calculation of BLUPs, QTL were mapped using BLUPs and individual year data in R using the package *qtl* [55]. Peaks were identified for each tested compound (Table 5, Figures 5 and S3). The identified peaks for CGA, ACQA1, and ACQA2 BLUPs clustered together on Vc02. The peaks for ACQA1 and ACQA2 BLUPs were identical, with a peak at 8.8 Mb with LOD scores of 26.8 and 23.6, and explaining 48.7% and 44.4% of the observed variation, respectively. The peak for CGA BLUPs was identified on Vc02 at 7.8 Mb with a LOD score of 17.7 and explaining 35.6% of the observed variation. Peaks for CA BLUPs were identified on Vc07 and Vc12, with LOD scores of 5.6 and 7.0, and explaining 12.9% and 16.0% of the observed variation, respectively. Mapping with phenotype data from individual years generally identified similar regions to the BLUPs, though sometimes with the peaks shifted up or downstream. One notable difference was that the 2020 CA data were associated with a peak on Vc02 with a LOD score of 5.15 explaining 9.7% of observed variation. The identified region on Vc02 overlaps with the region identified for CGA by Mengist et al. [48]. Complete marker-trait association LOD score data can be found in Table S3. Broad sense heritability, calculated using the BLUPs, for the tested traits was strong: 0.894 for CGA, 0.945 for ACQA1, 0.948 for ACQA2, and 0.891 for CA.

Table 5. Significant QTL identified. The chromosome on which the QTL is located, the marker interval of significant SNPs (above the LOD cutoff), the range and length of the significant interval in both cM and bp, the peak SNP, its location in cM and bp, the LOD score of the peak SNP, the LOD score cutoff for significance, and the percent observed phenotypic variation explained by the QTL are shown.

Trait	Chr	Start Marker	End Marker	Range (cM)	Range (bp)	Length (cM)	Length (bp)	Peak SNP	Peak (bp)	Peak (cM)	Peak LOD	LOD Cutoff	% Phenotypic Variation
acqa1_2019	2	91154:71:-	126708:471:-	0–51.1	242,415–24,241,116	51.13	23,998,701	105300:86:+	8,829,280	29.8	29.93	4.53	45.1
acqa1_2020	2	91154:71:-	126708:471:-	1–51.1	242,415–24,241,116	51.13	23,998,701	103495:459:-	7,840,716	26.1	24.27	4.60	38.1
acqa1_BLUP	2	91154:71:-	126708:471:-	2–51.1	242,415–24,241,116	51.13	23,998,701	105300:86:+	8,829,280	29.8	26.78	4.38	48.7
acqa2_2020	2	91154:71:-	126708:471:-	3–51.1	242,415–24,241,116	51.13	23,998,701	103495:459:-	7,840,716	26.1	20.10	4.60	32.8
acqa2_BLUP	2	91154:71:-	126708:471:-	4–51.1	242,415–24,241,116	51.13	23,998,701	105300:86:+	8,829,280	29.8	23.57	4.49	44.4
acqa2_2019	2	91210:108:-	126708:471:-	2.4–51.1	252,772–24,241,116	48.77	23,988,344	105300:86:+	8,829,280	29.8	25.50	4.61	40.0
cga_BLUP	2	91210:108:-	149574:48:-	2.4–73.1	252,772–38,430,662	70.75	38,177,890	103495:459:-	7,840,716	26.1	17.65	4.55	35.6
cga_2019	2	92629:11:-	138543:529:+	3.6–57.0	1,060,018–32,042,845	53.32	30,982,827	108869:604:+	11,684,755	35.8	15.90	4.51	27.3
cga_2020	2	94530:8:+	152995:69:-	6.3–75.9	2,175,510–40,705,298	69.61	38,529,788	103495:459:-	7,840,716	26.1	16.00	4.45	27.0
ca_2020	2	108869:604:+	113794:161:+	35.8–40.4	11,684,755–14,961,103	4.55	3,276,348	113794:161:+	14,961,103	40.4	5.15	4.33	9.7
ca_2019	7	562033:52:+	586588:110:+	42.0–58.3	11,162,618–28,613,768	16.34	17,451,150	569836:55:+	16,286,998	50.1	6.42	4.42	12.1
ca_2020	7	569836:55:+	591183:501:-	50.1–62.8	16,286,998–32,079,646	12.70	15,792,648	569836:55:+	16,286,998	50.1	5.48	4.33	10.3
ca_BLUP	7	569836:55:+	588225:292:+	50.1–60.2	16,286,998–29,752,448	10.10	13,465,450	569836:55:+	16,286,998	50.1	5.55	4.48	12.9
ca_2019	12	984802:89:+	1014237:66:+	57.9–98.6	18,054,908–38,178,446	40.62	20,123,538	997344:51:-	26,910,209	75.6	7.93	4.42	14.7
ca_2020	12	990220:105:+	1020637:545:-	68.9–111.1	21,657,929–42,473,585	42.27	20,815,656	997276:67:-	26,881,646	75.3	7.06	4.33	13.0
ca_BLUP	12	990220:632:+	1012122:503:+	69.1–94.3	21,658,456–36,900,682	25.21	15,242,226	997344:51:-	26,910,209	75.6	7.02	4.48	16.0

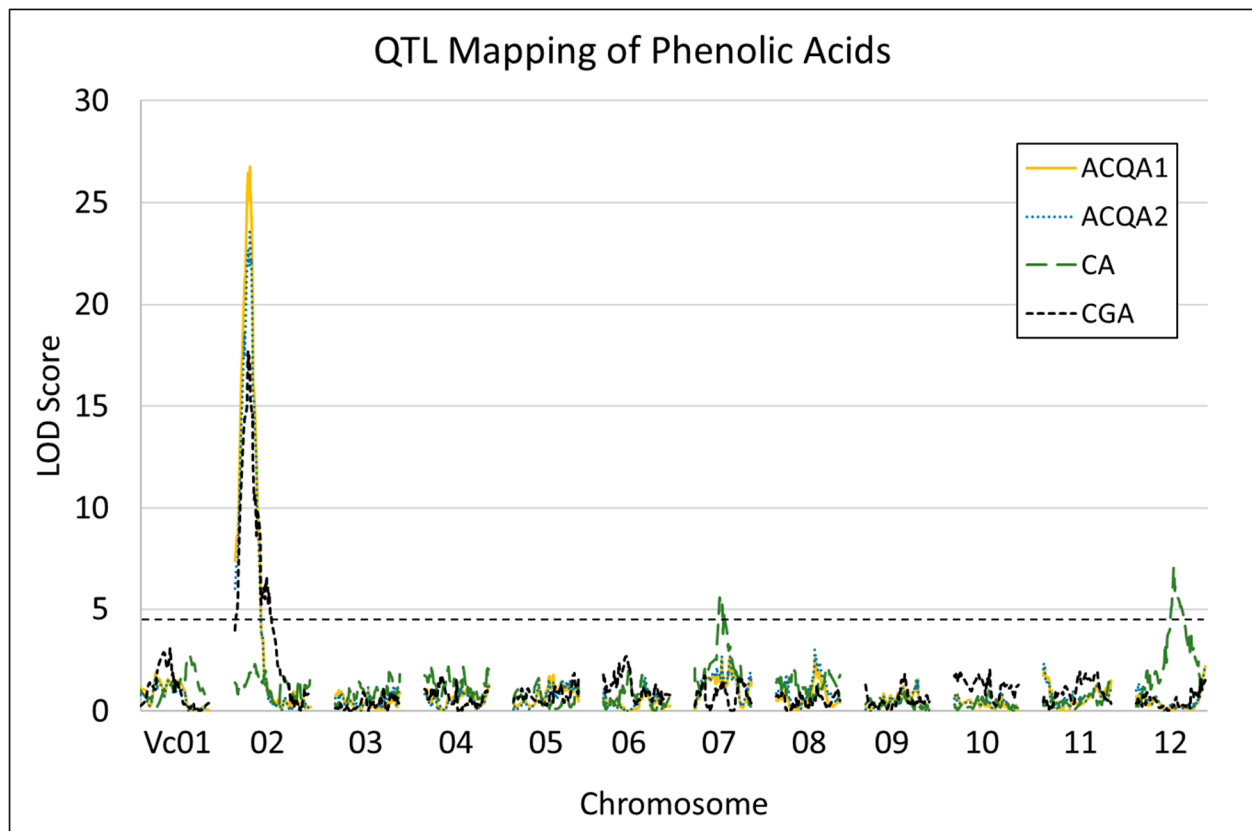


Figure 5. Trait mapping results for BLUPs of the phenolic acids. The horizontal dashed line indicates the significance cutoff (~4.5).

