

Article

Chromosome-level genome sequence assembly and genome-wide association study of *Muscadinia rotundifolia* reveal the genetics of 12 berry-related traits

Minkyu Park¹, Daniel Vera², Devaiah Kambrianda³, Pranavkumar Gajjar¹, Lance Cadle-Davidson⁴, Violeta Tsoolova¹ and Islam El-Sharkawy^{1,*}
¹Center for Viticulture and Small Fruit Research, College of Agriculture and Food Sciences, Florida A&M University, 6361 Mahan Dr., Tallahassee, FL 32308, USA

²Silico LLC, 23 Essex Street #761119, Melrose, MA 02176, USA

³Plant and Soil Sciences, Southern University Agricultural Research and Extension Center, 181 B. A. Little Dr., Baton Rouge, LA 70813, USA

⁴USDA-ARS, Grape Genetics Research Unit, 630 West W North St., Geneva, NY, 14456, USA

*Corresponding Author: E-mail: islam.elsharkawy@fam.u.edu

Abstract

Vitis has two subgenera: *Euvtis*, which includes commercially important *Vitis vinifera* and interspecific hybrid cultivars, and *Muscadinia*. Of note, the market for *Muscadinia* grapes remains small, and only *Muscadinia rotundifolia* is cultivated as a commercial crop. To establish a basis for the study of *Muscadinia* species, we generated chromosome-level whole-genome sequences of *Muscadinia rotundifolia* cv. Noble. A total of 393.8 Mb of sequences were assembled from 20 haploid chromosomes, and 26 394 coding genes were identified from the sequences. Comparative analysis with the genome sequence of *V. vinifera* revealed a smaller size of the *M. rotundifolia* genome but highly conserved gene synteny. A genome-wide association study of 12 *Muscadinia* berry-related traits was performed among 356 individuals from breeding populations of *M. rotundifolia*. For the transferability of markers between *Euvtis* and *Muscadinia*, we used 2000 core genome rhAmpSeq markers developed to allow marker transferability across *Euvtis* species. A total of 1599 (80%) rhAmpSeq markers returned data in *Muscadinia*. From the GWAS analyses, we identified a total of 52 quantitative trait nucleotides (QTNs) associated with the 12 berry-related traits. The transferable markers enabled the direct comparison of the QTNs with previously reported results. The whole-genome sequences along with the GWAS results provide a new basis for the extensive study of *Muscadinia* species.

Introduction

Muscadinia grapes, which are native to the southeastern US and constitute one of the two subgenera of *Vitis*, were the first American grapes to be cultivated [1]. The other subgenus, *Euvtis*, includes interspecific hybrids of *Vitis vinifera* and *Vitis labrusca* (e.g. “Concord”, “Niagara”) that are familiar European and American bunch grapes, respectively. *Muscadinia* is composed of three known species, *M. rotundifolia*, *M. munsoniana*, and *M. popenoei*, and is a much smaller group than *Euvtis*, which includes approximately 70 species [2]. The number of chromosomes is also different between *Euvtis* ($2n=38$) and *Muscadinia* ($2n=40$). Whereas bunch grapes are one of the most economically important fruit crops worldwide, the market for *Muscadinia* grapes remains small. However, because of the nutritional content of *Muscadinia* grapes, they are receiving increasing attention [3–5]. Among the three *Muscadinia* species, only *M. rotundifolia* is commercially important.

Compared to bunch grapes, *Muscadinia* grapes have many superior traits. *Muscadinia* grapes are tolerant to

many diseases that cause extensive economic losses in bunch grapes, such as downy and powdery mildew, Pierce's disease, and phylloxera [6–9]. However, introducing these disease resistance traits to *Euvtis* is not efficient due to the low success rate of interspecific crosses [10]. Only one resistance locus for powdery mildew and downy mildew, *RUN1/RPV1*, has been introduced to *V. vinifera* through grape breeding programs [11–15]. Along with the disease resistance traits, *Muscadinia* grapes are noted for their high nutritional content, including that of phenolic compounds [16]. The phytochemicals in *Muscadinia* grapes are effective against inflammation [17], in reducing obesity [18], in anticancer, and in antioxidant effects [19, 20]. *Muscadinia* grapes have a distinctive aroma and flavor and contain sweet juice with low acidity [1].

Although *Muscadinia* grapes have the potential to be grown as important commercial crops due to their beneficial traits, only a few genetic studies related to *Muscadinia* breeding have been conducted [21–23]. Because genetic studies of *Euvtis* are relatively well established

Received: June 25, 2021; Accepted: September 25, 2021; Published: 18 January 2022

©The Author(s) 2022. 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.

compared to those of *Muscadinia*, the use of genetic information from *Euvitis* as a reference would be an effective approach for establishing genetic studies of *Muscadinia*. For this purpose, molecular markers developed in *Euvitis* have been used in *Muscadinia* to develop genetic linkage maps [21, 23]. However, the rate of transferability of SNPs obtained by simple sequence comparisons between *V. vinifera* species was found to be 2.3% among non-*vinifera* *Vitis* species and 0% in *M. rotundifolia* [24]. To address this issue, a recent study targeting the *Euvitis* core genome utilized RNase H2 enzyme-dependent amplicon sequencing (rhAmpSeq) to develop molecular markers that were transferrable across *Euvitis* species [25]. A total of 2000 rhAmpSeq markers were developed in the study, and 91.9% were transferable among *Euvitis* species. This ~45-fold improvement in marker transferability is expected to be helpful in the genetic study of *Muscadinia* grapes.

To establish a basis for studies in *Muscadinia* grapes, we performed whole-genome sequencing and chromosome-level assembly in a *Muscadinia* grape cultivar (*M. rotundifolia* cv. Noble). For high-throughput trait-associated marker development, we also tested the 2000 rhAmpSeq markers in a genome-wide association study (GWAS) involving a total of 12 berry-related traits.

Results and discussion

Sequencing and comparative analysis of the *Muscadinia* genome

Genomic DNA of *M. rotundifolia* cv. Noble was used to generate a total of 50.7 Gb of 2×250 bp sequence reads. The reference-free profiling of the *Muscadinia* genome revealed diploid with 1.55–1.58% of heterozygosity and the estimated size of the haplotype genome was ~416.1 Mb (Supplementary Fig. S1). The total lengths of repetitive and unique sequences were estimated to be ~144.3 (34.7%) and ~271.8 Mb (65.3%), respectively. The sequences were assembled into contigs ($N_{50} = 107\,128$ bp) with 399.5 Mb of the total length (Supplementary Table S1). The scaffolding of the contigs was performed by Hi-C and PacBio sequencing, resulting in 20 haploid chromosomes with a total assembled haploid length of 393.8 Mb. To evaluate the completeness of the assembly, we realigned the raw Illumina genome sequencing reads to the assembly, and 98.6% of these reads were mapped to the genome, indicating a high degree of completeness. We also performed a survey of core eukaryotic genes using BUSCO [26], which identified the presence of 96.0% complete universal single-copy orthologs. A recent diploid assembly of *M. rotundifolia* cv. Trayshed identified a similar number of complete BUSCO orthologs (97.5%) and a similar genome size (400.4 Mb for Haplotype1 and 369.9 Mb for Haplotype 2), suggesting that this “Noble” assembly marginally underrepresented the haploid genome [27]. While the estimated genome size of *M. rotundifolia* cv. Trayshed by a flow cytometry method was 483.4 ± 3.1 Mb in the

previous study [27], our estimation for *M. rotundifolia* cv. Noble by a K-mer distribution analysis was 416.1 Mb. Considering that the sizes of the two independently assembled *Muscadinia* genome sequences with long-range sequence information are 393.8 and 400.4 Mb, respectively, the estimated genome size of *M. rotundifolia* would be close to 416.1 Mb. A total of 26 394 genes were annotated using a combination of PacBio Iso-Seq, Illumina RNA-seq data, and the amino acid sequences of *V. vinifera*, similar to gene counts obtained for “Trayshed” (25 706 for Haplotype 1 and 23 673 for Haplotype 2). The number of unique genes identified in our gene annotation that do not appear in *M. rotundifolia* cv. Trayshed was 2600. Compared to the genes identified in the *V. vinifera* PN40024 genome, the number of unique genes in *M. rotundifolia* cv. Noble was 1757 and that in *M. rotundifolia* cv. Trayshed was 1804, indicating that both *Muscadinia* genomes have similar quality of gene annotation. However, considering the close genetic distance of the two *Muscadinia* cultivars, the number of the unique genes might be due to the error of gene annotation in both *Muscadinia* genomes.

