

Leveraging foundation models to dissect the genetic basis of cluster compactness and yield in grapevine

Received: 2 April 2025

Accepted: 3 December 2025

Published online: 06 December 2025

Cite this article as: Sharma S., Munoz J.R., Torres-Lomas E. *et al.* Leveraging foundation models to dissect the genetic basis of cluster compactness and yield in grapevine. *Sci Rep* (2025). <https://doi.org/10.1038/s41598-025-31531-y>

Sadikshya Sharma, Jose R. Munoz, Efrain Torres-Lomas, Jerry Lin, Hollywood Banayad, Yaniv Lupo, Veronica Nunez, Ana Gaspar, Dario Cantu & Luis Diaz-Garcia

We are providing an unedited version of this manuscript to give early access to its findings. Before final publication, the manuscript will undergo further editing. Please note there may be errors present which affect the content, and all legal disclaimers apply.

If this paper is publishing under a Transparent Peer Review model then Peer Review reports will publish with the final article.

Leveraging foundation models to dissect the genetic basis of cluster compactness and yield in grapevine

Sadikshya Sharma^{1**}, Jose R. Munoz Jr.^{1**}, Efrain Torres-Lomas¹, Jerry Lin¹, Hollywood Banayad¹, Yaniv Lupo¹, Veronica Nunez¹, Ana Gaspar¹, Dario Cantu¹, Luis Diaz-Garcia^{1*}

¹Department of Viticulture and Enology, University of California Davis, CA 95616, USA

^{**}Contributed equally

^{*}Corresponding author: diazgarcia@ucdavis.edu

Abstract

Grape cluster compactness is a key trait that influences fruit quality, yield, and disease susceptibility. Understanding the genetic basis of this trait is essential for optimizing vineyard management and improving grapevine cultivars. In this study, we performed quantitative trait locus (QTL) mapping to identify genomic regions associated with cluster architecture and yield components in a bi-parental population derived from *Vitis vinifera* cv. Riesling × Cabernet Sauvignon. A total of 138 full-sibling progeny were evaluated over two growing seasons at Oakville, Napa Valley, California. Traditional yield-related traits were measured, including cluster number, total cluster weight, and average cluster weight. Additionally, an image-based phenotyping pipeline leveraging the foundation model Segment Anything Model (SAM) was employed to segment individual berries, measure their size and shape, and compute cluster compactness with minimal manual intervention. Trait correlations revealed that compact clusters tended to have a higher berry count but smaller berry size, highlighting the role of compactness in modulating cluster structure. Heritability estimates varied across traits, with berry dimensions and compactness displaying moderate to high heritability, indicating strong genetic control. Two parental linkage maps were constructed using a pseudo-test cross strategy. QTL mapping identified multiple loci associated with cluster architecture and yield components, with several stable QTLs detected across both years, with marker effects ranging from 7.6% to 22.1%. Notably, a QTL for cluster compactness was found in both seasons on chromosome 1 in Cabernet Sauvignon. Other stable QTLs were associated with berry size (chromosomes 6 and 17) and berry count (chromosome 5 in Cabernet Sauvignon and chromosome 7 in Riesling). Additional QTLs were detected in a single year, reflecting the influence of environmental variation. Our findings provide valuable insights into the application of foundation models requiring no prior training and minimal intervention for high-quality segmentation and enhance our understanding of the genetic architecture of cluster compactness and yield traits. The genomic regions

identified in this study offer promising targets for breeding programs aimed at improving grape quality and disease resistance.

Keywords: *Vitis vinifera*, QTL, image analysis, Segment Anything Model, cluster compactness, yield components

Introduction

Grape cluster architecture and compactness are two interrelated traits essential for vineyard management and the genetic improvement of grapes, as they influence yield, fruit quality, and susceptibility to pests and diseases [1-4]. Cluster architecture determines how the berries are arranged within a cluster and the distribution of free space [2,5]. Cluster compactness, on the other hand, is defined as the ratio of the volume of cluster components (such as the berries and the rachis) to the overall volume of the cluster [2,5]. Understanding the implications of cluster compactness is crucial, as its physical configuration can significantly impact berry development and the microenvironment within the cluster. When clusters are compact, berries are in close contact with one another, which can alter their physical and physiological properties. For instance, tight contact can inhibit the proper development of the waxy cuticle, compromising its function as a protective barrier against pathogens [6-8]. Furthermore, the inner portion of compact clusters has restricted airflow, resulting in higher humidity and temperature, creating a more favorable environment for pathogens [3,7]. In addition to reduced airflow, there is also reduced solar radiation reaching the inner berries of compact clusters, which affects berry ripening and composition [7].

Several studies have sought to identify the genetic factors underlying cluster architecture and compactness using a range of phenotyping methods, including both qualitative and quantitative approaches. One commonly used qualitative method relies on a set of descriptors that categorize clusters based on visual characteristics [3]. The most widely adopted system is the set of descriptors developed by the International Organization of Vine and Wine (OIV), which includes standardized criteria for evaluating various traits of the grapevine, including cluster morphology. For example, OIV descriptor 204 assesses cluster density, classifying clusters into five categories ranging from very loose, where berries are clearly separated with many visible pedicels, to very dense, where berries are visibly deformed due to compression. Additional descriptors, such as OIV 208 (cluster shape) and OIV 209 (number of cluster wings), can be used in combination to characterize overall cluster architecture. While OIV descriptors offer a relatively fast and low-cost means of phenotyping, their accuracy is often limited by the evaluator's subjectivity [3]. When multiple evaluators are involved, differences in interpretation can introduce inconsistency and increase error. In contrast, quantitative phenotyping is generally preferred for statistical and genetic analyses, as it provides

more precise, reproducible data that better supports the dissection of complex traits [2,9].

The analysis of 2D and 3D images of grapevine clusters offers a powerful approach for measuring cluster architecture and compactness. Techniques in this field range from traditional image thresholding for object identification to advanced machine learning models, some specifically trained on grapevine data, others based on more generalizable foundational models with little or no crop-specific training. These image-based methods enable the extraction of high-dimensional phenotypic traits, enhance the accuracy of characterizing complex traits like cluster architecture, and offer excellent scalability for large datasets. Examples of cluster imaging and image analysis include the study by Cubero et al. (2015), which developed a technique for measuring cluster compactness that was highly consistent with the OIV descriptor 204 [10]. Their approach utilized semi-automated image segmentation to remove the background from the images and segment the clusters into berries and rachis. These segmented images were subsequently used to collect morphological measurements and create models for predicting cluster compactness. Similarly, Underhill et al. (2020) employed high-throughput imaging to quantify cluster compactness and identified key quantitative trait loci (QTLs) related to cluster structure [3]. More recently, Torres-Lomas et al. (2024) demonstrated the application of the Segment Anything Model (SAM) for grape cluster image segmentation, enabling automated object identification without additional training [2]. Their study applied SAM to over 3,500 cluster images, generating more than 150,000 berry masks with spatial coordinates, which allowed comprehensive cluster architecture and compactness analyses. These examples illustrate the effectiveness of image-based phenotyping in providing precise, reproducible data that can enhance genetic analysis and breeding strategies for grapevine traits.

Cluster architecture and compactness are complex traits shaped by a combination of environmental conditions, genetic factors, and vineyard management practices [7]. Numerous studies have aimed to uncover their genetic basis, identifying a wide array of QTLs distributed across the grapevine genome. For example, Correa et al. (2014) identified several QTLs associated with rachis architecture, including a major locus for rachis length located near the microsatellite marker VMC2D9 on chromosome 9, using an F1 biparental population derived from a Ruby Seedless × Sultanina cross [11]. Similarly, Richter et al. (2019) mapped over two dozen QTLs related to cluster architecture traits in an F1 population from a GF.GA-47-42 × Villard Blanc cross [5]. This study identified QTLs for total berry volume (chromosome 1), cluster weight (chromosomes 2, 10, 12, 17, 18), compactness (chromosomes 1, 2, 15, 17), mean berry volume (chromosomes 3, 12, 17), and berry number (chromosomes 10, 17, 18). It also highlighted significant co-localization of QTLs for berry number and cluster weight on chromosomes 17 and 18. In another study, Underhill et al. (2020) combined manual and image-based phenotyping to identify QTLs associated with cluster compactness,

berry weight, and rachis length in a biparental population derived from a cross between MN1264 × MN1246, revealing key loci across multiple chromosomes [3]. Specifically, they identified QTLs for compactness (chromosomes 7, 12, 17, 18), berry weight (chromosomes 11, 17), cluster weight (chromosomes 4, 9, 16), and cluster area (chromosomes 9, 16). Tello et al. (2016), using a genome-wide association approach that included over 114 different varieties maintained in the Grapevine Collection of the Instituto de Ciencias de la Vid y del Vino, found QTLs for compactness across chromosomes 5, 11, 12, and 18, as well as for berry number on chromosomes 2, 5, 8, 11, and 12 [12]. More recently, Rist et al. (2022) employed 3D-based phenotyping in four different biparental populations, identifying multiple QTLs for berry number (chromosomes 4, 8, 10, 17, 18), berry diameter (chromosomes 6, 12, 17, 18), berry volume (chromosomes 6, 8, 12, 17, 18), and total berry volume (chromosomes 7, 10, 17, 18). Notably, chromosome 17 consistently harbored co-located QTLs for berry size traits—berry diameter, berry volume, and berry weight—suggesting a strong genetic control region for fruit morphology [13]. Together, these studies highlight the polygenic and highly complex genetic architecture underlying grapevine cluster traits, emphasizing the need for advanced phenotyping and mapping strategies to fully dissect their genetic control.

