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

for quantity and quality with the High Sensitivity chip in an Agilent 2100 Bioanalyzer (Agilent Technologies, CA) and was sequenced in 100bp, paired-end reads, using an Illumina HiSeq4000 sequencer (DNA Technology Core Facility, University of California, Davis). In addition, a cDNA SMRTbell library was prepared. First strand synthesis and cDNA amplification were accomplished using the NEBNext Single Cell/Low Input cDNA Synthesis & Amplification Module (New England, Ipswich, MA, US). These cDNAs were subsequently purified with ProNex magnetic beads (Promega, WI) following the instructions in the Iso-Seq Express Template Preparation for Sequel and Sequel II Systems protocol (Pacific Biosciences, Menlo Park, CA). ProNex magnetic beads (86 µl) were used to select amplified cDNA with a mode of 2 kb. At least 80 ng of the size-selected, amplified cDNA were used to prepare the cDNA SMRTbell library. This was followed by DNA damage repair and SMRTbell ligation using the SMRTbell Express Template Prep Kit 2.0 (Pacific Biosciences, Menlo Park, CA), following the manufacturer's protocol. One SMRT cell was sequenced on the PacBio Sequel I platform (DNA Technology Core Facility, University of California, Davis).

Genome annotation

The structural annotation of the *M. rotundifolia* genome was performed with a modified version of the pipeline used for the *Vitis vinifera* cv. Zinfandel genome (Vondras *et al.* 2019) and fully described here: https://github.com/andreaminio/AnnotationPipeline-EVM_based-DClab. Briefly, transcript evidence was obtained from external databases and by combining multiple strategies to build transcriptome assemblies from RNA-Seq sample using Stringtie v.1.3.4d (Pertea *et al.* 2015) and Trinity v.2.6.5 (Grabherr *et al.* 2011). In addition, PASA v.2.3.3 (Haas 2003) and high quality Iso-Seq data from Trayshed were used to produce high-quality gene models for training gene predictors. Public databases, transcriptome assemblies, and the Iso-Seq data described above were used as transcript and protein evidence. They were mapped on the genome using PASA v.2.3.3 (Haas 2003), MagicBLAST v.1.4.0 (Boratyn *et al.* 2019) and Exonerate v.2.2.0 (Slater and Birney 2005). *Ab initio* predictions were generated using BUSCO v.3.0.2 (Seppey *et al.* 2019), Augustus v.3.0.3 (Stanke *et al.* 2006), GeneMark v.3.47 (Lomsadze 2005) and SNAP v.2006-07-28 (Korf 2004). Repeats were annotated using RepeatMasker v.open-4.0.6 (Smit *et al.* 2013). Next, EvidenceModeler v.1.1.1 (Haas *et al.* 2008) used these predictions to generate consensus gene models. The final functional annotation was produced combining blastp v.2.2.28 (Camacho *et al.* 2009) hits against the Refseq plant protein database (ftp://ftp.ncbi.nlm.nih.gov/refseq, retrieved January 17th, 2019) and InterProScan v.5.28-67.0 (Jones *et al.* 2014) outputs using Blast2GO v.4.1.9 (Gotz *et al.* 2008).

Genome size quantification by flow cytometry

DNA content was estimated using flow cytometry (n = 3 individual leaves). *Lycopersicon esculentum* cv. Stupické polní tyčkové rané was selected as the internal DNA reference standard with a genome size of 2C = 1.96 pg; 1C = 958 Mb (Doležel *et al.* 1992). Nuclei extraction was performed using the Cystain PI absolute P kit (Sysmex America Inc). Approximately 5 mg (0.7 cm^2) of young healthy leaves from grapevine and tomato were cut finely with a razor blade in a Petri dish containing 500 µL of extraction buffer. The nuclei suspension was filtered through a 50 µm filter (CellTrics, Sysmex America Inc) and 2 mL of a propidium iodide staining solution was added (Doležel and Bartoš 2005; Bertier *et al.* 2013). Measurements were acquired using a Becton Dickinson FACScan (Franklin Lakes, NewJersey) equipped with a 488 nm laser. The data were analyzed using FlowJo v.10

(https://www.flowjo.com/solutions/flowjo). DNA content was inferred by linear regression using the tomato DNA reference standard. The genome size was comparable to PN40024 (Jaillon *et al.* 2007), at 487 Mb, but smaller than Cabernet Sauvignon (557.0 ± 2.4 Mbp; Supplemental Figure S1).

Comparison of genome assembly and synteny analysis

The Trayshed and Cabernet Sauvignon (Massonnet *et al.* In Press) genomes were compared with NUCmer (MUMmer v.4.0.00) (Marçais *et al.* 2018) and the --mum setting. Descriptive statistics of the alignment were obtained using the MUMmer "dnadiff" option. The delta file was filtered using delta-filter with these settings: -i 90 -l 7500. Blastp v.2.2.28 (Camacho *et al.* 2009) was used to align the annotated proteins of Trayshed and Cabernet Sauvignon. Pairwise protein information was associated with genes and processed with McScanX v.11.Nov.2013 (Wang *et al.* 2012) to identify syntenic regions.

