

Article

Stable QTL for malate levels in ripe fruit and their transferability across *Vitis* species

Noam Reshef^{1,*}, Avinash Karn², David C. Manns³, Anna Katharine Mansfield³, Lance Cadle-Davidson⁴, Bruce Reisch² and Gavin L. Sacks¹¹Department of Food Science, Cornell University, Ithaca, NY 14853, USA²Horticulture Section, School of Integrative Plant Science, Cornell University, Geneva, NY 14456, USA³Department of Food Science, Cornell AgriTech, Geneva, NY 14456, USA⁴USDA-ARS, Grape Genetics Research Unit, Geneva, NY 14456, USA

*Corresponding author. E-mail: nr444@cornell.edu

Abstract

Malate is a major contributor to the sourness of grape berries (*Vitis* spp.) and their products, such as wine. Excessive malate at maturity, commonly observed in wild *Vitis* grapes, is detrimental to grape and wine quality and complicates the introgression of valuable disease resistance and cold hardy genes through breeding. This study investigated an interspecific *Vitis* family that exhibited strong and stable variation in malate at ripeness for five years and tested the separate contribution of accumulation, degradation, and dilution to malate concentration in ripe fruit in the last year of study. Genotyping was performed using transferable rhAmpSeq haplotype markers, based on the *Vitis* collinear core genome. Three significant QTL for ripe fruit malate on chromosomes 1, 7, and 17, accounted for over two-fold and 6.9 g/L differences, and explained 40.6% of the phenotypic variation. QTL on chromosomes 7 and 17 were stable in all and in three out of five years, respectively. Variation in pre-veraison malate was the major contributor to variation in ripe fruit malate (39%), and based on two and five years of data, respectively, their associated QTL overlapped on chromosome 7, indicating a common genetic basis. However, use of transferable markers on a closely related *Vitis* family did not yield a common QTL across families. This suggests that diverse physiological mechanisms regulate the levels of this key metabolite in the *Vitis* genus, a conclusion supported by a review of over a dozen publications from the past decade, showing malate-associated genetic loci on all 19 chromosomes.

Introduction

Malate is ubiquitous across biological systems, including plants, where it is a key intermediate in essential biochemical pathways [1]. Malate is one of the predominant organic acids that accumulates in fleshy fruits [2] and is thus a major determinant of fruit sourness, titratable acidity, and pH [3, 4]. The sensory and chemical effects of malate can extend to products of fleshy fruits, such as wines produced from cultivated wine grapes, and malate is reported to be inversely correlated with wine grape quality [5].

In the European wine grape (*Vitis vinifera*), malate typically exhibits biphasic behavior over the growing season [6]. Prior to the onset of ripening (veraison), malate is synthesized mainly from sugars [7], and accumulates at concentrations reaching 40 mg/g fresh weight in the vacuoles of immature mesocarp cells [8], or 20 g/L in the juice of berries [9, 10]. After veraison, malate decreases by over 80% on a concentration basis [8–10] through its metabolism, via respiration and alternative pathways [7, 11], and by dilution, i.e. the increase in fruit volume.

Interestingly, malate concentration and its developmental pattern varies strongly among *Vitis* species. Unlike *vinifera*, wild *Vitis* species (e.g. *riparia* and *cinerea*) show a lack of or diminished dissimilation of malate during ripening [12], and malate in the juice of the mature berries of these species often exceeds 20 g/L [12, 13]. These wild species are of considerable interest to grape breeders due to their resistance to both biotic and abiotic stresses [14–16], but excessive malate (and thus sourness) in wild *Vitis* and their progeny have so far limited their commercial use.

To expedite and improve grape breeding, considerable efforts have been made in recent years to develop appropriate markers to permit marker assisted selection (MAS) [17–20] for desirable traits, and in the past decade over a dozen quantitative trait loci (QTL)-mapping studies have been performed in grape on acidity-related traits including malate (e.g. [21–24]). However, no outstanding locus that is both stable (across years) and conserved (across studies and genetic backgrounds) has been reported. One potential complication with previous QTL studies on malate is that phenotyping has typically only focused

Received: 9 July 2021; Accepted: 21 October 2021; Published: 28 February 2022; Corrected and Typeset: 28 February 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.

on malate concentrations at ripeness. However, as previously mentioned, malate concentration at ripeness is the consequence of three distinct processes (accumulation, dissimilation, and dilution). To date, only three studies attempted to separately detect loci associated with accumulation and final concentration [22, 25, 26] and no QTL studies have accounted separately for the effects of degradation and dilution. Additionally, several of the QTL studies reported limited segregation for malate in ripe fruit as compared to the sizable range that exists across the *Vitis* genus. This may have limited loci stability, especially due to the sensitivity of malate to harvest date and the environment [6, 27]. Finally, studies utilized distinct genetic marker systems including simple sequence repeat (SSR) [24, 28, 29], single nucleotide polymorphism (SNP) [22, 24], and genotyping by sequencing (GBS) [25, 30], hampering the ability to compare between associated loci, and introducing bias due to polymorphism within primer or probe sites, limited polymorphism across species, or genomic structural variations among germplasm [18, 20].

This study reports on an interspecific mapping family that exhibited strong and stable variation in malate at ripeness for five years. In the fifth year of the study, the separate contribution of accumulation, degradation, and dilution to malate in ripe fruit was used to understand the physiological basis of the identified QTL. A novel system of genetic markers targeting the core genome and reported to be transferable across the *Vitis* genus [20] was utilized on an additional family with close genetic relatedness, to assess their use as a tool to improve QTL consistency across families. Our results were integrated with the accumulated literature to report on the most stable and conserved loci associated with malate concentration in ripe grape berries.

Results

Progeny of two families covered a wide range of ripe fruit malate and were stable across multiple years

“Horizon” × Illinois 547–1 family

Between 93 and 162 individual vines were characterized for fruit malate levels at ripeness (g/L) in the “Horizon” × Illinois 547–1 family, during five years of study (Fig. 1 A). Distribution of the data collected in 2011, 2013, and 2018 were skewed to the right. The family exhibited a mean five-fold difference in malate levels at ripeness each year, with values ranging between 3 and 21 g/L across years. These values are comparable to malate concentrations measured in *V. vinifera* cultivars [31] and wild *Vitis cinerea* genotypes [12, 13], representing the low and high-end, respectively, both present in the family’s lineage. They cover approx. 70% of the entire range measured for malate in ripe fruit within the *Vitis* genus [13].

Differences in malate levels between progenies were stable from year to year and exhibited moderate-to-strong correlation between individual years ($r=0.60$

to 0.82), comparable to Chen, et al. [21], and strong correlation with the calculated multi-year best linear unbiased estimate (BLUE) ($r=0.84$ to 0.90) (Fig. 1 B). The broad sense heritability for malate concentration (g/L) at ripeness, calculated for five years of data, was 67%, similar to values reported by Bayo-Canha, et al. [24] and slightly lower than those reported by Duchene, et al. [23].

V. rupestris B38 × “Horizon” family

The *Vitis rupestris* B38 × “Horizon” family comprised 118 progenies, all characterized for fruit malate levels (g/L) at ripeness in 2012 and 2013, and a BLUE was calculated for the two years (Fig. 1 C). The data were also skewed to the right as in “Horizon” × Illinois 547–1. Malate concentration at ripeness in the family ranged between 1.9 and 13.2 g/L in the two studied years, a narrower range of fruit malate compared to that measured for “Horizon” × Illinois 547–1. Correlation was moderate to strong ($r=0.77$) between years and strong between years and the calculated BLUE ($r=0.93$ and 0.95) and indicated good year to year stability (Fig. 1 D). Broad sense heritability for malate concentrations at ripeness was 75%.

rhAmpSeq genetic maps provide high-density mapping and transferability across *Vitis* spp.

Out of 2000 rhAmpSeq core genome markers 1944 and 1948 markers returned data for *V. rupestris* B38 × “Horizon” ($N=215$) and “Horizon” × Illinois 547–1 ($N=346$) families, respectively. Thirteen F1 vines from “Horizon” × Illinois 547–1 and seven F1 vines from *V. rupestris* B38 × “Horizon” failed QC analyses and were removed from the genetic map construction as contaminants (Supplementary Files 1 and 2). Over 55% of the markers were placed on the genetic maps in 19 LG (Supplementary Files 3 and 4), providing an average genome wide coverage of over 96%. The other 45% of the markers failed to be placed on the genetic map because they were either monomorphic or distorted. Some physical gaps in the genome coverage of the rhAmpSeq markers (e.g. chromosomes 2, 3, and 11) were evident and were suggested by Zou, et al. [20] to represent mis-assembly or structural variation that seem to be more common in wild *Vitis* genotypes compared to *vinifera*. The total length of genetic maps ranged from 1050.7 to 1316.9 cM with an average marker density of 1.10 and 0.9 cM in *V. rupestris* B38 × “Horizon” and “Horizon” × Illinois 547–1, respectively. Similarly, marker orders on the genetic maps had a high correlation ($r > 0.94$) to their physical coordinates (Fig. 2).

Stable QTL on chromosomes 7 and 17 for malate in ripe “horizon” × Illinois 547–1 fruit

QTL analyses of fruit malate concentrations at ripeness were performed separately on data from each of the five years sampled and from the BLUE calculated for these measured values. The data exhibited non-linearity and skewness in all years except for 2019 and was therefore transformed accordingly (Supplementary Table 1). A significant QTL on chromosome 7 was found in all five

