Diploid chromosome-scale assembly of the *Muscadinia rotundifolia* genome supports chromosome fusion and disease resistance gene expansion during *Vitis* and *Muscadinia* divergence

Noé Cochetel, Andrea Minio, Amanda M. Vondras, Rosa Figueroa-Balderas, and Dario Cantu*
Department of Viticulture and Enology, University of California Davis, Davis, California, USA

Reference numbers for data available in public repositories:
- NCBI: PRJNA635946 and PRJNA593045
- Zenodo DOI: 10.5281/zenodo.3866087

Running title (35 char. max): The genome of *Muscadinia rotundifolia*

Key words: Muscadine grapes, Vitaceae evolution, Optical maps, Hybrid genome assembly, Full length cDNA sequencing, Disease resistance, Powdery mildew

Corresponding author: Dario Cantu, University of California, Davis, 95616 CA. Tel.: (530) 752-2929. Email: dacantu@ucdavis.edu

ABSTRACT (250 words max)

Muscadinia rotundifolia, the muscadine grape, has been cultivated for centuries in the southeastern United States. *M. rotundifolia* is resistant to many of the pathogens that detrimentally affect *Vitis vinifera,* the grape species commonly used for winemaking throughout Europe and in New World wine regions. For this reason, *M. rotundifolia* is a valuable genetic resource for breeding. Single molecule real-time reads were combined with optical maps to reconstruct the two haplotypes of each of the 20 *M. rotundifolia* cv. Trayshed (Trayshed, henceforth) chromosomes. Completeness and accuracy of the assembly was confirmed using a high-density linkage map of *M. rotundifolia*. Protein-coding genes were annotated using an integrated comprehensive approach that included full-length cDNA sequencing (Iso-Seq) to improve gene structure and hypothetical spliced variant predictions. Our data confirmed the fusion of chromosomes 7 and 20, which reduced the number of chromosomes in *Vitis* versus *Muscadinia* and pinpointed the location of the fusion in Cabernet Sauvignon and PN40024 chromosome 7. The numbers of nucleotide binding site leucine-rich repeats (NBS-LRR) in Trayshed and Cabernet Sauvignon were similar, but their locations were different. A dramatic expansion of the Toll/Interleukin-1 Receptor-like-X (TIR-X) class was detected on Trayshed chromosome 12 at the *Resistance to Uncinula necator 1* (*RUN1*)/ *Resistance to Plasmopara viticola 1* (*RPV1*) locus, which confers strong dominant resistance to powdery and downy mildew. A genome browser for Trayshed, its annotation, and an associated Blast tool are available at www.grapegenomics.com.

INTRODUCTION

Muscadinia rotundifolia is the foundation of the muscadine grape industry in the United States of America (USA), where its fruit, juice, and wine are produced (Olien 1990). *M. rotundifolia* is native to the warm and humid southeastern USA. Its natural habitat includes the states bordering the Gulf of Mexico and extends from northern Arkansas to Delaware (Bouquet 1978; Olmo 1986; Olien 1990; Heinitz *et al.* 2019). *Muscadinia* is a distinct genus closely related to the species best known for winemaking, *Vitis vinifera* (Small 1913; Bouquet 1978; Olmo 1986; Olien 1990).

Both *Vitis* and *Muscadinia* belong to the Vitaceae family, which contains 16 other genera and approximately 950 species (Wen *et al.* 2018). These genera diverged between 18 and 47 million years ago (Ma) (Wan *et al.* 2013; Liu *et al.* 2016; Ma *et al.* 2018), a process that involved several chromosome rearrangements (Karkamkar *et al.* 2010; Wen *et al.* 2018). *V. vinifera* has 19 chromosomes (2n = 38) (Olmo 1937), but like other genera in Vitaceae, including *Ampelopsis*, *Parthenocissus* and *Ampelocissus, M. rotundifolia* has 20 (2n = 40) (Branas 1932; Patil and Patil 1992; Karkamkar *et al.* 2010; Chu *et al.* 2018). Though successful crosses yielding fertile hybrids have been obtained, early attempts at producing hybrids of *V. vinifera* and *M. rotundifolia* often resulted in sterile progeny; this and the graft incompatibility observed between the two species (Patel and Olmo 1955) is thought to be caused by their difference in chromosome number (Ravaz 1902; Dearing 1917; Patel and Olmo 1955; Olmo 1971; Bouquet 1980; Walker *et al.* 1991). SSR markers and genetic linkage maps revealed that linkage groups LG7 and LG20 in *M. rotundifolia* correspond to the proximal and the distal regions of chromosome 7 in *V. vinifera* (Blanc *et al.* 2012; Delame *et al.* 2019; Lewter *et al.* 2019). Chromosome 7 in *Vitis* may be derived from the fusion of its ancestor's chromosomes 7 and 20. The presence of vestigial centromeric and/or telomeric repeats in chromosome 7 of *V. vinifera* (Lewter *et al.* 2019) and of chromosomes 7 and 20 in *M. rotundifolia* could provide additional support for this hypothesis, insight into the evolutionary history of Vitaceae, and understanding of the structure and function of the *M. rotundifolia* genome.

M. rotundifolia is a desirable partner with which to hybridize because it is resistant to stresses, pests, and diseases that adversely affect *V. vinifera* and cause substantial crop loss (Olmo 1971). *M. rotundifolia* is resistant to Pierce's disease (*Xylella fastidiosa*) (Ruel and Walker 2006), phylloxera (*Daktulosphaira vitifolia*) (Ravaz 1902; Davidis and Olmo 1964; Firoozabady and Olmo 1982), downy mildew (*Plasmopara viticola*) (Olmo 1971; Staudt and Kassemeyer 1995), powdery mildew (*Erysiphe necator* syn. *Uncinula necator*) (Olmo 1986; Merdinoglu *et al.* 2018), and other diseases and pests (Ravaz 1902; Olmo 1971; Walker *et al.* 2014). Several loci associated with resistance to pathogens affecting *V. vinifera* were identified in *M. rotundifolia,* including *Resistance to Uncinula necator 1* (*RUN1*) (Pauquet *et al.* 2001), *RUN2* (Riaz *et al.* 2011), *Resistance to Erysiphe Necator 5* (*REN5*) (Blanc *et al.* 2012), *Resistance to Plasmopara viticola 1* (*RPV1*) (Merdinoglu *et al.* 2003), and *RPV2* (Merdinoglu *et al.* 2018). However, the lack of a *M. rotundifolia* reference sequence and gene annotation limits gene discovery and the characterization of these and additional resistance loci.

A preliminary contig assembly of the *M. rotundifolia* cv. Trayshed (Trayshed, hereafter) genome using Single Molecule Real-Time (SMRT) sequencing was instrumental to resolving the genetic basis of sex determination in grapes (Massonnet *et al.* In Press). Here, we report the phased, chromosome-scale assembly of Trayshed, which was produced using a hybrid assembly of the

SMRT contigs with optical maps. Full-length cDNA isoforms (Iso-Seq) were also sequenced and used as transcriptional evidence for the annotation of protein coding genes. This assembly and its annotation were used to identify where *M. rotundifolia* chromosomes 7 and 20 fused to create *V. vinifera* chromosome 7 and to identify genes at disease resistance loci.

MATERIALS AND METHODS

Trayshed chromosome construction

Bionano's Next Generation Mapping (NGM) of Trayshed was generated using a Saphyr Genome Imaging Instrument using DLE-1 non-nicking enzyme (CTTAAG). High-molecular weight DNA (> 500 kb) was extracted by Amplicon Express (Pullman, WA). DNA was then nicked and labelled using the SaphyrPrep Kit (BioNano Genomics, San Diego, CA). Labelled DNA was loaded onto the SaphyrChip nanochannel array for imaging on the Saphyr system (BioNano Genomics). Imaged molecules longer than 150 kbp were kept. These molecules were then assembled using BioNano Solve v.3.3 (Lam *et al.* 2012) with parameters described in Supplemental file S1, generating a 1.18 Gbp consensus genome map with an N50 of 5.6 Mbp. PacBio contigs generated with FALCON-Unzip as described in Massonnet et al. (In Press) were assembled together with the NGM maps using HybridScaffold v.04122018 (Lam *et al.* 2012). The procedure was performed in four steps with varying conflict resolution approaches. The scaffolding parameters alternated from more to less conservative between steps 1 and 2 and between steps 3 and 4. In the first step, sequences and optical maps were broken ('-B 2 -N 2') when in conflict with one another (Supplemental file S2). In the second step, the scaffolds produced previously were compared to the optical maps with scaffolding parameters that were more permissive of conflict between the two (Supplemental file S3). During the third step, conflicts were resolved by breaking nucleotide sequences ('-B 1 -N 2') using the conservative set of parameters (Supplemental file S2) to correct haplotype switching and reduce fragmentation. The fourth step was like the third, but with less conservative scaffolding parameters (Supplemental file S3). The completeness of the gene space in the hybrid assembly was estimated by aligning the PN40024 V1 CDS against the Trayshed assembly using BLAT with default parameters (Kent 2002). PN40024 CDS were filtered prior to the alignment and included only single-copy genes (i.e., with unique mapping on its genome). The genomic sequences obtained were organized and sorted in two set of chromosomes with HaploSync v.1 (https://github.com/andreaminio/HaploSync) and based on synteny with *Vitis vinifera* PN40024 chromosomes.