Localization of PN40024 SNP markers and telomeric repeats analysis

Five hundred base-pair long genomic regions were extracted upstream and downstream of each PN40024 V0 SNP marker (Lewter *et al.,* 2019). GMAP v. 2015-11-20 (Wu and Watanabe 2005) was used to locate these sequences in Trayshed and Cabernet Sauvignon. The previously published consensus map of *M. rotundifolia* linkage groups was used to assess the completeness of the chromosome reconstruction (Lewter *et al.* 2019). Sliding windows (window size = 10 markers, sliding = 5 markers) were designed to move across the uniquely mapped markers on the genome. The percentage of relative marker positions consistent with their reference positions on the high-density genetic map was calculated per window. Telomeric repeats of "CCCTAA" and their reverse complement were searched along chromosome 7 of Cabernet Sauvignon and PN40024 V2 (Canaguier *et al.* 2017) using vmatchPattern from the R package Biostrings v.2.52.0 (H. Pagès *et al.* 2019). After peaks in the distribution of telomeric repeats were identified in both genomes, genomic regions of Cabernet Sauvignon (chr7:18,675,000-18,677,000 bp) and PN40024 (chr7:17571000-17573000 bp) were extracted. Motif enrichment analysis was performed using MEME v.5.1.0 (Bailey and Elkan 1994) and reported as "logo" using the ggseqlogo R package v.0.1 (Wagih 2017).

Identification of NBS-LRR genes

Predicted proteins from the Trayshed and Cabernet Sauvignon genomes were scanned using hmmsearch (HMMER v.3.3; http://hmmer.org/) with a sequence E-value threshold of 0.001 and Hidden Markov Models (HMM) corresponding to different Pfam (El-Gebali *et al.* 2019) domains: NB-ARC (Pfam PF00931), TIR (PF01582) and LRR (PF00560, PF07723, PF07725 and PF12799). Coiled coil (CC)-containing proteins were identified by COILS (Lupas *et al.* 1991) during the InterProScan annotation. This set of annotations was divided in six protein classes: CC-NBS-LRR, CC-NBS, TIR-NBS-LRR, TIR-NBS, NBS-LRR and NBS. To capture the largest number of potential NBS-LRR related genes, genes lacking the NB-ARC domain but with "NBS-LRR" functional annotations were also selected and divided in two classes: TIR-X and CC-X. For these eight protein classes, Multiple EM for Motif Elicitation (MEME) analysis was performed with the flags -mod anr -nmotifs 20 to identify conserved domains for each class. Then, FIMO v.5.1.0 (Find Individual Motif Occurences) (Grant *et al.* 2011) was ran on each protein class using the corresponding MEME results; proteins with at least five conserved domains were kept. NBS-LRR gene clusters were defined as groups of at least two NBS-LRR genes, each separated by no

more than eight non-NBS-LRR genes (Richly *et al.* 2002) in a region spanning a maximum of 200 kbp (Holub 2001).

RUN1/RPV1 locus analysis and NBS-LRR genes phylogeny

Boundaries of the *RUN1/RPV1* locus were defined by mapping two markers, VMC4f3.1 and VMC8g9 (Barker *et al.* 2005), on Trayshed and Cabernet Sauvignon using GMAP v. 2015-11-20 (Wu and Watanabe 2005). A tblastx (BLAST v.2.2.29) (Camacho *et al.* 2009) was performed using the genomic region corresponding to the resistance locus in Cabernet Sauvignon as a query and Trayshed chromosome 12, split in 50-kbp blocks, as a reference. The tblastx results were filtered, keeping hits with at least 90% of identity, a minimum length of 100 bp, and that were contained within the boundaries of the resistance locus in both the query and the reference. Lastly, hits matching LTR retrotransposons were discarded. Proteins corresponding to the first alternative spliced variant of each NBS-LRR gene of the *RUN1/RPV1* locus were aligned using MUSCLE v.3.8.31 (Edgar 2004) with default parameters and by including *GAPDH* (*VIT_17s0000g10430*) ortholog sequences as an outgroup. Distances between the alignments were extracted using the R package seqinr v.3.6.1 (Charif and Lobry 2007). The estimation of the corresponding neighbor-joining tree, its rooting, and the bootstrapping (1000 replicates) were performed using the R package ape v.5.3 (Paradis and Schliep 2019). The tree was drawn using ggtree v.2.0.2 (Yu *et al.* 2017). Coding sequences of the two Trayshed-specific gene clusters were aligned per class using MACSE v.2.03 (Ranwez *et al.* 2018) and trimmed using the option min_percent_NT_at_ends=0.80. Prior to the alignment, outliers were removed based on sequence length. Synonymous substitution rates (dS) were obtained using yn00 from PAML v.4.9f (Yang 2007).

Data analysis and visualization

Data were parsed with R v.3.6.0 (R Core Team 2019) in RStudio v.1.2.5033 (RStudio Team 2019) using tidyverse v.1.3.0 (Wickham *et al.* 2019), GenomicFeatures v.1.36.4 for genomic ranges manipulation (Lawrence *et al.* 2013), rtracklayer for GFF files v.1.44.4 (Lawrence *et al.* 2009) and viridis v.0.5.1 for the color gradients (Garnier 2018).

Data availability

Sequencing data are accessible at NCBI repository under the accession PRJNA635946 and PRJNA593045. Raw optical maps are available at Zenodo under the DOI 10.5281/zenodo.3866087. The supplemental files are available at Figshare. The pipeline for the gene annotation is available at https://github.com/andreaminio/AnnotationPipeline-EVM_based-DClab. Assembly and annotation files are available at www.grapegenomics.com, which also hosts a genome browser and a blast tool for Trayshed.