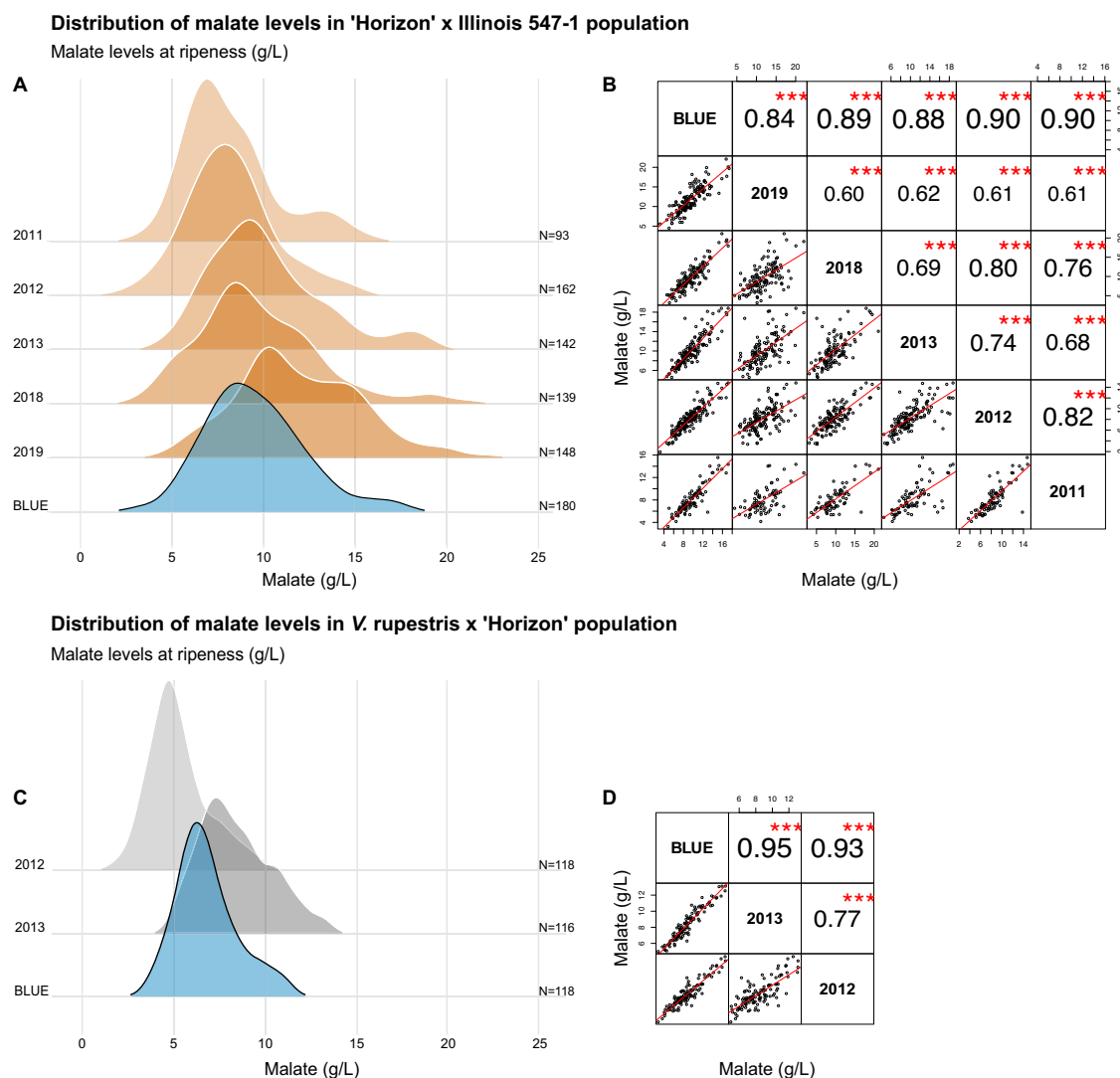


Figure 1. Malate concentrations (g/L) at ripeness in two interspecific *Vitis* families. The families were generated from a cross between “Horizon” and either Illinois 547-1 or *V. rupestris* B38. Data was collected for five years (2011–2013, 2018–2019) for “Horizon” × Illinois 547-1 and for two years for *V. rupestris* B38 × “Horizon” (2012 and 2013), and Best Linear Unbiased Estimates (BLUE) were calculated for each individual vine based on the multi-year data. A, C) Density plots representing the distribution of malate in each year and calculated multi-year BLUE for “Horizon” × Illinois 547-1 and *V. rupestris* B38 × “Horizon”, respectively. B, D) Pairwise correlations for malate concentrations (g/L) measured in each year and the calculated multi-year BLUE. The upper-right diagonal panel shows pairwise Pearson’s correlation coefficients with their corresponding p values represented by asterisks, where “****” stands for $p < 0.001$. Bottom-left panel shows pairwise scatterplots, red lines represent best linear fit.

years and the BLUE (Table 1, Fig. 3). The position of the QTL shifted slightly among years but peaked at 15.37 Mb (44.2 cM) in 2012, 2013, and the BLUE (Table 1). This marker was statistically significant also in 2011 and 2018, however, nearby markers yielded a superior association with fruit malate levels. An additional significant QTL on chromosome 17 was detected in 2012, 2013, 2018, and the BLUE, putatively overlapping with a malate QTL previously reported in a *vinifera* progeny [24]. The QTL on chromosome 7 explained a higher proportion of the phenotypic variance (mean = 17.8%) than chromosome 17 (mean = 12.6%). Two additional QTL were detected, on chromosome 10 in 2012 and on chromosome 1 in the BLUE but were not detected in other years. The significant QTL explained on average 31.3% of the total phenotypic variance measured in malate concentrations at

ripeness across all years, and over 40% of the variance in 2012 and the BLUE, where three loci were significant. For BLUE values, both the combined effect of all three QTL (40.6%), and that of the two stable loci on chromosomes 7 and 17 (34.5%) constituted most of the broad-sense heritability of malate at ripeness (67%). No significant QTL-QTL interaction was detected in any of the years. The models were additive, and average malate levels at ripeness in the different haplotypes ranged from 5.5 to 12.4 g/L ($n = 5$ and 6 , respectively) in the calculated BLUE, accounting for more than a 2-fold difference in malate concentrations (Fig. 3 E).

In the *V. rupestris* B38 × “Horizon” family, all data were skewed and therefore transformed accordingly (Supplementary Table 1). Two stable QTL were identified in both years and the BLUE on chromosomes 14 and

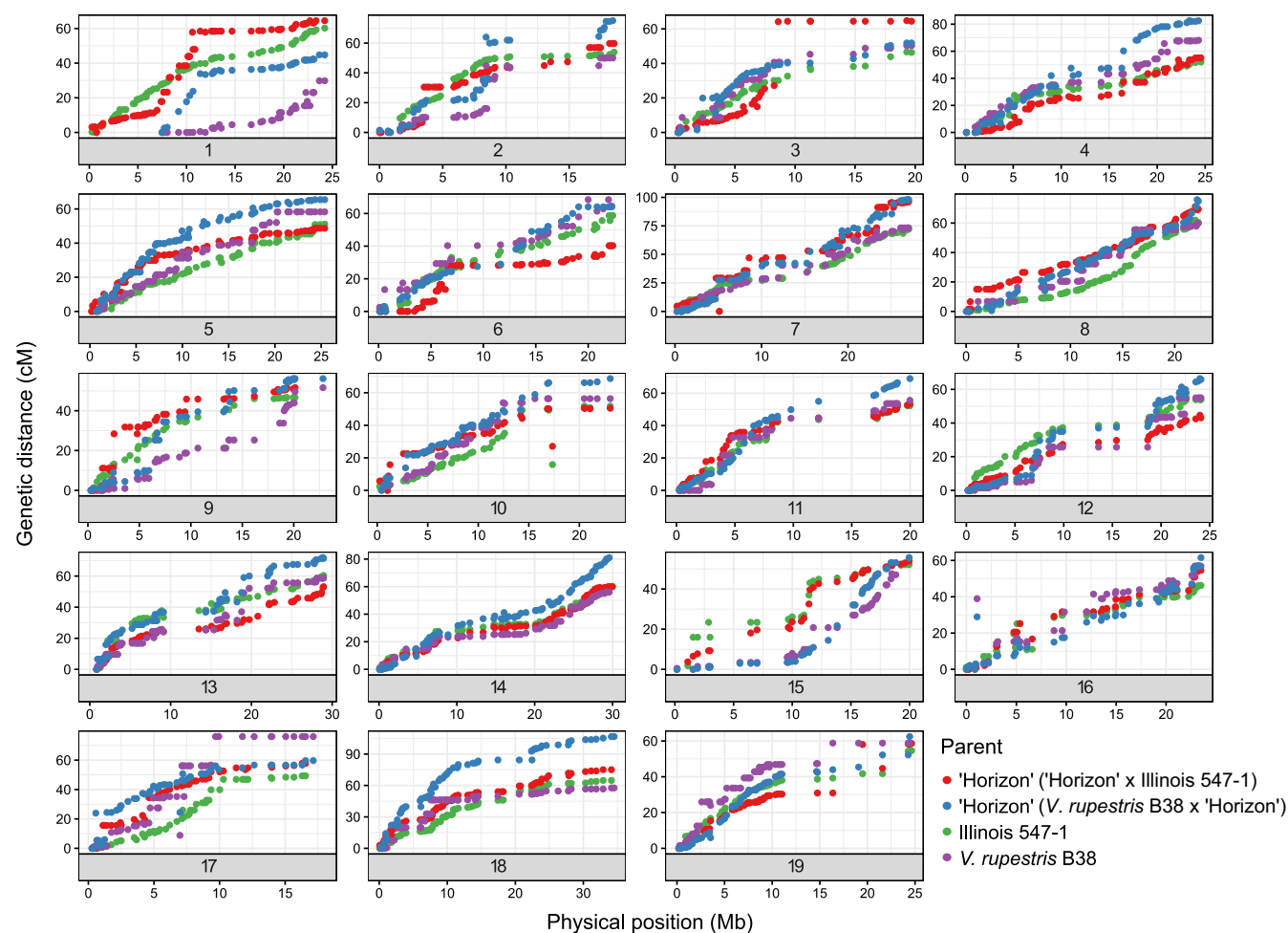


Figure 2. Relationship between the genetic and physical maps for the nineteen chromosomes of the parents of two interspecific grapevine families ("Horizon" × Illinois 547-1 and *V. rupestris* B38 × "Horizon"). Genetic distance: cM; Physical maps: Mb. The genetic distances of the markers were derived from the genetic map developed for each family and physical distances are from version 12X.2 of the PN40024 reference genome.

15 and peaked at 10.2 and 18.48 Mb, respectively, in 2013 and the BLUE (Table 1). No significant QTL-QTL interaction was detected, and the additive model of these QTL explained up to 26.9% of the variance in ripe fruit malate. Average malate levels at ripeness in the different haplotypes ranged from 5.4 to 8.8 g/L ($n=6$ and 11, respectively) in the calculated BLUE (Supplementary Fig. 1).

Malate accumulation pre-veraison is the major factor determining ripe fruit levels

A set of equations separating the contribution of accumulation ($[\text{Mal}]_{\text{green}}$), degradation ($[\text{Mal}]_{\text{degradation}}$), and dilution ($[\text{Mal}]_{\text{dilution}}$) on malate concentrations at ripeness ($[\text{Mal}]_{\text{ripe}}$), were applied to the time-resolved data obtained in 2019 from the "Horizon" × Illinois 547-1 family (materials and methods). Malate concentration at the onset of veraison was referred to as accumulation, as grape malate concentration reaches its peak at this stage prior to its typical dissimilation during ripening [10]. Dilution and degradation were calculated based on equation 2 and equation 3. Malate accumulation

ranged from 12.4 to 31.7 g/L (Fig 4 A); Degradation and dilution ranged from (−2.5) to 17.7 and (−4) to 11.9 g/L, respectively. Negative dilution values may represent a limited increase in fruit volume during ripening followed by dehydration of the fruit in the final stages of ripening, and is more often observed in wild *Vitis* species and related progeny than in *vinifera*. [12] (Supplementary Fig. 2). Negative degradation values indicate that some proportion of the progeny continues to accumulate malate during ripening. This phenomenon was previously reported in accessions of *V. cinerea*, which is part of the lineage of the Illinois 547-1 parent [12]. In a pairwise correlation analysis (Fig 4 B), final malate concentration in the fruit (i.e. 35–40 days post-veraison) were positively correlated with malate accumulation ($r=0.4$) and negatively correlated with degradation and dilution ($r=-0.21$ and -0.33 , respectively), as expected. In addition, degradation values were positively correlated with accumulation ($r=0.56$) and negatively correlated with dilution ($r=-0.45$). The former may indicate that high rates of malate degradation post-veraison are in part dependent on high pre-veraison malate concentrations.