2.5. Candidate Gene Identification

The minimal significant region on Vc02 (CGA, ACQA1, ACQA2), determined by where the mapped BLUP values overlapped with LOD scores above the threshold determined by permutation testing, stretched from 252,772 to 24,241,116 bp (23,988,344 bp). This region contains 1074 gene models. The region identified on Vc07 (CA) stretched from 16,286,998 to 29,752,448 bp (13,465,450 bp) and contained 474 gene models. The region identified on Vc12 (CA) stretched from 21,658,456 to 36,900,682 bp (15,242,226 bp) and contained 704 gene models. Within the significant region on Vc02, nine gene models had annotations matching known CGA biosynthesis pathway genes (Table 6). Seven had hits on hydroxycinnamoyl CoA shikimate/quinic acid hydroxycinnamoyltransferase (HCT), one from tea (*Camellia sinensis*) and six from coffee (*Coffea arabica*), while two had hits on UDP glucose:cinnamate glucosyl transferase (UGCT) from poplar (*Populus* spp.). Additionally, within the significant region on Vc02 were 15 gene models annotated as MYBs, a transcription factor family which has been associated with the phenylpropanoid biosynthesis pathway. The complete list of candidate gene models can be found in Table S4.

Table 6. Candidate genes on Vc02. Nine candidates with annotations related to the CGA biosynthesis pathway were located within the significant region on Vc02. The gene name, annotation, the species from which the annotation was derived, and physical location are shown.

Sequence Name	Description [Species]	Location (bp)
Vcev1_p0.Chr02.03043	UDP-glycosyltransferase 73B4-like [<i>Populus euphratica</i>]	2,761,603–2,764,485
Vcev1_p0.Chr02.03044	UDP-glycosyltransferase 73B4-like [<i>Populus euphratica</i>]	2,771,253–2,774,928
Vcev1_p0.Chr02.03666	hydroxycinnamoyl-CoA:shikimate/quinic acid hydroxycinnamoyltransferase [<i>Camellia sinensis</i>]	16,397,202–16,398,569

Table 6. Cont.

Sequence Name	Description [Species]	Location (bp)
Vcev1_p0.Chr02.03737	hydroxycinnamoyl-CoA shikimate/quinic acid hydroxycinnamoyl transferase [<i>Coffea arabica</i>]	18,658,014–18,660,014
Vcev1_p0.Chr02.03738	hydroxycinnamoyl-CoA shikimate/quinic acid hydroxycinnamoyl transferase [<i>Coffea arabica</i>]	18,666,233–18,668,469
Vcev1_p0.Chr02.03741	hydroxycinnamoyl-CoA shikimate/quinic acid hydroxycinnamoyl transferase [<i>Coffea arabica</i>]	18,749,921–18,753,906
Vcev1_p0.Chr02.03744	hydroxycinnamoyl-CoA shikimate/quinic acid hydroxycinnamoyl transferase [<i>Coffea arabica</i>]	18,912,803–18,914,710
Vcev1_p0.Chr02.03861	hydroxycinnamoyl-CoA shikimate/quinic acid hydroxycinnamoyl transferase [<i>Coffea arabica</i>]	22,777,759–22,779,946
Vcev1_p0.Chr02.03864	hydroxycinnamoyl-CoA shikimate/quinic acid hydroxycinnamoyl transferase [<i>Coffea arabica</i>]	22,898,343–22,901,060

3. Discussion

Phenolic acids are known to be potent antioxidants [10] with implications for human health. Breeding programs can make use of the marker-trait association presented in this study to develop marker-assisted breeding programs, potentially increasing the speed of the selection process and leading to the development of improved varieties with increased antioxidant potential.

The phenylpropanoid biosynthesis pathways have been extensively studied, with the pathways well understood [16,56–58], including both the enzymes involved, such as various transferases and hydroxylases and potential transcriptional regulators, such as those in the R2R3 MYB transcription factor family. Some potential pathways for the biosynthesis of ACQA1, ACQA2, and CA can be found in Figure 6.

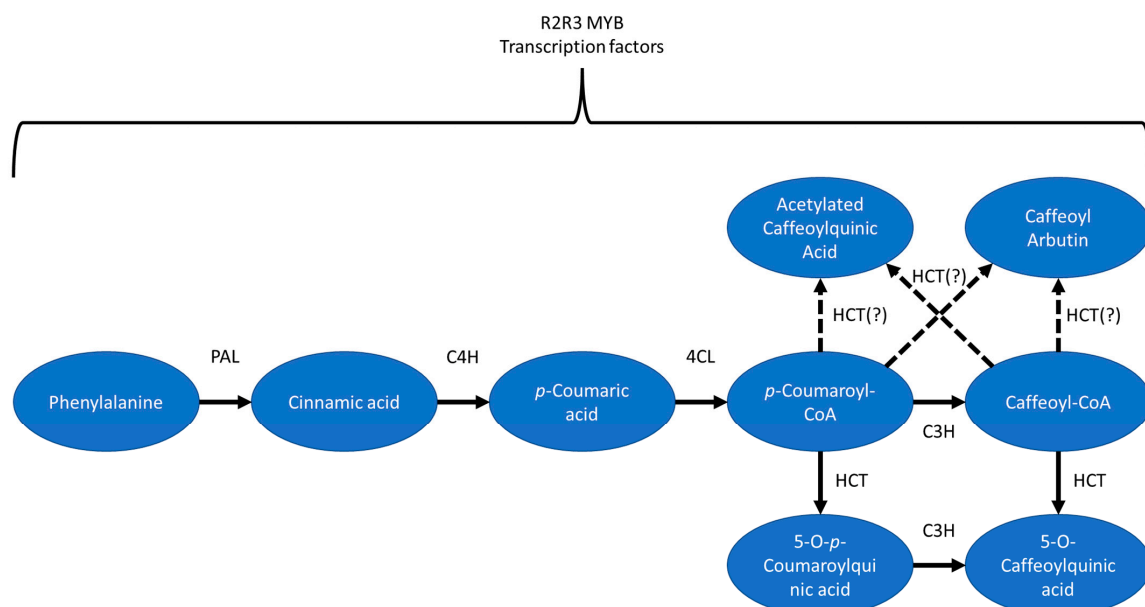


Figure 6. Some of phenolic acid biosynthesis pathways. Solid lines with arrows represent known relationships while dashed lines with arrows represent possible biosynthesis pathways based on the mapping data. Enzymes: phenylalanine ammonia lyase (PAL), cinnamate 4'-hydroxylase (C4H), 4-cinnamoyl CoA ligase (4CL), hydroxycinnamoyl CoA: quinate hydroxycinnamoyl transferase (HCT), and p-coumaroyl-3'-hydroxylase (C3H). Various R2R3 MYB transcription factors are believed to play regulatory roles in the various biosynthesis pathways.

Using the BLUPs, CGA, ACQA1, and ACQA2 had QTL located on the proximal arm of Vc02. Mapping for CA using BLUPs did not identify a peak on Vc02, though mapping with 2020 data did identify a QTL close to the others, and the 2019 data showed elevated correlation between CA and the ACQA compounds but one which did not rise to the level of significance. The overlapping QTL on Vc02 suggest that a single gene or several genes located in that region are involved in the biosynthesis of all four compounds. Notably, seven genes in the significant region on Vc02 had BLAST hits on HCT from tea (*Camellia sinensis*) and coffee (*Coffea arabica*). Previous studies have shown that HCT performs the final step in the main CGA biosynthesis pathway, producing caffeoyl-CoA, the penultimate compound in the path [16,56]. While this is the main pathway, others have been proposed [56]. Of relevance is the alternate pathway wherein cinnamic acid is bonded with glucose by UGCT. Two gene models in the significant region have BLAST hits on UGCT identified from poplar (*Populus* spp.). Unfortunately, while Mengist et al. [46] mapped CGA content and produced transcriptomes in tetraploid blueberry, they did not make a comparison between genotypes with varying levels of CGA and as yet, there is no searchable database with expression data in different tissues (J. L. Humann, personal communication).

R2R3 MYBs have been associated with the phenylpropanoid biosynthesis pathway [58,59]. Within the significant region identified on Vc02 are fifteen gene models annotated as MYB-encoding (Table S4), offering additional avenues of potential exploration. Future studies should focus on the potential activity of such transcription factors in the biosynthesis of CGA, ACQA1, ACQA2, and CA in *Vaccinium*, perhaps identifying markers which can be used for breeding purposes.