RESULTS AND DISCUSSION

Construction of the twenty phased chromosomes of *Muscadinia rotundifolia* cv. Trayshed

We reconstructed the twenty Trayshed chromosomes using optical maps produced with BioNano Genomics technology at 2,057x coverage. The BioNano consensus genome map was combined with the contigs produced from Single Molecule Real-Time (SMRT; Pacific Biosciences) long-reads to generate a hybrid genome assembly. The size of the hybrid assembly (898.5 Mbp) was

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

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

Analysis of chromosome structure supports a fusion of chromosomes 7 and 20 in *Vitis*

MUMMER was used to compare the structure of Trayshed with the chromosome-scale Cabernet Sauvignon assembly (Massonnet *et al.* In Press), which is the most complete and contiguous genome assembly of grapevine available to date. The two genomes were highly syntenic (Figure 1A). On average, colinear regions were 2,484 bp long and shared 92.46% identity (Supplementary Table S1). On average, insertions were 25-bp long and deletions were 15-bp long. The longest deletion (~50 kb) was detected on chromosome 8 of Cabernet Sauvignon while the longest insertion (78 kb) was observed on chromosome 11 (Supplemental Table S2).

Trayshed chromosomes 7 and 20 aligned at the beginning and end, respectively, of Cabernet Sauvignon chromosome 7. This confirms the fusion of these two chromosomes in *Vitis* that was previously reported using genetic maps (Blanc *et al.* 2012; Lewter *et al.* 2019) (Figure 1A).

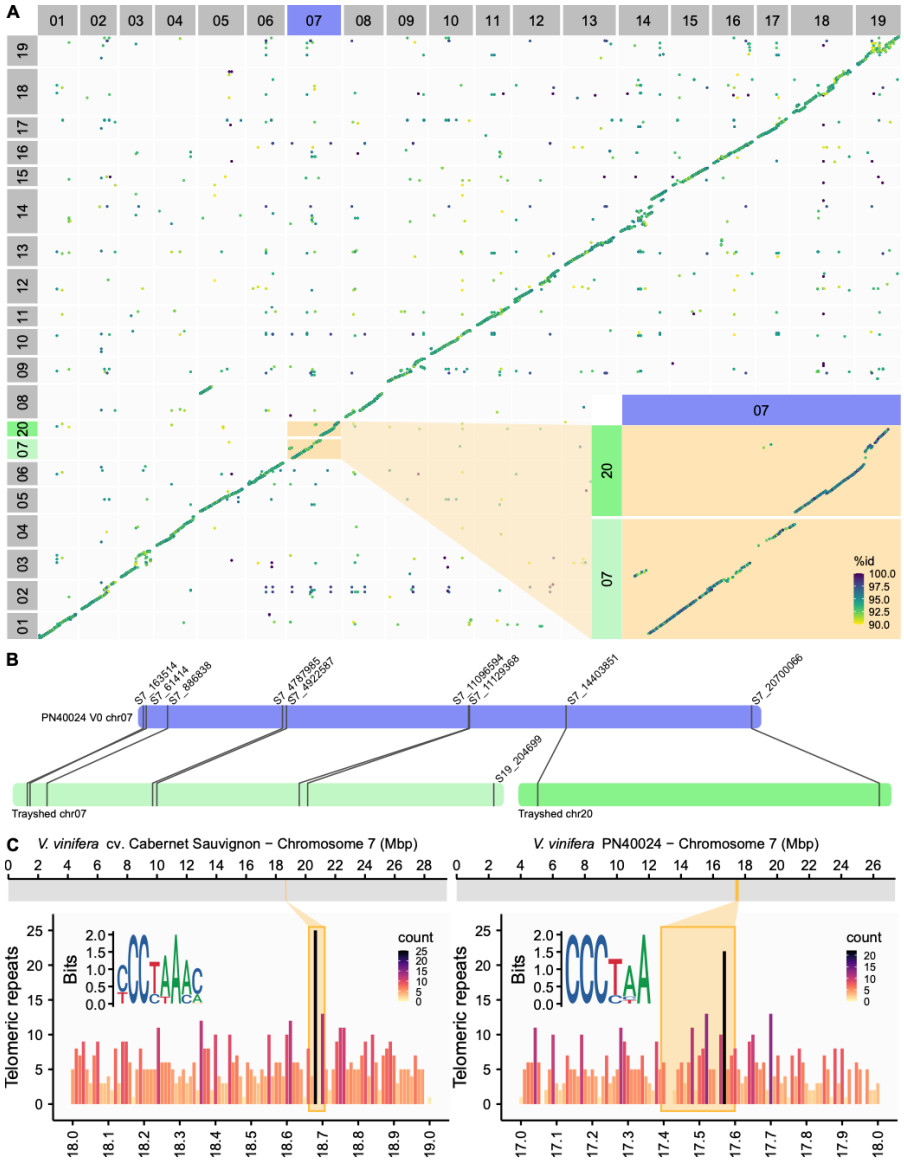

Figure 1: Chromosome fusion localization in *Vitis vinifera*
A. Whole genome alignment of Trayshed (y-axis) on Cabernet Sauvignon (x-axis). The putative fusion of Trayshed chromosomes 7 and 20 of Trayshed (green) and Cabernet Sauvignon chromosome 7th (purple) is inset. The percent identity (% id) between 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

Cabernet Sauvignon, we searched for telomeric repeats in the region of chromosome 7 that should contain the fusion point based on nucmer alignments. We examined the genomic region of Cabernet Sauvignon chromosome 7 that spanned Trayshed chromosome 7 and 20 (chr 7: 18,661,815 – 18,705,855 bp; Figure 1C, top panel) and found an enrichment of telomeric repeats in this region. (Figure 1C, bottom panel). A similar pattern was detected in the hypothetical fusion region of PN40024 (chr 7: 17,392,916 - 17,599,237 bp; Figure 1C). Together, these data support a chromosome number reduction by fusion in *V. vinifera* when compared with *M. rotundifolia*.