Table 1. Summary of QTL associated with fruit malate concentration at ripeness in “Horizon” × Illinois 547–1 and *V. rupestris* B38 × “Horizon” families

“Horizon” × Illinois 547–1							H ² = 0.67
Chr	Peak marker position (bp)	LOD Interval ^b (Mb)	Year	N	LOD score ^a	%variance QTL	%variance model ^c
1	8 371 555	8.02–10.61	BLUE	164	4.44**	6.12	40.6
7	10 066 816	7.5–23.4	2011	82	4.11*	20.6	20.6
	(15374402)		(2011)	(82)	(4.07*)	(20.5)	(20.5)
	15 374 402	11.96–17.38	2012	147	7.64***	13.6	43.4
	15 374 402	10.06–19.3	2013	130	5.77***	19.9	34.7
	15 374 402	11.96–20.3	BLUE	164	8.89***	17.8	40.6
	18 952 564	11.96–20.3	2018	128	7.41***	21.8	35.5
	(15374402)		(2018)	(128)	(6.28***)	(20.2)	(29.4)
	22 781 295	11.96–23.4	2019	137	4.2*	13.2	13.2
	(15374402)		(2019)	(137)	(3.38)	(10.7)	(10.7)
10	17 335 086	2.6–17.35	2012	147	4.19*	4.57	43.4
17	7 192 642	7.13–8.93	2012	147	5.89***	12	43.4
	8 226 581	2.44–8.93	2013	130	4.54**	16.3	34.7
	8 226 581	7.29–9.35	2018	128	4.09*	12.1	35.5
	8 226 581	6.75–8.93	BLUE	164	5.52***	10.1	40.6
<i>V. rupestris</i> B38 × “Horizon”							H ² = 0.75
14	10 197 424	3.3–24.75	2013	111	4.28*	9.03	24.98
	10 197 424	3.3–26.07	BLUE	113	4.16*	8.16	25.85
	24 462 111	7.17–26.07	2012	113	4.30**	10.6	26.9
15	17 991 582	16.81–19.96	2012	113	4.36**	10.9	26.9
	18 480 191	17.31–19.96	2013	111	4.19*	10.89	24.98
	18 480 191	17.22–19.96	BLUE	113	4.78**	12.36	25.85

^aLOD p values are represented by “*”, “**”, and “***” representing $p < 0.1$, $p < 0.05$, and $p < 0.01$, respectively, based on 1000 permutation tests. ^b 1.5-LOD drop-off confidence interval, ^c Additive QTL model combining all significant loci detected in the same year. Data in parenthesis describes the association between marker 7_15 374 402 and malate at ripeness in “Horizon” × Illinois 547–1, in years in which it was not the major marker.

The individual contributions of accumulation, degradation, and dilution to the variation in malate concentrations at ripeness in the family, were assessed using a multilinear regression based on equation 1 as a perfect model. The analysis indicated that accumulation, i.e. malate concentrations at the onset of veraison, explained the greatest proportion of the observed variation in malate at ripeness, but that the other two factors were also important contributors. Based on the multilinear regression, accumulation ([Mal]_{green}) explained 39% of the observed variation, followed by degradation ([Mal]_{degradation}, 33.1%), and dilution ([Mal]_{dilution}, 27.9%). The importance of all these factors to differences in grape malate at ripeness differs from observations on *Malus* (apple), where accumulation explains most variation in malate at ripeness [32].

QTL for malate at pre-veraison and ripeness overlap on chromosome 7

Genotype-specific sampling at the onset of veraison in 2019 and 2020, and further calculation based on the 2019 data, facilitated the discovery of a significant QTL associated with malate at the onset of veraison ([Mal]_{green}) and a weak QTL for malate degradation ([Mal]_{degradation}), based on equation 3 (g/L). A BLUE for [Mal]_{green} was calculated based on 2019 and 2020 data and was highly correlated to both years (Supplementary Fig. 3). Interestingly, it was associated with a single significant QTL that peaked at the same marker as the five-year QTL for [Mal]_{ripe} on chromosome 7 (7_15 374 402), explaining 18.7% of the variance (Fig. 5 A, Supplementary Table 2).

This result aligns with the important relative contribution of pre-veraison malate to ripe fruit malate in this family and supports a previously reported QTL for malate at the pre-veraison stage on chromosome 7 [22].

The amount of malate degraded during ripening was associated with a QTL on chromosome 5 (9.1 Mb) and explained 13% of the variance (Supplementary Table 2). However, the association was statistically weak, with a logarithm of the odds (LOD) score of 3.96 and p value of 0.097 based on 1000 permutation tests. The mean degraded malate during ripening in the different haplotypes ranged between 5.4 and 9.3 g/L accounting for approx. 4 g/L difference based on this QTL. Despite this marked effect, no QTL for malate at ripeness was found on chromosome 5 in any of the years of the current study, although a QTL on chromosome 5 for malate at ripeness has been reported for a *vinifera* family [24] (Table 2). In addition, no significant QTL was associated with the effect of dilution on the concentration of malate in the fruit ([Mal]_{dilution}).

QTL for malate and soluble solids at ripeness are adjacent on chromosome 17

Fruit total soluble solids content (TSS) is a common proxy for ripeness and negatively correlated with malate during fruit maturation in *vinifera* cultivars [31]. Therefore, the effect of phenological variation across the progeny on malate levels and its associated QTL was tested by studying the correlation between malate and TSS across the progeny, and testing for colocation of QTL for TSS and malate. The correlation between the five-year BLUE data

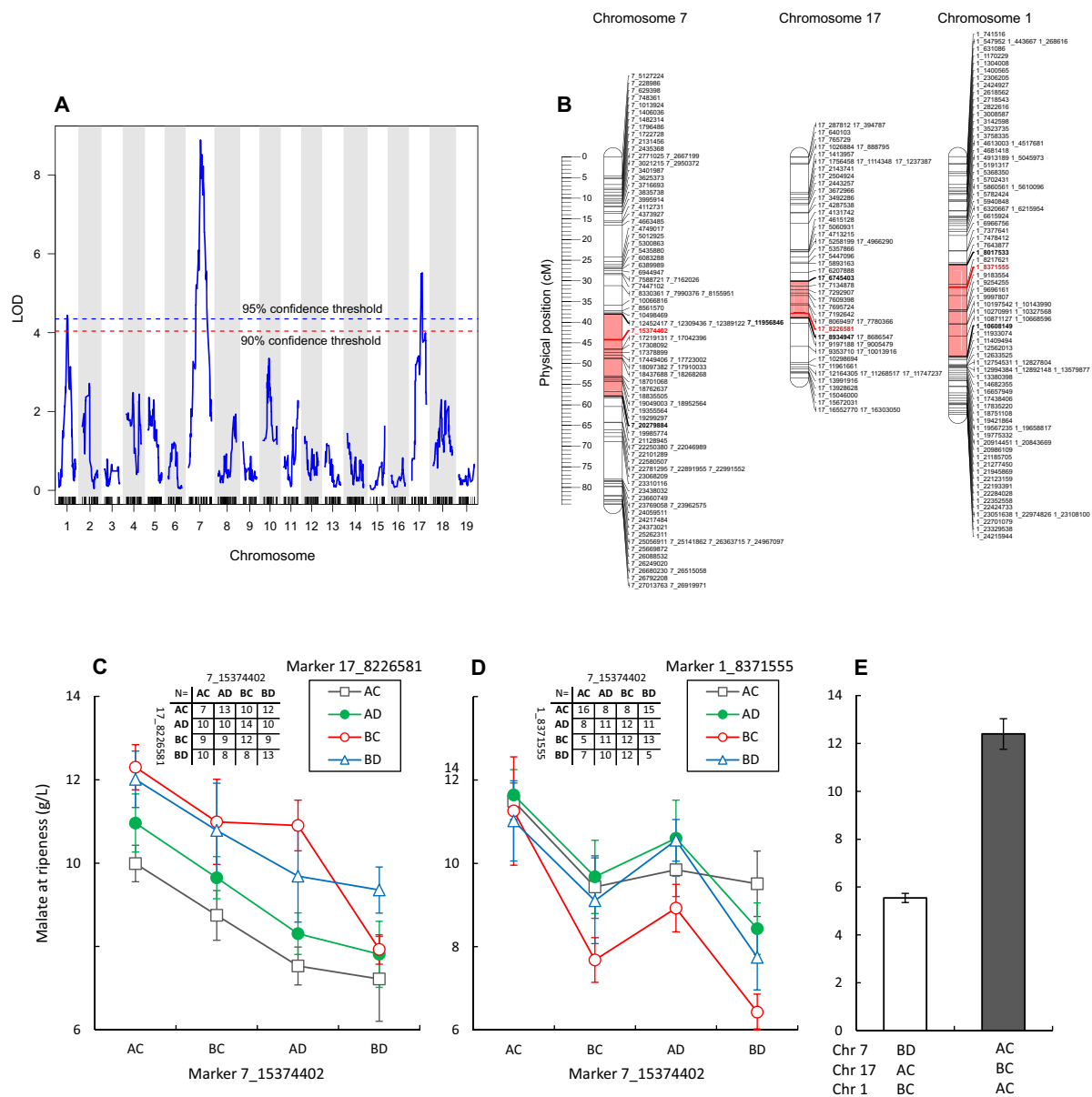


Figure 3. QTL for grape berry malate levels at ripeness. Data collected from an interspecific grapevine (*Vitis*) mapping family, generated from a cross between “Horizon” and Illinois 547–1. QTL analysis was performed using the best linear unbiased estimate (BLUE) calculated on five-years of data (2011–2013, 2018, and 2019). A) Logarithms of the odds (LOD) score for genetic markers distributed across the 19 chromosomes. B) The physical position of peak markers (bold red) and their LOD interval (interval is marked light red, edge markers are bold) on the three chromosomes found to significantly associate with the phenotype, ordered based on their relative contribution to the QTL model, from highest to lowest (7, 17, and 1, respectively). C, D) Effect of QTL marker haplotype combinations on fruit malate levels at ripeness, exhibiting the interaction between markers on chromosomes 7 and 17, and 7 and 1, in C and D, respectively. The number of genotypes per haplotype combination (N) is given in the inserted tables. E) Highest and lowest mean malate concentration at ripeness in haplotype combinations of the three QTL detected for malate in the family (for $n > 2$). $n = 6$ and 5, respectively, asterisk represents statistically significant differences between means ($p < 0.01$). Error bars are standard errors.

for TSS and malate at ripeness was weakly significant ($r = -0.167$, $p = 0.043$, Fig 5 B), indicating that the major QTL for malate, on chromosome 7, was largely unrelated to the variation in phenology. However, the multi-year QTL for TSS (Supplementary Table 3) was located adjacent to ripe fruit malate QTL on chromosome 17 (5.36 and 8.23 Mb, respectively), and their LOD intervals overlapped (Fig. 5 C). Based on a single and additive QTL models, the variance in malate explained by the QTL for TSS (9.1 and 5.4%, respectively) was lower than that explained by the QTL for malate (14.3 and 10.1%, respectively). When

using TSS as a covariate, malate QTL on chromosome 17 remained significant with a higher LOD score (5.64 compared to 5.52) but shifted 1 Mb to 7.19 Mb (35.6 cM). Based on these results we cannot clearly determine whether these adjacent QTL, on chromosome 17, represent a single locus that regulates both traits (pleiotropy) or two causal genes.