cDNA library preparation and sequencing

Total RNA from Trayshed leaves was isolated using a Cetyltrimethyl Ammonium Bromide (CTAB)-based extraction protocol as described in Blanco-Ulate *et al*. (2013). RNA purity was evaluated with a Nanodrop 2000 spectrophotometer (Thermo Scientific, Hanover Park, IL), quantity with a RNA broad range kit of the Qubit 2.0 Fluorometer (Life Technologies, Carlsbad, CA) and integrity using electrophoresis and an Agilent 2100 Bioanalyzer (Agilent Technologies, CA). Total RNA (300 ng, RNA Integrity Number > 8.0) was used for cDNA synthesis and library construction.

An RNA-Seq library was prepared using the Illumina TruSeq RNA sample preparation kit v.2 (Illumina, CA, USA) following Illumina's Low-throughput protocol. This library was evaluated

130 for quantity and quality with the High Sensitivity chip in an Agilent 2100 Bioanalyzer (Agilent
131 Technologies, CA) and was sequenced in 100bp, paired-end reads, using an Illumina HiSeq4000
132 sequencer (DNA Technology Core Facility, University of California, Davis). In addition, a cDNA
133 SMRTbell library was prepared. First strand synthesis and cDNA amplification were
134 accomplished using the NEBNext Single Cell/Low Input cDNA Synthesis & Amplification
135 Module (New England, Ipswich, MA, US). These cDNAs were subsequently purified with ProNex
136 magnetic beads (Promega, WI) following the instructions in the Iso-Seq Express Template
137 Preparation for Sequel and Sequel II Systems protocol (Pacific Biosciences, Menlo Park, CA).
138 ProNex magnetic beads (86 µl) were used to select amplified cDNA with a mode of 2 kb. At least
139 80 ng of the size-selected, amplified cDNA were used to prepare the cDNA SMRTbell library.
140 This was followed by DNA damage repair and SMRTbell ligation using the SMRTbell Express
141 Template Prep Kit 2.0 (Pacific Biosciences, Menlo Park, CA), following the manufacturer's
142 protocol. One SMRT cell was sequenced on the PacBio Sequel I platform (DNA Technology Core
143 Facility, University of California, Davis).
144
145 **Genome annotation**
146 The structural annotation of the *M. rotundifolia* genome was performed with a modified version
147 of the pipeline used for the *Vitis vinifera* cv. Zinfandel genome (Vondras *et al.* 2019) and fully
148 described here: https://github.com/andreaminio/AnnotationPipeline-EVM_based-DClab. Briefly,
149 transcript evidence was obtained from external databases and by combining multiple strategies to
150 build transcriptome assemblies from RNA-Seq sample using Stringtie v.1.3.4d (Pertea *et al.* 2015)
151 and Trinity v.2.6.5 (Grabherr *et al.* 2011). In addition, PASA v.2.3.3 (Haas 2003) and high quality
152 Iso-Seq data from Trayshed were used to produce high-quality gene models for training gene
153 predictors. Public databases, transcriptome assemblies, and the Iso-Seq data described above were
154 used as transcript and protein evidence. They were mapped on the genome using PASA v.2.3.3
155 (Haas 2003), MagicBLAST v.1.4.0 (Boratyn *et al.* 2019) and Exonerate v.2.2.0 (Slater and Birney
156 2005). *Ab initio* predictions were generated using BUSCO v.3.0.2 (Seppey *et al.* 2019), Augustus
157 v.3.0.3 (Stanke *et al.* 2006), GeneMark v.3.47 (Lomsadze 2005) and SNAP v.2006-07-28 (Korf
158 2004). Repeats were annotated using RepeatMasker v.open-4.0.6 (Smit *et al.* 2013). Next,
159 EvidenceModeler v.1.1.1 (Haas *et al.* 2008) used these predictions to generate consensus gene
160 models. The final functional annotation was produced combining blastp v.2.2.28 (Camacho *et al.*
161 2009) hits against the Refseq plant protein database (ftp://ftp.ncbi.nlm.nih.gov/refseq, retrieved
162 January 17th, 2019) and InterProScan v.5.28-67.0 (Jones *et al.* 2014) outputs using Blast2GO
163 v.4.1.9 (Gotz *et al.* 2008).
164
165 **Genome size quantification by flow cytometry**
166 DNA content was estimated using flow cytometry (n = 3 individual leaves).
167 *Lycopersicon esculentum* cv. Stupické polní tyčkové rané was selected as the internal DNA
168 reference standard with a genome size of 2C = 1.96 pg; 1C = 958 Mb (Doležel *et al.* 1992). Nuclei
169 extraction was performed using the Cystain PI absolute P kit (Sysmex America Inc).
170 Approximately 5 mg (0.7 cm^2) of young healthy leaves from grapevine and tomato were cut finely
171 with a razor blade in a Petri dish containing 500 µL of extraction buffer. The nuclei suspension
172 was filtered through a 50 µm filter (CellTrics, Sysmex America Inc) and 2 mL of a propidium
173 iodide staining solution was added (Doležel and Bartoš 2005; Bertier *et al.* 2013). Measurements
174 were acquired using a Becton Dickinson FACScan (Franklin Lakes, NewJersey) equipped with a
175 488 nm laser. The data were analyzed using FlowJo v.10

176 (https://www.flowjo.com/solutions/flowjo). DNA content was inferred by linear regression using
177 the tomato DNA reference standard. The genome size was comparable to PN40024 (Jaillon *et al.*
178 2007), at 487 Mb, but smaller than Cabernet Sauvignon (557.0 ± 2.4 Mbp; Supplemental Figure
179 S1).
180
181 **Comparison of genome assembly and synteny analysis**
182 The Trayshed and Cabernet Sauvignon (Massonnet *et al.* In Press) genomes were compared with
183 NUCmer (MUMmer v.4.0.00) (Marçais *et al.* 2018) and the --mum setting. Descriptive statistics
184 of the alignment were obtained using the MUMmer "dnadiff" option. The delta file was filtered
185 using delta-filter with these settings: -i 90 -l 7500. Blastp v.2.2.28 (Camacho *et al.* 2009) was used
186 to align the annotated proteins of Trayshed and Cabernet Sauvignon. Pairwise protein information
187 was associated with genes and processed with McScanX v.11.Nov.2013 (Wang *et al.* 2012) to
188 identify syntenic regions.
189
190 **Localization of PN40024 SNP markers and telomeric repeats analysis**
191 Five hundred base-pair long genomic regions were extracted upstream and downstream of each
192 PN40024 V0 SNP marker (Lewter *et al.,* 2019). GMAP v. 2015-11-20 (Wu and Watanabe 2005)
193 was used to locate these sequences in Trayshed and Cabernet Sauvignon. The previously published
194 consensus map of *M. rotundifolia* linkage groups was used to assess the completeness of the
195 chromosome reconstruction (Lewter *et al.* 2019). Sliding windows (window size = 10 markers,
196 sliding = 5 markers) were designed to move across the uniquely mapped markers on the genome.
197 The percentage of relative marker positions consistent with their reference positions on the high-
198 density genetic map was calculated per window. Telomeric repeats of "CCCTAA" and their
199 reverse complement were searched along chromosome 7 of Cabernet Sauvignon and PN40024 V2
200 (Canaguier *et al.* 2017) using vmatchPattern from the R package Biostrings v.2.52.0 (H. Pagès *et*
201 *al.* 2019). After peaks in the distribution of telomeric repeats were identified in both genomes,
202 genomic regions of Cabernet Sauvignon (chr7:18,675,000-18,677,000 bp) and PN40024
203 (chr7:17571000-17573000 bp) were extracted. Motif enrichment analysis was performed using
204 MEME v.5.1.0 (Bailey and Elkan 1994) and reported as "logo" using the ggseqlogo R package
205 v.0.1 (Wagih 2017).
206
207 **Identification of NBS-LRR genes**
208 Predicted proteins from the Trayshed and Cabernet Sauvignon genomes were scanned using
209 hmmsearch (HMMER v.3.3; http://hmmer.org/) with a sequence E-value threshold of 0.001 and
210 Hidden Markov Models (HMM) corresponding to different Pfam (El-Gebali *et al.* 2019) domains:
211 NB-ARC (Pfam PF00931), TIR (PF01582) and LRR (PF00560, PF07723, PF07725 and
212 PF12799). Coiled coil (CC)-containing proteins were identified by COILS (Lupas *et al.* 1991)
213 during the InterProScan annotation. This set of annotations was divided in six protein classes: CC-
214 NBS-LRR, CC-NBS, TIR-NBS-LRR, TIR-NBS, NBS-LRR and NBS. To capture the largest
215 number of potential NBS-LRR related genes, genes lacking the NB-ARC domain but with "NBS-
216 LRR" functional annotations were also selected and divided in two classes: TIR-X and CC-X. For
217 these eight protein classes, Multiple EM for Motif Elicitation (MEME) analysis was performed
218 with the flags -mod anr -nmotifs 20 to identify conserved domains for each class. Then, FIMO
219 v.5.1.0 (Find Individual Motif Occurences) (Grant *et al.* 2011) was ran on each protein class using
220 the corresponding MEME results; proteins with at least five conserved domains were kept. NBS-
221 LRR gene clusters were defined as groups of at least two NBS-LRR genes, each separated by no