Considerably less is known regarding the biosynthesis of other caffeoylquinic acid isomers. Our current understanding is that other caffeoylquinic acid isomers are derived from CGA, but there is little data on the enzymes that would be involved in such conversions [16]. In addition to the QTL identified on Vc02, CA also mapped to regions on Vc07 and Vc12, whereas the other compounds did not. This suggests that while CA biosynthesis is related to biosynthesis of the remaining compounds, it is more complicated. Further research is required to elucidate the biosynthesis pathways of the studied phenolic acids.

Comparing the results of trait mapping presented here with the previous findings of Mengist et al. [46] showcases the mapping power of the interspecific population. While Mengist et al. [46] used a population of 196 individuals and found QTL for CGA with LOD scores of 5.9–6.9, the present study used a comparable number of individuals and identified QTL for CGA with LOD scores of ~16.0. This difference is likely partially explained by the larger difference between the parents in the present study. While the parents in Mengist et al. [46] population, the cultivars Jewel and Draper-44392, differed by ~25mg/100g FW, BNJ05-237-8 and BNJ05-218-9 differed by ~110mg/100g FW. It is unlikely, however, that this would explain the 10-magnitude difference in probabilities. It is more likely that the diploid nature of the BNJ16-4 population allows for greater clarity. Amadeu et al. [60] demonstrates the increased accuracy of trait mapping in diploid as compared to tetraploid populations, suggesting that trait mapping efforts should be focused on diploid populations when possible. The BNJ16-4 population shows segregation for a wide range of traits, beyond those presented here. Future studies should make use of the population to map those traits with a similar degree of confidence.

In addition to its interspecific nature, the population used in this study offers advantages over the tetraploid populations used in other studies. This primarily stems from the diploid nature of the population, which greatly simplifies the genetic analysis. With just two copies of the genome, constructing maps is far simpler, as well as determining inheritance, dosage effects, and performance of QTL mapping [60]. Additionally, due to the relatively common failure of diploid male *Vaccinium* gametes to complete meiosis, ~13.5% of pollen grains have unreduced gametes [61], meaning that pollen from a diploid plant can be directly crossed with a tetraploid plant to bring the desired trait into the breeding population. Further, diploid blueberry can be induced to undergo polyploidization using oryzalin or colchicine [62]. Consequently, the bulk of breeding could potentially occur at

the diploid level and the findings of mapping studies done in diploid plants, such as the present one, could be directly applied to the breeding program.

The nature of the map conformed to expectations from the literature. The centromeric regions, as defined by a region with large physical distances and minimal genetic distances, and hence a slope near zero, was generally found at or near the center of the physical map (Figure S2). This accords with the karyotype data from Hall and Galleta [63], who reported that centromeres in *Vaccinium* were, when observed, median to submedian. The one exception was Vc06, where the centromere was shifted toward the proximal arm of the chromosome.

The present study identifies a strong QTL for the tested phenolic acids, and its overlap with the QTL region identified by Mengist et al. [46] lends credence to our finding. However, the lack of independent populations in which markers could be tested means that any genetic selection, such as in a breeding program, would require using an individual or parent from the BNJ16-4 population as the donor. Further analysis on phenolic acid content in independent populations, such as biparental, multiparental, or germplasm collections, is required to develop widely applicable markers which could be used in marker-assisted selection.

The observed near 1:1 correlation between ACQA1 and ACQA2, in conjunction with both traits showing highly significant associations with the same region on Vc02, suggests the possibility that biosynthesis of both compounds is carried out by a single enzyme, and which of the two compounds is produced is random. The minor positive correlation between CGA and CA observed in the BNJ16-4 population (0.26 in 2019 and 0.37 in 2020, Table 4) and the minor negative correlation between CGA and the ACQA compounds (Table 4) suggest a relationship in the biosynthetic pathways of these compounds.

Phenolic content in OPB-15 varied dramatically between 2019 and 2020, with nearly double the concentration or more of each compound observed in 2019 as in 2020 (Table 2). This pattern was not observed in any of the other parents or grandparents. Previous studies have found that blueberries collected later in the season show increased levels of phenols, anthocyanins, and antioxidants [3,64–66]. Indeed, harvesting began and ended later in 2019 than in 2020. Another possible cause is elevated UV radiation in 2019 compared to 2020, which may have affected the phenolic content, though UV radiation was not measured in the greenhouse in which the plants were grown. Previous studies in carrot [67] and blueberry [68] have indicated that CGA content increases in correlation with increased UV radiation. It should also be noted that OPB-15 had low fruit yields in 2020, perhaps contributing or relating to the observed difference in CGA content. Furthermore, it is worth noting that the BNJ16-4 population also showed a general drop in phenolic compound concentrations in 2020 compared to 2019, with the exception of ACQA1, though not to the same extent as observed in OPB-15 (Table 2). This general decrease was significant, as evident in the results from the Kruskal–Wallis test (Table 3). Environmental variation could be involved in the inconsistent identification of a peak on Vc02 for CA. Further investigation on the possible environmental factors affecting phenolic content in blueberry is required, such as additional years of data or trials in different locations.

Similarly, CA is thought to have potentially beneficial effects for human health. CA levels are high in the leaves of bearberry and *V. dunali*, and a tisane is made from the leaves of both in traditional medicine [39,69]. Arbutin and its derivatives, such as CA, are used as skin-whitening agents, due to their activity in inhibiting melanogenesis by inhibiting tyrosinase [39,70]; indeed, it has been shown to inhibit melanogenesis in zebrafish [71]. Arbutin has also been shown to have anticancer properties, likely due to its high level of antioxidant activity [72,73]. CA appears to be bioavailable, being present in urine following consumption [39]. This paper demonstrates the presence of CA in blueberry and suggests possible breeding avenues to increase CA content, with potential health benefits for consumers.

Each of the tested phenolic compounds contains a caffeoyl moiety (Figure 1A) as a component, suggesting that the significantly associated region encodes a gene responsible

for the esterification of caffeic acid with various R groups. Further investigation could be directed at identifying other esters of caffeic acid in blueberry and mapping them to see if they co-locate.

Cultivated blueberry consists mostly of tetraploid *V. corymbosum*, while the population tested here was developed from crosses between diploids, *V. corymbosum* var. *caesariense* and *V. darrowii* plants. The observed difference in the grandparents, where OPB-8 and OPB-15 (*V. corymbosum* var. *caesariense*) showed elevated ACQA1 and ACQA2 concentrations and NJ88-14-03 and NJ88-12-41 (*V. darrowii*) showed elevated CGA and CA concentrations, could represent different strategies by the species to combat oxidative stress. In addition, CGA is known to be bioavailable in humans [14,74,75] while to the authors' knowledge the bioavailability of ACQA1 and ACQA2 has not been tested. While further studies are required to determine the bioavailability of the ACQA compounds, should they be less bioavailable in humans or show lower antioxidant activity, breeding to increase CGA levels in blueberry cultivars could improve nutritional value of the berries. Wang et al. [51] compared various wild diploid species along with tetraploid cultivars of *V. corymbosum*, showing that the chemical composition of the tetraploids was distinct from that of the tested diploid species for anthocyanin and flavanol glycoside content. This could indicate that introgression from a variety collection, such as the grandparents of the BNJ16-4 population, would be of use in introducing variation to breeding programs.

4. Materials and methods

4.1. Plant Material

Population Development and Maintenance

A large F_1 population derived from crosses between *V. corymbosum* var. *caesariense* and *V. darrowii* was used for this research. Two wild *V. corymbosum* var. *caesariense* plants, OPB-15 and OPB-8, were collected from a native population in Burlington County, NJ, near the Phillip E. Marucci Center for Research and Extension (39.71° N, 74.51° W). Two wild *V. darrowii* plants, NJ88-12-41 and NJ88-14-03, were collected from native populations in Liberty County, Florida (30.24° N, 85.01° W) and along Route 98S in the Saint Joseph Bay area of Florida (29.78° N, 85.28° W), respectively. Crosses between these plants were made in 2005, NJ88-14-03 × OPB-15, and OPB-8 × NJ88-12-41. The resulting hybrid plants were BNJ05-218-9 and BNJ05-237-8, respectively.

Reciprocal crosses of the F_1 plants were made in 2016. Where BNJ05-218-9 was the female parent, plants were designated with the prefix "BNJ16-4". Where BNJ05-237-8 was the female parent, plants were designated with the prefix "BNJ16-5". The resulting F_1 population consists of 1025 full-sib individuals, 949 BNJ16-4's and 76 BNJ16-5's. A scheme representing the pedigree of the population is presented in Figure S4.