Overrepresented pathways and candidate genes

A total of 660 and 189 genes are found in the LOD intervals surrounding markers 7_15374402 and 17_8226581,

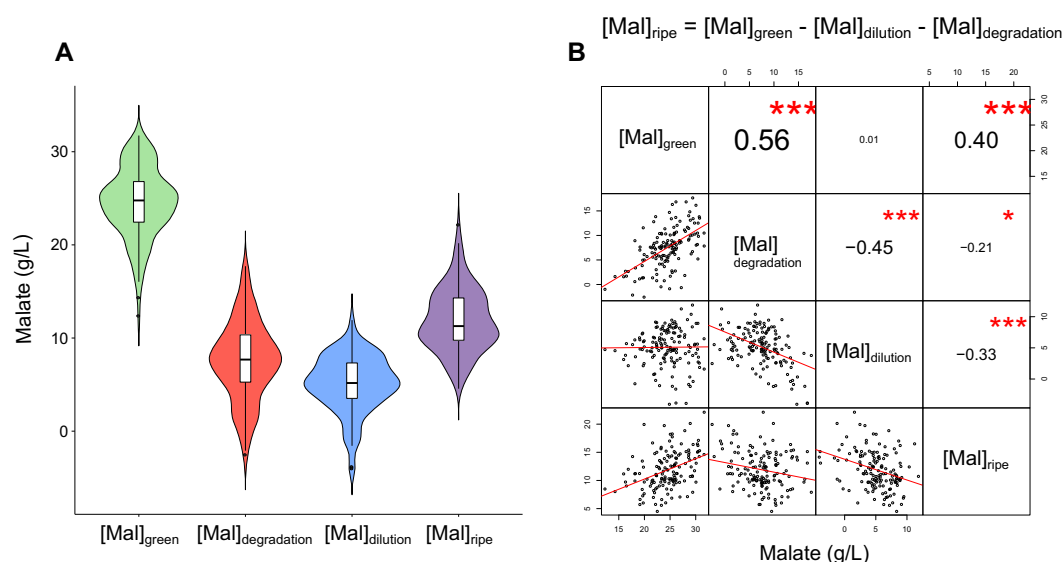


Figure 4. Measured and calculated components determining malate concentration in the fruit of an interspecific *Vitis* mapping family. The family was generated from a cross between “Horizon” and Illinois 547–1. A) Malate concentrations at the onset of veraison ($[\text{Mal}]_{\text{green}}$) and ripeness (35–40 days post-veraison, $[\text{Mal}]_{\text{ripe}}$), measured in 2019, and the calculated concentration change due to degradation and dilution during the ripening stage ($[\text{Mal}]_{\text{degradation}}$ and $[\text{Mal}]_{\text{dilution}}$, respectively). B) Pairwise correlations of measured ($[\text{Mal}]_{\text{green}}$ and $[\text{Mal}]_{\text{ripe}}$) and calculated ($[\text{Mal}]_{\text{degradation}}$ and $[\text{Mal}]_{\text{dilution}}$) values. Upper-right diagonal panel shows pairwise Pearson's correlation coefficients with their corresponding p values represented by asterisks, where “*”, “**”, and “***” stand for $p < 0.05$, $p < 0.01$ and $p < 0.001$, respectively. The bottom-left panel shows pairwise scatterplots, red lines represent best linear fit (n = 148).

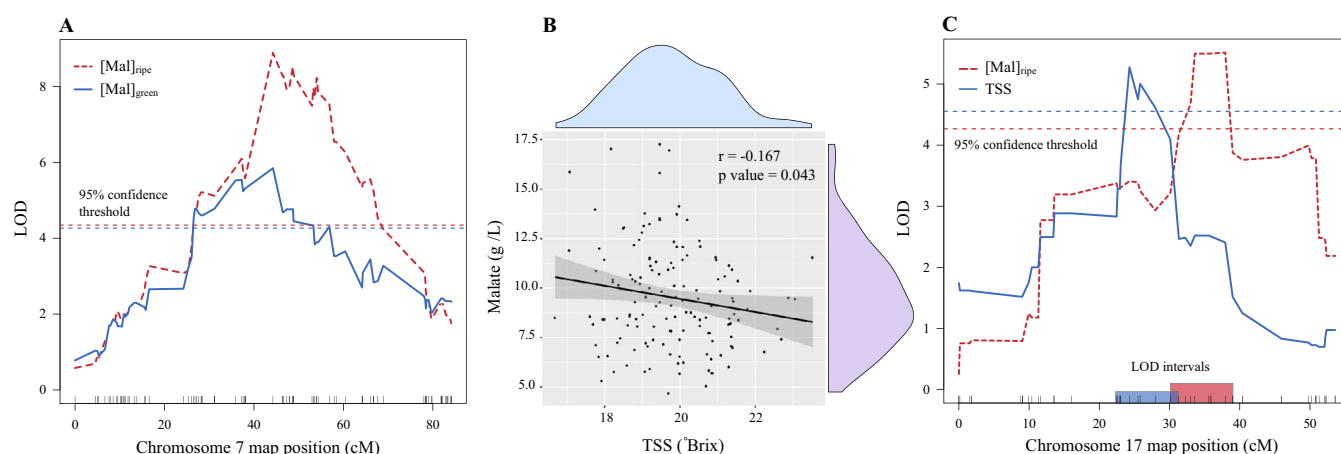


Figure 5. QTL analyses of malate and total soluble solids (TSS) in the fruit of an interspecific *Vitis* mapping family, generated from a cross between “Horizon” and Illinois 547–1. A) Logarithm of the Odds (LOD) score on chromosome 7 for a multi-year best linear unbiased estimate (BLUE) calculated for malate concentrations (g/L) at ripeness ($[\text{Mal}]_{\text{ripe}}$) and at the onset of veraison ($[\text{Mal}]_{\text{green}}$), based on five, and two years of data, respectively. N = 180 and 130 for $[\text{Mal}]_{\text{ripe}}$ and $[\text{Mal}]_{\text{green}}$, respectively. B) Scatterplot exhibiting the correlation between a five-year BLUE for TSS and malate at fruit ripeness. C) LOD scores on chromosome 17 for a multi-year BLUE calculated for malate and total soluble solids (TSS) content at ripeness ($[\text{Mal}]_{\text{ripe}}$ and TSS). N = 180 and 147 for $[\text{Mal}]_{\text{ripe}}$ and TSS, respectively. Colored rectangles on the x-axis and dashed horizontal lines, represent the 1.5–LOD support intervals and the 95% confidence threshold of the phenotype marked with the corresponding color, respectively.

respectively, stably associated with malate levels at ripeness. Pathway enrichment analysis of the LOD interval around marker 7_15 374 402 (Supplementary Fig. 4) highlighted overrepresented genes involved in ethylene and abscisic acid (ABA)-related signaling and associated transcription factors, including ethylene-responsive element binding protein (AP2/EREBP) and ABA-insensitive3/viviparous1 (ABI3/VP1), in addition to other stress-associated factors such as protein

phosphatase 2C (PP2C), zing finger type C3HC4, BHLH, and ubiquitin-mediated proteolysis genes. Additional overrepresented pathways involve mitochondrial porins and carrier-protein families (voltage-dependent anion channel 2 (VDAC2, marker position), a dicarboxylate/tricarboxylate carrier (DTC)), active transporters (ATP synthase and cation-transporting ATPases), and auxin transport. In addition, the glyoxylate and dicarboxylate, amino acid (e.g. alanine: glyoxylate aminotransferase

Table 2. Summary of significant QTL associated with malate in ripe grape – current and previous works

Chr/LG	Years found significant	Position (Mb or cM)	% Variance explained	Authors
1	1/5	8.37 Mb	6.1	Current study
	2/2	3.17–5.31 Mb	n.d.	Higginson [25]
2	1/2	6.0/15.46 Mb	n.d.	Marrano [47]
	1/6	31.9 cM	10.1	Bayo-Canha, et al. [24]
	1/3	n.d.	10.1	Schwander, et al. [48]
3	1/3	10.94–12.21 Mb	n.d.	Higginson [25]
	1/3	n.d.	9.8	Schwander, et al. [48]
4	1/2	22.97 Mb	n.d.	Marrano [47]
	1/3	n.d.	12.1	Schwander, et al. [48]
	1/3	69.6 cM	11	^b Houel, et al. [22]
5	1/6	41 cM	10.5	Bayo-Canha, et al. [24]
6	1/3	1.59 Mb	17.31	Chen, et al. [21]
	1/1	7.99 Mb	26.19	Yang, et al. [30]
7	5/5	15.37 Mb	Up to 21.8	Current study
	2/6	63.9/88.5 cM	Up to 11.9	Bayo-Canha, et al. [24]
	1/3	n.d.	8.4	Schwander, et al. [48]
	1/3	49/54 cM	Up to 15	^{b,c} Houel, et al. [22]
8	4/6	2.1–59.1 cM	Up to 20.4	Bayo-Canha, et al. [24]
	1/3	15.23–15.32 Mb	n.d.	Higginson [25]
9	1/6	74 cM	23.9	Bayo-Canha, et al. [24]
10	1/5	17.34 Mb	4.6	Current study
11	2/6	0/14.3–15 cM	Up to 14.4	Bayo-Canha, et al. [24]
	1/3	n.d.	10.8	Schwander, et al. [48]
12	1/2	3.45/7.13 Mb	n.d.	Marrano [47]
	2/3	n.d.	Up to 14.7	Schwander, et al. [48]
13	3/4	14–29 cM	up to 29.8	^a Ban, et al. [28]
14	2/2	10.2/24.54 Mb	Up to 10.64	Current study
	1/2	7.67 Mb	n.d.	Marrano [47]
	3/3	n.d.	Up to 25.5	Schwander, et al. [48]
	1/6	13.5 cM	14.7	Bayo-Canha, et al. [24]
15	2/2	17.99/18.48 Mb	Up to 12.36	Current study
	1/6	56.5 cM	11	Bayo-Canha, et al. [24]
16	1/3	n.d.	13.9	Schwander, et al. [48]
17	3/5	8.23 Mb	Up to 16.3	Current study
	4/6	0/26–30 cM	Up to 27.2	Bayo-Canha, et al. [24]
18	2/3	8.39/10.01/19.65 Mb	Up to 11.83	Chen, et al. [21]
19	1/2	6.33 Mb	n.d.	Marrano [47]

QTL associated with phenotypes other than malate concentrations are marked by the following letters: a) titratable acidity, b) malate/total acids ratio, and c) malate/tartrate ratio. Position preferentially given in Mega base pairs (Mb) where available and alternatively in centimorgan (cM). n.d. – not determined. QTL detected by the current study are highlighted in bold text.

2, and aspartate aminotransferase), carbon fixation, and glycerolipid pathways, were overrepresented. These include phosphoenolpyruvate carboxykinase (PEPCK) and aldehyde dehydrogenase (ALDH), involved in malate and pyruvate metabolism.