222 more than eight non-NBS-LRR genes (Richly *et al.* 2002) in a region spanning a maximum of 200
223 kbp (Holub 2001).
224
225 ***RUN1/RPV1* locus analysis and NBS-LRR genes phylogeny**
226 Boundaries of the *RUN1/RPV1* locus were defined by mapping two markers, VMC4f3.1 and
227 VMC8g9 (Barker *et al.* 2005), on Trayshed and Cabernet Sauvignon using GMAP v. 2015-11-20
228 (Wu and Watanabe 2005). A tblastx (BLAST v.2.2.29) (Camacho *et al.* 2009) was performed
229 using the genomic region corresponding to the resistance locus in Cabernet Sauvignon as a query
230 and Trayshed chromosome 12, split in 50-kbp blocks, as a reference. The tblastx results were
231 filtered, keeping hits with at least 90% of identity, a minimum length of 100 bp, and that were
232 contained within the boundaries of the resistance locus in both the query and the reference. Lastly,
233 hits matching LTR retrotransposons were discarded. Proteins corresponding to the first alternative
234 spliced variant of each NBS-LRR gene of the *RUN1/RPV1* locus were aligned using MUSCLE
235 v.3.8.31 (Edgar 2004) with default parameters and by including *GAPDH* (*VIT_17s0000g10430*)
236 ortholog sequences as an outgroup. Distances between the alignments were extracted using the R
237 package seqinr v.3.6.1 (Charif and Lobry 2007). The estimation of the corresponding neighbor-
238 joining tree, its rooting, and the bootstrapping (1000 replicates) were performed using the R
239 package ape v.5.3 (Paradis and Schliep 2019). The tree was drawn using ggtree v.2.0.2 (Yu *et al.*
240 2017). Coding sequences of the two Trayshed-specific gene clusters were aligned per class using
241 MACSE v.2.03 (Ranwez *et al.* 2018) and trimmed using the option
242 min_percent_NT_at_ends=0.80. Prior to the alignment, outliers were removed based on sequence
243 length. Synonymous substitution rates (dS) were obtained using yn00 from PAML v.4.9f (Yang
244 2007).
245
246 **Data analysis and visualization**
247 Data were parsed with R v.3.6.0 (R Core Team 2019) in RStudio v.1.2.5033 (RStudio Team 2019)
248 using tidyverse v.1.3.0 (Wickham *et al.* 2019), GenomicFeatures v.1.36.4 for genomic ranges
249 manipulation (Lawrence *et al.* 2013), rtracklayer for GFF files v.1.44.4 (Lawrence *et al.* 2009) and
250 viridis v.0.5.1 for the color gradients (Garnier 2018).
251
252 **Data availability**
253 Sequencing data are accessible at NCBI repository under the accession PRJNA635946 and
254 PRJNA593045. Raw optical maps are available at Zenodo under the DOI
255 10.5281/zenodo.3866087. The supplemental files are available at Figshare. The pipeline for the
256 gene annotation is available at https://github.com/andreaminio/AnnotationPipeline-EVM_based-
257 DClab. Assembly and annotation files are available at www.grapegenomics.com, which also hosts
258 a genome browser and a blast tool for Trayshed.
259
260
261 **RESULTS AND DISCUSSION**
262
263 **Construction of the twenty phased chromosomes of *Muscadinia rotundifolia* cv. Trayshed**
264 We reconstructed the twenty Trayshed chromosomes using optical maps produced with BioNano
265 Genomics technology at 2,057x coverage. The BioNano consensus genome map was combined
266 with the contigs produced from Single Molecule Real-Time (SMRT; Pacific Biosciences) long-
267 reads to generate a hybrid genome assembly. The size of the hybrid assembly (898.5 Mbp) was

268 nearly twice as large as the expected haploid genome size of Trayshed as determined by flow
269 cytometry (483.4 ± 3.1 Mbp; Supplemental Figure S1). To assess the gene content in the hybrid
270 assembly, we carried out a BUSCO analysis (Seppey *et al.* 2019) and mapped the PN40024 genes
271 on the scaffolds using BLAT (Kent 2002). About 66% of the BUSCO genes were found duplicated
272 and 2.16 ± 1.06 copies of the PN40024 genes aligned to the scaffolds. The size of the assembly,
273 the amount of duplicated BUSCO genes (Supplemental Table S1) and the two-fold PN20024 gene
274 content are strong evidence that the scaffolds obtained by hybrid assembly represent the diploid
275 genome of Trayshed. To separate homologous copies, we split the assembly into two phased sets
276 of scaffolds (Haplotype 1, Hap1, and Haplotype 2, Hap2; Table 1) using haplosync, which sort the
277 assembled sequences in chromosome structure while segregating them in non-overlapping sets of
278 alternative haplotypes by homology with a related species, *Vitis vinifera* cv. PN40024. The
279 structure of the Trayshed assembly was confirmed by comparing the positions along each
280 chromosome of the PN40024 V0 markers with their reported position on linkage groups (LG;
281 Lewter *et al*. (2019)) (Supplemental Figure S2 and S3). More than 80% of the markers' positions
282 were in agreement between the genetic and physical maps, with strong overall correlation ($R >$
283 0.90) along all chromosomes (Supplemental Figure S2).
284
285 **Table 1. Genome assembly statistics**

	Haplotype 1	**Haplotype 2**	**Unplaced**
Assembly length (bp)	460,586,882	302,699,902	136,115,388
Number of scaffolds	20	20	1,784
Average length (bp)	23,029,344	15,134,995	76,297
Maximum length (bp)	37,866,691	23,149,288	11,053,884
N50 length (bp)	24,189,316	15,419,175	314,857
Total gap length (bp)	24,598,680	29,789,394	19,410,932
Repetitive content (%)	40.52	37.80	42.51
Number of protein-coding genes	27,924	17,829	6,661
Complete BUSCO single copy (%)	95.1	68.6	18.6

286
287 **Analysis of chromosome structure supports a fusion of chromosomes 7 and 20 in *Vitis***
288 MUMMER was used to compare the structure of Trayshed with the chromosome-scale Cabernet
289 Sauvignon assembly (Massonnet *et al.* In Press), which is the most complete and contiguous
290 genome assembly of grapevine available to date. The two genomes were highly syntenic (Figure
291 1A). On average, colinear regions were 2,484 bp long and shared 92.46% identity (Supplementary
292 Table S1). On average, insertions were 25-bp long and deletions were 15-bp long. The longest
293 deletion (~50 kb) was detected on chromosome 8 of Cabernet Sauvignon while the longest
294 insertion (78 kb) was observed on chromosome 11 (Supplemental Table S2).
295
296 Trayshed chromosomes 7 and 20 aligned at the beginning and end, respectively, of Cabernet
297 Sauvignon chromosome 7. This confirms the fusion of these two chromosomes in *Vitis* that was
298 previously reported using genetic maps (Blanc *et al.* 2012; Lewter *et al.* 2019) (Figure 1A).
299