Plants were maintained in pots in a greenhouse located at the Philip E. Marucci Blueberry and Cranberry Center for Research and Extension in Chatsworth, New Jersey, USA (39.72° N, 74.51° W). The plants went through the normal cycle of seasonal growth and winter dormancy (greenhouse maintained in a "cold" state, allowing for winter chilling at minimum 0–4 °C). Bumblebees (Koppert Biological Systems, Howell, MI, USA) were brought into the greenhouse in late spring/early summer during flowering for open pollination for fruit set.

4.2. Fruit Collection

A subset of genotypes was selected arbitrarily to be included in the phenotyping as it was not feasible to analyze every genotype. Berry samples were collected from 185 genotypes in both 2019 and 2020, 45 just in 2019, and 48 just in 2020, as not every plant produced fruit in both years of the study. Samples were also collected from the parents and F_1 plants in both 2019 and 2020. Fully ripe, i.e., blue, fruit samples were harvested from each plant at 7–14-day intervals over the fruiting period. In 2019 fruit was harvested from the parents beginning on 9 April through 10 June, and from the BNJ16-4 subpopulation beginning on 4 June through 16 August, with a peak around 24 June. In

2020 fruit was harvested from the parents beginning on 1 May through 24 June, and from the 16-4 population beginning on 7 May through 2 July, with a peak around 4 June. Berry samples were placed in polyethylene bags and kept chilled until weight measurements were taken. After weighing, samples were stored at -80°C until analysis. Information on the number of berries collected from each plant in both 2019 and 2020 can be found in Table S5.

4.3. Genotyping and Map Construction

DNA Extraction and GBS

Young leaf samples from each genotype (BNJ16-4 and BNJ16-5 progeny, their parents, and grandparents) were collected in spring 2020 and DNA was extracted using a modified cetyltrimethylammonium bromide (CTAB) solution following Daverdin et al. (2017). DNA quantification was performed using a Qubit 3 fluorometer and the Qubit dsDNA BR assay kit (Invitrogen, Waltham, MA, USA).

Genotyping-by-sequencing (GBS) libraries were developed using a protocol derived from [76]. In brief, 200 ng of DNA was double digested using the restriction enzymes MspI and PstI-HF (New England Biolabs, Ipswich, MA, USA) at 37°C for 2 h. A common adaptor and a unique barcode adaptor for each accession were ligated to the digested genomic DNA. Following ligation, the solutions were cleaned using 0.7 volumes of Axyprep Mag PCR Clean-Up magnetic beads (Axygen, Union City, CA, USA). The cleaned fragments were then amplified using Taq 5X Master Mix (New England Biolabs) and PCR primers with specific sequences to allow Flowcell binding and Illumina sequencing. After amplification, DNA from each accession was quantified again using the Qubit 3 fluorometer, then diluted to 5 ng/ul. Several pools were made using the diluted DNA from 112 samples with different barcodes. These pools were then then cleaned again using the Axyprep magnetic beads and quantified to ensure a concentration greater than 5 ng/ul. Pools of purified and barcoded genomic DNA were sequenced by Genewiz (South Plainfield, NJ, USA) in a 2×150 bp configuration on an Illumina HiSeq to produce paired end reads.

4.4. Demultiplexing, Sequence Alignment, and Variant Calling

GBS sequencing data were processed using STACKS v2 [77]. The *process_radtags* function was used to filter for quality, demultiplex, and trim the raw reads. The following options were used: *-c* to clean the data, removing any reads with uncalled bases; *-q* to remove low quality reads ($\text{phred} \geq 10$); *-r* to rescue the barcodes and RAD-tags. Demultiplexed reads were then aligned to the reference genome of W85 [48], a wild diploid *V. corymbosum* var. *caesariense* accession collected in Ocean County, NJ, using SAMtools [78]. The obtained BAM files were processed in STACKS using the *ref_map.pl* command for SNP calling. To facilitate trait mapping, the outputs were formatted into VCF files using the *populations* command in STACKS with the following options: “*-min-samples-per-pop 0.90*” so that only SNPs present in at least 90% of the population would be retained; “*-min-maf 0.2*” so that only SNPs with a minor allele frequency $>20\%$ would be retained; “*-ordered-export*” to order the markers by physical location.

4.5. Map Construction

Map construction was performed using JoinMap v5 [79]. The VCF output file from STACKS was converted to JoinMap format in R. Prior to map construction, markers for which both parents were monomorphic or where either parent had missing data were filtered out. We examined progeny genotypes for patterns of coherent segregation and the presence of impossible genotypes, which were converted to missing data when aberrant genotypes were below a 5% threshold, otherwise loci above this threshold were removed from downstream analysis. A final step included the removal of markers with greater than 10% missing data in the progeny. The remaining markers were imported into JoinMap, with markers from each chromosome imported independently. In JoinMap, highly distorted ($p\text{-value} \leq 1 \times 10^{-5}$) and highly similar (≥ 0.95) markers were removed. Maps were

constructed using regression mapping with the Kosambi mapping function and with a fixed marker order based on the physical order within the reference genome. Default settings were used for ordering of markers within a linkage group, which include linkages with a recombination frequency <0.4000 , a LOD of >1.00 , a “goodness-of-fit” threshold of 5.00, followed by rippling after every added locus. The genotype data are available in Table S6, and the map information is available in Table S7.

4.6. Instantaneous Recombination Rate

A polynomial curve fitting the cM position as a function of physical location was generated in R version 4.1.3 (The R Foundation for Statistical Computing) for each of the 12 chromosomes. The linear model *lm* from the base *stats* package was used to determine the linear regression for polynomials up to the 25th degree. Then the output polynomials were examined to find the lowest degree polynomial with an R^2 value greater than 0.998. The recombination rate (first derivative) of this polynomial was calculated along the length of the chromosome using the *deriv* function, also from the base *stats* package.

4.7. Phenolic Acid Extraction and Quantification

Chemicals and Reagents

All solvents, including water, acetonitrile, methanol, and acetone, were purchased from EMD Millipore (Billercia, MA, USA) and were of HPLC grade. Acetic acid was purchased from Avantor Performance Materials (Center Valley, PA, USA), and formic acid was purchased from Mallinckrodt baker (Phillipsburg, NJ, USA). A commercial standard of chlorogenic acid (CGA) was purchased from Sigma-Aldrich, Inc. (St. Louis, MO, USA).

4.8. Extraction of Blueberry Phenolic Compounds

For chlorogenic acid quantification, depending on sample availability, 2–8 g of fruits were weighed, average berry weight (AW) was recorded, and samples were ground with a Precellys Evolution homogenizer (Bertin Corp., Rockville, MD, USA) using 2.8 mm ceramic beads at 7200 rpm for 1.5 min. 80% aqueous acetone with 0.1% acetic acid (1:4 sample to solvent *w/v*) was added to suspend ground fruit, and samples were extracted overnight at 4 °C in a standard refrigerator. Liquid extracts were then centrifuged at 13,300 rpm for 2 min, and 1 mL aliquots of clear supernatant were taken. Aliquots were dried with a SpeedVac vacuum concentrator (Savant SPD2010-220, Thermo Scientific, Waltham, MA, USA) under no heat and re-dissolved in 500 μ L of 100% methanol by sonication for 15 min. Samples were then centrifuged at 11,000 rpm for 5 min, and clear supernatants were analyzed with high-performance liquid chromatography (HPLC).

4.9. HPLC Apparatus and Conditions

Two HPLC systems were used for phenolic acid identification and quantification. The phenolic acids were analyzed in a Waters Alliance LC system composed of a Waters e2695 Separations Module and Waters 2998 PDA Detector (Waters Corp., Milford, MA, USA). A Gemini 150 $\times$ 4.6 mm 5 μ m C18 110 Å LC column (Phenomenex, Torrance, CA, USA) was used for separation, and compounds were detected at 366 nm. The injection volume was 10 μ L.

For compound identification of phenolic acids, the samples were analyzed using the method described in Wang et al. [51] with a Waters ACQUITY® UPLC I-Class system coupled with a Waters Vion Ion Mobility Quadrupole Time of Flight (IMS QToF) mass spectrometer (MS) (Waters Corp., Milford, MA, USA). The same column, solvent system, and elution gradient as described by Wang et al. [51] were used with the system for compound identification. In addition, a 1:3 splitter was used to direct one-fourth of the flow (0.25 mL/min) into the MS. Compounds were identified by liquid chromatography tandem mass spectrometry (LC-MS-MS) based on accurate masses, retention times, and UV absorbance at 305 to 390 nm. All solvent systems and elution gradients are summarized in Table 7.