The LOD interval around marker 17_8 226 581 includes a cytosolic precursor of the mitochondrial malate dehydrogenase (mMDH), a major gene involved in malate metabolism and previously suggested as a candidate gene [24]. Several overrepresented pathways in this locus are associated with hormonal signaling including jasmonate and gibberellin signaling and zeatin biosynthesis, as well as MYB-related and calcium signaling (Supplementary Fig. 4). Overrepresented pathways of primary metabolism include amino acid, and starch and sucrose metabolism. Other genes, not overrepresents but associated with genes highlighted in the QTL on chromosome 7, include a mitochondrial ATP synthase gamma chain, a mitochondrial carrier (exocyst subunit), and several

protein kinases (S6-kinase, part of Target of Rapamycin (TOR) signaling pathway).

Within the LOD interval for marker 7_10 066 816, associated with pre-veraison malate accumulation in 2019, we found 960 genes. These include the mitochondrial lipoamide dehydrogenase, a component shared among pyruvate dehydrogenase (PDH), 2-oxoglutarate dehydrogenase (OGDC), and branch-chained oxo-acid dehydrogenase (BCKDH) complexes, involved in the tricarboxylate cycle (TCA); two cytosolic malate dehydrogenase (cMDH), three NADH dehydrogenases, a vesicle-associated membrane transport protein (VAMP726); genes involved in hormonal signaling, including auxin-binding and auxin-sensitive (IAA12) proteins and ABA-insensitive 3, ethylene-responsive transcription factors, and cytokinin dehydrogenase, and several kinases including a group of Leucine-Rich Repeat receptor-like kinases (LRR-RLK). In addition, as the LOD interval of the pre-veraison malate (based on the BLUE) overlaps with

the QTL for malate at ripeness (marker 7_15374402), they share all the candidate genes aforementioned for ripeness QTL on chromosome 7.

Discussion

This six-year study of a grapevine family, expressing a broad range of malate concentrations, detected two stable loci associated with fruit malate concentration at ripeness. These loci, positioned on chromosomes 7 and 17, explained 40.6% of the phenotypic variation in this trait, and were detected in all years, and in three out of five years, of ripe fruit phenotyping, respectively. Additional loci were detected for both “Horizon” × Illinois 547–1 and the genetically related *V. rupestris* B38 × “Horizon” families, and haplotype sequences associated with a low malate phenotype (expanding 1–2 Mb from identified QTL), were identified in both families (Supplementary Files 6 and 7). Use of these sequences will facilitate MAS and the introgression of valuable traits from wild accessions to future cultivars, while controlling fruit malate levels at ripeness.

Owing to its typical biphasic behavior in grape, malate levels at ripeness represent the sum of temporally and physiologically distinct accumulation and dissimulation stages [6]. Therefore, in 2019 we performed time-resolved phenotyping at the onset of ripening and on ripe fruit to establish the relative contributions of accumulation, degradation, and berry expansion (dilution) on the variability of malate at ripeness. Differences among progenies in accumulation, degradation, and dilution reached 19.3, 20.2, and 15.9 g/L of malate, respectively, emphasizing the wide range of malate behavior across *Vitis* species, and specifically within the “Horizon” × Illinois 547–1 family. Malate concentration at the onset of veraison was the major factor determining malate at ripeness, and the associated multi-year loci for both traits overlapped on chromosome 7. This indicates that the chromosome 7 locus associated with malate at ripeness is likely involved in malate synthesis and accumulation.

The synthesis of malate occurs mainly from sugars via the coupled pathways of glycolysis and the mitochondrial TCA cycle [7]. Accordingly, the QTL on chromosome 7 detected here and in a previous study [22] codes for genes involved in mitochondrial activity, including major genes in the TCA cycle (PDH, OGDC), respiratory complex (NADH dehydrogenase, ATP synthase), and transport of energy and metabolites, including malate, across the mitochondrial membrane (VDAC2, DTC) [33, 34]. Furthermore, two genes directly involved in malate synthesis and metabolism in the cytosol, cMDH and PEPCK, are found in the QTL.

Once synthesized, the transport of malate into the vacuole is energy dependent. The activity of tonoplast proton pumps maintains an electrochemical gradient that supports the facilitated diffusion of malate anion across the tonoplast [11]. Recent findings in citrus and apple highlight a complex regulatory network involved

in this process. An ethylene-responsive factor (ERF) was found to interact with the vacuolar ATPase proton pump at the protein level, to modulate citric acid accumulation in citrus [35]. In apple, genotypic differences in malate accumulation were attributed to mutations in SAUR, PP2C, apple MYB transcription factor (MYB1), and aluminum-activated malate transporter (ALMT) [36, 37]. These factors are suggested to interact and regulate the transport of malate across the tonoplast. Intriguingly, the grape pre-veraison malate QTL on chromosome 7 codes for all these regulatory components identified in apple, excluding ALMT, suggesting that their functional role in grape should be investigated. Moreover, the expression of PP2C, auxin-related genes (auxin-efflux carrier, GH3.2, IAA12), and two protein kinases, positioned in this QTL, is reported to be downregulated in grape post-veraison, at which point malate accumulation ceases [38]. Two other loci, recently reported to be associated with grape malate at veraison, were not detected in our current study. These QTL were located on chromosomes 6 and 8, which contain ALMT2 and vacuolar ATPase proton pump, respectively [26].

Several genes located in the QTL detected for ripe fruit on chromosomes 7 and 17 are associated with the onset of ripening and with the dissimulation of malate. The expression level of APETALA2, positioned in the QTL on chromosome 7, was suggested as a key marker for the onset of grape ripening [39]. In *V. vinifera*, during that stage, malate is metabolized by several potential pathways, including gluconeogenesis [7]. This pathway is suggested to involve the transport of malate into the mitochondria by DTC [33], the oxidation of malate to oxaloacetate (OAA) by mitochondrial malate dehydrogenase (mMDH), and the cytosolic conversion of OAA to phosphoenolpyruvate by PEPCK [40]. Accordingly, the expression pattern of these enzymes in *V. vinifera* grape agrees with their role in the dissimulation of malate [33, 40–42], and that of mMDH located in the QTL on chromosome 17, specifically, was found to be upregulated during ripening [42]. The co-localization of genes involved in malate biosynthesis and degradation, in a QTL that is associated with malate levels at the pre-veraison stage, may be regarded as confounding. However, as suggested by Sweetman, et al. [7] and references therein, it is likely that both metabolic trends are active to some extent at every stage of berry development, and that changes in their rates determine the net change in malate during the pre- and post-veraison stages. In support of that, while PEPCK expression was upregulated at the dissimulation (i.e. post-veraison) stage, both its transcripts and enzyme activity were found to be present in malate-accumulating berries at the pre-veraison stage [7, 40, 43].

Grape ripening (post-veraison) is typically characterized by a concomitant accumulation of soluble solids (e.g. sugars) and dissimulation of malate. As a result, variation in phenology is a possible source of variation in TSS and malate within the progeny and may affect the identification of their associated QTL. Specifically,

QTL associated with grapevine phenology, i.e. flowering time and the degree-day interval from budbreak to flowering, were previously detected on chromosomes 14 and 17, and 7 and 14, respectively [44, 45]. Therefore, the effect of phenological variation on malate levels and their associated QTL identified here, were evaluated by using TSS as a proxy for fruit ripening and by performing genotype-specific sampling in the fifth year of the study (2019). The correlation between TSS and malate and the use of TSS as a covariate in our analyses indicated that phenological variation did not have a major effect on the stable QTL identified here. This may be attributed to the extensive range of malate degradation and malate levels at ripeness, covered in this progeny. However, the multi-year QTL for TSS and malate were adjacent on chromosome 17 and their LOD intervals overlapped. This suggests that either both phenotypes are controlled by a single locus (pleiotropy) or that the QTL on chromosome 17 for malate was affected, to a certain extent, by variation in phenology. The latter hypothesis is supported by the observation that while malate QTL on chromosome 7 was still evident, neither the TSS nor malate QTL on chromosome 17 were significant in 2019, when more rigorous control of phenology was employed. Nonetheless, pleiotropy has been previously suggested to take part in the genetic control of sugars and malate in grape [18] and requires further investigation.

QTL for ripe fruit malate reported in *Vitis* spp. are scattered and unstable

The past decade has seen many investigations into the genetic regulation of acidity in grape, especially malate. Table 2 summarizes the findings of twelve studies conducted between the years 2013 and 2020, including the current one. Eight of these studies reported significant QTL for grape titratable acidity or malate at ripeness. These works included various genetic backgrounds, including a survey of *V. vinifera sylvestris* and *V. vinifera sativa* genotypes, pure *vinifera* families, and interspecific *Vitis* families. In summary, QTL associated with fruit malate or malate-dependent phenotypes (e.g. titratable acidity, malate/total acids ratio) have been reported in all 19 chromosomes of the *Vitis* genus. However, only 11 out of 37 QTL presented here (29.7%) were reported to be stable over at least two years and in at least 50% of the years tested. Other quality-related traits do not exhibit this lack of consistency. For instance, the presence of retrotransposons in *VvMYB1* has been associated with non-pigmented skin color in multiple *Vitis* species [46]. Multiple acidity-related QTL have also been detected in other fruit crops, such as apple. In that case, two QTL emerge as highly consistent across studies and genetic backgrounds [37]. Nonetheless, there is difficulty in comparing the number and consistency of identified QTL between fruit species, as these can be affected by factors independent of the physiological nature of the regulation of malate. This includes the genetic diversity

being studied, occurrence of independent selection and domestication events, tendency for out crossing, etc.

Potentially, some detected QTL could be artifacts resulting from an environmental effect, sampling bias, and other technical issues. However, considering that many of the reported QTL explain only a small amount of the variance in malate (<30%, and often much less; Table 2), it seems more likely that the instability of QTL across studies is a consequence of the large number of enzymes involved in the accumulation and dissimilation of malate [7] and the likely involvement of a complex regulatory network [36, 37].

Genetic markers are not a major source of inconsistency in malate QTL among families

Another possible explanation for the inconsistency of QTL across studies is that they vary in the technology and specific set of genetic markers used. We therefore tested whether using transferable *rhAmpSeq* markers would highlight common QTL associated with malate levels at ripeness in other interspecific families.

We observed significant QTL located at different chromosomes, even in families with strong genetic relatedness. The *V. rupestris* B38 × “Horizon” and “Horizon” × Illinois 547–1 families share “Horizon” and *V. rupestris* B38 in both ancestries. The families were also grown at the same location (Geneva, NY). Nonetheless, significant QTL for multi-year values of malate at ripeness (BLUE) differed between the families and were located on chromosomes 14 and 15 for *V. rupestris* B38 × “Horizon” compared to 7, 17, and 1 for “Horizon” × Illinois 547–1 (Fig. 6). In addition, we tested for minor effects of allelic differences in the markers identified in one family on ripe fruit malate levels in the other. In both cases, these differences were not statistically significant (Supplementary Fig. 5). One possibility is that physical gaps in the genome coverage of the *rhAmpSeq* haplotype markers have limited the identification of QTL in these regions, that may have been common to both families. However, based on the haplotypes of the peak QTL markers in the parents and grandparents of the two families (Supplementary Fig. 6), the identification of unique QTL in the two related families stems from genotypic differences between Illinois 547–1 and *V. rupestris* B38. This is evident in the case of the peak markers in the QTL on chromosomes 7, 10, 14, and 15, but not in the case of the peak markers on chromosome 17, where Illinois 547–1 and *V. rupestris* have the same haplotype. Taken together, this further suggests that the large variation in malate (~10-fold) in ripe fruit of the *Vitis* genus is a result of multiple enzymes directly involved in accumulation and dissimilation of malate and/or in a complex regulatory network. As such, the large number of QTL and inconsistency among previous studies likely represent different routes for control of malate levels in grape.