To investigate the assembly accuracy of the *Muscadinia* chromosome sequences, we compared the collinear chromosome sequences of *M. rotundifolia* cv. Noble, *M. rotundifolia* cv. Trayshed [27], and *V. vinifera* PN40024 [28]. For the comparison, we used nucleotide BLAST results between *M. rotundifolia* cv. Noble and *M. rotundifolia* cv. Trayshed and the position information of orthologous gene sets between *M. rotundifolia* cv. Noble and *V. vinifera* PN40024 (Fig. 1). Between the two *Muscadinia* cultivars, at least ten partial inversions, three translocations, and five inversions with translocation were observed, indicating that one of the two chromosome assemblies might contain assembly errors or genomic variations. The assembled genome size of *V. vinifera* PN40024 was 498 Mb [28], about 21% larger than that of *M. rotundifolia* cv. Noble. Gene synteny was highly conserved between *M. rotundifolia* cv. Noble and *V. vinifera* PN40024, but at least 18 partial inversions were also observed. Chromosome 7 of *V. vinifera* was observed to be divided into chromosomes 7 and 20 in *M. rotundifolia*, as previously shown [27].

Genotyping of the *Muscadinia* population with *Euvitis* rhAmpSeq markers

The chromosome-level assembly of the *Muscadinia* genome enabled the use of GWAS for the genetic analysis of *Muscadinia* traits. Of the 2000 rhAmpSeq markers developed to target the core *Euvitis* genomes [25], a total of 1599 markers showed polymorphisms in 356 individuals (Supplementary Table S2). After filtering and imputation, a total of 1283 markers and 348 individuals were used for the subsequent analyses. The imputation accuracy was 0.94. Consistent with the development of rhAmpSeq markers mainly in gene-rich regions [25], the 1283 markers primarily aligned to gene-rich regions of *Muscadinia* chromosome sequences (Fig. 2A,

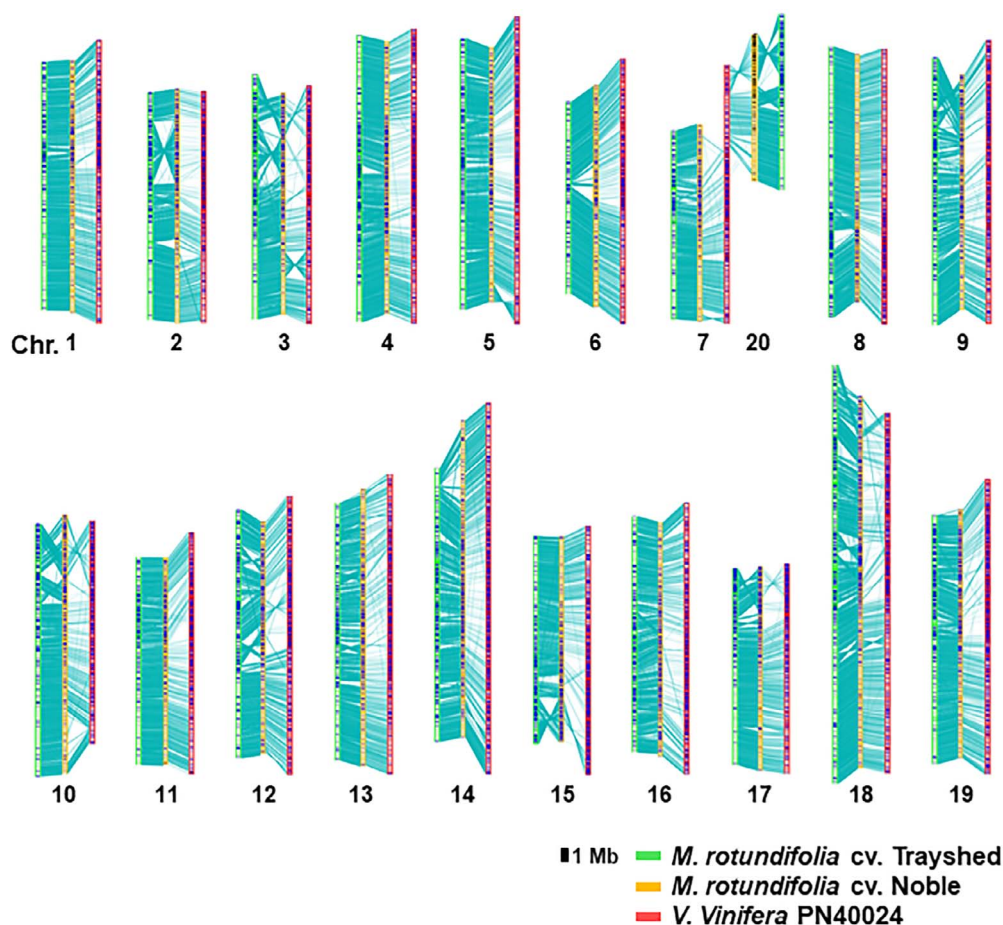


Figure 1. Comparative analysis of the collinear chromosome sequences of *Muscadinia rotundifolia* cv. Noble, *M. rotundifolia* cv. Trayshed, and *Vitis vinifera* PN40024. The vertical bars indicate chromosomes of *M. rotundifolia* cv. Trayshed (left green bar), *M. rotundifolia* cv. Noble (middle yellow bar), and *V. vinifera* PN40024 (right red bar). The positions of collinear regions between the two *Muscadinia* cultivars and the orthologous gene sets between *M. rotundifolia* cv. Noble and *V. vinifera* PN40024 are linked with lines, respectively. The blue and white blocks on each chromosome indicate the distribution of LTR retrotransposons and genes. The chromosome numbers are presented at the bottom of the graph.

upper panel). The distribution of the distances between adjacent markers was consistent for each chromosome, with median values ranging from ~120 to 250 kb and interquartile values ranging from ~70 to 510 kb (Fig. 2A, lower panel).

Linkage disequilibrium (LD) between adjacent markers averaged 0.36 (r [2]), and the distance over which LD decayed to half of the initial value was ~2.3 Mb (Fig. 2B). This represents far slower LD decay than that reported in *Euveitis*, in which rapid LD decay has been observed within 11.6 and 16.6 kb [29, 30]. The reason for the slow LD decay might be the origin of the population which consists of breeding populations with multiple parent sets (Supplementary Table S2). Similar slow LD decay was also observed in a GWAS of breeding lines of apple derived from crossing [31].

To identify the number of subpopulations (k) of the *Muscadinia* population, we performed a parametric clustering analysis implemented by the ADMIXTURE software [32]. Subpopulations from $k=1$ to $k=30$ were tested, and the minimum of the cross-validation error was reached at $k=15$ (Fig. 2C and Supplementary Fig. S2). Therefore, the optimum k value

was determined to be 15, and the value was used in the subsequent GWAS analysis.