Table 7. Solvent system and elution gradient for HPLC analysis of phenolic acids in blueberry. Time point indicates time since sample begins running.

Time Point (min)	Flow (mL/min)	0.1% Formic Acid in Water (% Solution)	0.1% Formic Acid in Acetonitrile (% Solution)
0	1	100	0
1	1	86	14
5	1	85	15
14	1	85	15
20	1	84.6	15.4
30	1	83	17
35	1	73	27
38	1	60	40
40	1	20	80
43	1	100	0
50	1	100	0
80	0	100	0

4.10. Compound Identification with MS Spectrometry of Samples

Ion-Mobility High-Resolution Mass Spectrometry data were acquired in high-definition MS^E mode, with the following parameters: ion source, ESI negative ion; analyzer type, sensitivity; source temperature, 100 °C; desolvation temperature, 400 °C; cone gas flow, 50 L/h; desolvation gas flow, 850 L/h; capillary voltage, 2.50 kV; low collision energy, 6.0 eV; high collision energy, 15.0–45.0 eV; mass range, 50–2000 *m/z*; scan rate, 0.25 s. Leucine encephalin (50 pg/mL, 10 µL/min) was used for lock mass correction at 0.25 min intervals. MS and ion mobility data were acquired and processed in UNIFI (Waters Corp., Milford, MA, USA).

4.11. Compound Characterization and Quantification

Phenolic acid characterization was carried out by comparing LC retention times, UV spectra and/or MS/MS data with standards (Table 1). For quantification of phenolic compounds, chromatograms were viewed at absorbance 366 nm and quantified as equivalents of their available standard, chlorogenic acid. The concentration of each compound was expressed in milligrams of its equivalent external standard per gram of fresh weight sample.

4.12. Phenolic Acid Data Analysis

Statistical analyses were performed using R version 4.1.1 (The R Foundation for Statistical Computing) and Microsoft Excel for Microsoft 365 MSO (New York, NY, USA). The *lme4* [80], *lmerTest* [81], and *emmeans* [82] packages were used to fit a linear regression model to the data, determine whether a significant difference exists between genotypes with Satterthwaite's method, and run post-hoc analyses on differences between pairs of genotypes using Tukey multiple comparison tests with Kenward-Roger's degree of freedom method. The *corrplot* [83] package was used to generate a Pearson's correlation matrix among the phenolic concentrations. Correlations were classified as strong if $r > |0.7|$, moderate if $|0.3| > r > |0.7|$, or weak if $r < |0.3|$. Kruskal–Wallis H tests were used to evaluate harvest year and localization effects on individual compound concentrations in the parents and F₁ plants. Excel was used to generate tables using mean and standard deviation values derived from the data. Excel was also used to generate frequency distribution histograms and tables using F₁ data.

Best linear unbiased estimates (BLUPs) were calculated using the *lme4* package in R. Only genotypes with phenotypic data from both 2019 and 2020 were used to calculate

BLUPs, with harvest year considered as a fixed variable. Broad-sense heritability was calculated using the variance components as follows, modified from Mengist et al. [46]:

$$H^2 = \frac{\sigma_g^2}{\sigma_g^2 + \frac{\sigma_{gy}^2}{y} + \frac{\sigma_e^2}{s}} \quad (1)$$

where σ_g^2 , σ_{gy}^2 , and σ_e^2 , are the variance components of genotypes, genotype-by-environment interactions, and environment, respectively; y is the number of years the plants were phenotyped (2 years of the study) and s is the total number of samples in the data set.

4.13. Trait Mapping

Trait mapping was done using the R package *qtl* [55]. Data were imported using the “read.cross” function, treating the population as a four-way cross and incorporating information about marker phase, and the “jittermap” function was used to separate markers at the same cM location. The probability value of each SNP was determined with the function “calc.genoprob (data, step = 0).” Afterward, to map the QTL probabilities, both a standard interval mapping using the EM algorithm: “scanone (data)” and a Haley-Knott regression: “scanone (data, method = “hk”, n.cluster = 2)” were used. Both algorithms showed similar results. To test for significance, 1000 permutations were performed on the Haley-Knott regression: “scanone (data, method = “hk”, n.perm = 1000)”. Only SNPs with LOD scores greater than the significance threshold determined through the permutation test were considered significant. Trait mapping was performed on calculated BLUPs and individual year data. Phenotypic data used in this study are available in Table S8. SNPs with LOD scores above the significance threshold were considered to be significant.

4.14. Candidate Gene Identification

The QTL mapping results for each trait and for each year were compared to identify the region containing overlap between significant regions as determined by the LOD scores. The list of annotated gene models from the W85 (*V. corymbosum* var. *caesariense*) sequence [48] was surveyed to identify lists of genes located in the overlapping significant regions as determined by the permutation test. The gene annotations on this list were then examined for similarity with genes known from the literature to be involved in the CGA biosynthesis pathway.

5. Conclusions

This study analyzed phenolic acid content and identified marker-trait associations in an interspecific population developed from crosses between *V. corymbosum* var. *caesariense* and *V. darrowii*. The QTL identified explained 9.7–48.7% of the observed variation. We identified overlapping peaks on Vc02 for all the tested compounds, with additional peaks for CA on Vc07 and Vc12. This suggests that located within the significant region on Vc02 is a gene, or cluster of genes, which plays a major role in the phenolic acid biosynthesis pathway. This study demonstrates the applicability of interspecific populations to trait mapping. The identified QTL can be used in breeding programs to improve the nutritional value of newly developed cultivars.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/plants12061346/s1>. Table S1: Sex-specific map information; Table S2: Polynomial descriptions of the chromosomes; Table S3: LOD score results from QTL mapping; Table S4: List of candidate genes based on minimal overlapping significant regions; Table S5: Location of phenylpropanoid bio-synthesis pathway genes. Gene models were identified using Pathway Tools on vaccinium.org; Table S6: Berry sample collection and extraction. Table S7: Genotype data of the parents and F1 plants; Table S8: Genetic map of the F1 population; Table S9: Phenotype data; Figure S1: Bar chart of linear mixed effects model estimated average concentrations (mg/g FW) of phenolic compounds for four wild diploid blueberry genotypes, *V. corymbosum* var.

caesariense (OPB-8 and OPB-15, in blue), and *V. darrowii* (NJ88-12-41 and NJ88-14-03, in red) and two hybrids (BNJ05-237-8 and BNJ05-218-9, in purple); Figure S2: Genetic map against physical map, with line of best fit calculated from the lowest degree polynomial with R2 value above 0.998; Figure S3: Trait mapping results for the tested phenolic acids in 2019 and 2020. The dashed line indicates the significance cutoff. Figure S4. BNJ16-4 pedigree. The BNJ16-5 population is the result of a reciprocal cross between the hybrids.

Author Contributions: Conceptualization: N.V., I.A.H. and Y.K.; methodology: N.V., Y.W., Y.K. and I.A.H.; software: I.A.H., Y.K. and J.S.H.; validation: I.A.H.; formal analysis: I.A.H. and Y.K.; investigation: I.A.H. and Y.K.; resources: N.V., J.J.-C. and M.I.; writing—original draft: I.A.H. and Y.K.; writing—review and editing: I.A.H., Y.K., J.S.H., J.J.-C., G.J.M., M.I. and N.V.; visualization: I.A.H. and Y.K.; supervision: G.J.M., N.V. and J.J.-C.; project administration: N.V., J.J.-C. and I.A.H.; funding acquisition: M.I., N.V. and J.J.-C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded through USDA-NIFA-AFRI Grant 2017-67013-26215, and New Jersey Blueberry and Cranberry Research Council. The project was partially supported by the National Institute of Food and Agriculture, United States Department of Agriculture, under award number 2019–51181–30015, project “VacciniumCAP: Leveraging genetic and genomic resources to enable development of blueberry and cranberry cultivars with improved fruit quality attributes”.

Data Availability Statement: The data presented in this study are available in the supplementary material.