In this study, we performed QTL mapping using an F1 mapping population derived from a cross between two pure *Vitis vinifera* cultivars with contrasting cluster architectures: Riesling and Cabernet Sauvignon. Our goal was to identify genomic regions associated with yield components and cluster architecture by leveraging the Segment Anything Model, or SAM. Using SAM, we processed over 7,000 2D cluster images, generating nearly 330,000 individual berry masks, enabling precise measurements of berry morphology, spatial distribution, and cluster compactness, all without requiring additional model training. This methodology allowed us to successfully identify multiple QTLs linked to cluster architecture, compactness, and yield, with several QTLs consistently detected across two growing seasons. These findings offer valuable insights into the genetic determinants of cluster architecture and compactness, paving the way for breeding strategies aimed at enhancing fruit quality and reducing disease susceptibility.

Results

Digital phenotyping for cluster compactness

To characterize grape cluster architecture and compactness, we phenotyped 414 vines from 138 progenies of a Riesling × Cabernet Sauvignon cross over two consecutive growing seasons, 2023 and 2024. This mapping population was grown in Oakville, Napa Valley, California, a premier winegrowing region in the United States.

We collected two sets of phenotypic data: traditional yield-related traits, including the number of clusters per vine, total cluster weight per vine, and average cluster weight; and digitally phenotyped traits, including berry number, projected area, length, and width. From these, we calculated the compactness index, which is defined as the ratio of the sum of individual berry projected areas to the convex hull surrounding the entire cluster.

Across both seasons, a total of 7,483 cluster images were acquired, with each cluster imaged from two angles to maximize berry visibility. Using the Segment Anything Model along with a previously developed algorithm for berry segmentation [2], we generated 329,469 berry masks (an example of a processed cluster is shown in Figures 1A and 1B). While a single image underestimates the true berry count due to occlusion, as demonstrated in our previous work and this study, a linear regression-based correction factor was applied to improve the accuracy of berry number estimates (Figure 1C). To quantify compactness, we calculated the ratio between the sum of all berry masks (i.e., total projected berry area) and the convex hull, defined as the smallest convex polygon encompassing all berry masks (Figure 1D). To visually validate this approach, we ranked all clusters by compactness index and selected the 30 least compact (top) and 30 most compact (bottom) clusters for comparison (Figure 1E). The gradient from less to more compact clusters is evident, with loosely arranged berries in clusters at the top of the panel and densely packed berries in those at the bottom.

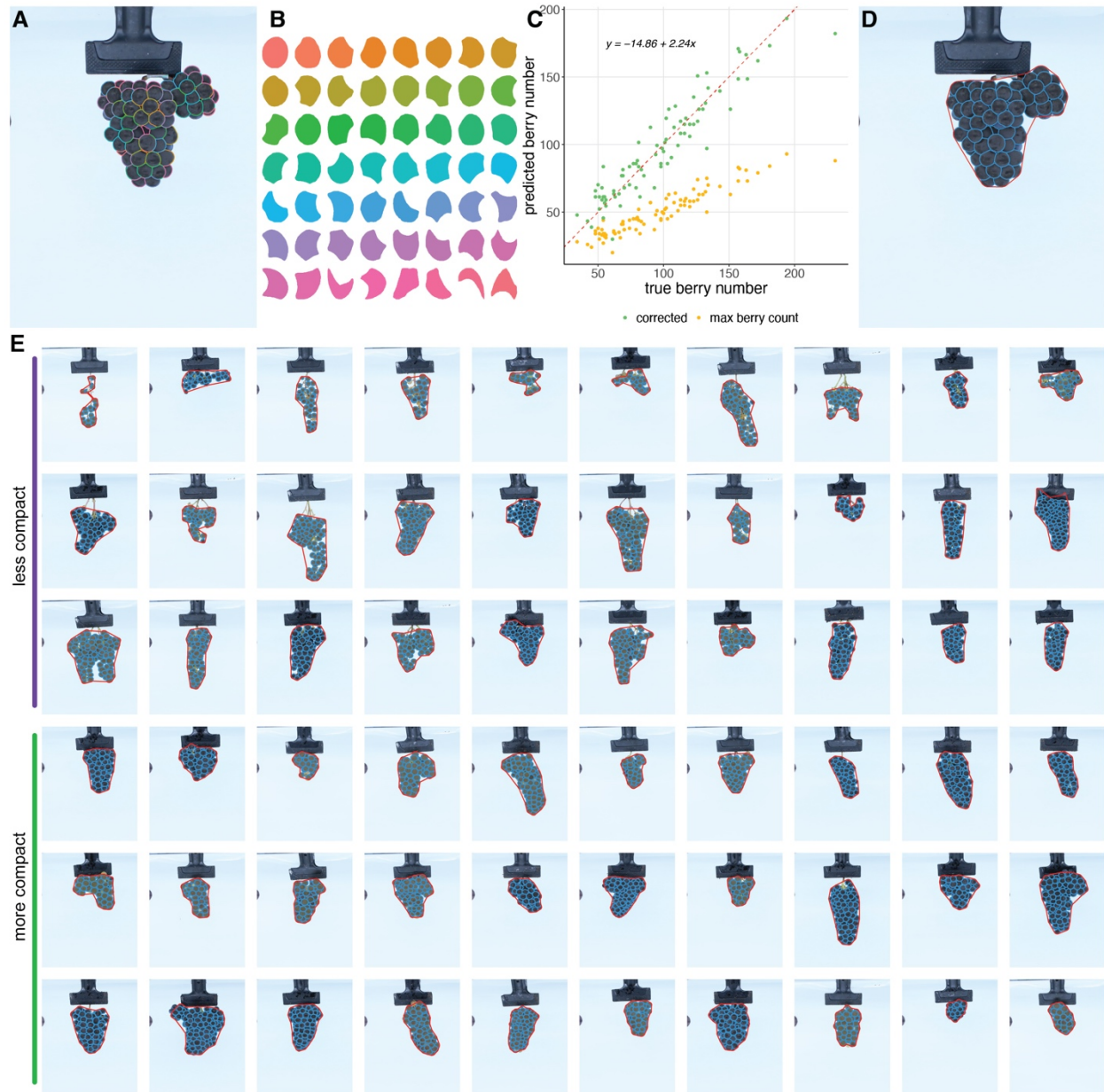


Figure 1. Identification of berry segments using the Segment Anything Model (SAM). (A) Example of individual berry identification from a grape cluster. (B) Berry masks generated from the cluster image in panel A. (C) Correlation between actual and predicted berry counts from SAM, using 84 clusters. Green points represent corrected counts, adjusted using a linear model based on the angle with the highest berry count. (D) Cluster compactness is calculated as the ratio of the total berry mask area to the area of the convex hull (outlined in red). (E) Examples of clusters with low and high compactness indices.

Variation in yield and cluster traits across years

Best linear unbiased predictors (BLUPs) for both yield components and image-derived (digital) traits displayed continuous distributions with strong year-to-year consistency ($R^2 = 0.43$ to 0.65 ; Figure 2A). Furthermore, transgressive segregation was observed across all traits. The two parental cultivars displayed expected phenotypic patterns: Riesling produced smaller, more compact clusters with fewer and smaller berries, while Cabernet Sauvignon produced larger, looser clusters with more and larger berries. All raw data distributions were normally distributed, except for cluster weight, which showed slight skewness toward zero (Supplementary Figure 1).