Full length cDNA sequencing and annotation of the protein-coding genes

Together with the public data and RNA-Seq data already available, high-quality full-length cDNA sequences (Iso-Seq; Pacific Biosciences, Supplemental Figure S4) were generated to be used as transcriptional evidence for gene model predictions. From leaf tissues, 336,932 raw reads were sequenced, clustered, and error corrected with LSC (Au *et al.* 2012) to generate 34,558 high-quality isoforms, 260 low-quality isoforms, and 111,672 singletons (Supplemental Figure S4). After obtaining consensus gene models, alternative splice variants were predicted if supported by Iso-Seq data, with 1.58 transcripts per gene on average (Supplemental Figure S4). BUSCO genes (97.6%) were well-represented in the diploid Trayshed transcriptome. The structural annotation included 52,414 protein coding genes (Table 1) and 83,010 proteins (including the alternative forms). There were 27,924 and 17,829 genes on haplotypes 1 and 2, respectively (Table 1). The repetitive content composed 39.90% of the genome. On average, colinear blocks of homologous genes were ~ 28-genes long (Supplemental Table S3) based on McScanX results. The longest block of colinear genes was found on chromosome 6 and contained 695 genes (Supplemental Table S3). Based on blastp results with an inclusive filtering threshold set at 70% for both query and target coverage and identity, 50,670 (96.7%) Trayshed genes are homologous with 52,684 (91.78%) genes of Cabernet Sauvignon. The 1,890 proteins that did not have a potential ortholog in Cabernet Sauvignon were significantly enriched in functional categories related to compound metabolic process ($P < 1e-04$; Fisher's exact test).

NBS-LRR intracellular receptors encoded in the Trayshed genome

Trayshed is attractive to grape breeding programs because it is resistant to pathogens that affect *V. vinifera*. In plants, resistance to pathogens is primarily attributed to the activity of resistance genes (R-genes). Nucleotide binding site leucine-rich repeat (NBS-LRR) genes constitute the largest family of R-genes. NBS-LRR genes are associated with resistance to powdery and downy mildew pathogens in several species, including *Arabidopsis* (Warren *et al.* 1998), wheat (He *et al.* 2018), barley (Wei *et al.* 1999; Zhou *et al.* 2001), pepper (Jo *et al.* 2017), and grapes (Riaz *et al.* 2011; Zini *et al.* 2019).

We divided the NBS-LRR genes in eight different classes depending on the domain detected in the proteins: CC-NBS-LRR, CC-NBS, CC-X, TIR-NBS-LRR, TIR-NBS, TIR-X, NBS-LRR, NBS (called hereafter NBS-LRR when described as a whole). Overall, the Trayshed genome has slightly more NBS-LRR genes (1,075) than Cabernet Sauvignon (1,009). Most NBS-LRR genes were found on chromosomes 12 and 13 (91 genes on each; 1.54 and 1.21x more than Cabernet Sauvignon, respectively), followed by chromosome 19 (58 genes; 1.49x more than Cabernet Sauvignon) (Figure 2A, Supplementary Table S4). In both species, chromosome 4 contained the fewest NBS-LRR genes (2). In Cabernet Sauvignon (Figure 2B, Supplementary Table S4), most NBS-LRR genes were detected on chromosome 13 (75 genes), 9 (72 genes; 1.57x more than

Trayshed) and 18 (65 genes; 1.38x more than Trayshed). Within the genome, NBS-LRR genes form clusters that could be critical to their function as resistance loci (Hulbert *et al.* 2001; Meyers *et al.* 2003). The number of clusters detected in Trayshed (123) was slightly higher than in Cabernet Sauvignon (121), but their arrangement in the genome varied substantially (Supplemental Figure S5). Chromosomes 12 and 13 contain the largest number of clusters (23 and 22 clusters, respectively) in Trayshed. In Cabernet Sauvignon, clusters were the most numerous on chromosome 13 and 18 (20 and 18 clusters, respectively). These differences could plausibly reflect to disparities in disease resistance.

Expansion of NBS-LRR genes at the *RUN1/RPV1* locus in Trayshed

The clusters of NBS-LRR genes on chromosome 12 in Trayshed are mostly absent from Cabernet Sauvignon (Figure 2A and 2B). This chromosome contains a known powdery mildew resistance locus, *RUN1/RPV1* (Pauquet *et al.* 2001), with boundaries defined by the VMC4f3.1 and VMC8g9 SSR markers (Barker *et al.* 2005). This region was studied in Trayshed (chr12: 19,181,246 – 26,483,447 bp) and Cabernet Sauvignon (chr12: 18,746,148 – 23,728,139 bp); it included 355 annotated genes in Trayshed and 287 genes in Cabernet Sauvignon. Within the *RUN1/RPV1* locus, 57 NBS-LRR were identified in Trayshed and 33 in Cabernet Sauvignon.

Notably, most of the NBS-LRR classes are present for both species in this region, but the TIR-X class is only present in Trayshed. A phylogenetic analysis of the NBS-LRR proteins at the *RUN1/RPV1* resistance locus identified two Trayshed-specific clusters (Figure 2C). The timing of the divergence of these sequences was estimated as described in Massonnet *et al*. (In Press). This involved aligning corresponding gene coding sequences with MACSE (Ranwez *et al.* 2018) and calculating the rate of synonymous mutations with yn00 from PAML (Yang 2007). This method timed the divergence of TIR-X genes at nearly 16 Ma, after the divergence of the *Muscadinia* and *Vitis* genera (about 18 Ma) (Wan *et al.* 2013). Thus, a recent expansion event of the TIR-X class genes might have occurred in only *Muscadinia* and other non-TIR classes would have expanded prior.