Phenotype analysis of 12 berry-related traits of *Muscadinia* grapes

For GWAS, we used phenotype data for 12 berry-related traits of *Muscadinia* grapes (Supplementary Fig. S3). To investigate the correlation coefficient among the 12 traits, we performed the Pearson test with the phenotype data. Two groups of the traits revealed significant positive correlations (Fig. 3). One group included three fruit color-related traits like visible color, chroma, and lightness, revealing ≥ 0.69 of the correlation coefficient (r). The other group included the traits like length, width, weight per berry, and firmness ($r \geq 0.54$). The length and width revealed a strong positive correlation ($r=0.92$), indicating *Muscadinia* berries tend to have a round shape. Interestingly, the firmness of berries also showed a positive correlation with the length, width, and weight per berry ($r \geq 0.54$), indicating that larger berries tend to be firmer. The number of berries per cluster negatively correlated with the length, width, and weight per berry ($r \leq -0.27$).

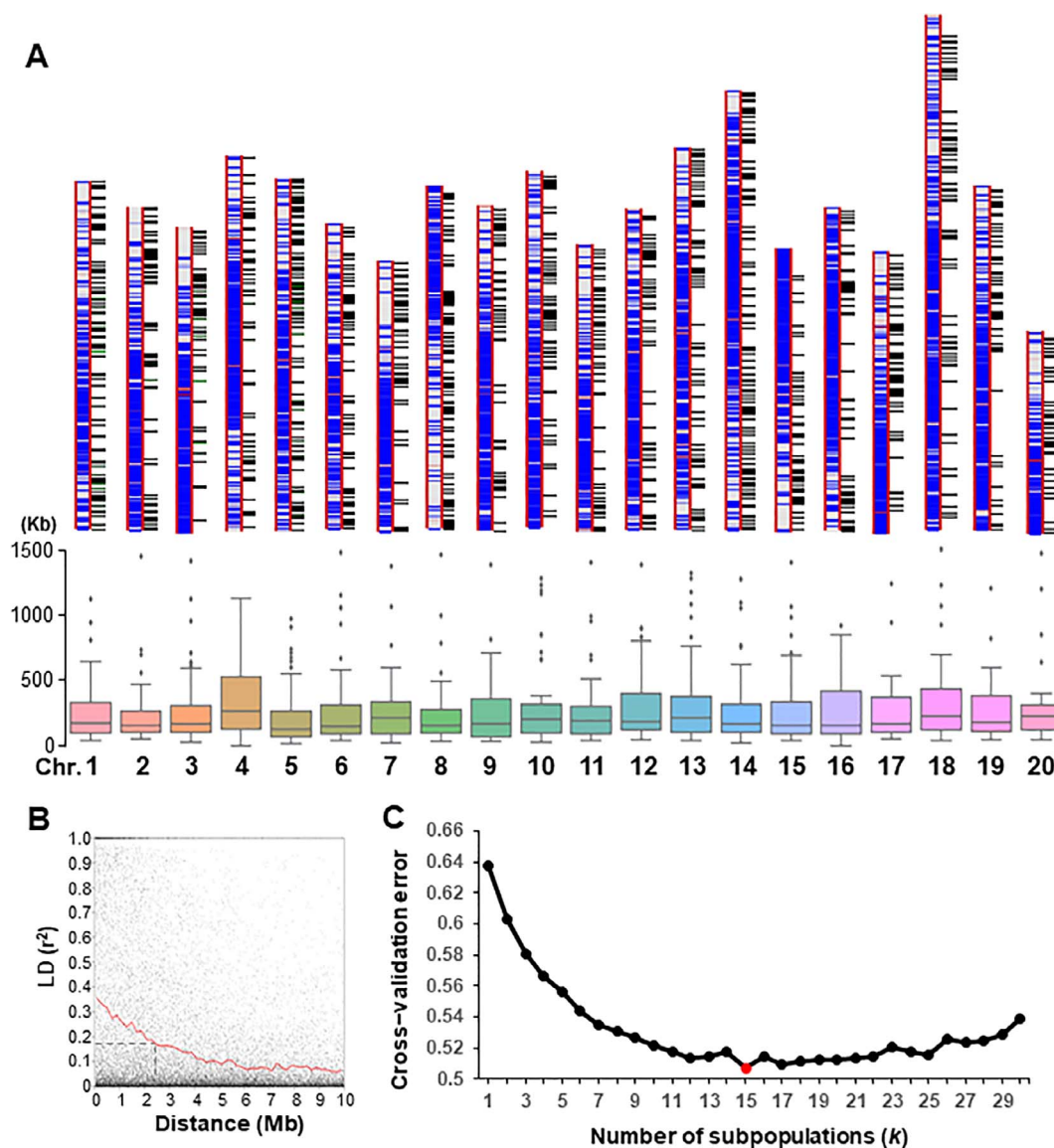


Figure 2. Marker distribution, linkage disequilibrium (LD) and population structure analysis. **A** Distribution of the 1283 rhAmpSeq markers on the chromosome sequences of *M. rotundifolia*. Vertical bars indicate *Muscadinia* chromosomes. Blue and white boxes in each bar indicate the distribution of LTR retrotransposons and genes, respectively. Black lines to the right of the bars represent the distribution of markers. The box plots under the chromosomes represent the distribution of distances between two adjacent markers. The chromosome numbers are presented at the bottom of the box plot. **B** LD values (r^2) by distance (black dots) and mean values (red line). **C** Cross-validation error values for $k=1-30$.

Quantitative genetic analysis of the 12 berry-related traits with multi-locus GWAS approaches

We performed multi-locus GWAS for the genetic analysis of quantitative traits. Because the quantitative trait nucleotides (QTNs) detected by multiple multi-locus approaches are known to be more reliable [33], we tested a total of six multi-locus methods, including mrMLM, FASTmrMLM, FASTmrEMMA, pLARmEB, pkWmEB, and ISIS EM-BLASSO [34]. To balance the high power of the multi-locus approaches for QTN detection and low false-positive rate, we used ≥ 3 of a logarithm of the odds (LOD) threshold value [33]. If multiple methods detect the same QTN, we chose the result with the highest $-\log_{10}(P)$ value. With this approach, we identified a total of 52 QTNs that were associated with the 12 berry-

related traits (Fig. 4, Table 1, Supplementary Fig. S4, and Supplementary Table S3). Of the 52 QTNs, 26 were detected by multiple methods (Fig. 4, indicated by pink dots), indicating that these QTNs are more reliable than those were detected by a single method (Fig. 4, indicated by blue dots). Among the six methods, pLARmEB revealed the most sensitive detection of QTNs by detecting a total of 35 QTNs (Table 1).