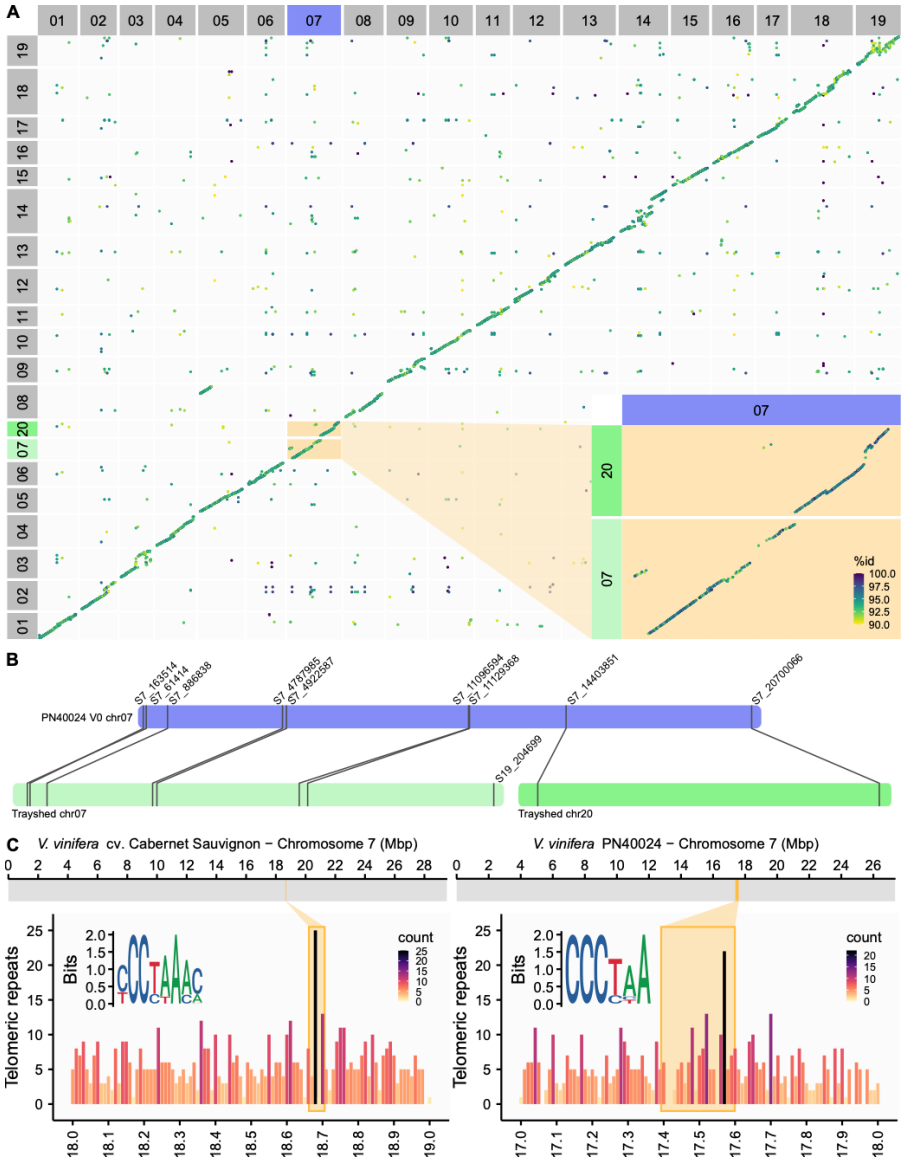

Figure 1: Chromosome fusion localization in *Vitis vinifera*

A. Whole genome alignment of Trayshed (y-axis) on Cabernet Sauvignon (x-axis). The putative fusion of Trayshed chromosomes 7 and 20 of Trayshed (green) and Cabernet Sauvignon chromosome 7th (purple) is inset. The percent identity (% id) between the alignments is displayed as a color gradient. **B**. The positions of markers on Trayshed chromosomes 7 and 20 (green) and their corresponding location on PN40024 V0 chromosome 7. **C**. The distribution of telomeric repeats in 1 Mbp containing the expected chromosome fusion in Cabernet Sauvignon (left panel) and PN40024 (right panel). The frequency of telomere repeats is represented as a color gradient. The most enriched motif in the 2-kbp region surrounding the peak of telomeric repeats is inset for each genotype.

The marker-containing regions used for Trayshed chromosome structure validation also support this hypothesis; all markers aligning Trayshed LG 7 and 20 were associated with *V. vinifera* chromosome 7 (Figure 1B). Telomere repeats can be used as indicators of chromosome number reduction rearrangements (Sousa and Renner 2015). An enrichment of these sequences is expected in a genomic region if a fusion occurred. To determine the position of chromosome fusion in

317 Cabernet Sauvignon, we searched for telomeric repeats in the region of chromosome 7 that should
318 contain the fusion point based on nucmer alignments. We examined the genomic region of
319 Cabernet Sauvignon chromosome 7 that spanned Trayshed chromosome 7 and 20 (chr 7:
320 18,661,815 – 18,705,855 bp; Figure 1C, top panel) and found an enrichment of telomeric repeats
321 in this region. (Figure 1C, bottom panel). A similar pattern was detected in the hypothetical fusion
322 region of PN40024 (chr 7: 17,392,916 - 17,599,237 bp; Figure 1C). Together, these data support
323 a chromosome number reduction by fusion in *V. vinifera* when compared with *M. rotundifolia*.
324
325 **Full length cDNA sequencing and annotation of the protein-coding genes**
326 Together with the public data and RNA-Seq data already available, high-quality full-length cDNA
327 sequences (Iso-Seq; Pacific Biosciences, Supplemental Figure S4) were generated to be used as
328 transcriptional evidence for gene model predictions. From leaf tissues, 336,932 raw reads were
329 sequenced, clustered, and error corrected with LSC (Au *et al.* 2012) to generate 34,558 high-
330 quality isoforms, 260 low-quality isoforms, and 111,672 singletons (Supplemental Figure S4).
331 After obtaining consensus gene models, alternative splice variants were predicted if supported by
332 Iso-Seq data, with 1.58 transcripts per gene on average (Supplemental Figure S4). BUSCO genes
333 (97.6%) were well-represented in the diploid Trayshed transcriptome. The structural annotation
334 included 52,414 protein coding genes (Table 1) and 83,010 proteins (including the alternative
335 forms). There were 27,924 and 17,829 genes on haplotypes 1 and 2, respectively (Table 1). The
336 repetitive content composed 39.90% of the genome. On average, colinear blocks of homologous
337 genes were ~ 28-genes long (Supplemental Table S3) based on McScanX results. The longest
338 block of colinear genes was found on chromosome 6 and contained 695 genes (Supplemental Table
339 S3). Based on blastp results with an inclusive filtering threshold set at 70% for both query and
340 target coverage and identity, 50,670 (96.7%) Trayshed genes are homologous with 52,684
341 (91.78%) genes of Cabernet Sauvignon. The 1,890 proteins that did not have a potential ortholog
342 in Cabernet Sauvignon were significantly enriched in functional categories related to compound
343 metabolic process (P < 1e-04; Fisher's exact test).
344
345 **NBS-LRR intracellular receptors encoded in the Trayshed genome**
346 Trayshed is attractive to grape breeding programs because it is resistant to pathogens that affect *V.*
347 *vinifera*. In plants, resistance to pathogens is primarily attributed to the activity of resistance genes
348 (R-genes). Nucleotide binding site leucine-rich repeat (NBS-LRR) genes constitute the largest
349 family of R-genes. NBS-LRR genes are associated with resistance to powdery and downy mildew
350 pathogens in several species, including *Arabidopsis* (Warren *et al.* 1998), wheat (He *et al.* 2018),
351 barley (Wei *et al.* 1999; Zhou *et al.* 2001), pepper (Jo *et al.* 2017), and grapes (Riaz *et al.* 2011;
352 Zini *et al.* 2019).
353
354 We divided the NBS-LRR genes in eight different classes depending on the domain detected in
355 the proteins: CC-NBS-LRR, CC-NBS, CC-X, TIR-NBS-LRR, TIR-NBS, TIR-X, NBS-LRR, NBS
356 (called hereafter NBS-LRR when described as a whole). Overall, the Trayshed genome has slightly
357 more NBS-LRR genes (1,075) than Cabernet Sauvignon (1,009). Most NBS-LRR genes were
358 found on chromosomes 12 and 13 (91 genes on each; 1.54 and 1.21x more than Cabernet
359 Sauvignon, respectively), followed by chromosome 19 (58 genes; 1.49x more than Cabernet
360 Sauvignon) (Figure 2A, Supplementary Table S4). In both species, chromosome 4 contained the
361 fewest NBS-LRR genes (2). In Cabernet Sauvignon (Figure 2B, Supplementary Table S4), most
362 NBS-LRR genes were detected on chromosome 13 (75 genes), 9 (72 genes; 1.57x more than