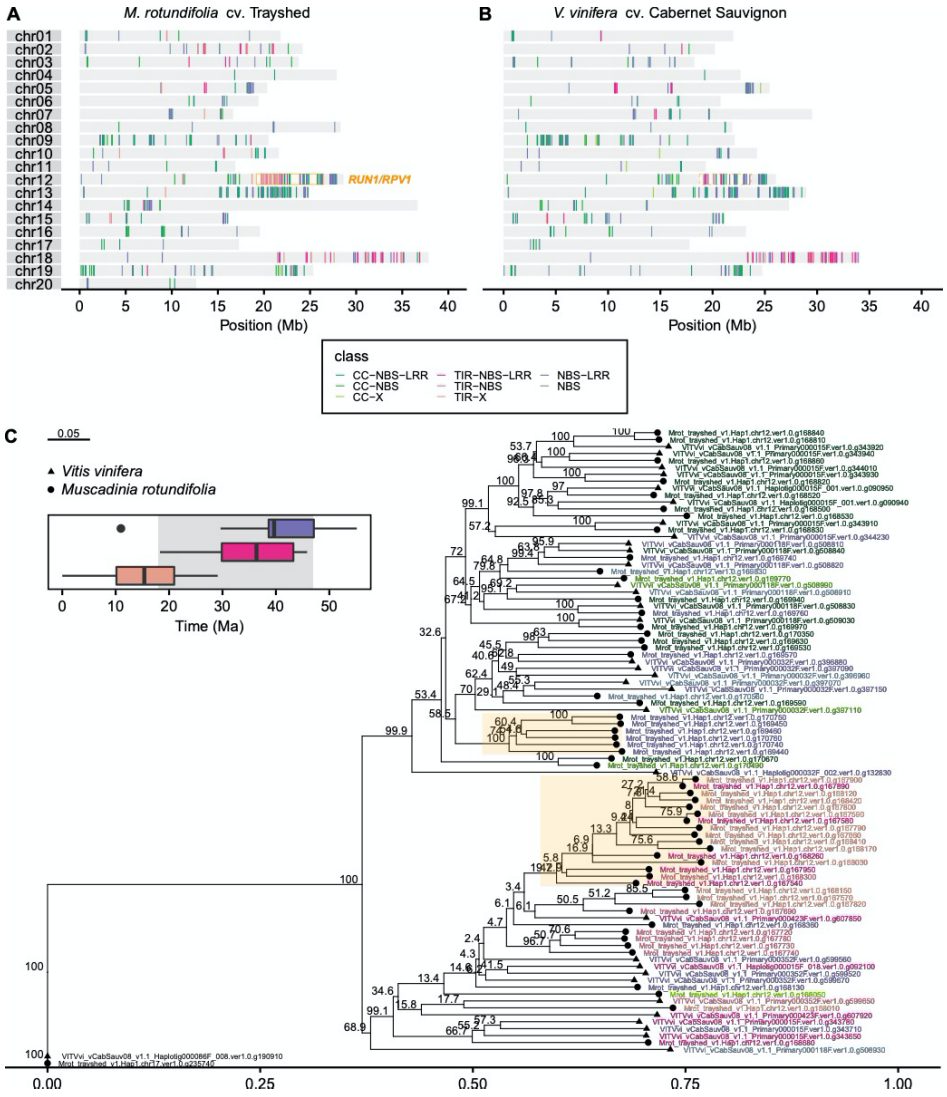

Figure 2: Distribution of NBS-LRR genes in Trayshed and Cabernet Sauvignon genomes
Location of the different classes of NBS-LRR along the different chromosomes of Trayshed (**A**) and Cabernet Sauvignon (**B**). The locations of the resistance locus *RUN1/RPV1* is indicated with an orange box. **C**. 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.

A tblastx was done to characterize gene duplication in the *RUN1/RPV1* locus in both Trayshed and Cabernet Sauvignon. Within the resistance locus boundaries of both genomes (and excluding LTR retrotransposons), 9,672 genomic regions of Cabernet Sauvignon aligned to 12,346 regions in Trayshed. These regions intersected 217 Cabernet Sauvignon genes and 250 Trayshed genes. Genic regions were included in 64.74% of the Cabernet Sauvignon sequences that successfully aligned to Trayshed, and gene hits made up 67.10% of the total regions hit in Trayshed. On average, one region in Cabernet Sauvignon had 1.78 hits in Trayshed. For the most duplicated regions, a single sequence from Cabernet Sauvignon matched up to 15 different regions in Trayshed (Figure 3). The most duplicated regions in Trayshed (with at least 10 hits matching a single Cabernet Sauvignon region) were positioned between ~19.7 and 22 Mbp on chromosome 12 (Figure 3) and corresponded to the genomic region identified by Feechan *et al*. (2013). It includes UDP-glucosyltransferase coding genes followed by resistance gene analogs. In Cabernet Sauvignon, highly duplicated regions were mostly composed of kinetochore NDC80 complex genes; these matched genes with similar or NBS-LRR functions in Trayshed.

In contrast to Cabernet Sauvignon, Trayshed has more functionally annotated disease-related genes and an expansion of TIR-X that likely occurred after the divergence of their genera. Similar research could be undertaken to characterize the genes or features at other known resistance loci in *Muscadinia rotundifolia*, like *RUN2* and *REN5* (Riaz *et al.* 2011; Blanc *et al.* 2012). Acquiring this understanding could be expedited with the availability of high-quality reference sequences for resistant selections, like Trayshed, and susceptible cultivars, like Cabernet Sauvignon and others (Massonnet *et al.* In Press).