The number of clusters per vine increased from a mean of 31.58 ± 5.8 (SD) in 2023 to 45.62 ± 7.14 in 2024. Riesling and Cabernet Sauvignon followed this trend, increasing from 38.36 to 60.92, and from 42.98 to 54.09, respectively. In contrast, average cluster weight decreased slightly, from $74.94 \text{ g} \pm 25.98$ in 2023 to $65.05 \text{ g} \pm 16.93$ in 2024. Cabernet Sauvignon consistently produced heavier clusters than Riesling, with weights of 115.67 g versus 76.67 g in 2023, and 88.42 g versus 59.58 g in 2024. Total cluster weight per vine increased overall from $2.42 \text{ kg} \pm 0.86$ in 2023 to $3.05 \text{ kg} \pm 0.77$ in 2024. However, the parental cultivars exhibited different trends: Riesling increased from 3.00 to 3.73 kg, while Cabernet Sauvignon declined slightly from 5.11 to 4.79 kg. These changes are likely attributable to differences in environmental conditions and slight variations in management practices, such as pruning intensity, resulting from different personnel performing these tasks.

Among the image-derived traits, berry number per cluster averaged 46.02 ± 8.78 in 2023 and 43.65 ± 7.49 in 2024. Riesling produced fewer berries (37.2 in 2023, 34.37 in 2024), while Cabernet Sauvignon had higher counts (58.56 and 55.66, respectively). Berry projected area showed a slight decline from $222.41 \text{ mm}^2 \pm 24.53$ in 2023 to $215.99 \text{ mm}^2 \pm 18.9$ in 2024. Riesling maintained consistent berry sizes ($\sim 248 \text{ mm}^2$), while Cabernet Sauvignon showed a marked decrease from 257.75 mm^2 in 2023 to 197.60 mm^2 in 2024. Similar patterns were observed for berry length and width. Berry length averaged $17.08 \text{ mm} \pm 1.00$ in 2023 and $16.89 \text{ mm} \pm 0.77$ in 2024; berry width decreased from $17.02 \text{ mm} \pm 0.99$ to $16.84 \text{ mm} \pm 0.79$. Riesling remained relatively stable for these traits, while Cabernet Sauvignon showed mild reductions. Cluster compactness remained stable across years, with mean values of 0.18 ± 0.009 in both seasons. However, the parental lines were distinct: Riesling consistently exhibited greater compactness (0.19) than Cabernet Sauvignon (0.17), aligning with known phenotypic differences between these cultivars.

Trait correlations and implications for cluster architecture

As expected, berry size-related traits were highly correlated (Figure 2B). For instance, berry projected area and berry length showed a very strong correlation (Pearson's $r = 0.99$) in both years. Similarly, berry projected area was highly correlated with berry width (Pearson's $r > 0.98$) in both years. Berry count exhibited a strong correlation with cluster weight (Pearson's $r = 0.81$ in 2023 and 0.72 in 2024) and total cluster weight (Pearson's $r = 0.76$ in 2023 and 0.62 in 2024). Additionally, the number of clusters per vine was positively correlated with total cluster weight (Pearson's r is 0.60 in 2023 and 0.65 in 2024). A negative correlation was observed between cluster aspect ratio and the compactness index in 2023 (Pearson's $r = -0.22$), suggesting that their aspect ratio decreases as clusters become more compact, meaning they become wider and less elongated. Conversely, when clusters are more elongated, with a higher aspect ratio, berries are more loosely packed, reducing compactness. Furthermore, the compactness index for both years was negatively correlated with the number of clusters per vine (Pearson's $r = -0.15$ in 2023 and -0.19 in 2024). This might suggest that when the number of clusters per vine decreases, individual clusters receive more resources, such as water, nutrients, and carbohydrates, leading to an increased number of berries per cluster and tighter packing, which results in a higher compactness index.

Heritability of yield and digital traits

Heritability estimates varied across traits and years (Figure 2C). Berry number had a moderate heritability of 0.52 in 2023 and 0.51 in 2024. The berry's projected area had higher heritability (0.73 in 2023, dropping to 0.59 in 2024), reflecting greater environmental influence in the second year. Berry length and width showed similar patterns, with heritability decreasing from 0.76 to 0.58 for length and from 0.75 to 0.61 for width. Cluster compactness showed moderate heritability (0.57 in 2023, 0.56 in 2024), while the number of clusters per vine had lower estimates (0.39 in 2023, 0.37 in 2024). Cluster weight showed a decline in heritability from 0.7 in 2023 to 0.49 in 2024, and total cluster weight followed a similar trend (0.5 in 2023, 0.38 in 2024), indicating increased environmental influence in 2024.

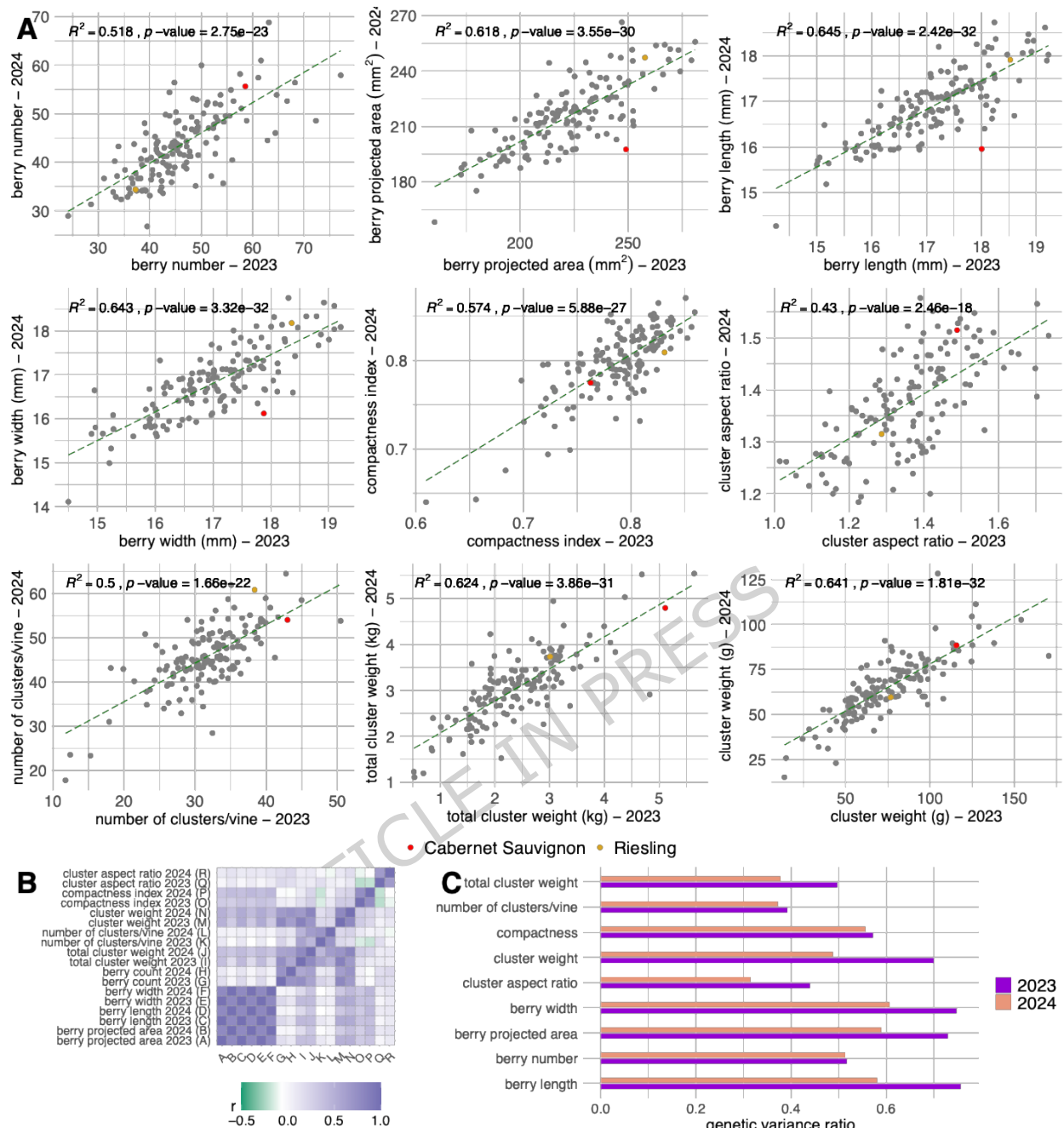


Figure 2. Phenotypic variability, trait correlations, and heritability. (A) Comparison of the 2023 and 2024 seasons for yield-related and digital traits based on the general mean + BLUPs. The dashed green line represents the linear regression between the two seasons, while red and orange points correspond to the parental cultivars, Cabernet Sauvignon and Riesling, respectively. **(B)** Correlations between BLUPs for all traits were analyzed separately for each season. **(C)** Proportion of genetic variance relative to the total variance.