363 Trayshed) and 18 (65 genes; 1.38x more than Trayshed). Within the genome, NBS-LRR genes
364 form clusters that could be critical to their function as resistance loci (Hulbert *et al.* 2001; Meyers
365 *et al.* 2003). The number of clusters detected in Trayshed (123) was slightly higher than in
366 Cabernet Sauvignon (121), but their arrangement in the genome varied substantially
367 (Supplemental Figure S5). Chromosomes 12 and 13 contain the largest number of clusters (23 and
368 22 clusters, respectively) in Trayshed. In Cabernet Sauvignon, clusters were the most numerous
369 on chromosome 13 and 18 (20 and 18 clusters, respectively). These differences could plausibly
370 reflect to disparities in disease resistance.
371
372 **Expansion of NBS-LRR genes at the *RUN1/RPV1* locus in Trayshed**
373 The clusters of NBS-LRR genes on chromosome 12 in Trayshed are mostly absent from Cabernet
374 Sauvignon (Figure 2A and 2B). This chromosome contains a known powdery mildew resistance
375 locus, *RUN1/RPV1* (Pauquet *et al.* 2001), with boundaries defined by the VMC4f3.1 and VMC8g9
376 SSR markers (Barker *et al.* 2005). This region was studied in Trayshed (chr12: 19,181,246 –
377 26,483,447 bp) and Cabernet Sauvignon (chr12: 18,746,148 – 23,728,139 bp); it included 355
378 annotated genes in Trayshed and 287 genes in Cabernet Sauvignon. Within the *RUN1/RPV1* locus,
379 57 NBS-LRR were identified in Trayshed and 33 in Cabernet Sauvignon.
380
381 Notably, most of the NBS-LRR classes are present for both species in this region, but the TIR-X
382 class is only present in Trayshed. A phylogenetic analysis of the NBS-LRR proteins at the
383 *RUN1/RPV1* resistance locus identified two Trayshed-specific clusters (Figure 2C). The timing of
384 the divergence of these sequences was estimated as described in Massonnet *et al*. (In Press). This
385 involved aligning corresponding gene coding sequences with MACSE (Ranwez *et al.* 2018) and
386 calculating the rate of synonymous mutations with yn00 from PAML (Yang 2007). This method
387 timed the divergence of TIR-X genes at nearly 16 Ma, after the divergence of the *Muscadinia* and
388 *Vitis* genera (about 18 Ma) (Wan *et al.* 2013). Thus, a recent expansion event of the TIR-X class
389 genes might have occurred in only *Muscadinia* and other non-TIR classes would have expanded
390 prior.

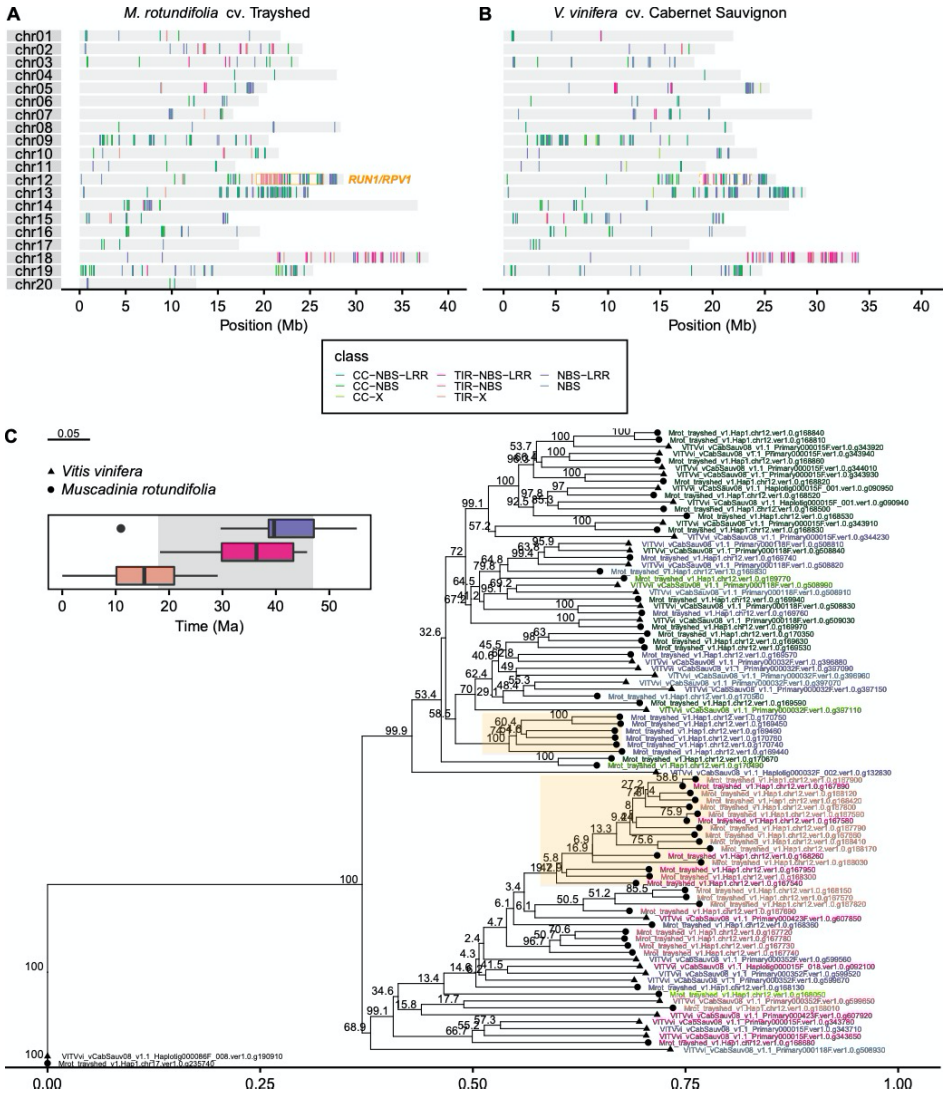

Figure 2: Distribution of NBS-LRR genes in Trayshed and Cabernet Sauvignon genomes

Location of the different classes of NBS-LRR along the different chromosomes of Trayshed (**A**) and Cabernet Sauvignon (**B**). The locations of the resistance locus *RUN1/RPV1* is indicated with an orange box. **C**. Phylogenetic tree of the NBS-LRR proteins in the *RUN1/RPV1* locus of both Trayshed and Cabernet Sauvignon. The boxplot represents an approximation of the expansion time of the gene clusters highlighted in orange. The approximation of the divergence time between *Vitis* and *Muscadinia* genera is highlighted in grey.

401 A tblastx was done to characterize gene duplication in the *RUN1/RPV1* locus in both Trayshed and
402 Cabernet Sauvignon. Within the resistance locus boundaries of both genomes (and excluding LTR
403 retrotransposons), 9,672 genomic regions of Cabernet Sauvignon aligned to 12,346 regions in
404 Trayshed. These regions intersected 217 Cabernet Sauvignon genes and 250 Trayshed genes.
405 Genic regions were included in 64.74% of the Cabernet Sauvignon sequences that successfully
406 aligned to Trayshed, and gene hits made up 67.10% of the total regions hit in Trayshed. On
407 average, one region in Cabernet Sauvignon had 1.78 hits in Trayshed. For the most duplicated
408 regions, a single sequence from Cabernet Sauvignon matched up to 15 different regions in
409 Trayshed (Figure 3). The most duplicated regions in Trayshed (with at least 10 hits matching a
410 single Cabernet Sauvignon region) were positioned between ~19.7 and 22 Mbp on chromosome
411 12 (Figure 3) and corresponded to the genomic region identified by Feechan *et al*. (2013). It
412 includes UDP-glucosyltransferase coding genes followed by resistance gene analogs. In Cabernet
413 Sauvignon, highly duplicated regions were mostly composed of kinetochore NDC80 complex
414 genes; these matched genes with similar or NBS-LRR functions in Trayshed.
415
416 In contrast to Cabernet Sauvignon, Trayshed has more functionally annotated disease-related
417 genes and an expansion of TIR-X that likely occurred after the divergence of their genera. Similar
418 research could be undertaken to characterize the genes or features at other known resistance loci
419 in *Muscadinia rotundifolia*, like *RUN2* and *REN5* (Riaz *et al.* 2011; Blanc *et al.* 2012). Acquiring
420 this understanding could be expedited with the availability of high-quality reference sequences for
421 resistant selections, like Trayshed, and susceptible cultivars, like Cabernet Sauvignon and others
422 (Massonnet *et al.* In Press).