Berry visible color, chroma, and lightness

We performed GWAS with three types of berry color: visible color, chroma, and lightness. A total of nine QTNs were detected for these traits, and three of them, including chr4_9358331, chr4_11238026, and chr16_23164495, were pleiotropic QTNs identified in multiple berry

Table 1. List of QTNs for the 12 berry-related traits identified by six multi-locus GWAS methods

Trait ID	Trait name	QTNs*	Chr.	Position (bp)	LOD	-log ₁₀ (p)	PVE** (%)	Methods***					
								1	2	3	4	5	6
1	Visible color	chr4_9358331	4	7 904 506	14.97	16.00	10.46				•		
		chr4_11238026	4	10 084 946	3.49	4.22	19.93					•	
2	Chroma	chr2_6241099	2	7 109 208	3.51	4.23	10.23					•	
		chr4_2563214	4	2 437 667	3.99	4.74	9.13	○	○		•		
		chr4_3801978	4	3 548 913	3.51	4.23	5.78					•	○
		chr4_9358331	4	7 904 506	4.33	5.10	8.37	○	○	•	○		○
		chr4_11238026	4	10 084 946	12.53	13.51	28.15	○	○	○	○	•	○
		chr13_29070085	13	22 494 261	7.99	8.89	3.03					•	
3	Lightness	chr16_23164495	16	18 709 306	6.43	7.28	3.88					•	
		chr4_9358331	4	7 904 506	4.35	5.12	8.06	○	○	•	○		
		chr4_11238026	4	10 084 946	8.32	9.22	10.66	○	○	○	•	○	○
		chr16_23164495	16	18 709 306	4.78	5.56	4.48				○	•	○
		chr17_1629617	17	1 456 443	3.89	4.64	1.73				•		
		chr17_4889129	17	4 791 763	3.80	4.55	3.35				•		
4	Length	chr2_4825658	2	4 458 307	5.73	6.55	10.26	○	○	○	•	○	○
		chr2_13537760	2	13 137 932	3.44	4.17	1.63	○	•				
		chr5_24329039	5	19 702 139	3.03	3.73	1.50				•		
		chr10_7345058	10	7 353 876	3.46	4.18	1.30				•		
		chr10_10547760	10	9 977 603	4.20	4.97	4.80	○	○		•		○
		chr15_13698923	15	10 956 919	4.71	5.50	1.33	○	○	○	•	○	○
5	Width	chr18_23887396	18	20 585 980	3.93	4.68	3.46		○		•		
		chr19_7169706	19	6 350 998	3.81	4.55	6.12				○		•
		chr1_21821786	1	18 219 115	3.26	3.97	1.13	○	○		•	•	
		chr2_4825658	2	4 458 307	6.97	7.83	13.22	○	○	○	•	○	○
		chr2_13537760	2	13 137 932	3.63	4.36	1.19				•		
		chr5_2429823	5	1 720 612	4.56	5.34	3.11				•		
6	Weight/berry	chr5_24329039	5	19 702 139	4.09	4.84	2.46				•		
		chr10_7345058	10	7 353 876	3.10	3.80	1.05				•		
		chr15_13698923	15	10 956 919	3.45	4.18	1.03			○		○	•
		chr18_23887396	18	20 585 980	4.44	5.21	3.85				•		
		chr2_3474971	2	3 072 740	3.53	4.26	2.01			•			
		chr2_4825658	2	4 458 307	6.00	6.84	11.16	○	○		•	○	○
7	Firmness	chr3_5507072	3	4 718 447	3.09	3.79	0.96				•		
		chr6_506716	6	355 390	5.21	6.01	5.39	○	○		•	○	○
		chr7_2667199	7	2 513 742	3.39	4.11	1.69				•		
		chr9_9426423	9	7 742 852	4.27	5.04	1.61				•		
		chr9_19317456	9	16 267 491	4.35	5.11	3.12				•		
		chr13_393712	13	413 927	3.29	4.01	2.60	•	○		○		○
8	No. of berries/cluster	chr2_4825658	2	4 458 307	5.70	6.52	9.41			○	•		
		chr3_15938806	3	11 742 919	3.04	3.74	2.83			•			
		chr8_2784591	8	2 800 563	5.11	5.91	4.74		○	○	•		○
9	Weight/cluster	chr5_5884957	5	4 895 628	3.04	3.74	7.85				•		
		chr7_26249020	20	9 464 815	3.01	3.71	1.66					•	
		chr14_21740224	14	17 526 760	3.13	3.84	6.95	○	○			•	
10	No. of seeds/berry	chr17_11747237	17	10 580 775	3.99	4.74	16.97	•	○		○		○
		chr7_22781295	20	4 818 819	3.76	4.50	12.89	○	•				○
11	Scar pattern	chr17_7695724	17	7 309 793	4.07	4.82	10.65	○	○	○			•
		chr10_650959	10	458 147	3.23	3.94	4.40					•	
		chr11_5215446	11	5 016 853	3.71	4.44	11.78					•	
		chr14_17450562	14	13 939 411	3.70	4.44	9.21	•					
12	Maturation date	chr14_23376067	14	18 959 313	3.12	3.82	1.98		•				
		chr18_7185348	18	6 784 087	3.67	4.41	6.48	○	○	○	•	○	○

•: The most significant method; ○: Significant methods *QTN names in bold font are pleiotropic QTNs that are detected associate with multiple traits. **PVE: Phenotypic variation explained ***1: mrMLM; 2: FASTmrMLM; 3: FASTmrEMMA; 4: pLARM; 5: pKwMEB; 6: ISIS EM-BLASSO

color traits (Table 1). Among them, chr4_9358331 and chr4_11238026 were identified as the major QTNs in the three traits revealing 8.06–28.15% of phenotypic variation explained (PVE). Because the markers used in this study were named after the chromosome and position in the

reference genome sequence of *V. vinifera* [28], the marker position in the *V. vinifera* genome could be identified from the marker name. Therefore, the positions of the two major QTNs for berry color in the *V. vinifera* genome were 9.36 and 11.24 Mb on chromosome 4.

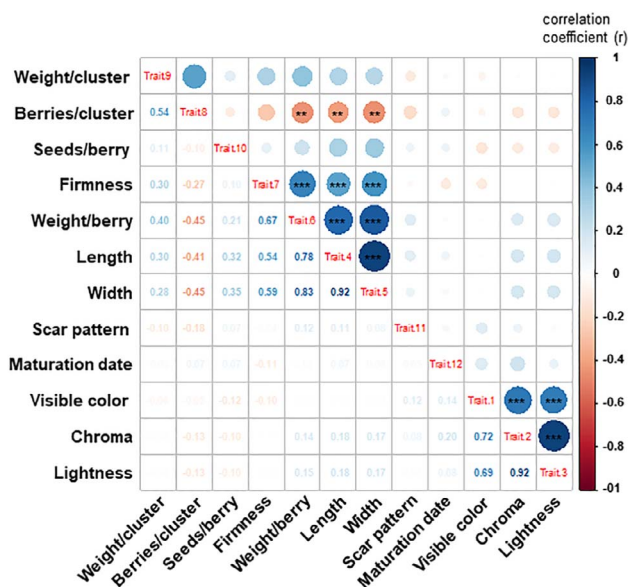


Figure 3. Correlation coefficient analysis of 12 berry-related traits. The correlation coefficient values are presented by circles at the top of the diagonal and by numbers at the bottom. Positive and negative correlations are presented by blue and red colors, respectively. The circle size and color intensity are proportional to the correlation coefficient values. The significance levels are presented with stars. **: $p \leq 0.01$ and ***: $p \leq 0.001$.

The position of chr4_11238026 was consistent with the previous finding obtained from a high-density linkage map of *Muscadinia* species [22]. In the *Muscadinia* linkage map, the berry color locus was predicted to be located between 11.09 and 11.99 Mb on *Euvitis* chromosome 4, which includes the position of chr4_11238026. Another report identified a candidate gene for the determination of *Muscadinia* berry color based on the high-density linkage map of *Muscadinia* [35]. In the study, a glutathione S-transferase gene (*VvGST4*) associated with the anthocyanin biosynthetic pathway [36, 37] was suggested as the candidate gene [35]. To determine whether our QTN is physically associated with the candidate gene, we investigated annotated genes between the two flanking markers of the QTN in the *Muscadinia* genome. The positions of the flanking markers were 9.96–10.75 Mb on *Muscadinia* chromosome 4, and a total of 29 genes were identified (Supplementary Table S4). Of the 29 genes, four glutathione S-transferase genes were identified. Among them, the two genes between 10.58 and 10.61 Mb of *Muscadinia* chromosome 4 (Supplementary Table S4, *evm.model.chr4.1308* and *evm.model.chr4.1310*) corresponded to the candidate gene (*VvGST4*) with $\geq 98\%$ nucleotide sequence identity. This indicates that the chr4_11238026 QTN for berry color in our study is associated with the known candidate gene for the trait. In the previous genetic study, the berry color locus was identified by qualitative trait scoring (black or bronze), indicating the locus is for qualitative trait determination [22]. Therefore, the chr4_11238026 QTN corresponding to the same locus was also expected to be the qualitative trait locus for *Muscadinia* berry color.