QTL mapping for cluster compactness and yield components

Two genetic maps were constructed using a pseudo-test cross linkage mapping strategy [14]. The Cabernet Sauvignon parental map comprised 3,030 markers distributed across 982 bins (unique genetic positions), while the Riesling parental map contained 3,360 markers across 889 bins. Overall, physical and genetic distances showed strong collinearity (Spearman's $r > 0.98$), although some regions showed local discrepancies (Figure 3). After SNP calling, correction and imputation of marker data were applied to mitigate the impact of potentially incorrect genotype calls. Notably, none of the QTLs identified in this study were located within genomic regions showing inconsistencies between genetic and physical marker positions.

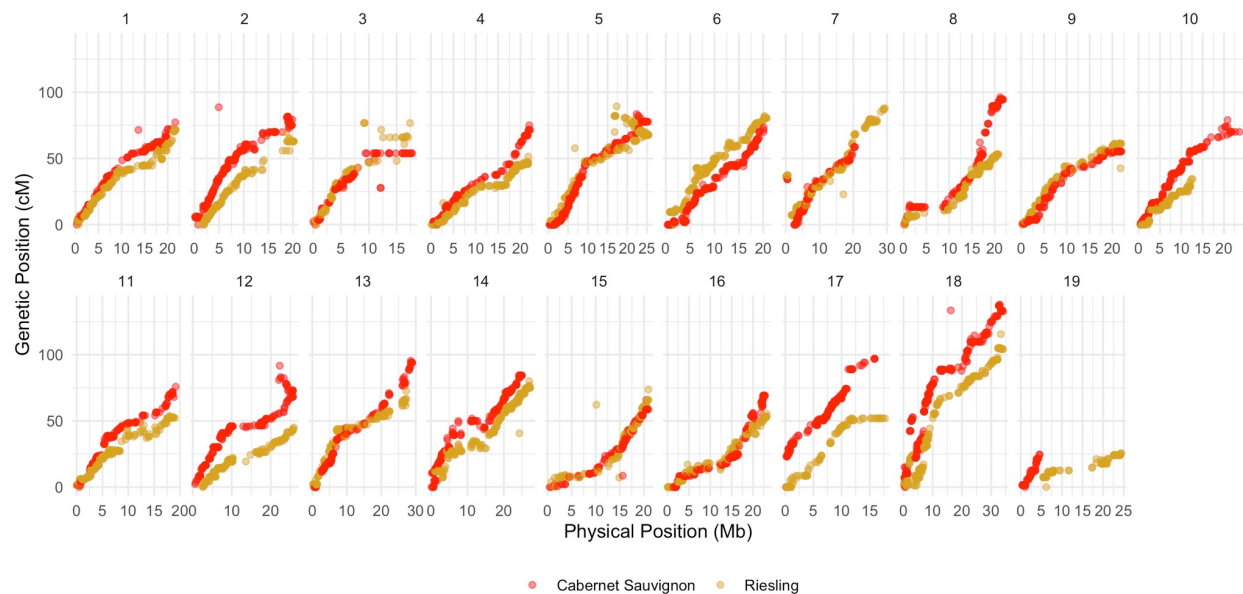


Figure 3. Relationship between genetic and physical positions for each chromosome in the parental maps.

QTL mapping identified ten consistent QTLs in both growing seasons (Figure 4). Nine QTLs were found on the Cabernet Sauvignon parental map with markers segregating as AB×AA or Im×II, while one QTL was detected on the Riesling parental map with markers segregating as AA×AB or nn×np. For each trait, significance thresholds were established using a permutation test, with 1,000 random permutations of genotype and phenotype labels. The genome-wide threshold was set at the 95th percentile of the maximum test statistic distribution. These QTLs were associated with berry area, compactness, berry count, length, and width. Specifically, QTLs for berry length, width, and area were consistently detected on chromosomes 6 and 17 in Cabernet Sauvignon, with marker effects ranging from 7.6% to 17.8%. The high correlation between these traits may be explained by the co-localization of their QTLs. A

QTL for compactness was identified on chromosome 1 with marker effects of 10.3% in 2023 and 12.3% in 2024. Three QTLs on chromosomes 4, 5, and 7 were found for berry count in Cabernet Sauvignon and Riesling, respectively, with marker effects between 9.2% and 22.1%.

An additional 26 QTLs were identified using data from a single year, with marker effects of up to 22.9% (for number of clusters in the Riesling genetic map). The effect sizes for all QTLs, including those detected in one or both years, are summarized in Table 1.

We additionally conducted QTL mapping using BLUPs derived from a multi-year model. Notably, except for one QTL, the multi-year analysis recovered all QTLs identified in both years of the original year-by-year analysis. This outcome was expected, given the moderate to high correlations between years (Figure 2A). The QTL mapping results from multi-year analysis is presented in the Supplementary Table 1.

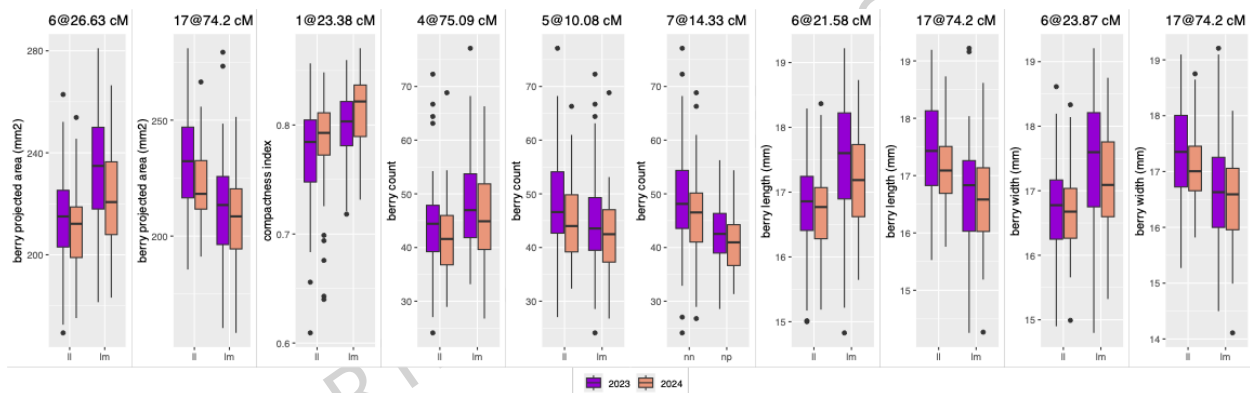


Figure 4. Marker effects for QTLs identified in two growing seasons. QTL mapping was conducted independently for both parents using a backcross model (pseudo-testcross QTL mapping). Therefore, markers are either (i) $lm \times ll$, where Cabernet Sauvignon is heterozygous and Riesling is homozygous, or (ii) $nn \times np$, where Cabernet Sauvignon is homozygous, and Riesling is heterozygous.

Table 1. QTLs for yield-related traits and cluster compactness identified in a Riesling $\times$ Cabernet Sauvignon population using a pseudo-test cross linkage mapping strategy. QTLs detected in both years, based on the 1.5 LOD interval, are displayed in the top part of the table. Physical positions are based on the Cabernet Sauvignon cl.08 v1.1 haplotype 1 genome. Model variance refers to the percentage of phenotypic variance explained by the QTL model. 1.5 LOD interval limits represent the flanking genetic positions (in cM) around the QTL peak.