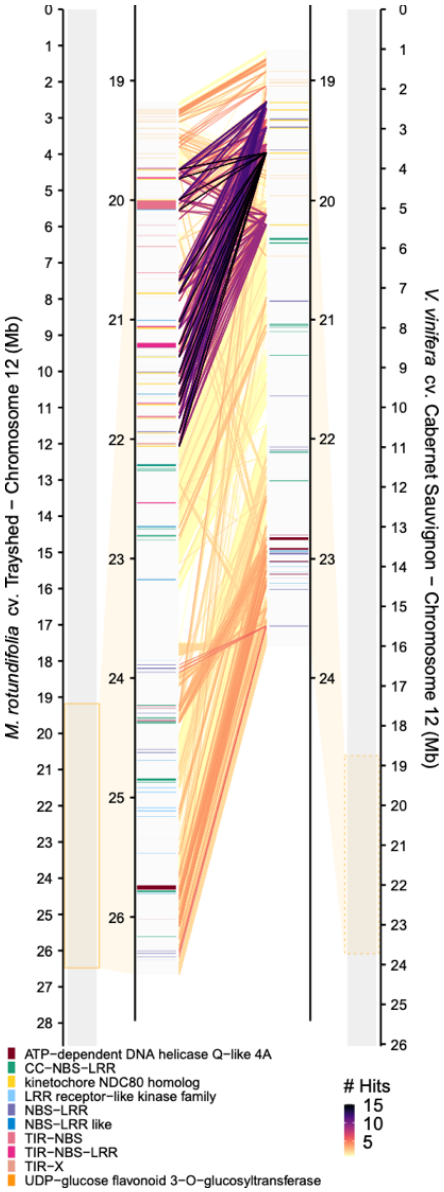

Figure 3: Synteny between the *RUN1/RPV1* locus in Trayshed and Cabernet Sauvignon
Chromosome 12 of Trayshed (left) and Cabernet Sauvignon (right) are shown. The *RUN1/RPV1* resistance locus is highlighted with an orange box for both genotypes. Syntenic gene content within the locus is shown in the center of the figure. Connections were labeled based on tblastx results. Duplications of Cabernet Sauvignon sequences are represented by the color gradient. Highly duplicated sequences in Trayshed are dark purple. The functional annotations depicted include descriptions represented by at least 10 genes or related to NBS-LRR.

ACKNOWLEDGMENTS
This work was funded by the NSF grant #1741627 and partially supported by funds to D.C. from the E. & J. Gallo Winery and the Louis P. Martini Endowment in Viticulture.

436 **LITERATURE CITED**

437 Au, K. F., J. G. Underwood, L. Lee, and W. H. Wong, 2012 Improving PacBio Long Read
438 Accuracy by Short Read Alignment (Y. Xing, Ed.). PLoS ONE 7: e46679.
439 Bailey, T. L., and C. Elkan, 1994 Fitting a mixture model by expectation maximization to
440 discover motifs in biopolymers. Proc Int Conf Intell Syst Mol Biol 2: 28–36.
441 Barker, C. L., T. Donald, J. Pauquet, M. B. Ratnaparkhe, A. Bouquet *et al.*, 2005 Genetic and
442 physical mapping of the grapevine powdery mildew resistance gene, Run1, using a
443 bacterial artificial chromosome library. Theor Appl Genet 111: 370–377.
444 Bertier, L., L. Leus, L. D'hondt, A. W. A. M. de Cock, and M. Höfte, 2013 Host Adaptation and
445 Speciation through Hybridization and Polyploidy in Phytophthora. PLOS ONE 8:
446 e85385.
447 Blanc, S., S. Wiedemann-Merdinoglu, V. Dumas, P. Mestre, and D. Merdinoglu, 2012 A
448 reference genetic map of Muscadinia rotundifolia and identification of Ren5, a new major
449 locus for resistance to grapevine powdery mildew. Theor Appl Genet 125: 1663–1675.
450 Blanco-Ulate, B., E. Vincenti, A. L. T. Powell, and D. Cantu, 2013 Tomato transcriptome and
451 mutant analyses suggest a role for plant stress hormones in the interaction between fruit
452 and Botrytis cinerea. Front. Plant Sci. 4: 142.
453 Boratyn, G. M., J. Thierry-Mieg, D. Thierry-Mieg, B. Busby, and T. L. Madden, 2019 Magic-
454 BLAST, an accurate RNA-seq aligner for long and short reads. BMC Bioinformatics 20:
455 405.
456 Bouquet, A., 1978 La muscadine (Vitis rotundifolia michx) et sa culture aux Etats-Unis. OENO
457 One 12: 1–20.
458 Bouquet, A., 1980 Vitis x Muscadinia hybridization: A new way in grape breeding for disease
459 resistance in France, pp. 42–61 in *Proceedings of the 3rd International Symposium on*
460 *Grape Breeding*, University of California, Davis, Department of Viticulture and Enology.
461 Branas, M. M., 1932 Sur la caryologie des Ampélidées. C R Acad Sci Paris 194: 121–123.
462 Camacho, C., G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos *et al.*, 2009 BLAST+:
463 architecture and applications. BMC Bioinformatics 10: 421.
464 Canaguier, A., J. Grimplet, G. Di Gaspero, S. Scalabrin, E. Duchene *et al.*, 2017 A new version
465 of the grapevine reference genome assembly (12X.v2) and of its annotation (VCost.v3).
466 Genom Data 14: 56–62.
467 Charif, D., and J. R. Lobry, 2007 SeqinR 1.0-2: A Contributed Package to the R Project for
468 Statistical Computing Devoted to Biological Sequences Retrieval and Analysis, pp. 207–
469 232 in *Structural Approaches to Sequence Evolution: Molecules, Networks, Populations*,
470 edited by U. Bastolla, M. Porto, H. E. Roman, and M. Vendruscolo. Springer Berlin
471 Heidelberg, Berlin, Heidelberg.
472 Chin, C.-S., D. H. Alexander, P. Marks, A. A. Klammer, J. Drake et al., 2013 Nonhybrid,
473 finished microbial genome assemblies from long-read SMRT sequencing data. Nat
474 Methods 10: 563–569.
475 Chin, C.-S., P. Peluso, F. J. Sedlazeck, M. Nattestad, G. T. Concepcion et al., 2016 Phased
476 diploid genome assembly with single-molecule real-time sequencing. Nat Methods
477 13:474 1050–1054.
478 Chu, Z.-F., J. Wen, Y.-P. Yang, Z.-L. Nie, and Y. Meng, 2018 Genome size variation and
479 evolution in the grape family Vitaceae: Genome size variation in Vitaceae. Jnl of
480 Sytematics Evolution 56: 273–282.