Because we used quantitative phenotype scores for the berry color-related traits, other QTNs for these traits were expected to determine quantitative traits. The pleiotropic QTN, chr4_9358331, revealed 8.37–10.46% of PVE in the three berry color-related traits and is considered a major QTN for *Muscadinia* berry-color traits. The flanking markers for this QTN were located at 7.49 and 7.97 Mb on *Muscadinia* chromosome 4, and a total of 22 genes were identified between the two markers (Supplementary Table S4).

The major quantitative trait loci (QTLs) of berry color in *Euvitis* species have been identified as *VvMYBA1* and *VvMYBA2* [29, 38]. The orthologs of these genes in *Muscadinia* were identified around 19.01 Mb of *Muscadinia* chromosome 2. However, no QTN was detected around that region in this study. A previous study reported additional QTLs for the *Euvitis* berry color trait based on a high-density linkage map [39]. One of the additional QTLs was identified as being associated with *VvGST4*, which is related to the qualitative trait of *Muscadinia* berry color [35]. Among the rest QTLs, only one QTL, col-2-2, showed a similar location with one of our QTNs, chr2_6241099. One of the candidate genes identified by col-2-2 was *VIT_02s0012g00630*, and its *Muscadinia* ortholog was located at 7.46 Mb of *Muscadinia* chromosome 2 which is about 0.35 Mb away from the position of chr2_6241099. The flanking markers of chr2_6241099 were identified at 6.99 and 8.57 Mb on *Muscadinia* chromosome 2. Therefore, the chr2_6241099 QTN was expected to be the same QTL as col-2-2. Except for the chr2_6241099 QTN, the rest QTNs revealed different positions from the previous results in *Euvitis*. These results may indicate that the QTNs of berry color traits in *Muscadinia* are different from those in *Euvitis*; therefore, the berry color of *Muscadinia* might be regulated by a different genetic background from that of *Euvitis*.

Berry length, berry width, weight per berry, berry firmness, and number of berries per cluster

A total of eight QTNs were detected in berry length and width, respectively (Table 1). Of the eight QTNs, six were pleiotropic QTNs identified in both traits. This might explain the strong correlation (Fig. 3, $r = 0.92$) between the two traits. Among the six pleiotropic QTNs, chr2_4825658 revealed the highest PVE in both traits as 10.26% in length and 13.22% in width. The chr2_4825658 QTN was also detected in weight per berry and firmness, revealing 11.16% and 9.41% of PVE. The PVE of the rest QTNs was $\leq 6.12\%$. Therefore, chr2_4825658 was the major QTN for the four traits and might explain their correlation (Fig. 3, $r \geq 0.54$). The flanking markers of chr2_4825658 were located at 4.29 and 6.72 Mb of *Muscadinia* chromosome 2. Between the two flanking markers, a total of 211 genes were identified (Supplementary Table S4). Because of the long distance between the flanking markers, it would also be possible that multiple genes in this region are involved in the four traits.

There has been no genetic study of *Muscadinia* berries related to size, weight, and firmness. In contrast, a recent

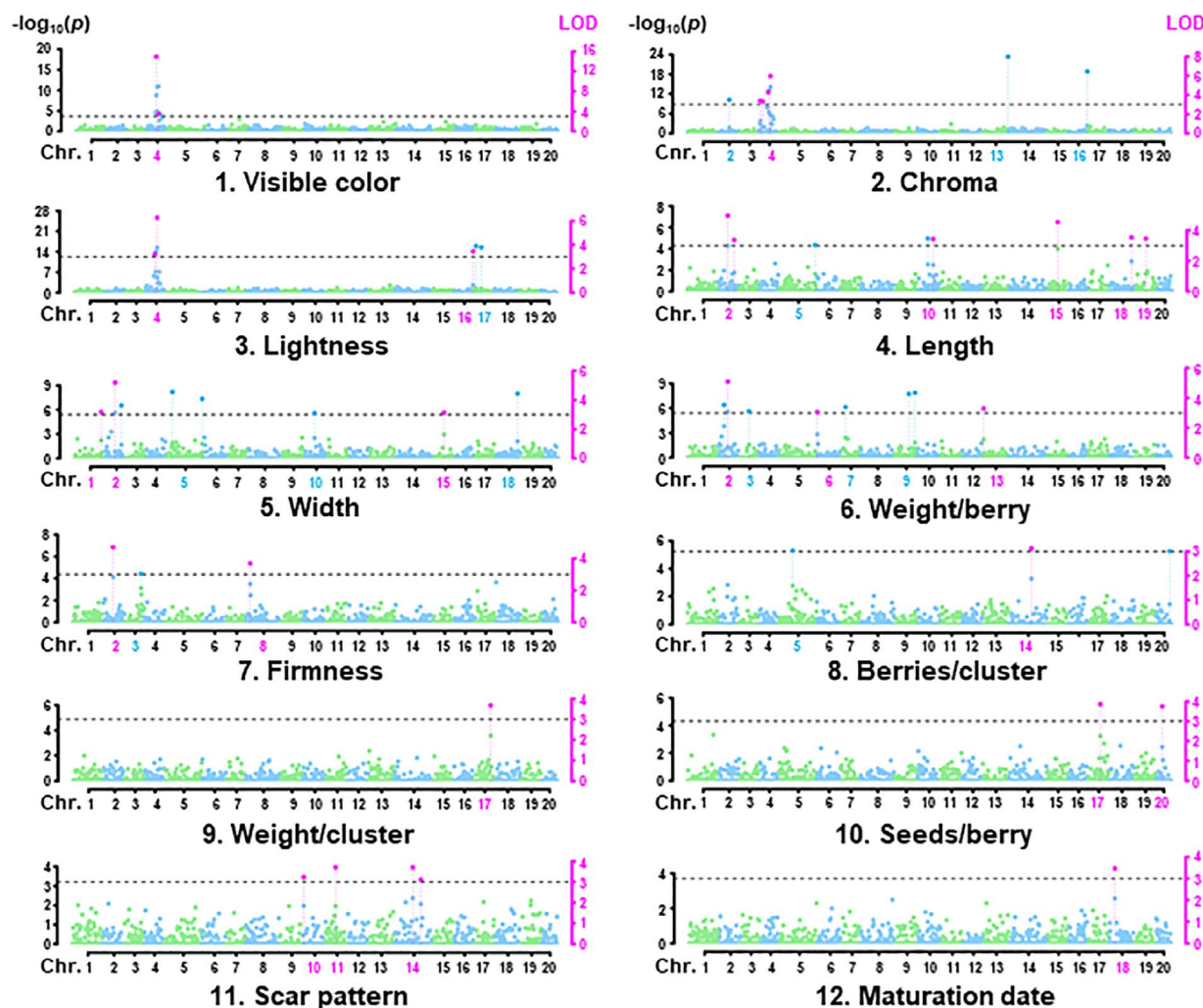


Figure 4. GWAS results for the 12 berry-related traits. Numbers 1–12 to the left of each chart correspond to the traits listed below the figure. The dotted black lines indicate the significance threshold at LOD = 3.0. The significant QTNs detected by multiple and single GWAS methods are marked with pink and blue dots with vertical dotted lines, respectively. Chromosome numbers containing the significant QTNs are marked with the same color. For the results by multiple methods, the median value of $-\log_{10}(p)$ from the mrMLM, FASTmrMLM, FASTmrEMMA, and pKWmeB methods was used in plotting.

study identified four markers associated with berry weight in *Euvitis* species [29]. However, the four identified markers were on *Euvitis* chromosomes 17, 18, and 19, different from our findings in *Muscadinia*. In our study, chr2_4825658 was the major QTN for berry weight, and the next was chr6_506716 with 5.39% of PVE. The two QTNs were located on chromosomes 2 and 6. The rest QTNs for berry weight revealed $\leq 3.12\%$ of PVE.