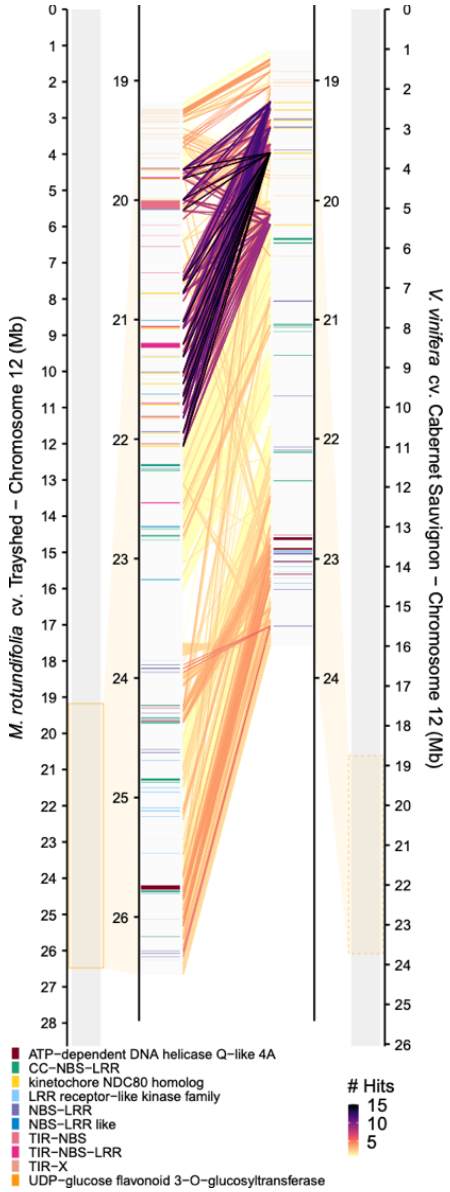

Figure 3: Synteny between the *RUN1/RPV1* locus in Trayshed and Cabernet Sauvignon
Chromosome 12 of Trayshed (left) and Cabernet Sauvignon (right) are shown. The *RUN1/RPV1* resistance locus is highlighted with an orange box for both genotypes. Syntenic gene content within the locus is shown in the center 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.

LITERATURE CITED

Au, K. F., J. G. Underwood, L. Lee, and W. H. Wong, 2012 Improving PacBio Long Read Accuracy by Short Read Alignment (Y. Xing, Ed.). PLoS ONE 7: e46679.

Bailey, T. L., and C. Elkan, 1994 Fitting a mixture model by expectation maximization to discover motifs in biopolymers. Proc Int Conf Intell Syst Mol Biol 2: 28–36.

Barker, C. L., T. Donald, J. Pauquet, M. B. Ratnaparkhe, A. Bouquet *et al.*, 2005 Genetic and physical mapping of the grapevine powdery mildew resistance gene, Run1, using a bacterial artificial chromosome library. Theor Appl Genet 111: 370–377.

Bertier, L., L. Leus, L. D'hondt, A. W. A. M. de Cock, and M. Höfte, 2013 Host Adaptation and Speciation through Hybridization and Polyploidy in Phytophthora. PLOS ONE 8: e85385.

Blanc, S., S. Wiedemann-Merdinoglu, V. Dumas, P. Mestre, and D. Merdinoglu, 2012 A reference genetic map of Muscadinia rotundifolia and identification of Ren5, a new major locus for resistance to grapevine powdery mildew. Theor Appl Genet 125: 1663–1675.

Blanco-Ulate, B., E. Vincenti, A. L. T. Powell, and D. Cantu, 2013 Tomato transcriptome and mutant analyses suggest a role for plant stress hormones in the interaction between fruit and Botrytis cinerea. Front. Plant Sci. 4: 142.

Boratyn, G. M., J. Thierry-Mieg, D. Thierry-Mieg, B. Busby, and T. L. Madden, 2019 Magic-BLAST, an accurate RNA-seq aligner for long and short reads. BMC Bioinformatics 20: 405.

Bouquet, A., 1978 La muscadine (Vitis rotundifolia michx) et sa culture aux Etats-Unis. OENO One 12: 1–20.

Bouquet, A., 1980 Vitis x Muscadinia hybridization: A new way in grape breeding for disease resistance in France, pp. 42–61 in *Proceedings of the 3rd International Symposium on Grape Breeding*, University of California, Davis, Department of Viticulture and Enology.

Branas, M. M., 1932 Sur la caryologie des Ampélidées. C R Acad Sci Paris 194: 121–123.

Camacho, C., G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos *et al.*, 2009 BLAST+: architecture and applications. BMC Bioinformatics 10: 421.

Canaguier, A., J. Grimplet, G. Di Gaspero, S. Scalabrin, E. Duchene *et al.*, 2017 A new version of the grapevine reference genome assembly (12X.v2) and of its annotation (VCost.v3). Genom Data 14: 56–62.

Charif, D., and J. R. Lobry, 2007 SeqinR 1.0-2: A Contributed Package to the R Project for Statistical Computing Devoted to Biological Sequences Retrieval and Analysis, pp. 207–232 in *Structural Approaches to Sequence Evolution: Molecules, Networks, Populations*, edited by U. Bastolla, M. Porto, H. E. Roman, and M. Vendruscolo. Springer Berlin Heidelberg, Berlin, Heidelberg.