Acknowledgments: The authors would like to thank Ajay Singh for guidance on the chemical analysis and Zhenyu Jia for assistance with quantitative analysis.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

CGA	chlorogenic acid
5-CQA	5-O-caffeoylquinic acid
3-CQA	3-O-caffeoylquinic acid
ACQA1	acetyl-caffeoylquinic acid isomer 1
ACQA2	acetyl-caffeoylquinic acid isomer 2
CA	caffeoylarbutin

References

1. FAOSTAT. Crops. 2022. Available online: <http://www.fao.org/faostat/en/#data/QC> (accessed on 17 December 2022).
2. Howell, A.; Kalt, W.; Duy, J.; Forney, C.; McDonald, J. Horticultural Factors Affecting Antioxidant Capacity of Blueberries and other Small Fruit. *Horttechnology* **2001**, *11*, 523–528. [\[CrossRef\]](#)
3. Prior, R.L.; Cao, G.; Martin, A.; Sofic, E.; McEwen, J.; O'Brien, C.; Lischner, N.; Ehlenfeldt, M.; Kalt, W.; Krewer, G.; et al. Antioxidant Capacity as Influenced by Total Phenolic and Anthocyanin Content, Maturity, and Variety of *Vaccinium* Species. *J. Agric. Food Chem.* **1998**, *46*, 2686–2693. [\[CrossRef\]](#)
4. Wang, H.; Cao, G.; Prior, R.L. Total Antioxidant Capacity of Fruits. *J. Agric. Food Chem.* **1996**, *44*, 701–705. [\[CrossRef\]](#)
5. Arnason, T.; Hebda, R.J.; Johns, T. Use of plants for food and medicine by Native Peoples of eastern Canada. *Can. J. Bot.* **1981**, *59*, 2189–2325. [\[CrossRef\]](#)
6. Haddad, P.S.; Depot, M.; Settaf, A.; Chabli, A.; Cherrah, Y. Comparative Study on the Medicinal Plants Most Recommended by Traditional Practitioners in Morocco and Canada. *J. Herbs Spices Med. Plants* **2003**, *10*, 25–45. [\[CrossRef\]](#)
7. Ramakrishna, A.; Ravishankar, G.A. Influence of abiotic stress signals on secondary metabolites in plants. *Plant Signal. Behav.* **2011**, *6*, 1720–1731. [\[CrossRef\]](#) [\[PubMed\]](#)
8. Mei, Y.; Sun, H.; Du, G.; Wang, X.; Lyu, D. Exogenous chlorogenic acid alleviates oxidative stress in apple leaves by enhancing antioxidant capacity. *Sci. Hortic.* **2020**, *274*, 109676. [\[CrossRef\]](#)
9. McCord, J.M. The evolution of free radicals and oxidative stress. *Am. J. Med.* **2000**, *108*, 652–659. [\[CrossRef\]](#) [\[PubMed\]](#)
10. Dumanović, J.; Nepovimova, E.; Natić, M.; Kuća, K.; Jačević, V. The Significance of Reactive Oxygen Species and Antioxidant Defense System in Plants: A Concise Overview. *Front. Plant Sci.* **2021**, *11*, 2106. [\[CrossRef\]](#)
11. Clifford, M.N. Chlorogenic acids and other cinnamates—Nature, occurrence, dietary burden, absorption and metabolism. *J. Sci. Food Agric.* **2000**, *80*, 1033–1043. [\[CrossRef\]](#)

12. Lu, H.; Tian, Z.; Cui, Y.; Liu, Z.; Ma, X. Chlorogenic acid: A comprehensive review of the dietary sources, processing effects, bioavailability, beneficial properties, mechanisms of action, and future directions. *Compr. Rev. Food Sci. Food Saf.* **2020**, *19*, 3130–3158. [\[CrossRef\]](#) [\[PubMed\]](#)
13. Moeenfar, M.; Rocha, L.; Alves, A. Quantification of Caffeoylquinic Acids in Coffee Brews by HPLC-DAD. *J. Anal. Methods Chem.* **2014**, *2014*, 965353. [\[CrossRef\]](#) [\[PubMed\]](#)
14. Naveed, M.; Hejazi, V.; Abbas, M.; Kamboh, A.A.; Khan, G.J.; Shumzaid, M.; Ahmad, F.; Babazadeh, D.; FangFang, X.; Modarresi-Ghazani, F.; et al. Chlorogenic acid (CGA): A pharmacological review and call for further research. *Biomed. Pharmacother.* **2018**, *97*, 67–74. [\[CrossRef\]](#)
15. Magaña, A.A.; Kamimura, N.; Soumyanath, A.; Stevens, J.F.; Maier, C.S. Caffeoylquinic acids: Chemistry, biosynthesis, occurrence, analytical challenges, and bioactivity. *Plant J.* **2021**, *107*, 1299–1319. [\[CrossRef\]](#)
16. Clifford, M.N.; Jaganath, I.B.; Ludwig, I.A.; Crozier, A. Chlorogenic acids and the acyl-quinic acids: Discovery, biosynthesis, bioavailability and bioactivity. *Nat. Prod. Rep.* **2017**, *34*, 1391–1421. [\[CrossRef\]](#)
17. Panico, R.; Powell, W.; Richer, J.-C. *A Guide to IUPAC Nomenclature of Organic Compounds*; Blackwell Scientific Publications: Oxford, UK, 1993; Volume 2.
18. Kader, F.; Rovel, B.; Girardin, M.; Metche, M. Fractionation and identification of the phenolic compounds of Highbush blueberries (*Vaccinium corymbosum*, L.). *Food Chem.* **1996**, *55*, 35–40. [\[CrossRef\]](#)
19. Drewnowski, A.; Gomez-Careros, C. Bitter taste, phytonutrients, and the consumer: A review. *Am. J. Clin. Nutr.* **2000**, *72*, 1424–1435. [\[CrossRef\]](#)
20. Song, G.-Q.; Hancock, J.F. *Vaccinium*. In *Wild Crop Relatives: Genomic and Breeding Resources Temperate Fruits*; Springer: Berlin/Heidelberg, Germany, 2011; pp. 197–221. [\[CrossRef\]](#)
21. Schuster, B.; Herrmann, K. Hydroxybenzoic and hydroxycinnamic acid derivatives in soft fruits. *Phytochemistry* **1985**, *24*, 2761–2764. [\[CrossRef\]](#)
22. Kalt, W.; McDonald, J.E. Chemical Composition of Lowbush Blueberry Cultivars. *J. Am. Soc. Hortic. Sci.* **1996**, *121*, 142–146. [\[CrossRef\]](#)
23. Kang, J.; Thakali, K.M.; Jensen, G.S.; Wu, X. Phenolic Acids of the Two Major Blueberry Species in the US Market and Their Antioxidant and Anti-inflammatory Activities. *Plant Foods Hum. Nutr.* **2015**, *70*, 56–62. [\[CrossRef\]](#)
24. Markakis, P.; Jarczyk, A.; Krishna, S. Foodstuffs Analysis, Nonvolatile Acids of Blueberries. *J. Agric. Food Chem.* **1963**, *11*, 8–11. [\[CrossRef\]](#)
25. Gao, L.; Mazza, G. Quantitation and Distribution of Simple and Acylated Anthocyanins and Other Phenolics in Blueberries. *J. Food Sci.* **1994**, *59*, 1057–1059. [\[CrossRef\]](#)
26. Kundu, A.; Vadassery, J. Chlorogenic acid-mediated chemical defence of plants against insect herbivores. *Plant Biol.* **2019**, *21*, 185–189. [\[CrossRef\]](#) [\[PubMed\]](#)
27. Meng, S.; Cao, J.; Feng, Q.; Peng, J.; Hu, Y. Roles of Chlorogenic Acid on Regulating Glucose and Lipids Metabolism: A Review. *Evidence-Based Complement. Altern. Med.* **2013**, *2013*, 11. [\[CrossRef\]](#)
28. De Sotillo, D.V.R.; Hadley, M. Chlorogenic acid modifies plasma and liver concentrations of: Cholesterol, triacylglycerol, and minerals in (fa/fa) Zucker rats. *J. Nutr. Biochem.* **2002**, *13*, 717–726. [\[CrossRef\]](#)
29. De Sotillo, D.V.R.; Hadley, M.; Sotillo, J.E. Insulin receptor exon 11+/- is expressed in Zucker (fa/fa) rats, and chlorogenic acid modifies their plasma insulin and liver protein and DNA. *J. Nutr. Biochem.* **2006**, *17*, 63–71. [\[CrossRef\]](#)
30. Upadhyay, R.; Mohan Rao, L.J. An Outlook on Chlorogenic Acids—Occurrence, Chemistry, Technology, and Biological Activities. *Crit. Rev. Food Sci. Nutr.* **2013**, *53*, 968–984. [\[CrossRef\]](#)
31. Kweon, M.-H.; Hwang, H.-J.; Sung, H.-C. Identification and Antioxidant Activity of Novel Chlorogenic Acid Derivatives from Bamboo (*Phyllostachys edulis*). *J. Agric. Food Chem.* **2001**, *49*, 4646–4655. [\[CrossRef\]](#)
32. Liang, N.; Kitts, D.D. Role of Chlorogenic Acids in Controlling Oxidative and Inflammatory Stress Conditions. *Nutrients* **2015**, *8*, 16. [\[CrossRef\]](#)
33. Ong, K.W.; Hsu, A.; Tan, B.K.H. Anti-diabetic and anti-lipidemic effects of chlorogenic acid are mediated by ampk activation. *Biochem. Pharmacol.* **2013**, *85*, 1341–1351. [\[CrossRef\]](#)
34. NPaynter, N.P.; Yeh, H.-C.; Voutilainen, S.; Schmidt, M.I.; Heiss, G.; Folsom, A.R.; Brancati, F.L.; Kao, W.H.L. Coffee and Sweetened Beverage Consumption and the Risk of Type 2 Diabetes Mellitus: The Atherosclerosis Risk in Communities Study. *Am. J. Epidemiol.* **2006**, *164*, 1075–1084. [\[CrossRef\]](#)
35. Wan, C.-W.; Wong, C.N.-Y.; Pin, W.-K.; Wong, M.H.-Y.; Kwok, C.-Y.; Chan, R.Y.-K.; Yu, P.H.-F.; Chan, S.-W. Chlorogenic Acid Exhibits Cholesterol Lowering and Fatty Liver Attenuating Properties by Up-regulating the Gene Expression of PPAR- α in Hypercholesterolemic Rats Induced with a High-Cholesterol Diet. *Phytotherapy Res.* **2013**, *27*, 545–551. [\[CrossRef\]](#) [\[PubMed\]](#)
36. Larsson, S.C.; Virtamo, J.; Wolk, A. Coffee Consumption and Risk of Stroke in Women. *Stroke* **2011**, *42*, 908–912. [\[CrossRef\]](#) [\[PubMed\]](#)
37. Tajik, N.; Tajik, M.; Mack, I.; Enck, P. The potential effects of chlorogenic acid, the main phenolic components in coffee, on health: A comprehensive review of the literature. *Eur. J. Nutr.* **2017**, *56*, 2215–2244. [\[CrossRef\]](#) [\[PubMed\]](#)
38. Jaiswal, R.; Matei, M.F.; Golon, A.; Witt, M.; Kuhnert, N. Understanding the fate of chlorogenic acids in coffee roasting using mass spectrometry based targeted and non-targeted analytical strategies. *Food Funct.* **2012**, *3*, 976–984. [\[CrossRef\]](#) [\[PubMed\]](#)