In berry firmness, chr2_4825658 was the major QTN. In multiple fruit crops, the firmness of fruits is known to be determined by the activity of a polygalacturonase gene [40–42]. Among the genes identified by chr2_4825658, a polygalacturonase gene was identified at 6.23 Mb of *Muscadinia* chromosome 2 (Supplementary Table S4, evm.model.chr2.980). A QTL study with apple (*Malus domestica* Borkh.) identified that the polygalacturonase gene is one of the QTLs that determines the apple fruit firmness [43]. These may indicate that the polygalacturonase identified by chr2_4825658 is a candidate gene for the berry firmness of *Muscadinia* grapes.

The number of berries per cluster revealed three QTNs. Of the three QTNs, chr5_5884957 revealed the highest PVE at 7.85%, and a total of 42 genes were identified between the flanking markers of the QTN (Supplementary Table S4). Although it revealed a significant negative correlation with the three berry size-related traits, it did not share any pleiotropic QTN with the three traits.

Weight per cluster, number of seeds per berry, scar pattern, and maturation date

Because berry size and the number of berries per cluster were negatively correlated, weight per cluster was expected to be related to yield. Only one QTN (chr17_11747237) was detected for the weight per cluster with 16.97% of PVE. A total of 96 genes were identified between flanking markers of the QTN (Supplementary Table S4). In *Euvitis* grapes, the QTL for cluster weight was analyzed in three different years and revealed unstable QTLs [44], indicating that the QTN for weight per cluster of *Muscadinia* may also need

further study to obtain reliable results. Two QTNs were detected for the number of seeds per berry, including chr7_22781295 and chr17_7695724 with 12.89 and 10.65% of PVE, respectively. A total of 33 and 15 genes were identified by chr7_22781295 and chr17_7695724, respectively (Supplementary Table S4). The scar pattern was measured to the extent of the wound on berry skin when separating a mature berry from the pedicel. A total of four QTNs were detected for scar pattern. Among them, chr11_5215446 revealed the highest PVE with 11.78%. A total of 36 genes were identified between the flanking markers of chr11_5215446 (Supplementary Table S4). In the maturation date, one QTN (chr18_7185348) was detected with 6.48% of PVE. The QTN was supported by all the six methods, indicating that it is statistically reliable. A total of 72 genes were identified between the flanking markers of the QTN (Supplementary Table S4).

GWAS of *Muscadinia* grapes

We performed a GWAS of *Muscadinia* berry-related traits with the 1283 rhAmpSeq markers developed from *Euveitis* genome sequences [25]. Although the number of markers used was much smaller than that employed in the genotyping by sequencing (GBS) method, the rhAmpSeq markers provide more information for tracking ancestral haplotypes, as each amplicon is a local haplotype spanning multiple SNPs [25]. The rhAmpSeq marker resolution was sufficient here in part because our population showed slow LD decay. However, some chromosomes harbored empty blocks that did not contain markers in gene-rich regions (Fig. 2A: chromosomes 2, 4, 6, 8, 10, 18, and 20). This is an issue that should be solved to use transferable markers across a wide range of *Vitis* species. Although the transferability of the *Vitis* rhAmpSeq markers was shown to be 91.9% when applied to *Euveitis* species, it was 61.9% to the *Muscadinia* species examined in our study. This difference is partly due to the genetic distance between *Euveitis* and *Muscadinia* species and the low diversity among the evaluated *Muscadinia* individuals, which consisted of progeny from multiple parental sets (Supplementary Table S2). The low diversity of the *Muscadinia* population might also be related to the slow LD decay observed. Therefore, the development of a proper core *Muscadinia* collection will be necessary to perform an efficient GWAS with a high resolution.

Among the 12 berry traits analyzed in our study, the three traits related to berry color could be compared to the results of the previous genetic analysis in *Muscadinia* species [22]. The two independent genetic analyses of berry color in *Muscadinia* grapes revealed the same results, indicating that the results of our GWAS are reliable. The multi-locus GWAS methods are more powerful than the single-locus method in detecting QTNs for complex traits [33]. Using six different multi-locus GWAS methods, we could obtain reliable QTNs for the 12 *Muscadinia* berry-related traits. In particular, the pleiotropic QTNs identified from the significantly correlated traits

would provide strong evidence for further study to identify pleiotropic genes that control the traits.

Conclusions

In this study, we report a chromosome-level reference genome sequence for *Muscadinia* grapes. Based on the assembled chromosome sequences, the genome size of *M. rotundifolia* was approximately 21% smaller than that of *V. vinifera*, but gene synteny was highly conserved between the two genomes. Our multi-locus GWAS analysis of *Muscadinia* grapes with transferable markers between *Muscadinia* and *Euveitis* revealed reliable results in the identification of quantitative trait-associated markers. The transferability of the markers used in our GWAS will provide new opportunities to study comparative genetics between *Muscadinia* and *Euveitis* species. Along with the obtained whole-genome sequence, the GWAS of *M. rotundifolia* will help expand the study of *Muscadinia* grapes.

Materials and methods

Genome sequencing and assembly

Plant samples were collected from a 5-year-old *M. rotundifolia* cv. Noble grapevine. Flash-frozen leaf was ground with a mortar and pestle under liquid nitrogen, and nuclear DNA was isolated using the Dellaporta method. Sequencing libraries were prepared using 1 µg of DNA using the Illumina TruSeq PCR-free kit (Illumina), and size selection for ~250 bp fragments was performed using SPRIselect beads (Beckman Coulter). Libraries were sequenced at ~121× depth on an Illumina HiSeq 2500 in rapid run mode, and a total of 101428643 paired-end reads were generated. The total length of the reads was 50.7 Gb. The ploidy level was determined with Smudgeplot software [45]. The heterozygosity and haploid genome length of the *Muscadinia* genome were estimated with GenomeScope 2.0 [45]. For Smudgeplot and GenomeScope analyses, K-mers were counted with KMC software at K = 21 [46]. We assembled the reads using DISCOVAR de novo (Broad Institute, Cambridge, MA) to generate sequence contigs. To scaffold the resulting contigs, we performed HiRise genome scaffolding using both in situ and in vitro Hi-C. To fill the remaining assembly gaps, we generated ~10X coverage PacBio sequences with an average fragment size of 10 kb. The gap-filling was performed with PBJelly (ver. 14.9.9).

Gene annotation

To annotate genes, we first annotated transposable elements with RepeatModeler v2.0.1 and masked repeat sequences with RepeatMasker v4.1.0. To identify and characterize genes, we used muscadine transcriptome data. For this purpose, we generated RNA-seq data on the Illumina HiSeq 2500 platform from triplicate samples of 13 tissues (adult leaves, closed flowers, fine roots, fruit set, mature fruits, open flowers, pollen, preveraison, ripening seeds, ripening skin/flesh, tendrils, veraison,

and young leaves). In addition, we performed PacBio Iso-Seq on pooled RNA using the Sequel II platform. A total of 4 834 933 circular consensus reads were generated, with a total length of 11.9 Gb and an average read length of 2.5 kb. By combining Illumina RNA-seq, PacBio Iso-Seq, and protein sequences of the *V. vinifera* genome, we performed gene prediction with BRAKER [47].