Year	Parent	Trait	Chr	Genetic position (cM)	Physical position (bp)	Marker variance (%)	LOD score	Model variance (%)	1.5 LOD interval (cM)
Consistent QTL discovered in two years									
2023	Cab Sauv	berry area	6	26.6	8,170,801	16.0	6.1	31.6	15.58-37.68
2024			6	21.6	6,208,069	13.2	9.0	64.1	15.63-27.53
2023	Cab Sauv	berry area	17	74.2	10,728,284	17.8	6.7	31.6	64.65-83.75
2024			17	62.7	8,241,035	15.9	10.6	64.1	53.49-71.89
2023	Cab Sauv	compactness	1	23.4	4,848,398	10.3	6.2	10.3	5.48-41.28
2024			1	21.5	4,241,270	12.3	3.4	12.3	5.72-37.22
2023	Cab Sauv	berry count	4	75.1	21,951,164	18.2	6.1	22.2	63.6-86.6
2024			4	67.2	21,001,269	20.1	6.7	23.4	56.4-78
2023	Cab Sauv	berry count	5	10.4	4,649,500	17.3	5.9	22.2	5.74-15.14
2024			5	10.1	4,379,291	19.3	6.5	23.4	5.38-14.78
2023	Riesling	berry count	7	13.6	3,144,729	22.1	8.9	38.0	8.26-18.96
2024			7	14.3	4,617,742	9.2	3.8	34.8	1.08-27.58
2023	Cab Sauv	berry length	6	26.6	8,170,801	15.2	5.9	32.1	15.58-37.68
2024			6	21.6	6,208,069	9.9	3.8	30.4	5.13-38.03
2023	Cab Sauv	berry length	17	74.2	10,728,284	19.1	7.2	32.1	64.8-83.6
2024			17	74.2	10,728,284	15.1	5.7	30.4	56-92.4
2023	Cab Sauv	berry width	6	26.6	8,170,801	14.9	5.6	29.7	16.08-37.18
2024			6	23.9	6,969,197	7.6	2.9	29.0	6.92-40.82
2023	Cab Sauv	berry width	17	74.2	10,728,284	16.8	6.2	29.7	64.1-84.3
2024			17	74.2	10,728,284	15.7	5.8	29.0	56-92.4
Single year QTLs									
2024	Cab Sauv	berry area	1	30.1	5,783,345	5.6	4.2	64.1	6.8-53.3
2024	Cab Sauv	berry area	5	58.3	16,726,371	10.7	7.6	64.1	55.83-60.83
2024	Cab Sauv	berry area	8	76.3	18,524,489	5.2	3.9	64.1	54.06-98.46
2024	Cab Sauv	berry area	10	66.1	18,712,836	5.1	3.8	64.1	50.35-81.75
2024	Cab Sauv	berry area	11	49.5	11,515,400	14.2	9.6	64.1	42.91-56.01
2024	Cab Sauv	berry area	12	53.0	20,509,714	9.9	7.0	64.1	37.25-68.75
2024	Cab Sauv	berry area	14	22.8	2,731,011	6.8	5.0	64.1	16.52-29.12
2024	Riesling	aspect ratio	3	40.9	6,732,209	12.0	6.4	12.0	20.09-61.79
2023	Cab Sauv	cluster weight	5	2.9	2,842,257	10.0	3.5	21.2	1.01-4.75
2023	Riesling	cluster weight	7	12.2	3,144,650	19.0	7.2	32.5	6.83-17.53
2023	Riesling	cluster weight	13	55.4	20,915,884	16.3	6.3	32.5	44.01-69.01
2024	Riesling	cluster weight	14	10.6	3,135,853	11.6	6.4	11.6	42.43-83.23
2023	Cab Sauv	cluster weight	17	48.3	5,387,153	8.9	3.1	21.2	1.06-20.16
2024	Cab Sauv	cluster weight	17	74.2	10,728,284	11.6	3.4	11.6	3.8-92.8
2023	Riesling	cluster weight	18	7.0	5,524,525	9.2	3.7	32.5	46.3-102.1
2023	Riesling	compactness	11	23.0	5,263,713	12.6	4.6	12.6	1.56-14.06

2024	Riesling	berry count	6	60.1	14,808,244	7.3	3.1	34.8	0.41-45.61
2023	Riesling	berry count	13	55.4	20,915,884	17.7	7.3	38.0	69.34-80.84
2024	Riesling	berry count	14	10.6	3,135,853	16.2	6.4	34.8	56.37-77.97
2023	Riesling	berry count	18	11.7	6,123,609	10.6	4.6	38.0	21.75-98.35
2024	Cab Sauv	berry length	18	66.7	8,376,291	8.7	3.4	30.4	48.29-62.59
2023	Riesling	number of clusters	5	25.3	5,685,740	14.9	5.2	23.5	0.29-20.93
2023	Riesling	number of clusters	17	51.6	11,986,380	22.9	7.6	23.5	0.13-23.31
2023	Riesling	total weight cluster	7	9.3	3,001,260	10.4	3.6	22.3	48.47-84.97
2023	Riesling	total weight cluster	13	57.2	22,415,752	12.6	4.4	22.3	16.4-34.1
2024	Cab Sauv	berry width	18	65.8	8,317,021	8.9	3.4	29.0	47.96-55.16

Discussion

This study offers novel insights into the genetic determinants of grape cluster architecture and yield components through QTL mapping in a bi-parental population derived from Riesling × Cabernet Sauvignon. Leveraging image-based phenotyping powered by the Segment Anything Model [15], or SAM, we acquired high-resolution morphological data for over 7,000 cluster images, leading to the identification of multiple QTLs associated with berry size and cluster compactness. Notably, our findings reveal stable genetic loci across multiple years, underscoring their potential applications in breeding programs aimed at enhancing grape quality and disease resistance.

Compact grape clusters have been widely associated with increased susceptibility to fungal diseases, particularly *Botrytis* bunch rot, due to restricted airflow, increased humidity retention, and elevated temperatures within the cluster [16]. Moreover, dense clusters inhibit the development of the protective waxy cuticle, a key barrier against pathogen invasion [17-18], and also reduces sunlight exposure within the clusters, further limiting UV-induced skin thickening and wax deposition [19-20]. Furthermore, heterogeneous ripening in compact clusters adversely affects sugar accumulation [21-22], anthocyanin biosynthesis [23-24], and tannin content [25-26], leading to inconsistent fruit maturity and diminished wine quality [27], ultimately resulting in economic losses for growers. To mitigate the adverse effects of compact clusters, growers commonly apply gibberellic acid GA to loosen clusters, especially in the table grape industry [28-29]. However, integrating genetic solutions through breeding presents a more sustainable and cost-effective approach.

A key contribution of this work is the application of foundational models like SAM to process images and generate high-quality masks for all berries within a cluster. Traditional methods for assessing grape cluster architecture include visual scoring, manual measurements, and conventional computer vision techniques, each with significant limitations. Visual scoring is subjective and prone to evaluator bias, manual

measurements are labor-intensive and impractical for large-scale studies, and conventional computer vision models require extensive training and often fail to generalize across different grape varieties and imaging conditions. Cluster compactness is a critical trait that lacks easy and scalable measurement methods, yet it significantly influences grape quality, disease susceptibility, and yield. Several limitations commonly encountered when phenotyping cluster architecture and compactness can be effectively addressed with the cluster analysis pipeline presented in this study. For example, Underhill et al. (2020) highlighted how variation in peduncle length across genotypes affects the positioning of clusters when clipped for imaging [3]. Short peduncles can cause the tops of clusters to extend outside the camera's field of view, leading to incomplete data capture. By using SAM, each object is segmented separately and then filtered based on size or shape, ensuring that only grape berries are retained for analysis. Since rachis and clippers have distinct morphometric features, they can be easily discarded, allowing clusters to be suspended centrally within the image frame to ensure that the entire cluster remains within the field of view. This is especially useful for highly shouldered clusters with short peduncles. Moreover, berry segmentation with SAM is equally effective regardless of berry color, addressing a common challenge in traditional threshold-based methods, which often struggle to separate objects of interest from the background. Unlike conventional methods, SAM, a zero-shot model, requires no task-specific training and is generalizable enough to perform effectively on grapevine clusters. This study demonstrates SAM's high accuracy in identifying individual berries in two-dimensional cluster images, achieving a strong correlation with human-identified berries (Pearson's $r = 0.96$). Using this model, we processed thousands of cluster images, generating nearly 330,000 berry masks, each linked to spatial coordinates within their clusters. This capability enables comprehensive, high-throughput analysis of cluster architecture, eliminating the need for time-consuming manual annotation while ensuring high reproducibility across different imaging conditions [2]. By leveraging SAM's automated and scalable segmentation capabilities, we were able to accurately phenotype grape clusters and identify a moderate-effect QTL for cluster compactness across multiple harvest seasons in a pure *Vitis vinifera* population. This highlights the potential of foundation models in plant phenotyping and their ability to streamline genetic studies of complex traits.

Comparison with previous studies reveals both shared and new QTLs, while also emphasizing the difficulties in directly comparing results across studies. For example, our identification of stable QTLs for berry size traits such as berry area, length, and width on chromosomes 6 and 17 aligns with findings from Rist et al. (2022), who also reported QTLs for berry diameter and volume on these chromosomes [13]. This supports the use of berry area as a proxy for overall berry size in image-based phenotyping approaches. Similarly, our detection of QTLs for berry number on chromosomes 5 and 7 partially overlaps with loci reported by Tello et al. (2016) [12].