481 Davidis, U. X., and H. P. Olmo, 1964 The Vitis vinifera x V. rotundifolia hybrids as phylloxera
482 resistant rootstocks. Vitis 4: 129–143.
483 Dearing, C., 1917 Muscadine grape breeding. J. Hered. 8: 409–424.
484 Delame, M., E. Prado, S. Blanc, G. Robert-Siegwald, C. Schneider *et al.*, 2019 Introgression
485 reshapes recombination distribution in grapevine interspecific hybrids. Theor Appl Genet
486 132: 1073–1087.
487 Doležel, J., and J. Bartoš, 2005 Plant DNA Flow Cytometry and Estimation of Nuclear Genome
488 Size. Ann. Bot. 95: 99–110.
489 Doležel, J., S. Sgorbati, and S. Lucretti, 1992 Comparison of three DNA fluorochromes for flow
490 cytometric estimation of nuclear DNA content in plants. Physiol. Plant. 85: 625–631.
491 Edgar, R. C., 2004 MUSCLE: multiple sequence alignment with high accuracy and high
492 throughput. Nucleic Acids Res. 32: 1792–1797.
493 El-Gebali, S., J. Mistry, A. Bateman, S. R. Eddy, A. Luciani *et al.*, 2019 The Pfam protein
494 families database in 2019. Nucleic Acids Res. 47: D427–D432.
495 English, A. C., S. Richards, Y. Han, M. Wang, V. Vee et al., 2012 Mind the Gap: Upgrading
496 Genomes with Pacific Biosciences RS Long-Read Sequencing Technology. PLoS ONE
497 7: e47768.
498 English, A. C., W. J. Salerno, and J. G. Reid, 2014 PBHoney: identifying genomic variants via
499 long-read discordance and interrupted mapping. BMC Bioinformatics 15: 180.
500 Feechan, A., C. Anderson, L. Torregrosa, A. Jermakow, P. Mestre *et al.*, 2013 Genetic dissection
501 of a TIR-NB-LRR locus from the wild North American grapevine species *Muscadinia*
502 *rotundifolia* identifies paralogous genes conferring resistance to major fungal and
503 oomycete pathogens in cultivated grapevine. Plant J 76: 661–674.
504 Firoozabady, E., and H. P. Olmo, 1982 Resistance to grape phylloxera in Vitis vinifera x V.
505 rotundifolia grape hybrids. Vitis 21: 1–4.
506 Garnier, S., 2018 *viridis: Default Color Maps from "matplotlib."*
507 Gotz, S., J. M. Garcia-Gomez, J. Terol, T. D. Williams, S. H. Nagaraj *et al.*, 2008 High-
508 throughput functional annotation and data mining with the Blast2GO suite. Nucleic Acids
509 Res. 36: 3420–3435.
510 Grabherr, M. G., B. J. Haas, M. Yassour, J. Z. Levin, D. A. Thompson *et al.*, 2011 Full-length
511 transcriptome assembly from RNA-Seq data without a reference genome. Nat.
512 Biotechnol. 29: 644–652.
513 Grant, C. E., T. L. Bailey, and W. S. Noble, 2011 FIMO: scanning for occurrences of a given
514 motif. Bioinformatics 27: 1017–1018.
515 Haas, B. J., 2003 Improving the Arabidopsis genome annotation using maximal transcript
516 alignment assemblies. Nucleic Acids Res. 31: 5654–5666.
517 Haas, B. J., S. L. Salzberg, W. Zhu, M. Pertea, J. E. Allen *et al.*, 2008 Automated eukaryotic
518 gene structure annotation using EVidenceModeler and the Program to Assemble Spliced
519 Alignments. Genome Biol 9: R7.
520 He, H., S. Zhu, R. Zhao, Z. Jiang, Y. Ji *et al.*, 2018 Pm21, Encoding a Typical CC-NBS-LRR
521 Protein, Confers Broad-Spectrum Resistance to Wheat Powdery Mildew Disease. Mol.
522 Plant 11: 879–882.
523 Heinitz, C. C., J. Uretsky, J. C. Dodson Peterson, K. G. Huerta-Acosta, and M. A. Walker, 2019
524 Crop Wild Relatives of Grape (Vitis vinifera L.) Throughout North America, pp. 329–
525 351 in *North American Crop Wild Relatives, Volume 2: Important Species*, edited by S.

526 L. Greene, K. A. Williams, C. K. Khoury, M. B. Kantar, and L. F. Marek. Springer
527 International Publishing, Cham.
528 Holub, E. B., 2001 The arms race is ancient history in Arabidopsis, the wildflower. Nat Rev
529 Genet 2: 516–527.
530 Hulbert, S. H., C. A. Webb, S. M. Smith, and Q. Sun, 2001 RESISTANCE GENE
531 COMPLEXES: Evolution and Utilization. Annu. Rev. Phytopathol. 39: 285–312.
532 Jaillon, O., J. M. Aury, B. Noel, A. Policriti, C. Clepet *et al.*, 2007 The grapevine genome
533 sequence suggests ancestral hexaploidization in major angiosperm phyla. Nature 449:
534 463–468.
535 Jo, J., J. Venkatesh, K. Han, H.-Y. Lee, G. J. Choi *et al.*, 2017 Molecular Mapping of PMR1, a
536 Novel Locus Conferring Resistance to Powdery Mildew in Pepper (Capsicum annuum).
537 Front. Plant Sci. 8: 2090.
538 Jones, P., D. Binns, H.-Y. Chang, M. Fraser, W. Li *et al.*, 2014 InterProScan 5: genome-scale
539 protein function classification. Bioinformatics 30: 1236–1240.
540 Karkamkar, S. P., S. G. Patil, and S. C. Misra, 2010 Cyto–morphological studies and their
541 significance in evolution of family Vitaceae. Nucleus 53: 37–43.
542 Kent, W. J., 2002 BLAT : The Blast-Like Alignment Tool. Genome Res. 12: 656–664.
543 Korf, I., 2004 Gene finding in novel genomes. BMC Bioinformatics 5: 59.
544 Lam, E. T., A. Hastie, C. Lin, D. Ehrlich, S. K. Das *et al.*, 2012 Genome mapping on
545 nanochannel arrays for structural variation analysis and sequence assembly. Nat
546 Biotechnol 30: 771–776.
547 Lawrence, M., R. Gentleman, and V. Carey, 2009 rtracklayer: an R package for interfacing with
548 genome browsers. Bioinformatics 25: 1841–1842.
549 Lawrence, M., W. Huber, H. Pagès, P. Aboyoun, M. Carlson *et al.*, 2013 Software for
550 Computing and Annotating Genomic Ranges (A. Prlic, Ed.). PLoS Comput Biol 9:
551 e1003118.
552 Lewter, J., M. L. Worthington, J. R. Clark, A. V. Varanasi, L. Nelson *et al.*, 2019 High-density
553 linkage maps and loci for berry color and flower sex in muscadine grape (Vitis
554 rotundifolia). Theor Appl Genet 132: 1571–1585.
555 Liu, X.-Q., S. M. Ickert-Bond, Z.-L. Nie, Z. Zhou, L.-Q. Chen *et al.*, 2016 Phylogeny of the
556 Ampelocissus–Vitis clade in Vitaceae supports the New World origin of the grape genus.
557 Mol Phylogenet Evol 95: 217–228.
558 Lomsadze, A., 2005 Gene identification in novel eukaryotic genomes by self-training algorithm.
559 Nucleic Acids Res. 33: 6494–6506.
560 Lupas, A., M. Van Dyke, and J. Stock, 1991 Predicting coiled coils from protein sequences.
561 Science 252: 1162–1164.
562 Ma, Z.-Y., J. Wen, S. M. Ickert-Bond, Z.-L. Nie, L.-Q. Chen *et al.*, 2018 Phylogenomics,
563 biogeography, and adaptive radiation of grapes. Mol Phylogenet Evol 129: 258–267.
564 Marçais, G., A. L. Delcher, A. M. Phillippy, R. Coston, S. L. Salzberg *et al.*, 2018 MUMmer4: A
565 fast and versatile genome alignment system (A. E. Darling, Ed.). PLoS Comput Biol 14:
566 e1005944.
567 Massonnet, M., N. Cochetel, A. Minio, A. M. Vondras, A. Muyle *et al.*, In Press The genetic
568 basis of sex determination in grapevines (*Vitis spp.*). Nature Communications. doi:
569 10.1101/2019.12.11.861377.