The unique genes in *M. rotundifolia*, cv. Noble and *M. rotundifolia*, cv. Trayshed were identified by nucleotide BLAST search (single high-scoring segment pair (HSP), score: ≥ 200 , identity: ≥ 80 , and e-value: $\leq e-10$) with coding sequences (CDSs).

Comparative genome analysis

Comparative genome analysis was performed on the chromosome sequences of *M. rotundifolia* cv. Noble, *M. rotundifolia* cv. Trayshed, and *V. vinifera* PN40024. Colinear regions between the two *Muscadinia* cultivars were identified by nucleotide BLAST search (score: ≥ 2000 , identity: $\geq 97\%$, and e-value: $\leq e-10$). To identify colinear regions between *M. rotundifolia* cv. Noble and *V. vinifera* PN40024, orthologous CDSs were identified according to reciprocal best hits from a BLAST search (single HSP, score: ≥ 150 , and e-value: $\leq e-8$) using all the CDSs of the two species. The reverse transcriptase (RT) domain sequences of LTR retrotransposons were isolated with HMMER (default options) using customized HMM profiles for the Gypsy and Copia superfamilies (Supplementary Files 1 and 2). The position information of the RT sequences was used to display the distribution of LTR retrotransposons in the two *Muscadinia* cultivars and *V. vinifera* chromosome sequences. The visualization of the comparative analysis results was carried out with an in-house Python script.

Genotyping, population structure, and LD

The genomic DNA of the 356 *Muscadinia* individuals was extracted with Qiagen DNeasy plant kits by following the provided manual. The genotyping of the 356 individuals was performed with the 2000 rhAmpSeq markers by following the method described in Zou et al. [25]. A total of 1599 markers revealed polymorphisms among the 356 individuals. Markers with a minor allele frequency of ≤ 0.01 and individuals with a minimum heterozygosity proportion of ≤ 0.05 were filtered, leaving 1283 markers and 348 individuals, respectively. The missing genotype data were imputed with the LD KNNI method implemented in TASSEL 5.0, and the imputation accuracy was calculated with the same software. The marker density on the *Muscadinia* chromosome sequences was analyzed by aligning the 1283 marker sequences with BLAST (single hit, single HSP, score: ≥ 40 , and e-value: $\leq e-8$, identity: ≥ 0.95). The genotype data of the 348 individuals was used to analyze the population structure. The number of subpopulations (k) was inferred by ADMIXTURE analysis [48]. The bar plot of the 348 individuals was presented with STRUCTURE 2.3.4 [49, 50] for $k = 15$. The LD (r [2]) values among the 1283 markers were calculated with the

software implemented in TASSEL 5.0, and the graph of LD decay was generated with an in-house Python script.

Phenotyping and GWAS analysis

All phenotype data of the 348 *Muscadinia* individuals were obtained from previous studies by Campbell et al. [51–53]. The correlation coefficient analysis of the phenotypes was conducted using the R package “corrplot”. The associations between phenotypes and markers were analyzed with six multi-locus GWAS methods implemented in the R package “mrMLM.GUI”. The Q-matrix for $k = 15$ was generated with STRUCTURE software and used in the GWAS. Kinship (K) matrix was calculated with the software implemented in mrMLM.GUI. Default options were used for the rest. The Manhattan plots and Q–Q plots were generated with the software implemented in mrMLM.GUI.

Acknowledgments

This work was supported by the 1890 Institution Teaching, Research, and Extension Capacity Building Grants (CBG) Program (grant no. 2019-38821-29150; project accession no. 1018161) from the USDA National Institute of Food and Agriculture and the USDA-NIFA-Evans Allen, Research Program (grant no. FAMUVITI-2023; project accession no. 1019285). We thank Linda Cote, Peter Schweitzer, Cheng Zou, and Qi Sun at the Biotechnology Resource Center of Cornell University for their support of rhAmpSeq genotyping, developed and facilitated by the US Department of Agriculture (USDA)-National Institute of Food and Agriculture (NIFA) Specialty Crop Research Initiative. Award No. 2017-51181-26829.

Contributions

M.P., I.E., and D.K. designed the research. M.P. performed the genome and GWAS analysis. D.V. assembled the genome sequences and annotated genes. P.G. performed phenotyping. L.C. performed genotyping. M.P. visualized data and wrote the manuscript. I.E., D.K., V.T., and L.C. revised the manuscript. All authors read and approved the final manuscript.

Data availability

The *Muscadinia* genome sequences that support the findings of this study have been deposited in GenBank (NCBI), and the accession number is PRJNA738474.

Conflict of interest

The authors declare no competing interests.

Supplementary data

Supplementary data is available at *Horticulture Research Journal* online.