Chin, C.-S., D. H. Alexander, P. Marks, A. A. Klammer, J. Drake et al., 2013 Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. Nat Methods 10: 563–569.

Chin, C.-S., P. Peluso, F. J. Sedlazeck, M. Nattestad, G. T. Concepcion et al., 2016 Phased diploid genome assembly with single-molecule real-time sequencing. Nat Methods 13:474 1050–1054.

Chu, Z.-F., J. Wen, Y.-P. Yang, Z.-L. Nie, and Y. Meng, 2018 Genome size variation and evolution in the grape family Vitaceae: Genome size variation in Vitaceae. Jnl of Sytematics Evolution 56: 273–282.

Davidis, U. X., and H. P. Olmo, 1964 The Vitis vinifera x V. rotundifolia hybrids as phylloxera resistant rootstocks. Vitis 4: 129–143.

Dearing, C., 1917 Muscadine grape breeding. J. Hered. 8: 409–424.

Delame, M., E. Prado, S. Blanc, G. Robert-Siegwald, C. Schneider *et al.*, 2019 Introgression reshapes recombination distribution in grapevine interspecific hybrids. Theor Appl Genet 132: 1073–1087.

Doležel, J., and J. Bartoš, 2005 Plant DNA Flow Cytometry and Estimation of Nuclear Genome Size. Ann. Bot. 95: 99–110.

Doležel, J., S. Sgorbati, and S. Lucretti, 1992 Comparison of three DNA fluorochromes for flow cytometric estimation of nuclear DNA content in plants. Physiol. Plant. 85: 625–631.

Edgar, R. C., 2004 MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32: 1792–1797.

El-Gebali, S., J. Mistry, A. Bateman, S. R. Eddy, A. Luciani *et al.*, 2019 The Pfam protein families database in 2019. Nucleic Acids Res. 47: D427–D432.

English, A. C., S. Richards, Y. Han, M. Wang, V. Vee et al., 2012 Mind the Gap: Upgrading Genomes with Pacific Biosciences RS Long-Read Sequencing Technology. PLoS ONE 7: e47768.

English, A. C., W. J. Salerno, and J. G. Reid, 2014 PBHoney: identifying genomic variants via long-read discordance and interrupted mapping. BMC Bioinformatics 15: 180.

Feechan, A., C. Anderson, L. Torregrosa, A. Jermakow, P. Mestre *et al.*, 2013 Genetic dissection of a TIR-NB-LRR locus from the wild North American grapevine species *Muscadinia rotundifolia* identifies paralogous genes conferring resistance to major fungal and oomycete pathogens in cultivated grapevine. Plant J 76: 661–674.

Firoozabady, E., and H. P. Olmo, 1982 Resistance to grape phylloxera in Vitis vinifera x V. rotundifolia grape hybrids. Vitis 21: 1–4.

Garnier, S., 2018 *viridis: Default Color Maps from "matplotlib."*

Gotz, S., J. M. Garcia-Gomez, J. Terol, T. D. Williams, S. H. Nagaraj *et al.*, 2008 High-throughput functional annotation and data mining with the Blast2GO suite. Nucleic Acids Res. 36: 3420–3435.

Grabherr, M. G., B. J. Haas, M. Yassour, J. Z. Levin, D. A. Thompson *et al.*, 2011 Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat. Biotechnol. 29: 644–652.

Grant, C. E., T. L. Bailey, and W. S. Noble, 2011 FIMO: scanning for occurrences of a given motif. Bioinformatics 27: 1017–1018.

Haas, B. J., 2003 Improving the Arabidopsis genome annotation using maximal transcript alignment assemblies. Nucleic Acids Res. 31: 5654–5666.

Haas, B. J., S. L. Salzberg, W. Zhu, M. Pertea, J. E. Allen *et al.*, 2008 Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. Genome Biol 9: R7.

He, H., S. Zhu, R. Zhao, Z. Jiang, Y. Ji *et al.*, 2018 Pm21, Encoding a Typical CC-NBS-LRR Protein, Confers Broad-Spectrum Resistance to Wheat Powdery Mildew Disease. Mol. Plant 11: 879–882.

Heinitz, C. C., J. Uretsky, J. C. Dodson Peterson, K. G. Huerta-Acosta, and M. A. Walker, 2019 Crop Wild Relatives of Grape (Vitis vinifera L.) Throughout North America, pp. 329–351 in *North American Crop Wild Relatives, Volume 2: Important Species*, edited by S.

L. Greene, K. A. Williams, C. K. Khoury, M. B. Kantar, and L. F. Marek. Springer International Publishing, Cham.

Holub, E. B., 2001 The arms race is ancient history in Arabidopsis, the wildflower. Nat Rev Genet 2: 516–527.

Hulbert, S. H., C. A. Webb, S. M. Smith, and Q. Sun, 2001 RESISTANCE GENE COMPLEXES: Evolution and Utilization. Annu. Rev. Phytopathol. 39: 285–312.

Jaillon, O., J. M. Aury, B. Noel, A. Policriti, C. Clepet *et al.*, 2007 The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. Nature 449: 463–468.

Jo, J., J. Venkatesh, K. Han, H.-Y. Lee, G. J. Choi *et al.*, 2017 Molecular Mapping of PMR1, a Novel Locus Conferring Resistance to Powdery Mildew in Pepper (Capsicum annuum). Front. Plant Sci. 8: 2090.

Jones, P., D. Binns, H.-Y. Chang, M. Fraser, W. Li *et al.*, 2014 InterProScan 5: genome-scale protein function classification. Bioinformatics 30: 1236–1240.