Importantly, we found a stable QTL for cluster compactness on chromosome 1, a region not commonly associated with compactness in previous studies, pointing to a possible new locus. Several studies have reported overlapping QTLs for compactness on chromosomes 12, 17, and 18 [3,5,12]. However, they also reported unique, non-overlapping loci, likely due to differences in populations, phenotyping strategies, and trait definitions. Moreover, unless studies perform synteny analyses at the sequence level between the reference genomes used for each mapping population, comparisons of QTL positions based solely on genetic distance should be approached with caution. Differences in both physical and genetic distances, as well as genomic architecture across genotypes, can lead to misalignments and misinterpretations.

To minimize phenological variation, we employed a single harvest date for all genotypes, carefully timed to match the late-season maturity of most individuals. Although *véraison* was not formally tracked in this study, this approach ensured consistent handling of all genotypes during sampling. We believe this minimized bias related to ripening traits such as cluster weight and compactness, which remain relatively stable late in development when berry expansion slows, and volume may plateau or even decrease slightly due to evaporative water loss [30]. It is important to note that some early-ripening offspring may have experienced partial over-ripening and berry shriveling by the single harvest date. This could be a potential limitation of the study, as excessive berry shriveling might affect cluster compactness. However, no excessive shriveling was observed that would likely significantly impact cluster compactness. Additionally, many traits analyzed, like cluster number and berry count, are established earlier in the season and are less affected by harvest timing. From a logistical and statistical perspective, using a single harvest date also ensured consistent handling across genotypes and reduced environmental variation. To further refine this approach, future studies could determine the harvest date by using a fixed number of days after the onset of *véraison*, which would help minimize the impact of berry shriveling on compactness. Overall, while precisely overlapping with QTLs identified in previous studies is challenging due to differences in genetic backgrounds, marker systems, and reference genomes, our findings both confirm previously identified regions such as chromosome 17 for berry size and yield, and suggest new loci on chromosome 1 for compactness, providing valuable targets for future validation and breeding.

The lack of correlation between berry count and cluster compactness observed in our study has important implications for breeding strategy. While berry count was strongly and positively correlated with both cluster weight and total yield, it showed no significant association with compactness. This decoupling suggests that, within this genetic background, it is feasible to select for increased berry number as a yield-enhancing trait without simultaneously increasing cluster compactness, a trait often linked with disease susceptibility and reduced fruit quality due to limited airflow. These findings are particularly valuable because compactness is notoriously difficult to

manage in breeding due to its complex architecture. Our results, enabled with high-throughput phenotyping via the SAM pipeline, offer a practical and scalable path forward. Furthermore, co-localization of QTLs for berry number, cluster weight, and yield reinforces the genetic basis for such selection gains. While these findings are specific to our population, they raise the question of whether similar decoupling patterns exist in other genetic backgrounds. Reports from other populations are limited, but initial evidence from some interspecific crosses suggests the possibility of disentangling yield and compactness in certain contexts [3,7,13]. Future work should explore the genetic and phenotypic independence of these traits across diverse *Vitis* populations to validate the generality of this strategy. Doing so could help refine breeding targets for balanced improvement of yield and fruit quality.

These two parental cultivars, Cabernet Sauvignon and Riesling, differ markedly across a range of biological and agronomic traits, including their geographical origins, historical patterns of cultivation, phenology, growth habits, fruit chemistry, and enological characteristics. Most relevant to this study, however, are their contrasting cluster architectures—Riesling typically forms smaller, more compact clusters, while Cabernet Sauvignon produces larger, looser ones. With a strong consumer preference for pure *V. vinifera* cultivars and wines, identifying trait donors within *V. vinifera* is crucial for breeding strategies. Unlike breeding programs that rely on non-*vinifera* genetic material, which often necessitate multiple rounds of backcrossing to eliminate undesirable traits, the use of pure *V. vinifera* donors allows for more efficient gene introgression. While the individual marker effects identified in this study may not be strong enough to justify marker-assisted selection as a standalone strategy, our findings underscore the genetic variability for cluster architecture that can be unlocked through a single cross. This highlights the potential for recurrent selection, coupled with novel tools such as genomic prediction and AI-enabled computer vision, as a promising breeding approach to improve grape cluster structure and adaptability.

While this study provides valuable insights into the genetic architecture of grape cluster architecture and yield components, it is important to acknowledge some limitations, such as the mapping population size of 138 individuals, which although sufficient for detecting moderate-effect QTLs, limits the resolution of QTL intervals and statistical power for identifying small-effect loci. Consequently, some QTL regions span relatively large genomic intervals, making it challenging to pinpoint individual candidate genes for each trait with confidence. Further studies incorporating larger populations and complementary approaches such as fine mapping and transcriptomics could help narrow these intervals and refine gene discovery. Nonetheless, the stable QTLs detected across years in this study serve as promising targets for further validation and breeding efforts.

Material and methods

Plant material

The study utilized a genetic population of 138 full-sibling progeny derived from a cross between *Vitis vinifera* cv. Riesling and Cabernet Sauvignon. This cross was initially made by M.A. Walker and C.P. Meredith in 1994, and the progeny were initially planted at the Wolfskill Experimental Orchards in Winters, CA. In 2017, the plants were propagated, grafted onto 3309C, and transplanted to the UC Davis Viticulture and Enology Department Experimental Station in Oakville, Napa County, CA, USA (38°25'45.4"N; 122°24'36.4"W). This same population and planting have been used to understand the genetic basis of microbiome recruitment in grapevine [31]. The experimental design consisted of three randomized blocks, each containing three vines per genotype, including the parents, leading to a total of nine replicates per genotype and 1,242 vines. Within each block, the three replicate plants of a genotype were placed consecutively in a row, with their positions randomized across the three blocks. The vineyard was established with 1.8 m × 2.4 m vine and row spacing, with rows oriented in a northwest-southeast direction. The vines were trained to a modified vertical shoot positioning system and pruned to two-bud spurs during winter dormancy. Drip irrigation was applied using two pressure-compensating emitters per plant. Cabernet Sauvignon vines were planted as a buffer surrounding the trial to isolate the experimental population from the vineyard edges.

Traditional phenotyping

Grape clusters were harvested during two growing seasons, 2023 and 2024. In 2023, harvest took place on October 4; in 2024, it was conducted on September 20. All clusters from the middle vine of the three consecutive vines per genotype were collected for each genotype. The total number of clusters per vine was recorded, and the clusters were weighed to determine the total cluster weight per vine and average cluster weight. Additionally, five clusters were randomly selected from each middle vine for imaging, as described in the following section. Harvest was completed in a single day for each season to maintain consistency. Ripening time (i.e., technological maturity) varied across the population but was not recorded. Fruit was harvested when Cabernet Sauvignon, one of the parental varieties, reached approximately 24.5-25.5 °Brix.

Digital phenotyping

To assess grape cluster architecture, imaging and image processing were performed following a previously developed methodology [2]. Images from the fruit

harvested in 2023 were also used in Torres-Lomas et al. (2024) for the development of the image analysis pipeline and the application of the Segment Anything Model (SAM) for rapid and accurate cluster segmentation. For each vine, five clusters were randomly selected and photographed using a photo light box against a white background using a Canon EOS 70D camera equipped with a 24-mm prime lens. Grape clusters were illuminated using dimmable LED lights and were adjusted to between 5,000 and 5,500 K. All images were captured at a fixed distance of 70.5 cm, and a reference object was included in each frame to allow for size normalization and correction of any variations in camera-to-cluster distance. Camera settings were kept constant (aperture f/5, exposure time 1/500 s) to ensure image consistency. Based on previous findings, we captured two images per cluster, one from the front (0°) and one from the side (90°), as this approach has been shown to reliably estimate berry count and overall cluster structure [2]. Before segmentation, a region of interest (ROI) was automatically defined to isolate the grape cluster from the background. This step ensures that the segmentation algorithm processes only the relevant parts of the image. In this context, segmentation refers to the process of identifying and delineating each individual berry within the cluster image. This is a critical step that enables downstream analysis of cluster architecture based on detailed berry-level data.

For segmentation, we used the Segment Anything Model v1, a state-of-the-art computer vision tool that allows automated object identification. Within the defined ROI, a grid of 32 × 32 evenly spaced prompts guided the model to identify potential objects. The SAM framework, equipped with a ViT-H (Vision Transformer – Huge) encoder, generated precise segmentation masks that delineated individual berries within each image.