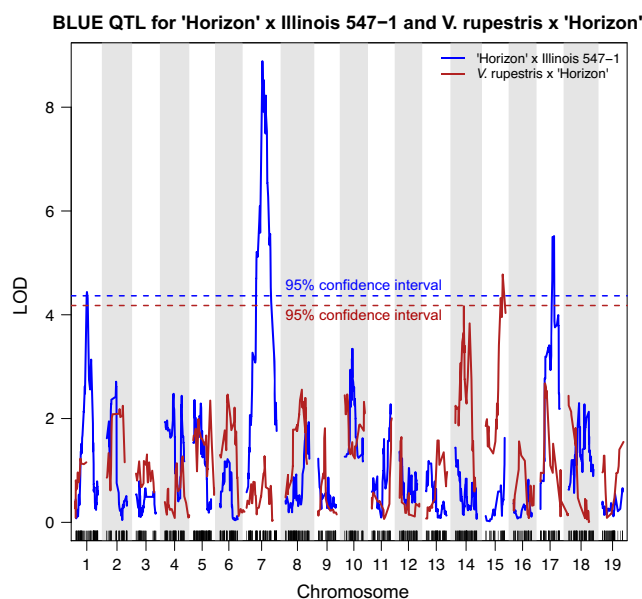


Figure 6. QTL for grape berry malate levels at ripeness in two genetically related interspecific grapevine (*Vitis*) mapping families. The families were generated from a cross between “Horizon” and Illinois 547-1 and *V. rupestris* B38 and “Horizon”. QTL analyses were performed on the best linear unbiased estimate (BLUE) calculated on five and two years of data collected in these families, respectively, using the same set of transferable genetic markers. The figure shows the Logarithms of the odds (LOD) score for genetic markers distributed across the 19 chromosomes. Blue and red dashed horizontal lines represent the 95% confidence interval calculated for “Horizon” × Illinois 547-1 and *V. rupestris* B38 × “Horizon”, respectively.

Consistent and stable malate QTL across studies and diverse genetic backgrounds

To support the efforts for breeding cultivars with desired malate concentration, there is a need to identify loci that are stable from year to year, conserved between species, and have a strong impact on malate at ripeness.

Based on current literature, QTL located on chromosomes 1, 7, 8, 11, 12, 14, 15, and 17 were found stable, i.e. detected for several years in the same study, and highlighted by at least two independent studies. Of these, QTL on chromosomes 7 and 14 were highlighted by four independent studies, yet since the genetic locations of the markers have not been specified in most cases, it is not possible to determine if these loci indeed overlap. This lists the current putatively conserved QTL associated with grape malate at ripeness. Candidate genes suggested in these loci code for metabolic enzymes involved in the synthesis and degradation of malate. Pyruvate kinase, on chromosome 8 (*VIT_08s0056g00190*) [24], participates in the synthesis of malate from sugars through glycolysis [7], isocitrate lyase and malate synthase on chromosomes 12 (*VIT_12s0059g02350*) [47] and 17 (*VIT_17s0000g01820*) [24], respectively, catalyze the synthesis of malate from isocitrate in the glyoxylate cycle, and mitochondrial malate dehydrogenase on chromosome 17 (*VIT_17s0000g06270*), suggested by Bayo-Canha, et al. [24] and the current study, catalyzes the interconversion of malate and OAA in the TCA cycle.

Nonetheless, candidate gene identification is biased and involves the selection of genes with clear functional relevance out of a list of up to thousands of genes. Further functional research is required to improve our understanding of non-metabolic regulation of malate in grape and reveal additional candidate genes.

To distill major QTL of interest, it is essential to consider the impact of these QTL on fruit malate concentration. To the best of our knowledge, the markers detected here, offer the strongest combined effect on grape malate concentration at ripeness (6.9 g/L). As stable markers with high impact, they represent potential targets for further research and breeding efforts.

This six-year study revealed two stable QTL associated with fruit malate levels at ripeness in an interspecific *Vitis* family. Combined, these QTL explain 40.6% of the phenotypic variation in the trait. Based on differences in malate levels between haplotypes in this family, utilization of the reported genetic markers in MAS is expected to accelerate the breeding of cultivars with improved palatability and provide source of variation for further fine mapping and molecular studies. Nonetheless, our analysis and the synthesis of existing literature highlights a lack of consistency and transferability of detected genetic markers, associated with malate, among *Vitis* families. This forms a major obstacle for the control of berry sourness through MAS. We hypothesize that the large number of molecular markers identified in this study and previous work, and their distribution across the grape genome, is a consequence of malate’s central role in primary metabolism. Therefore, control of grape malate and fruit sourness prompts better understanding of the intricate biological network and physiological processes involved in the regulation of malate. Studies of gene co-expression networks and alternative approaches to marker identification, such as eQTLs, may be more effective than conventional genomic QTL to provide the next breakthrough.

Materials and methods

Plant material

A complex interspecific F_1 family of 346 genotypes was generated by crossing the cultivars “Horizon” (“Seyval blanc” × “Schuyler”, with *V. vinifera*, *Vitis labrusca*, *Vitis aestivalis* and *V. rupestris* background) and breeding selection Illinois 547-1 (*V. rupestris* B38 × *V. cinerea* B9) [49, 50]. The family was first established in 1988 and further enlarged in 1996 [17], and the vineyard maintained by Cornell University in Geneva, NY. This family segregates for flower sex [50] and ca. 180 individuals have perfect flowers and therefore bear fruit. An additional interspecific mapping F_1 family was generated by crossing “Horizon” with *V. rupestris* B38 in 2008 resulting in 215 perfect-flowered vines, of which 118 were fruitful. The *V. rupestris* B38 × “Horizon” and “Horizon” × Illinois 547-1 families share “Horizon” and *V. rupestris* B38 in both

ancestries. In addition, the families are maintained at the same farm.

DNA extraction and rhAmpSeq genotyping

DNA from each vine in this study was extracted as described in Hyma, et al. [17]. rhAmpSeq sequencing and marker design were previously conducted using multiple species as described in Zou, et al. [20] and the families in this study were genotyped using the 2000 rhAmpSeq marker panel. rhAmpSeq sequencing data for the mapping family was initially analyzed using the Perl script *analyze_amplicon.pl* (https://github.com/avinaashkarn/analyze_amplicon/blob/master/analyze_amplicon.pl), and later re-analyzed with an upgraded pipeline optimized for rhAmpSeq data analysis (https://bitbucket.org/cornell_bioinformatics/amplicon). Finally, using a custom Perl script, *haplotype_to_VCF.pl* (https://github.com/avinashkarn/analyze_amplicon/blob/master/haplotype_to_VCF.pl), the four most frequent haplotype alleles for each marker in the *hapgeno* file were converted to a VCF file, where each haplotype allele of a marker was converted to a pseudo A, C, G or T allele. The raw converted VCF file was imported in TASSEL [51] (Trait Analysis by association, Evolution and Linkage) 5.2.51 software and the genotypes were imputed using the LD-kNNi imputation plugin also known as LinkImpute (v1.1.4) using the default parameters (High LD Sites = 30, Number of nearest neighbors = 10, and max distance between site to find LD = 10 000 000).

Genotype quality control

Multidimensional scaling (MDS) and Mendelian error detection were performed as quality control (QC) analyses to identify vines in both families with genotyping errors due to self-pollination, pollen contamination, or mislabeling. For MDS analysis, a genome-wide pairwise identity-by-state (IBS) distance matrix was first calculated in TASSEL using 1 – IBS, followed by the MDS analysis. The first two principal coordinates (PCOs) from the MDS results were graphically depicted in the R statistical software using the “ggplot2” package [52]. Similarly, the Mendelian error detection analysis was performed using the *mendelian* plugin in BCFtools [53], where a VCF and pedigree file from each family were used as inputs. Vines that failed these two QC analyses were removed from the genetic map construction and QTL analyses.

Genetic map construction

After imputation and QC, pedigree information and VCF files of each individual vine were used as inputs in Lep-MAP3 [54] v.0.2 (LM3) to construct the genetic maps. The following LM3 modules and steps were used to construct the genetic maps: (1) *ParentCall2* module of Lep-MAP3 was used to call parental genotypes; (2) *Filtering2* module was used to filter distorted markers based on a χ^2 (chi-squared) test and monomorphism (non-segregating); (3) *SeparateChromosomes2* module was used to split the

markers into 19 linkage groups (LG) (4) *OrderMarkers2* module was used to order the markers within each LG using 20 iterations per group, and computing parental (sex specific) and sex averaged genetic distances. The marker order of the genetic maps was evaluated for the consistency, genome organization, and structural variation by correlating with their physical coordinates on the *V. vinifera* “PN40024” reference genome (version 12X.v2). Finally, the phased output data from the *OrderMarkers2* step were converted into phased genotype data using the *map2genotypes.awk* script, which was further converted into 4way-cross format (where “1 1” = AC = 1, “1 2” = AD = 2, “2 1” = BC = 3, and “2 2” = BD = 4) for QTL analysis in “R/qtl” package [55], R statistical software [56].

Sampling

The “Horizon” × Illinois 547–1 family was sampled during six growing seasons, between 2011–2013 and 2018–2020, consisting of 93, 162, 142, 139, 148, and 130 fruiting genotypes, respectively. Sampling at the ripe fruit stage was performed in all years except for 2020. Year to year variation in the number of fruiting genotypes was largely caused by bird damage, most prominently in 2011. In 2011–2013 and 2018, ripe fruit sampling was performed on the same date across all fruiting genotypes. The sampling date for each year was determined based on fruit color, in-field measurement of TSS on a sub-sample of the progeny, and the sanitary state of the vineyard. The median TSS content at ripeness was 19.3, 21.3, 20.2, 17.7, and 19.9°Brix, in 2011, 2012, 2013, 2018, and 2019, respectively.

In 2019 and 2020, the “Horizon” × Illinois 547–1 family was sampled using a modified protocol to facilitate stage-specific analysis of fruit malate content. The date of onset of veraison was monitored across the family by manual and visual inspection of fruit softening and appearance of red color. Softening green berries were sampled from each fruiting genotype to represent the peak of malate concentration in the fruit. In addition, ripe fruits were sampled in 2019, 35–40 days following the onset of veraison, to separately calculate malate accumulation and dissimulation during fruit maturation in each genotype.