39. Ștefănescu, B.E.; Szabo, K.; Mocan, A.; Crișan, G. Phenolic Compounds from Five Ericaceae Species Leaves and Their Related Bioavailability and Health Benefits. *Molecules* **2019**, *24*, 2046. [\[CrossRef\]](#)
40. Haslam, E.; Naumann, M.O.; Britton, G. 1082. Phenolic constituents of *Vaccinium vitis idaea* L. *J. Chem. Soc.* **1964**, 5649–5654. [\[CrossRef\]](#)
41. Hokkanen, J.; Mattila, S.; Jaakola, L.; Pirttilä, A.M.; Tolonen, A. Identification of Phenolic Compounds from Lingonberry (*Vaccinium vitis-idaea* L.), Bilberry (*Vaccinium myrtillus* L.) and Hybrid Bilberry (*Vaccinium x intermedium* Ruthe L.) Leaves. *J. Agric. Food Chem.* **2009**, *57*, 9437–9447. [\[CrossRef\]](#)
42. Zhao, P.; Tanaka, T.; Hirabayashi, K.; Zhang, Y.-J.; Yang, C.-R.; Kouno, I. Caffeoyl arbutin and related compounds from the buds of *Vaccinium dunalianum*. *Phytochemistry* **2008**, *69*, 3087–3094. [\[CrossRef\]](#)
43. Li, N.; Zeng, W.-L.; Luo, X.-L.; Yang, C.-R.; Zhang, Y.-J.; Ding, Y.; Zhao, P. A new arbutin derivative from the leaves of *Vaccinium dunalianum* wight. *Nat. Prod. Res.* **2018**, *32*, 65–70. [\[CrossRef\]](#)
44. Colle, M.; Leisner, C.P.; Wai, C.M.; Ou, S.; Bird, K.A.; Wang, J.; Wisecaver, J.H.; Yocca, A.E.; Alger, E.I.; Tang, H.; et al. Haplotype-phased genome and evolution of phytonutrient pathways of tetraploid blueberry. *Gigascience* **2019**, *8*, giz012. [\[CrossRef\]](#) [\[PubMed\]](#)
45. Yu, J.; Hulse-Kemp, A.M.; Babiker, E.; Staton, M. High-quality reference genome and annotation aids understanding of berry development for evergreen blueberry (*Vaccinium darrowii*). *Hortic. Res.* **2021**, *8*, 228. [\[CrossRef\]](#) [\[PubMed\]](#)
46. Mengist, M.F.; Grace, M.H.; Mackey, T.; Munoz, B.; Pucker, B.; Bassil, N.; Luby, C.; Ferruzzi, M.; Lila, M.A.; Iorizzo, M. Dissecting the genetic basis of bioactive metabolites and fruit quality traits in blueberries (*Vaccinium corymbosum* L.). *Front. Plant Sci.* **2022**, *13*, 964656. [\[CrossRef\]](#) [\[PubMed\]](#)
47. Mengist, M.F.; Bostan, H.; Young, E.; Kay, K.L.; Gillitt, N.; Ballington, J.; Kay, C.D.; Ferruzzi, M.G.; Ashrafi, H.; Lila, M.A.; et al. High-density linkage map construction and identification of loci regulating fruit quality traits in blueberry. *Hortic. Res.* **2021**, *8*, 169. [\[CrossRef\]](#)
48. Mengist, M.F.; Bostan, H.; De Paola, D.; Teresi, S.J.; Platts, A.E.; Cremona, G.; Qi, X.; Mackey, T.; Bassil, N.V.; Ashrafi, H.; et al. Autopolyploid inheritance and a heterozygous reciprocal translocation shape chromosome genetic behavior in tetraploid blueberry (*Vaccinium corymbosum*). *New Phytol.* **2022**, *237*, 1024–1039. [\[CrossRef\]](#)
49. Cappai, F.; Amadeu, R.R.; Benevenuto, J.; Cullen, R.; Garcia, A.; Grossman, A.; Ferrão, L.F.V.; Munoz, P. High-Resolution Linkage Map and QTL Analyses of Fruit Firmness in Autotetraploid Blueberry. *Front. Plant Sci.* **2020**, *11*, 562171. [\[CrossRef\]](#)
50. Hummer, K.E.; Williams, K.A.; Bushakra, J.M. North American Crop Wild Relatives of Temperate Berries (*Fragaria* L., *Ribes* L., *Rubus* L., and *Vaccinium* L.). In *North American Crop Wild Relatives*; Springer: Cham, Switzerland, 2019; Volume 2, pp. 283–327. [\[CrossRef\]](#)
51. Wang, Y.; Fong, S.K.; Singh, A.P.; Vorsa, N.; Johnson-Cicalese, J. Variation of Anthocyanins, Proanthocyanidins, Flavonols, and Organic Acids in Cultivated and Wild Diploid Blueberry Species. *Hortscience* **2019**, *54*, 576–585. [\[CrossRef\]](#)
52. Wang, Y.; Fong, S.K.; Singh, A.P.; Vorsa, N.; Johnson-Cicalese, J. Blueberries and Cranberries. In *Temperate Fruit Crop Breeding Germplasm to Genomics*; Springer: Berlin/Heidelberg, Germany, 2008; pp. 115–150. [\[CrossRef\]](#)
53. Ranger, C.M.; Singh, A.P.; Johnson-Cicalese, J.; Polavarapu, S.; Vorsa, N. Intraspecific Variation in Aphid Resistance and Constitutive Phenolics Exhibited by the Wild Blueberry *Vaccinium darrowii*. *J. Chem. Ecol.* **2007**, *33*, 711–729. [\[CrossRef\]](#)
54. Kim, C.; Guo, H.; Kong, W.; Chandnani, R.; Shuang, L.-S.; Paterson, A.H. Application of genotyping by sequencing technology to a variety of crop breeding programs. *Plant Sci.* **2016**, *242*, 14–22. [\[CrossRef\]](#)
55. Broman, K.W.; Wu, H.; Sen, S.; Churchill, G.A. R/qtl: QTL mapping in experimental crosses. *Bioinformatics* **2003**, *19*, 889–890. [\[CrossRef\]](#)
56. Niggeweg, R.; Michael, A.J.; Martin, C. Engineering plants with increased levels of the antioxidant chlorogenic acid. *Nat. Biotechnol.* **2004**, *22*, 746–754. [\[CrossRef\]](#) [\[PubMed\]](#)
57. Luo, J.; Butelli, E.; Hill, L.; Parr, A.; Niggeweg, R.; Bailey, P.; Weisshaar, B.; Martin, C. AtMYB12 regulates caffeoyl quinic acid and flavonol synthesis in tomato: Expression in fruit results in very high levels of both types of polyphenol. *Plant J.* **2008**, *56*, 316–326. [\[CrossRef\]](#) [\[PubMed\]](#)
58. Yang, B.; Li, Y.; Song, Y.; Wang, X.; Guo, Q.; Zhou, L.; Xue, X.; Zhang, C. The R2R3-MYB transcription factor VcMYB4a inhibits lignin biosynthesis in blueberry (*Vaccinium corymbosum*). *Tree Genet. Genomes* **2022**, *18*, 27. [\[CrossRef\]](#)
59. Pratyusha, D.S.; Sarada, D.V.L. MYB transcription factors—Master regulators of phenylpropanoid biosynthesis and diverse developmental and stress responses. *Plant Cell Rep.* **2022**, *41*, 2245–2260. [\[CrossRef\]](#) [\[PubMed\]](#)
60. Amadeu, R.R.; Muñoz, P.R.; Zheng, C.; Endelman, J.B. QTL mapping in outbred tetraploid (and diploid) diallel populations. *Genetics* **2021**, *219*, iyab124. [\[CrossRef\]](#)
61. Ortiz, R.; Vorsa, N.; Bruederle, L.P.; Laverty, T. Occurrence of unreduced pollen in diploid blueberry species, *Vaccinium* sect. *Cyanococcus*. *Theor. Appl. Genet.* **1992**, *85*, 55–60. [\[CrossRef\]](#)
62. Tsuda, H.; Kojima, S.; Tetsumura, T.; Komatsu, H.; Kunitake, H. Induction of Polyploids in *Vaccinium* Using Oryzalin and Colchicine Treatments. *Hortic. Res.* **2012**, *11*, 205–212. [\[CrossRef\]](#)
63. Hall, S.H.; Galletta, G.J. Comparative Chromosome Morphology of Diploid *Vaccinium* Species1. *J. Am. Soc. Hortic. Sci.* **1971**, *96*, 289–292. [\[CrossRef\]](#)
64. Mallik, A.U.; Hamilton, J. Harvest date and storage effect on fruit size, phenolic content and antioxidant capacity of wild blueberries of NW Ontario, Canada. *J. Food Sci. Technol.* **2017**, *54*, 1545–1554. [\[CrossRef\]](#)