Post-processing steps were applied to refine the initial predictions. Segmented objects were filtered based on geometric properties such as area, perimeter, aspect ratio (the ratio of length to width), and intersection-over-union (IoU) thresholds to remove incorrectly labeled structures. To further improve accuracy, elliptical Fourier descriptors were computed for each segmented shape, followed by principal component analysis (PCA) to distinguish true berries from non-berry elements like rachis fragments or background artifacts. This resulted in a clean and accurate set of berry masks suitable for phenotypic analysis. Final berry-level measurements, including berry count, size, shape, and spatial distribution, were extracted and analyzed using the R package Momocs 1.4.1 [32]. A linear regression-based correction factor was applied to improve the accuracy of berry number estimates due to underestimations of the actual berry count caused by occlusion. All image processing and analysis steps were implemented in Python 3.11.

Statistical analysis

For traditional yield-related traits (number of clusters per vine, total cluster weight per vine, and average cluster weight), we used a mixed model with block as a fixed effect and genotype as a random effect. For the digital traits, including berry number, berry projected area, length, width, and cluster compactness, we first averaged the values across the five imaged clusters per genotype per block, then applied a similar mixed model with block and genotype as fixed and random effects, respectively. Best linear unbiased predictors were estimated for each trait and used for QTL mapping. In addition to single-year analyses, we also implemented a multi-year mixed model to account for data across both 2023 and 2024. In this model, year and block nested within year were treated as fixed effects, and genotype and genotype-by-year interaction were treated as random effects. This approach allowed for the estimation of across-year BLUPs, which were subsequently used in QTL analysis. Heritability was calculated using the formula: $H^2 = \frac{V_g}{V_g + V_e}$. All statistical models were implemented using R 4.2 and the R package *lme4* 1.1-36 [33].

Genotyping and linkage mapping

The mapping population, including both progeny and parents, was genotyped using genotyping-by-sequencing (GBS) following the method of Hyma et al. (2015) [34]. DNA was extracted from young leaves using the DNeasy 96-well DNA extraction kit (Qiagen, Valencia, CA, USA). Library preparation involved digestion with ApeKI, followed by sequencing on a HiSeq2000 platform (Illumina Inc., San Diego, CA, USA) at the Institute of Biotechnology, Genomics Facility (Cornell University, Ithaca, NY). Raw sequencing reads were processed using the TASSEL 5 GBS 5.0 v2 pipeline with default parameters [35]. Single nucleotide polymorphisms (SNPs) were identified by aligning sequence reads to the Cabernet Sauvignon cl.08 v1.1 Hap1 [36] genome assembly using BWA 0.7.12-r1039 [37-38]. The TASSEL 5 standalone program was then used to filter SNPs based on coverage depth (minimum read depth >6) and minor allele frequency (>0.1). SNPs and genotypes with more than 10% missing data were removed.

Two genetic maps, one for Cabernet Sauvignon and another for Riesling, were constructed using a pseudo-test cross-linkage mapping strategy [14] with the R package BatchMap 1.0.3.0 [39]. QTL mapping was performed using the R/qtl 1.7 function `stepwiseqtl` (method = "hk", additive.only = FALSE), which applies forward and backward selection to identify a multiple QTL model [40]. Model selection was based on a penalized LOD score, with separate penalties for main effects and interactions. Penalties were calculated using the `calc.penalties()` function, which used the output of `scantwo()` with 1,000 permutations for each trait.

Acknowledgments

The authors would like to thank Guillermo Garcia-Zamora, Dan Ng, Mikayla Bailey, Gonzalo Ruiz, and Jovita Lopez for their support.

Competing interests

The authors declare no conflicts of interest.

Funding

This project was partially funded by the American Vineyard Foundation (projects 2023-2781 and 2025-2953), the California Grape Rootstock Improvement Commission (project A24-0828), and the USDA-NIFA Specialty Crop Research Initiative (projects 2022-51181-38240 and 2024-51181-43236).

Data Availability

The BLUPs and parental linkage maps used for QTL mapping are provided in Supplementary Table 2 and 3. Raw sequencing data for mapping population are available as an NCBI BioProject under the accession number PRJNA1249123.

Author information

Authors and Affiliations

¹Department of Viticulture and Enology, University of California Davis, CA 95616, USA

**Contributed equally

*Corresponding author: diazgarcia@ucdavis.edu

Contributions

S.S, J.M, and L.D.G designed the study and wrote the original manuscript draft. S.S, J.M, H.B, Y.L, V.N, A.G, and L.D.G collected the data and conducted preliminary analysis. E.T-L and L.D.G performed image analysis. S.S, J.M, and L.D.G conducted the genetic analyses. All authors read and approved the final manuscript.

References

1. Tello, J., Aguirrezábal, R., Hernáiz, S., Larreina, B., Montemayor, M. I., Vaquero, E., & Ibáñez, J. (2015). Multicultivar and multivariate study of the natural variation for grapevine bunch compactness. *Australian Journal of Grape and Wine Research*, 21(2), 277-289. <https://doi.org/10.1111/ajgw.12121>
2. Torres-Lomas, E., Lado-Bega, J., Garcia-Zamora, G., & Diaz-Garcia, L. (2024). Segment Anything for comprehensive analysis of grapevine cluster architecture and berry properties. *Plant Phenomics*, 6, 0202. DOI: [10.34133/plantphenomics.0202](https://doi.org/10.34133/plantphenomics.0202)

3. Underhill, A., Hirsch, C., & Clark, M. (2020). Image-based phenotyping identifies quantitative trait loci for cluster compactness in grape. *Journal of the American Society for Horticultural Science*, 145(6), 363-373.
<https://doi.org/10.21273/JASHS04932-20>
4. Vail, M. E., Wolpert, J. A., Gubler, W. D., & Rademacher, M. R. (1998). Effect of cluster tightness on Botrytis bunch rot in six Chardonnay clones. *Plant Disease*, 82(1), 107-109. <https://doi.org/10.1094/PDIS.1998.82.1.107>
5. Richter, R., Gabriel, D., Rist, F., Töpfer, R., & Zyprian, E. (2019). Identification of co-located QTLs and genomic regions affecting grapevine cluster architecture. *Theoretical and Applied Genetics*, 132, 1159-1177.
<https://doi.org/10.1007/s00122-018-3269-1>
6. Herzog, K., Wind, R., & Töpfer, R. (2015). Impedance of the grape berry cuticle as a novel phenotypic trait to estimate resistance to Botrytis cinerea. *Sensors*, 15(6), 12498-12512. [10.3390/s150612498](https://doi.org/10.3390/s150612498)
7. Tello, J., & Ibáñez, J. (2018). What do we know about grapevine bunch compactness? A state-of-the-art review. *Australian Journal of Grape and Wine Research*, 24(1), 6-23. <https://doi.org/10.1111/ajgw.12310>
8. Zyprian, E., Richter, R., Rossmann, S., Theres, K., & Töpfer, R. (2018, July). Molecular analysis of bunch architecture in grapevine. In *XII International Conference on Grapevine Breeding and Genetics 1248* (pp. 327-330). [10.17660/ActaHortic.2019.1248.47](https://doi.org/10.17660/ActaHortic.2019.1248.47)
9. Azevedo, C. F., Ferrão, L. F. V., Benevenuto, J., de Resende, M. D. V., Nascimento, M., Nascimento, A. C. C., & Munoz, P. R. (2024). Using visual scores for genomic prediction of complex traits in breeding programs. *Theoretical and Applied Genetics*, 137(1), 9. <https://doi.org/10.1007/s00122-023-04512-w>
10. Cubero, S., Diago, M. P., Blasco, J., Tardaguila, J., Prats-Montalbán, J. M., Ibanez, J., & Aleixos, N. (2015). A new method for assessment of bunch compactness using automated image analysis. *Australian Journal of Grape and Wine Research*, 21(1), 101-109. <https://doi.org/10.1111/ajgw.12118>
11. Correa, J., Mamani, M., Munoz-Espinoza, C., Laborie, D., Muñoz, C., Pinto, M., & Hinrichsen, P. (2014). Heritability and identification of QTLs and underlying candidate genes associated with the architecture of the grapevine cluster (*Vitis vinifera* L.). *Theoretical and Applied Genetics*, 127, 1143-1162.
<https://doi.org/10.1007/s00122-014-2286-y>
12. Tello, J., Torres-Pérez, R., Grimplet, J., & Ibáñez, J. (2016). Association analysis of grapevine bunch traits using a comprehensive approach. *Theoretical and Applied Genetics*, 129, 227-242. [http://dx.doi.org/10.1007/s00122-015-2623-9](https://doi.org/10.1007/s00122-015-2623-9)
13. Rist, F., Schwander, F., Richter, R., Mack, J., Schwandner, A., Hausmann, L., ... & Herzog, K. (2022). Relieving the phenotyping bottleneck for grape bunch architecture in grapevine breeding research: implementation of a 3D-based