The *V. rupestris* B38 × “Horizon” family was sampled and analyzed as described for “Horizon” × Illinois 547–1, in the 2012 and 2013 growing seasons. 118 and 116 genotypes were sampled in 2012 and 2013 and the median TSS content at sampling was 22.8 and 18.6°Brix, respectively.

All samples were collected in Ziploc bags, transported in a cooler box (approx. 10°C) to the lab and stored at –80°C until processing.

Analysis of fruit composition

From each sampled genotype, 50 berries were weighed, crushed, and left for 10 min at room temperature to allow for skin tissue to macerate with the juice. The juice was filtered using a double layer of cheesecloth

into 50 mL falcon tubes and centrifuged for five min at 5000 RPM (Labnet Z400, Hermle Labortechnik, Wehingen, Germany). Total soluble solids were measured during 2011–2013 and 2018 using a digital refractometer (Sper Scientific, Ltd., model #300053, Scottsdale, AZ). In 2011–2013, malate, tartrate, and citrate were quantified on an Agilent 1260 HPLC system (Santa Clara, CA) fitted with a Bio-Rad Aminex HPX-87H ion exclusion column (Hercules, CA) fitted with a micro-guard cation-H refill cartridge (Bio-Rad cat #125–0129) and UV/VIS diode array detector monitoring at 210 nm, based on a previously described method [57]. Standards (containing 1 g/L of each organic acid) and blanks were run every 10 analyses. In 2018–2020, glucose, fructose, and malate were quantified enzymatically (GF 8363 for sugars and ML 8343 for malic acid) on an RX Monaco autoanalyzer (Randox Laboratories Ltd.). Control samples were repeated every 30 to 40 samples and values had an average coefficient of variance of 3.7, 1.8, and 2.4% for glucose, total hexose sugars, and malate, respectively.

Quantitative contributions of accumulation, dilution, and degradation to malate concentration during ripening

Data collected during 2019 from the “Horizon” × Illinois 547–1 family included berry weight and malate concentration measured at the onset of veraison (i) and at ripeness, i.e. 35–40 days post-veraison (ii). Based on these data, an equation was constructed to separately express the contribution of malate accumulation, degradation, and berry dilution to malate concentrations measured at ripeness (ii), for each genotype. The concentration of malate at ripeness (ii) was expressed as the loss in malate concentration from the end of the accumulation stage, i.e. the onset of veraison (i), to ripeness (ii), caused by degradation and dilution as follows:

$$[\text{Mal}]_{\text{ripeness}} = [\text{Mal}]_{\text{green}} - [\text{Mal}]_{\text{dilution}} - [\text{Mal}]_{\text{degradation}} \quad (1)$$

Where $[\text{Mal}]_{\text{green}}$ and $[\text{Mal}]_{\text{ripeness}}$ represent the concentration of malate in the fruit (mg/mL) at sampling points (i) and (ii), respectively.

$[\text{Mal}]_{\text{dilution}}$ represents the change in malate concentration between sampling points (i) and (ii), resulting only from the change in fruit volume:

$$[\text{Mal}]_{\text{dilution}} = \frac{\text{Mal}_{\text{green}}}{V_{\text{green}}} - \frac{\text{Mal}_{\text{green}}}{V_{\text{ripeness}}} \quad (2)$$

Where: V_{green} and V_{ripeness} represent berry volume (mL) at sampling points (i) and (ii), respectively, and $\text{Mal}_{\text{green}}$ represent malate content (mg/berry) at sampling point (i). Berry volume was calculated from the berry weight and specific gravity (S.G. 20/20), where S.G. was calculated from TSS content (°Brix) using standard sucrose conversion tables.

$[\text{Mal}]_{\text{degradation}}$ represents the change in malate concentration between sampling points (i) and (ii) due to a

change in malate content (in mg) per berry:

$$[\text{Mal}]_{\text{degradation}} = \frac{\text{Mal}_{\text{green}} - \text{Mal}_{\text{ripeness}}}{V_{\text{ripeness}}} \quad (3)$$

Where $\text{Mal}_{\text{ripeness}}$ represents malate content (mg/berry) at sampling point (ii).

QTL analysis

Sex-averaged 4way-cross genetic maps were used in QTL mapping using the “R/qtl” package [55] of R statistical software [56]. The genotype probabilities were calculated using *calc.genoprob* with *step=0* (probabilities were calculated only at the marker locations) and assumed genotyping error rate of 1.0×10^{-4} . The *scanone* function and Haley-Knott (HK) method was used for QTL scan, and 1000 permutation tests at alpha of 0.1 and 0.05 were performed to determine the significant LOD thresholds for each trait analyzed. Phenotypic data found to be non-linear or skewed were power transformed using *boxcoxnc* function in “AID” package. Further, QTL support intervals were determined by calculating 1.5 – LOD support intervals using *lodint* function, and *Fitqtl* function was used to test for qtl-qtl interactions and calculate the total phenotypic variation explained by the QTL and its effect. Similarly, to obtain the phenotypic data for the QTL analysis, best linear unbiased estimates (BLUEs) were calculated in TASSEL across years for all genotypes by treating the effects of vines as fixed.

Statistical analyses

The individual contribution of accumulation, degradation, and dilution to final malate concentrations was assessed by performing a multilinear regression using equation 1 as a perfect model. For this, the package “hier.part” [58], compatible with perfect models, was used. The relative contribution of each component was calculated by performing all regressions on the model components using *all.regs* function, followed by the function *partition*, which uses those regressions to partition the individual contribution of each component.

The broad-sense heritability (H^2) was estimated by REML (restricted maximum likelihood) variance components method [44, 59, 60] using linear mixed effects models implemented in the R package *lme4* [61] as $H^2 = \frac{\sigma_G^2}{\sigma_G^2 + \sigma_e^2}$ where σ_G^2 and σ_e^2 are the variance of the genetic and residuals variance, respectively, and n is the number of experiments over the years, respectively.

Tukey HSD test was performed to compare the means of different haplotypes, using the built in *anova* function and *HSD.test* function from package “agricolae” [62]. Density plots were visualized using R package “ggplot2” [52] and correlation matrices were based on Pearson’s correlation, performed using the base *cor.test* function and *chart.Correlation* in package “PerformanceAnalytics”

[63]. MapChart [64] was used to construct the physical map of chromosomes of interest.

Pathway enrichment analysis was performed using VitisPathway [65] with VitisNet accessions, using 1000 permutations.

Acknowledgments

This research was funded by the US Department of Agriculture (USDA)-National Institute of Food and Agriculture (NIFA) Specialty Crop Research Initiative, award No. 2017-51181-26829. We thank the Cornell University Biotechnology Resource Center for providing the necessary equipment and technical support for the sequencing of DNA samples.

Author Contribution

GLS, BR, LCD, and NR took part in planning and designing the experiment. All genetic material was bred by BR. LCD and AK performed genotyping, genetic map construction, and identification of desired alleles. GLS, AKM, DCM, and NR performed fruit phenotyping. AK and NR performed QTL and data analyses. GLS, BR, LCD, AK, and NR drafted the manuscript. All authors approved the final version.

Data availability

Genetic maps and rhAmpSeq haplotype sequences are available as supplementary files. Perl scripts used for the analysis are provided as links and can be found in the methods section.

Conflict of interest

The authors declare that they have no conflict of interest.

Supplementary data

Supplementary data is available at Horticulture Research Journal online.

References

1. Fernie AR, Martinoia EM. Jack of all trades or master of a few? *Phytochemistry*. 2009;**70**:828–32. <https://doi.org/10.1016/j.phytochem.2009.04.023>.
2. Walker, R. P. & Famiani, F. in *Horticultural Reviews* Vol. 45 (ed I. Warrington) Ch. Chapter 8. Organic Acids in Fruits: metabolism, functions and contents, 371–430 John Wiley & Sons, Inc., 2018.
3. Neta ERD, Johanningsmeier SD, McFeeters RF. The chemistry and physiology of sour taste - a review. *J Food Sci*. 2007;**72**:R33–8. <https://doi.org/10.1111/j.1750-3841.2007.00282.x>.
4. Klee HJ. Improving the flavor of fresh fruits: genomics, biochemistry, and biotechnology. *New Phytol*. 2010;**187**:44–56. <https://doi.org/10.1111/j.1469-8137.2010.03281.x>.
5. Cleary M et al. *Advances in Wine Research*. In: *Advances in Wine Research*. American Chemical Society, 2015,365–78.
6. Ruffner HP. Metabolism of tartaric and malic-acids in vitis - a review. *B Vitis*. 1982;**21**:346–58.
7. Sweetman C, Deluc LG, Cramer GR et al. Regulation of malate metabolism in grape berry and other developing fruits. *Phytochemistry*. 2009;**70**:1329–44. <https://doi.org/10.1016/j.phytochem.2009.08.006>.
8. Coombe BG. Distribution of solutes within the developing grape berry in relation to its morphology. *Am J Enol Vitic*. 1987;**38**:120–7.
9. Famiani F, Farinelli D, Pallioti A et al. Is stored malate the quantitatively most important substrate utilised by respiration and ethanolic fermentation in grape berry pericarp during ripening? *Plant Physiol Biochem*. 2014;**76**:52–7. <https://doi.org/10.1016/j.plaphy.2013.12.017>.
10. Terrier N, Romieu C. Molecular Biology & Biotechnology of the Grapevine. In: Kallioopi A, Roubelakis-Angelakis, eds. Springer: Netherlands, 2001,35–57.
11. Etienne A, Genard M, Lobit P et al. What controls fleshy fruit acidity? A review of malate and citrate accumulation in fruit cells. *J Exp Bot*. 2013;**64**:1451–69. <https://doi.org/10.1093/jxb/ert035>.
12. Burzynski-Chang EA, Brown EJ, Reshef N et al. Malate content in wild Vitis spp. demonstrates a range of Behaviors during berry maturation. *Am J Enol Vitic*. 2020;**71**:80–7. <https://doi.org/10.5344/ajev.2019.19015>.
13. Kliever WM. Concentration of tartrates, Malates, glucose and fructose in the fruits of the genus Vitis. *Am J Enol Vitic*. 1967;**18**: 87–96.
14. Reisch BI, Owens CL, Cousins PS. Fruit Breeding. In: Badenes M, Byrne D, eds. *Ch. GrapeFruit Breeding*. Springer, 2012.
15. Cadle-Davidson L, Chicoine DR, Consolue NH. Variation within and among Vitis spp. for foliar resistance to the powdery mildew pathogen Erysiphe necator. *Plant Dis*. 2011;**95**:202–11. <https://doi.org/10.1094/PDIS-02-10-0092>.
16. Staudt G, Kassemeyer HH. Evaluation of downy mildew resistance in various accessions of wild Vitis species. *Vitis*. 1995;**34**: 225–8.
17. Hyma KE, Barba P, Wang M et al. Heterozygous mapping strategy (HetMappS) for high resolution genotyping-by-sequencing markers: a case study in grapevine. *PLoS One*. 2015;**10**. <https://doi.org/10.1371/journal.pone.0134880>.
18. Yang SS, Fresnedo-Ramirez J, Wang M et al. A next-generation marker genotyping platform (AmpSeq) in heterozygous crops: a case study for marker-assisted selection in grapevine. *Horticulture Research*. 2016;**3**. <https://doi.org/10.1038/hortres.2016.2>.
19. Doligez A, Adam-Blondon AF, Cipriani G et al. An integrated SSR map of grapevine based on five mapping populations. *Theor Appl Genet*. 2006;**113**:369–82. <https://doi.org/10.1007/s00122-006-0295-1>.
20. 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**:11. <https://doi.org/10.1038/s41467-019-14280-1>.
21. Chen J, Wang N, Fang LC et al. Construction of a high-density genetic map and QTLs mapping for sugars and acids in grape berries. *BMC Plant Biol*. 2015;**15**. <https://doi.org/10.1186/s12870-015-0428-2>.
22. Houel C, Chatbanyong R, Doligez A et al. Identification of stable QTLs for vegetative and reproductive traits in the microvine (Vitis vinifera L.) using the 18 K Infinium chip. *BMC Plant Biol*. 2015;**15**. <https://doi.org/10.1186/s12870-015-0588-0>.
23. Duchene E, Dumas V, Jaegli N et al. Genetic variability of descriptors for grapevine berry acidity in Riesling, Gewurztraminer and their progeny. *Aust J Grape Wine Res*. 2014;**20**:91–9. <https://doi.org/10.1111/ajgw.12051>.
24. Bayo-Canha A, Costantini L, Fernandez-Fernandez JI et al. QTLs related to berry acidity identified in a wine grapevine popula-