References

- Olien WC. The muscadine grape: botany, viticulture, history, and current industry. *HortScience*. 1990;**25**:732–9.
- Brizicky GK. The genera of Vitaceae in the southeastern United States. *J Arnold Arbor*. 1965;**46**:48–67.
- Conner PJ. A century of muscadine grape (*Vitis rotundifolia* Michx.) breeding at the University of Georgia. *Acta Hortic*. 2009;**827**:481–4.
- Yi W, Fischer J, Akoh CC. Study of anticancer activities of muscadine grape phenolics in vitro. *J Agric Food Chem*. 2005;**53**: 8804–12.
- Brown JC, Wang J, Kasman L. et al. Activities of muscadine grape skin and quercetin against *helicobacter pylori* infection in mice. *J Appl Microbiol*. 2011;**110**:139–46.
- Merdinoglu D, Wiedeman-Merdinoglu S, Coste P. et al. Genetic analysis of downy mildew resistance derived from *Muscadinia rotundifolia*. *Acta Hortic*. 2003;**603**:451–6.
- Staudt G. Evaluation of resistance to grapevine powdery mildew (*Uncinula necator* [SCHW.] Burr., anamorph *Oidium tuckeri* BERK.) in accessions of *Vitis* species. *Vitis*. 1997;**36**:151–4.
- Olmo HP. *Vinifera rotundifolia* hybrids as wine grapes. *Am J Enol Vitic*. 1971;**22**:87.
- Firoozabady E, Olmo HP. Resistance to grape phylloxera in *Vitis vinifera* x *V. rotundifolia* grape hybrids. *Vitis*. 1982;**21**:1–4.
- Olmo HP. The potential role of (*vinifera* x *rotundifolia*) hybrids in grape variety improvement. *Experientia*. 1986;**42**:921–6.
- Pauquet J, Bouquet A, This P. et al. Establishment of a local map of AFLP markers around the powdery mildew resistance gene *Run1* in grapevine and assessment of their usefulness for marker assisted selection. *Theor Appl Genet*. 2001;**103**:1201–10.
- Barker CL, Donald T, Pauquet J. et al. Genetic and physical mapping of the grapevine powdery mildew resistance gene, *Run1*, using a bacterial artificial chromosome library. *Theor Appl Genet*. 2005;**111**:370–7.
- Feechan A, Anderson C, Torregrosa L. et al. Genetic dissection of a TIR-NB-LRR locus from the wild north American grapevine species *Muscadinia rotundifolia* identifies paralogous genes conferring resistance to major fungal and oomycete pathogens in cultivated grapevine. *Plant J*. 2013;**76**:661–74.
- Feechan A, Kocsis M, Riaz S. et al. Strategies for *RUN1* deployment using *RUN2* and *REN2* to manage grapevine powdery mildew informed by studies of race specificity. *Phytopathology*. 2015;**105**: 1104–13.
- Agurto M, Schlechter RO, Armijo G. et al. *RUN1* and *REN1* pyramiding in grapevine (*Vitis vinifera* cv. Crimson seedless) displays an improved defense response leading to enhanced resistance to powdery mildew (*Erysiphe necator*). *Front Plant Sci*. 2017;**8**: 758.
- Stringer SJ, Marshall DA, Perkins-Veazie P. Nutriceutical compound concentrations of muscadine (*Vitis rotundifolia* michx.) grape cultivars and breeding lines. *Acta Hortic*. 2009;**841**: 553–6.
- Bralley EE, Hargrove JL, Greenspan P. et al. Muscadine skin nutraceutical extract is highly anti-inflammatory in the TPA model of topical inflammation. *FASEB J*. 2006;**20**:A604–5.
- Gourineni V, Shay N, Chung S. et al. Nutraceutical values of muscadine against obesity and metabolic complications in-vivo. *FASEB J*. 2012;**26**:818.6–6.
- Hudson TS, Hartle DK, Hursting SD. et al. Inhibition of prostate cancer growth by muscadine grape skin extract and resveratrol through distinct mechanisms. *Cancer Res*. 2007;**67**: 8396.
- Mendonca P, Darwish AG, Tsoleva V. et al. The anticancer and antioxidant effects of muscadine grape extracts on racially different triple-negative breast cancer cells. *Anticancer Res*. 2019;**39**:4043.
- Riaz S, Hu R, Walker MA. A framework genetic map of *Muscadinia rotundifolia*. *Theor Appl Genet*. 2012;**125**:1195–210.
- Lewter J, Worthington ML, Clark JR. et al. High-density linkage maps and loci for berry color and flower sex in muscadine grape (*Vitis rotundifolia*). *Theor Appl Genet*. 2019;**132**:1571–85.
- Blanc S, Wiedemann-Merdinoglu S, Dumas V. et al. A reference genetic map of *Muscadinia rotundifolia* and identification of *Ren5*, a new major locus for resistance to grapevine powdery mildew. *Theor Appl Genet*. 2012;**125**:1663–75.
- Vezzulli S, Micheletti D, Riaz S. et al. A SNP transferability survey within the genus *Vitis*. *BMC Plant Biol*. 2008;**8**:128.
- Zou C, Karn A, Reisch B. et al. Haplotyping the *Vitis* collinear core genome with rhAmpSeq improves marker transferability in a diverse genus. *Nat Commun*. 2020;**11**:413.
- Simão FA, Waterhouse RM, Ioannidis P. et al. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*. 2015;**31**:3210–2.
- Cochetel N, Minio A, Massonnet M. et al. Diploid chromosome-scale assembly of the *Muscadinia rotundifolia* genome supports chromosome fusion and disease resistance gene expansion during *Vitis* and *Muscadinia* divergence. *G3 (Bethesda)*. 2021;**11**:jkab033.
- Jaillon O, Aury J-M, Noel B. et al. The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature*. 2007;**449**:463–7.
- Guo D-L, Zhao H-L, Li Q. et al. Genome-wide association study of berry-related traits in grape [*Vitis vinifera* L.] based on genotyping-by-sequencing markers. *Hortic Res*. 2019;**6**:11.
- Zhang H, Fan X, Zhang Y. et al. Identification of favorable SNP alleles and candidate genes for seedlessness in *Vitis vinifera* L. using genome-wide association mapping. *Euphytica*. 2017;**213**:136.
- Minamikawa MF, Kunihiya M, Noshita K. et al. Tracing founder haplotypes of Japanese apple varieties: application in genomic prediction and genome-wide association study. *Hortic Res*. 2021;**8**:49.
- Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res*. 2009;**19**: 1655–64.
- Zhang Y-M, Jia Z, Dunwell JM. Editorial: the applications of new multi-locus GWAS methodologies in the genetic dissection of complex traits. *Front Plant Sci*. 2019;**10**:100.
- Zhang Y-W, Tamba CL, Wen Y-J. et al. mrMLM v4.0.2: an R platform for multi-locus genome-wide association studies. *Genom Proteom Bioinform*. 2020;**18**:481–7.
- Varanasi A. et al. Glutathione S-transferase: a candidate gene for berry color in muscadine grapes (*Vitis rotundifolia*). *bioRxiv*2020.07.14.202903. 2020.
- Alfenito MR, Souer E, Goodman CD. et al. Functional complementation of anthocyanin sequestration in the vacuole by widely divergent glutathione S-transferases. *Plant Cell*. 1998;**10**: 1135–49.
- Conn S, Curtin C, Bézier A. et al. Purification, molecular cloning, and characterization of glutathione S-transferases (GSTs) from pigmented *Vitis vinifera* L. cell suspension cultures as putative anthocyanin transport proteins. *J Exp Bot*. 2008;**59**: 3621–34.
- Walker AR, Lee E, Bogs J. et al. White grapes arose through the

- mutation of two similar and adjacent regulatory genes. *Plant J.* 2007;**49**:772–85.
39. Sun L, Li S, Jiang J. et al. New quantitative trait locus (QTLs) and candidate genes associated with the grape berry color trait identified based on a high-density genetic map. *BMC Plant Biol.* 2020;**20**:302.
 40. Hobson GE. The firmness of tomato fruit in relation to polygalacturonase activity. *J Hortic Sci.* 1965;**40**:66–72.
 41. Tijskens LMM, Rodis PS, Hertog MLATM. et al. Kinetics of polygalacturonase activity and firmness of peaches during storage. *J Food Eng.* 1998;**35**:111–26.
 42. Paniagua C, Ric-Varas P, García-Gago JA. et al. Elucidating the role of polygalacturonase genes in strawberry fruit softening. *J Exp Bot.* 2020;**71**:7103–17.
 43. Costa F, Peace CP, Stella S. et al. QTL dynamics for fruit firmness and softening around an ethylene-dependent polygalacturonase gene in apple (*malus×domestica* Borkh.). *J Exp Bot.* 2010;**61**:3029–39.
 44. Fanizza G, Lamaj F, Costantini L. et al. QTL analysis for fruit yield components in table grapes (*Vitis vinifera*). *Theor Appl Genet.* 2005;**111**:658–64.
 45. Ranallo-Benavidez TR, Jaron KS, Schatz MC. GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. *Nat Commun.* 2020;**11**:1432.
 46. Kokot M, Długosz M, Deorowicz S. KMC 3: counting and manipulating k-mer statistics. *Bioinformatics.* 2017;**33**:2759–61.
 47. Hoff KJ, Lange S, Lomsadze A. et al. BRAKER1: unsupervised RNA-seq-based genome annotation with GeneMark-ET and AUGUSTUS. *Bioinformatics.* 2016;**32**:767–9.
 48. Alexander DH, Lange K. Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. *BMC Bioinformatics.* 2011;**12**:246.
 49. Porras-Hurtado L, Ruiz Y, Santos C. et al. An overview of STRUCTURE: applications, parameter settings, and supporting software. *Front Genet.* 2013;**4**:98–8.
 50. Evanno G, Regnaut S, Goudet J. Detecting the number of clusters of individuals using the software structure: a simulation study. *Mol Ecol.* 2005;**14**:2611–20.
 51. Campbell J, Sarkhosh A, Habibi F. et al. Evaluation of biochemical juice attributes and color-related traits in muscadine grape population. *Foods.* 2021;**10**:1101.
 52. Campbell J, Sarkhosh A, Habibi F. et al. Biometrics assessment of cluster- and berry-related traits of muscadine grape population. *Plan Theory.* 2021;**10**:1067.
 53. Campbell J, Gajjar P, Ismail A. et al. Determination of fertility-related traits in muscadine grape population. *Plan Theory.* 2021;**10**:1175.