65. Castrejón, A.D.R.; Eichholz, I.; Rohn, S.; Kroh, L.W.; Huyskens-Keil, S. Phenolic profile and antioxidant activity of highbush blueberry (*Vaccinium corymbosum* L.) during fruit maturation and ripening. *Food Chem.* **2008**, *109*, 564–572. [CrossRef]
66. Connor, A.M.; Luby, J.J.; Tong, C.B.; Finn, C.E.; Hancock, J.F. Genotypic and Environmental Variation in Antioxidant Activity, Total Phenolic Content, and Anthocyanin Content among Blueberry Cultivars. *J. Am. Soc. Hortic. Sci.* **2002**, *127*, 89–97. [CrossRef]
67. Pertuzatti, P.B.; Barcia, M.T.; Gómez-Alonso, S.; Godoy, H.T.; Hermosin-Gutierrez, I. Phenolics profiling by HPLC-DAD-ESI-MSn aided by principal component analysis to classify Rabbiteye and Highbush blueberries. *Food Chem.* **2021**, *340*, 127958. [CrossRef]
68. Surjadinata, B.B.; Jacobo-Velázquez, D.A.; Cisneros-Zevallos, L. UVA, UVB and UVC Light Enhances the Biosynthesis of Phenolic Antioxidants in Fresh-Cut Carrot through a Synergistic Effect with Wounding. *Molecules* **2017**, *22*, 668. [CrossRef]
69. Zhao, T.; Sun, M.; Kong, L.; Xue, Q.; Wang, Y.; Wang, Y.; Khan, A.; Cao, J.; Cheng, G. Bioactivity-Guided Isolation of Phytochemicals from *Vaccinium dunalianum* Wight and Their Antioxidant and Enzyme Inhibitory Activities. *Molecules* **2021**, *26*, 2075. [CrossRef]
70. Zhu, W.; Gao, J. The Use of Botanical Extracts as Topical Skin-Lightening Agents for the Improvement of Skin Pigmentation Disorders. *J. Investig. Dermatol. Symp. Proc.* **2008**, *13*, 20–24. [CrossRef] [PubMed]
71. Xu, M.; Lao, Q.-C.; Zhao, P.; Zhu, X.-Y.; Zhu, H.-T.; Luo, X.-L.; Yang, C.-R.; He, J.-H.; Li, C.-Q.; Zhang, Y.-J. 6'-O-Caffeoylarbutin inhibits melanogenesis in zebrafish. *Nat. Prod. Res.* **2014**, *28*, 932–934. [CrossRef] [PubMed]
72. Tai, A.; Ohno, A.; Ito, H. Isolation and Characterization of the 2,2'-Azinobis(3-ethylbenzothiazoline-6-sulfonic acid) (ABTS) Radical Cation-Scavenging Reaction Products of Arbutin. *J. Agric. Food Chem.* **2016**, *64*, 7285–7290. [CrossRef] [PubMed]
73. Vilckickyte, G.; Raudone, L.; Petrikaite, V. Phenolic Fractions from *Vaccinium vitis-idaea* L. and Their Antioxidant and Anticancer Activities Assessment. *Antioxidants* **2020**, *9*, 1261. [CrossRef]
74. Liu, T.; Ortiz, J.A.; Taing, L.; Meyer, C.A.; Lee, B.; Zhang, Y.; Shin, H.; Wong, S.S.; Ma, J.; Lei, Y.; et al. Cistrome: An integrative platform for transcriptional regulation studies. *Genome Biol.* **2011**, *12*, R83. [CrossRef]
75. Clifford, M.N.; Kerimi, A.; Williamson, G. Bioavailability and metabolism of chlorogenic acids (acyl-quinic acids) in humans. *Compr. Rev. Food Sci. Food Saf.* **2020**, *19*, 1299–1352. [CrossRef]
76. Elshire, R.J.; Glaubitz, J.C.; Sun, Q.; Poland, J.A.; Kawamoto, K.; Buckler, E.S.; Mitchell, S.E. A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species. *PLoS ONE* **2011**, *6*, e19379. [CrossRef] [PubMed]
77. Rivera-Colón, A.G.; Catchen, J.M. Population genomics analysis with RAD, reprised: Stacks 2. *bioRxiv* **2021**. preprint. [CrossRef]
78. Danecek, P.; Bonfield, J.K.; Liddle, J.; Marshall, J.; Ohan, V.; Pollard, M.O.; Whitwham, A.; Keane, T.; McCarthy, S.A.; Davies, R.M.; et al. Twelve years of SAMtools and BCftools. *GigaScience* **2021**, *10*, giab008. [CrossRef] [PubMed]
79. Van Ooijen, J.W. *JoinMap 5 @Software for the Calculation of Genetic Linkage Maps in Experimental Populations of Diploid Species*; Kyazma, B.V.: Wageningen, The Netherlands, 2019.
80. Bates, D.; Maechler, M.; Bolker, B.; Steven, W. *Package 'lme4'*; R Foundation: Indianapolis, IN, USA, 2022.
81. Kuznetsova, A.; Brockhoff, P.B.; Christensen, R.H.B.; Jensen, S.P. *Package 'lmerTest'*. 2020. Available online: <https://github.com/runehaubo/lmerTestR> (accessed on 5 March 2023).
82. Russell, V.L.; Buerkner, P.; Herve, M.; Love, J.; Singmann, H. *Package 'Emmeans'*; R Foundation: Indianapolis, IN, USA, 2022. [CrossRef]
83. Wei, T.; Simko, V.; Levy, M.; Xie, Y.; Jin, Y.; Zemla, J.; Freidank, M.; Cai, J.; Protivinsky, T. *Package 'Corrplot'*. 2022. Available online: <https://github.com/taiyun/corrplot> (accessed on 1 March 2023).

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.