- tion grown in warm weather. *Plant Mol Biol Rep.* 2019;**37**:157–69. <https://doi.org/10.1007/s11105-019-01145-6>.
25. Higginson EG. The genetic basis of acid composition in developing berries of the cultivated grapevine *Vitis vinifera* Ph.D. thesis. In: University of Adelaide, 2016.
 26. Duchene E, Dumas V, Butterlin G et al. Genetic variations of acidity in grape berries are controlled by the interplay between organic acids and potassium. *Theor Appl Genet.* 2020;**133**: 993–1008. <https://doi.org/10.1007/s00122-019-03524-9>.
 27. Kliever WM. Berry composition of vitis-vinifera cultivars as influenced by photo-temperatures and nycto-temperatures during maturation. *J Am Soc Hortic Sci.* 1973;**98**:153–9.
 28. Ban Y, Mitani N, Sato A et al. Genetic dissection of quantitative trait loci for berry traits in interspecific hybrid grape (*Vitis labruscana* x *Vitis vinifera*). *Euphytica.* 2016;**211**:295–310. <https://doi.org/10.1007/s10681-016-1737-8>.
 29. Viana AP, Riaz S, Walker MA. Genetic dissection of agronomic traits within a segregating population of breeding table grapes. *Genet Mol Res.* 2013;**12**:951–64. <https://doi.org/10.4238/2013.April.2.11>.
 30. Yang SS, Fresnedo-Ramirez J, Sun Q et al. Next generation mapping of Enological traits in an F-2 interspecific grapevine hybrid family. *PLoS One.* 2016;**11**. <https://doi.org/10.1371/journal.pone.0149560>.
 31. Waterhouse AL, Sacks GL, Jeffery DW. Understanding Wine Chemistry. In: John Wiley & Sons, Ltd, 2016.
 32. Bai Y, Dougherty L, Li M et al. A natural mutation-led truncation in one of the two aluminum-activated malate transporter-like genes at the ma locus is associated with low fruit acidity in apple. *Mol Gen Genomics.* 2012;**287**:663–78. <https://doi.org/10.1007/s00438-012-0707-7>.
 33. Regalado A, Leonardo C, Bitetto M et al. Characterization of mitochondrial dicarboxylate/tricarboxylate transporters from grape berries. *Planta.* 2013;**237**:693–703. <https://doi.org/10.1007/s00425-012-1786-8>.
 34. Homble F, Krammer EM, Prevost M. Plant VDAC: facts and speculations. *Biochimica Et Biophysica Acta-Biomembranes.* 2012;**1818**: 1486–501. <https://doi.org/10.1016/j.bbmem.2011.11.028>.
 35. Li SJ, Yin XR, Xie XL et al. The citrus transcription factor, CitERF13, regulates citric acid accumulation via a protein-protein interaction with the vacuolar proton pump, CitVHA-c4. *Sci Rep.* 2016;**6**. <https://doi.org/10.1038/srep20151>.
 36. Hu DG, Li YY, Zhang QY et al. The R2R3-MYB transcription factor MdMYB73 is involved in malate accumulation and vacuolar acidification in apple. *Plant J.* 2017;**91**:443–54. <https://doi.org/10.1111/tpj.13579>.
 37. Jia D, Shen F, Wang Y et al. Apple fruit acidity is genetically diversified by natural variations in three hierarchical epistatic genes: MdSAUR37, MdPP2CH and MdALMTII. *Plant J.* 2018;**95**: 427–43. <https://doi.org/10.1111/tpj.13957>.
 38. Fortes AM, Agudelo-Romero P, Silva MS et al. Transcript and metabolite analysis in Trincadeira cultivar reveals novel information regarding the dynamics of grape ripening. *BMC Plant Biol.* 2011;**11**:149. <https://doi.org/10.1186/1471-2229-11-149>.
 39. Fasoli M, Richter CL, Zenoni S et al. Timing and order of the molecular events marking the onset of berry ripening in grapevine. *Plant Physiol.* 2018;**178**:1187–206. <https://doi.org/10.1104/pp.18.00559>.
 40. Terrier N, Glissant D, Grimplet J et al. Isogene specific oligo arrays reveal multifaceted changes in gene expression during grape berry (*Vitis vinifera* L.) development. *Planta.* 2005;**222**:832–47. <https://doi.org/10.1007/s00425-005-0017-y>.
 41. OR E, Baybik J, Sadka A et al. Isolation of mitochondrial malate dehydrogenase and phosphoenolpyruvate carboxylase cDNA clones from grape berries and analysis of their expression pattern throughout berry development. *J Plant Physiol.* 2000;**157**: 527–34. [https://doi.org/10.1016/S0176-1617\(00\)80108-X](https://doi.org/10.1016/S0176-1617(00)80108-X).
 42. Deluc LG, Grimplet J, Wheatley MD et al. Transcriptomic and metabolite analyses of cabernet sauvignon grape berry development. *BMC Genomics.* 2007;**8**:429. <https://doi.org/10.1186/1471-2164-8-429>.
 43. Ruffner HP, Hawker JS, Hale CR. Temperature and enzymic control of malate metabolism in berries of vitis-vinifera. *Phytochemistry.* 1976;**15**:1877–80. [https://doi.org/10.1016/S0031-9422\(00\)88835-4](https://doi.org/10.1016/S0031-9422(00)88835-4).
 44. Duchene E, Butterlin G, Dumas V et al. Towards the adaptation of grapevine varieties to climate change: QTLs and candidate genes for developmental stages. *Theor Appl Genet.* 2012;**124**:623–35. <https://doi.org/10.1007/s00122-011-1734-1>.
 45. Kamal N, Schwandner A, Viehover P et al. Characterization of genes and alleles involved in the control of flowering time in grapevine. *PLoS One.* 2019;**14**:e0214703. <https://doi.org/10.1371/journal.pone.0214703>.
 46. Kobayashi S, Goto-Yamamoto N, Hirochika H. Retrotransposon-induced mutations in grape skin color. *Science.* 2004;**304**. <https://doi.org/10.1126/science.1095011>.
 47. Marrano A. Genome-Wide Patterns of Genetic Variation among Wild and Cultivated Grapevines (*V. vinifera* L.) Ph.D. thesis. Università degli Studi di Padova; 2013.
 48. Schwander F, Zyprian E, Topfer R. Xi International Conference on Grapevine Breeding and Genetics Vol. 1082. In: Li SH, Archbold D, London J, eds. *Acta Horticulturae.* 2015;315–9.
 49. Reisch B, Robinson WB, Kimball K et al. Horizon grape. *HortScience.* 1983;**18**:108–9.
 50. Dalbo MA, Ye GN, Weeden NF et al. A gene controlling sex in grapevines placed on a molecular marker-based genetic map. *Genome.* 2000;**43**:333–40. <https://doi.org/10.1139/gen-43-2-333>.
 51. Bradbury PJ, Zhang Z, Kroon DE et al. TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics.* 2007;**23**:2633–5. <https://doi.org/10.1093/bioinformatics/btm308>.
 52. Wickham H. *ggplot2: Elegant Graphics for Data Analysis.* Springer; 2009.
 53. Danecek P, Auton A, Abecasis G et al. The variant call format and VCFtools. *Bioinformatics.* 2011;**27**:2156–8. <https://doi.org/10.1093/bioinformatics/btr330>.
 54. Rastas P. Lep-MAP3: robust linkage mapping even for low-coverage whole genome sequencing data. *Bioinformatics.* 2017;**33**: 3726–32. <https://doi.org/10.1093/bioinformatics/btx494>.
 55. Broman KW, Wu H, Sen S et al. R/QTL: QTL mapping in experimental crosses. *Bioinformatics.* 2003;**19**:889–90. <https://doi.org/10.1093/bioinformatics/btg112>.
 56. R Development Core Team. R: A language and environment for statistical computing. In: *R Foundation for Statistical Computing.* Vienna, Austria. URL <https://www.R-project.org/>, 2019.
 57. Castellari M, Versari A, Spinabelli U et al. An improved HPLC method for the analysis of organic acids, carbohydrates, and alcohols in grape musts and wines. *J Liq Chromatogr Relat Technol.* 2000;**23**:2047–56. <https://doi.org/10.1081/jlc-100100472>.
 58. Nally RM, Walsh CJ. Hierarchical partitioning public-domain software. *Biodivers Conserv.* 2004;**13**:659–60. <https://doi.org/10.1023/B:BIOC.0000009515.11717.0b>.
 59. Holland JB, Nyquist WE, Cervantes-Martínez CT. *Plant Breeding Reviews.* 2002:9–112.

60. Kawamura K, Hibrand-Saint Oyant L, Crespel L et al. Quantitative trait loci for flowering time and inflorescence architecture in rose. *Theor Appl Genet.* 2011;**122**:661–75. <https://doi.org/10.1007/s00122-010-1476-5>.
61. Bates D, Maechler M, Bolker B et al. Fitting linear mixed-effects models using lme4. *J Stat Softw.* 2015;**67**:1–48. <https://doi.org/10.18637/jss.v067.i01>.
62. Mendiburu, F. agricolae: Statistical Procedures for Agricultural Research. R package version 1.2–4. <https://CRAN.R-project.org/package=agricolae> (2016).
63. Peterson, B. G. & Carl, P. PerformanceAnalytics: Econometric Tools for Performance and Risk Analysis. R package version 2.0.4. <https://CRAN.R-project.org/package=PerformanceAnalytics> 2020.
64. Voorrips RE. MapChart: software for the graphical presentation of linkage maps and QTLs. *J Hered.* 2002;**93**:77–8. <https://doi.org/10.1093/jhered/93.1.77>.
65. Osier MV. VitisPathways: gene pathway analysis for V-vinifera. *Vitis.* 2016;**55**:129–33. <https://doi.org/10.5073/vitis.2016.55.129-133>.