- phenotyping approach for quantitative trait locus mapping. *Horticulturae*, 8(10), 907. <https://doi.org/10.3390/horticulturae8100907>
14. Grattapaglia, D., & Sederoff, R. (1994). Genetic linkage maps of *Eucalyptus grandis* and *Eucalyptus urophylla* using a pseudo-testcross: mapping strategy and RAPD markers. *Genetics*, 137(4), 1121. <https://doi.org/10.1093/genetics/137.4.1121>
 15. Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., & Girshick, R. (2023). Segment anything. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 4015-4026).
 16. Hed, B., Ngugi, H. K., & Travis, J. W. (2009). Relationship between cluster compactness and bunch rot in Vignoles grapes. *Plant Disease*, 93(11), 1195-1201. <https://doi.org/10.1094/PDIS-93-11-1195>
 17. Marois, J. J., Nelson, J. K., Morrison, J. C., Lile, L. S., & Bledsoe, A. M. (1986). The influence of berry contact within grape clusters on the development of *Botrytis cinerea* and epicuticular wax. *American Journal of Enology and Viticulture*, 37(4), 293-296. [10.5344/ajev.1986.37.4.293](https://doi.org/10.5344/ajev.1986.37.4.293)
 18. Gabler, F. M., Smilanick, J. L., Mansour, M., Ramming, D. W., & Mackey, B. E. (2003). Correlations of morphological, anatomical, and chemical features of grape berries with resistance to *Botrytis cinerea*. *Phytopathology*, 93(10), 1263-1273. <https://doi.org/10.1094/PHYTO.2003.93.10.1263>
 19. Percival, D. C., Sullivan, J. A., & Fisher, K. H. (1993). Effect of cluster exposure, berry contact and cultivar on cuticular membrane formation and occurrence of bunch rot(*Botrytis cinerea* PERS.: FR.) with 3 *Vitis vinifera* L. cultivars. *Vitis*, 32(2), 87-97.
 20. Martínez-Lüscher, J., Torres, N., Hilbert, G., Richard, T., Sánchez-Díaz, M., Delrot, S., & Gomès, E. (2014). Ultraviolet-B radiation modifies the quantitative and qualitative profile of flavonoids and amino acids in grape berries. *Phytochemistry*, 102, 106-114. <https://doi.org/10.1016/j.phytochem.2014.03.014>
 21. Pieri, P., Zott, K., Gomès, E., & Hilbert, G. (2016). Nested effects of berry half, berry and bunch microclimate on biochemical composition in grape. *Oeno One*, 50(3), 23-41. [10.20870/oeno-one.2016.50.3.52](https://doi.org/10.20870/oeno-one.2016.50.3.52)
 22. Sweetman, C., Sadras, V. O., Hancock, R. D., Soole, K. L., & Ford, C. (2014). Metabolic effects of elevated temperature on organic acid degradation in ripening *Vitis vinifera* fruit. *Journal of Experimental Botany*, 65(20), 5975-5988. <https://doi.org/10.1093/jxb/eru343>
 23. Sadras, V. O., & Moran, M. A. (2012). Elevated temperature decouples anthocyanins and sugars in berries of Shiraz and Cabernet Franc. *Australian Journal of Grape and Wine Research*, 18(2), 115-122. <https://doi.org/10.1111/j.1755-0238.2012.00180.x>

24. Rienth, M., Torregrosa, L., Luchaire, N., Chatbanyong, R., Lecourieux, D., Kelly, M. T., & Romieu, C. (2014). Day and night heat stress trigger different transcriptomic responses in green and ripening grapevine (*Vitis vinifera*) fruit. *BMC Plant Biology*, 14, 1-18. <https://doi.org/10.1186/1471-2229-14-108>
25. Cohen, S. D., Tarara, J. M., & Kennedy, J. A. (2008). Assessing the impact of temperature on grape phenolic metabolism. *Analytica Chimica Acta*, 621(1), 57-67. <https://doi.org/10.1016/j.aca.2007.11.029>
26. Koyama, K., Ikeda, H., Poudel, P. R., & Goto-Yamamoto, N. (2012). Light quality affects flavonoid biosynthesis in young berries of Cabernet Sauvignon grape. *Phytochemistry*, 78, 54-64. <https://doi.org/10.1016/j.phytochem.2012.02.026>
27. Armstrong, C. E. J., Ristic, R., Boss, P. K., Pagay, V., & Jeffery, D. W. (2021). Effect of grape heterogeneity on wine chemical composition and sensory attributes for *Vitis vinifera* cv. Cabernet Sauvignon. *Australian Journal of Grape and Wine Research*, 27(2), 206-218. <https://doi.org/10.1111/ajgw.12469>
28. Domingos, S., Nobrega, H., Raposo, A., Cardoso, V., Soares, I., Ramalho, J. C., & Goulao, L. F. (2016). Light management and gibberellic acid spraying as thinning methods in seedless table grapes (*Vitis vinifera* L.): Cultivar responses and effects on the fruit quality. *Scientia Horticulturae*, 201, 68-77. <https://doi.org/10.1016/j.scienta.2016.01.034>
29. Wegher, M., Faralli, M., & Bertamini, M. (2022). Cluster-zone leaf removal and GA3 application at early flowering reduce bunch compactness and yield per vine in *Vitis vinifera* cv. Pinot Gris. *Horticulturae*, 8(1), 81. <https://doi.org/10.3390/horticulturae8010081>
30. Keller, M. *The Science of Grapevines: Anatomy and Physiology*. (Academic Press, 2010).
31. Floerl, L., Lin, J., Griggs, R. G., Massonnet, M., Cochetel, N., Figueroa-Balderas, R., Cantu, D., & Bokulich, N. A. (2025). The genetic basis of microbiome recruitment in grapevine and its association with fermentative and pathogenic taxa. *bioRxiv*. <https://doi.org/10.1101/2025.03.31.646331>
32. Bonhomme, V., Picq, S., Gaucherel, C., & Claude, J. (2014). Momocs: outline analysis using R. *Journal of Statistical Software*, 56, 1-24. <https://doi.org/10.18637/jss.v056.i13>
33. Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67, 1-48. <https://doi.org/10.18637/jss.v067.i01>
34. Hyma, K. E., Barba, P., Wang, M., Londo, J. P., Acharya, C. B., Mitchell, S. E., & Cadle-Davidson, L. (2015). Heterozygous mapping strategy (HetMappS) for high resolution genotyping-by-sequencing markers: a case study in grapevine. *PLoS One*, 10(8), e0134880. <https://doi.org/10.1371/journal.pone.0134880>

35. Glaubitz, J. C., Casstevens, T. M., Lu, F., Harriman, J., Elshire, R. J., Sun, Q., & Buckler, E. S. (2014). TASSEL-GBS: a high capacity genotyping by sequencing analysis pipeline. *PloS one*, 9(2), e90346.
<https://doi.org/10.1371/journal.pone.0090346>
36. Massonnet, M., Cochetel, N., Minio, A., Vondras, A. M., Lin, J., Muyle, A., & Cantu, D. (2020). The genetic basis of sex determination in grapes. *Nature communications*, 11(1), 2902. <https://doi.org/10.1038/s41467-020-16700-z>
37. Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*, 25(14), 1754-1760.
<https://doi.org/10.1093/bioinformatics/btp324>
38. Chin, C. S., Peluso, P., Sedlazeck, F. J., Nattestad, M., Concepcion, G. T., Clum, A., & Schatz, M. C. (2016). Phased diploid genome assembly with single-molecule real-time sequencing. *Nature Methods*, 13(12), 1050-1054.
<https://doi.org/10.1038/nmeth.4035>
39. Schiffthaler, B., Bernhardsson, C., Ingvarsson, P. K., & Street, N. R. (2017). BatchMap: a parallel implementation of the OneMap R package for fast computation of F1 linkage maps in outcrossing species. *PloS One*, 12(12), e0189256. <https://doi.org/10.1371/journal.pone.0189256>
40. Broman, K. W., Wu, H., Sen, S., & Churchill, G. A. (2003). R/qtl: QTL mapping in experimental crosses. *Bioinformatics*, 19(7), 889-890.
<https://doi.org/10.1093/bioinformatics/btg112>