570 Merdinoglu, D., C. Schneider, E. Prado, S. Wiedemann-Merdinoglu, and P. Mestre, 2018
571 Breeding for durable resistance to downy and powdery mildew in grapevine. OENO One
572 52: 203–209.
573 Merdinoglu, D., S. Wiedeman-Merdinoglu, P. Coste, V. Dumas, S. Haetty *et al.*, 2003 Genetic
574 analysis of downy mildew resistance derived from Muscadinia rotundifolia. Acta Hortic.
575 603: 451–456.
576 Meyers, B. C., A. Kozik, A. Griego, H. Kuang, and R. W. Michelmore, 2003 Genome-Wide
577 Analysis of NBS-LRR–Encoding Genes in Arabidopsis. Plant Cell 15: 809–834.
578 Minio, A., M. Massonnet, R. Figueroa-Balderas, A. Castro, and D. Cantu, 2019 Diploid Genome
579 Assembly of the Wine Grape Carménère. G3: Genes|Genomes|Genetics 9: 1331–1337.
580 Myers, G., 2014 Efficient Local Alignment Discovery amongst Noisy Long Reads, pp. 52–67 in
581 Algorithms in Bioinformatics, edited by D. Brown and B. Morgenstern. Springer Berlin
582 Heidelberg, Berlin, Heidelberg.
583 Olien, W. C., 1990 The Muscadine Grape: Botany, Viticulture, History, and Current Industry.
584 HortSci 25: 732–739.
585 Olmo, H. P., 1937 Chromosome Numbers in the European Grape (*Vitis vinifera*). CYTOLOGIA
586 Fujii Jubilae: 606–613.
587 Olmo, H. P., 1986 The potential role of (vinifera x rotundifolia) hybrids in grape variety
588 improvement. Experientia 42: 921–926.
589 Olmo, H. P., 1971 Vinifera Rotundifolia Hybrids as Wine Grapes. Am. J. Enol. Vitic. 22: 87–91.
590 Pagès, H., P. Aboyoun, R. Gentleman, and S. DebRoy, 2019 Biostrings: Efficient manipulation
591 of biological strings.
592 Paradis, E., and K. Schliep, 2019 ape 5.0: an environment for modern phylogenetics and
593 evolutionary analyses in R (R. Schwartz, Ed.). Bioinformatics 35: 526–528.
594 Patel, G. I., and H. P. Olmo, 1955 Cytogenetics of Vitis: I. The hybrid V. vinifera x V.
595 rotundifolia. Am. J. Bot. 42: 141–159.
596 Patil, S. G., and V. P. Patil, 1992 Karyomorphology of Vitis vinifera, V. rotundifolia and their
597 hybrid. Cytologia 57: 91–95.
598 Pauquet, J., A. Bouquet, P. This, and A.-F. Adam-Blondon, 2001 Establishment of a local map
599 of AFLP markers around the powdery mildew resistance gene Run1 in grapevine and
600 assessment of their usefulness for marker assisted selection. Theor Appl Genet 103:
601 1201–1210.
602 Pertea, M., G. M. Pertea, C. M. Antonescu, T.-C. Chang, J. T. Mendell *et al.*, 2015 StringTie
603 enables improved reconstruction of a transcriptome from RNA-seq reads. Nat Biotechnol
604 33: 290–295.
605 R Core Team, 2019 *R: A language and environment for statistical computing*. R Foundation for
606 Statistical Computing, Vienna, Austria.
607 Ranwez, V., E. J. P. Douzery, C. Cambon, N. Chantret, and F. Delsuc, 2018 MACSE v2: Toolkit
608 for the Alignment of Coding Sequences Accounting for Frameshifts and Stop Codons (C.
609 Wilke, Ed.). Mol Biol Evol 35: 2582–2584.
610 Ravaz, L., 1902 *Les Vignes américaines. Porte greffes et producteurs directs, caractères,*
611 *aptitudes*.
612 Riaz, S., A. C. Tenscher, D. W. Ramming, and M. A. Walker, 2011 Using a limited mapping
613 strategy to identify major QTLs for resistance to grapevine powdery mildew (Erysiphe
614 necator) and their use in marker-assisted breeding. Theor Appl Genet 122: 1059–1073.

615 Richly, E., J. Kurth, and D. Leister, 2002 Mode of Amplification and Reorganization of
616 Resistance Genes During Recent Arabidopsis thaliana Evolution. Mol Biol Evol 19: 76–
617 84.
618 RStudio Team, 2019 *RStudio: Integrated Development Environment for R*. RStudio, Inc.,
619 Boston, MA.
620 Ruel, J. J., and M. A. Walker, 2006 Resistance to Pierce's Disease in Muscadinia rotundifolia
621 and Other Native Grape Species. Am. J. Enol. Vitic. 57: 158–165.
622 Seppey, M., M. Manni, and E. M. Zdobnov, 2019 BUSCO: Assessing Genome Assembly and
623 Annotation Completeness, pp. 227–245 in *Gene Prediction: Methods and Protocols*,
624 edited by M. Kollmar. Springer New York, New York, NY.
625 Slater, G., and E. Birney, 2005 Automated generation of heuristics for biological sequence
626 comparison. BMC Bioinformatics 6: 31.
627 Small, J. K., 1913 *Flora of the southeastern United States: being descriptions of the seed-plants,*
628 *ferns and fern-allies growing naturally in North Carolina, South Carolina, Georgia,*
629 *Florida, Tennessee, Alabama, Mississippi, Arkansas, Louisiana and in Oklahoma and*
630 *Texas east of the one hundredth meridian*. J.K. Small, New York.
631 Smit, A., R. Hubley, and P. Green, 2013 *RepeatMasker Open-4.0*.
632 Sousa, A., and S. S. Renner, 2015 Interstitial telomere-like repeats in the monocot family
633 Araceae: Chromosome Evolution in the Araceae. Bot J Linn Soc 177: 15–26.
634 Stanke, M., O. Keller, I. Gunduz, A. Hayes, S. Waack *et al.*, 2006 AUGUSTUS: ab initio
635 prediction of alternative transcripts. Nucleic Acids Res. 34: W435–W439.
636 Staudt, G., and H. H. Kassemeyer, 1995 Evaluation of downy mildew resistance in various
637 accessions of wild Vitis species. Vitis 34: 225–228.
638 Vondras, A. M., A. Minio, B. Blanco-Ulate, R. Figueroa-Balderas, M. A. Penn *et al.*, 2019 The
639 genomic diversification of grapevine clones. BMC Genomics 20: 972.
640 Wagih, O., 2017 *ggseqlogo: A "ggplot2" Extension for Drawing Publication-Ready Sequence*.
641 Walker, M. A., L. A. Lider, A. C. Goheen, and H. P. Olmo, 1991 VR 039-16 Grape Rootstock.
642 HortSci 26: 1224–1225.
643 Walker, M. A., K. Lund, C. Agüero, S. Riaz, K. Fort *et al.*, 2014 Breeding grape rootstocks for
644 resistance to phylloxera and nematodes - it's not always easy. Acta Hortic. 1045: 89–97.
645 Wan, Y., H. R. Schwaninger, A. M. Baldo, J. A. Labate, G.-Y. Zhong *et al.*, 2013 A
646 phylogenetic analysis of the grape genus (Vitis L.) reveals broad reticulation and
647 concurrent diversification during neogene and quaternary climate change. BMC Evol
648 Biol 13: 141.
649 Wang, Y., H. Tang, J. D. DeBarry, X. Tan, J. Li *et al.*, 2012 MCScanX: a toolkit for detection
650 and evolutionary analysis of gene synteny and collinearity. Nucleic Acids Res. 40: e49.
651 Warren, R. F., A. Henk, P. Mowery, E. Holub, and R. W. Innes, 1998 A Mutation within the
652 Leucine-Rich Repeat Domain of the Arabidopsis Disease Resistance Gene RPS5 Partially
653 Suppresses Multiple Bacterial and Downy Mildew Resistance Genes. Plant Cell 10:
654 1439–1452.
655 Wei, F., K. Gobelman-Werner, S. M. Morroll, J. Kurth, and L. Mao, 1999 The Mla (Powdery
656 Mildew) Resistance Cluster Is Associated With Three NBS-LRR Gene Families and
657 Suppressed Recombination Within a 240-kb DNA Interval on Chromosome 5S (1HS) of
658 Barley. Genetics 153: 1929–1948.

659 Wen, J., L.-M. Lu, Z.-L. Nie, X.-Q. Liu, N. Zhang *et al.*, 2018 A new phylogenetic tribal
660 classification of the grape family (Vitaceae): Tribal classification of Vitaceae. Jnl of
661 Sytematics Evolution 56: 262–272.
662 Wickham, H., M. Averick, J. Bryan, W. Chang, L. McGowan *et al.*, 2019 Welcome to the
663 Tidyverse. Journal of Open Source Software 4: 1686.
664 Wu, T. D., and C. K. Watanabe, 2005 GMAP: a genomic mapping and alignment program for
665 mRNA and EST sequences. Bioinformatics 21: 1859–1875.
666 Yang, Z., 2007 PAML 4: Phylogenetic Analysis by Maximum Likelihood. Mol Biol Evol 24:
667 1586–1591.
668 Yu, G., D. K. Smith, H. Zhu, Y. Guan, and T. T.-Y. Lam, 2017 ggtree: an r package for
669 visualization and annotation of phylogenetic trees with their covariates and other
670 associated data. Methods Ecol. Evol. 8: 28–36.
671 Zhou, F., J. Kurth, F. Wei, C. Elliott, G. Valè *et al.*, 2001 Cell-Autonomous Expression of Barley
672 Mla1 Confers Race-Specific Resistance to the Powdery Mildew Fungus via a Rar1-
673 Independent Signaling Pathway. Plant Cell 13: 337–350.
674 Zini, E., C. Dolzani, M. Stefanini, V. Gratl, P. Bettinelli *et al.*, 2019 R-Loci Arrangement Versus
675 Downy and Powdery Mildew Resistance Level: A Vitis Hybrid Survey. IJMS 20: 3526.