Karkamkar, S. P., S. G. Patil, and S. C. Misra, 2010 Cyto–morphological studies and their significance in evolution of family Vitaceae. Nucleus 53: 37–43.

Kent, W. J., 2002 BLAT : The Blast-Like Alignment Tool. Genome Res. 12: 656–664.

Korf, I., 2004 Gene finding in novel genomes. BMC Bioinformatics 5: 59.

Lam, E. T., A. Hastie, C. Lin, D. Ehrlich, S. K. Das *et al.*, 2012 Genome mapping on nanochannel arrays for structural variation analysis and sequence assembly. Nat Biotechnol 30: 771–776.

Lawrence, M., R. Gentleman, and V. Carey, 2009 rtracklayer: an R package for interfacing with genome browsers. Bioinformatics 25: 1841–1842.

Lawrence, M., W. Huber, H. Pagès, P. Aboyoun, M. Carlson *et al.*, 2013 Software for Computing and Annotating Genomic Ranges (A. Prlic, Ed.). PLoS Comput Biol 9: e1003118.

Lewter, J., M. L. Worthington, J. R. Clark, A. V. Varanasi, L. Nelson *et al.*, 2019 High-density linkage maps and loci for berry color and flower sex in muscadine grape (Vitis rotundifolia). Theor Appl Genet 132: 1571–1585.

Liu, X.-Q., S. M. Ickert-Bond, Z.-L. Nie, Z. Zhou, L.-Q. Chen *et al.*, 2016 Phylogeny of the Ampelocissus–Vitis clade in Vitaceae supports the New World origin of the grape genus. Mol Phylogenet Evol 95: 217–228.

Lomsadze, A., 2005 Gene identification in novel eukaryotic genomes by self-training algorithm. Nucleic Acids Res. 33: 6494–6506.

Lupas, A., M. Van Dyke, and J. Stock, 1991 Predicting coiled coils from protein sequences. Science 252: 1162–1164.

Ma, Z.-Y., J. Wen, S. M. Ickert-Bond, Z.-L. Nie, L.-Q. Chen *et al.*, 2018 Phylogenomics, biogeography, and adaptive radiation of grapes. Mol Phylogenet Evol 129: 258–267.

Marçais, G., A. L. Delcher, A. M. Phillippy, R. Coston, S. L. Salzberg *et al.*, 2018 MUMmer4: A fast and versatile genome alignment system (A. E. Darling, Ed.). PLoS Comput Biol 14: e1005944.

Massonnet, M., N. Cochetel, A. Minio, A. M. Vondras, A. Muyle *et al.*, In Press The genetic basis of sex determination in grapevines (*Vitis spp.*). Nature Communications. doi: 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.

Richly, E., J. Kurth, and D. Leister, 2002 Mode of Amplification and Reorganization of Resistance Genes During Recent Arabidopsis thaliana Evolution. Mol Biol Evol 19: 76–84.

RStudio Team, 2019 *RStudio: Integrated Development Environment for R*. RStudio, Inc., Boston, MA.

Ruel, J. J., and M. A. Walker, 2006 Resistance to Pierce's Disease in Muscadinia rotundifolia and Other Native Grape Species. Am. J. Enol. Vitic. 57: 158–165.

Seppey, M., M. Manni, and E. M. Zdobnov, 2019 BUSCO: Assessing Genome Assembly and Annotation Completeness, pp. 227–245 in *Gene Prediction: Methods and Protocols*, edited by M. Kollmar. Springer New York, New York, NY.

Slater, G., and E. Birney, 2005 Automated generation of heuristics for biological sequence comparison. BMC Bioinformatics 6: 31.

Small, J. K., 1913 *Flora of the southeastern United States: being descriptions of the seed-plants, ferns and fern-allies growing naturally in North Carolina, South Carolina, Georgia, Florida, Tennessee, Alabama, Mississippi, Arkansas, Louisiana and in Oklahoma and Texas east of the one hundredth meridian*. J.K. Small, New York.

Smit, A., R. Hubley, and P. Green, 2013 *RepeatMasker Open-4.0*.

Sousa, A., and S. S. Renner, 2015 Interstitial telomere-like repeats in the monocot family Araceae: Chromosome Evolution in the Araceae. Bot J Linn Soc 177: 15–26.

Stanke, M., O. Keller, I. Gunduz, A. Hayes, S. Waack *et al.*, 2006 AUGUSTUS: ab initio prediction of alternative transcripts. Nucleic Acids Res. 34: W435–W439.

Staudt, G., and H. H. Kassemeyer, 1995 Evaluation of downy mildew resistance in various accessions of wild Vitis species. Vitis 34: 225–228.

Vondras, A. M., A. Minio, B. Blanco-Ulate, R. Figueroa-Balderas, M. A. Penn *et al.*, 2019 The genomic diversification of grapevine clones. BMC Genomics 20: 972.

Wagih, O., 2017 *ggseqlogo: A "ggplot2" Extension for Drawing Publication-Ready Sequence*.

Walker, M. A., L. A. Lider, A. C. Goheen, and H. P. Olmo, 1991 VR 039-16 Grape Rootstock. HortSci 26: 1224–1225.

Walker, M. A., K. Lund, C. Agüero, S. Riaz, K. Fort *et al.*, 2014 Breeding grape rootstocks for resistance to phylloxera and nematodes - it's not always easy. Acta Hortic. 1045: 89–97.

Wan, Y., H. R. Schwaninger, A. M. Baldo, J. A. Labate, G.-Y. Zhong *et al.*, 2013 A phylogenetic analysis of the grape genus (Vitis L.) reveals broad reticulation and concurrent diversification during neogene and quaternary climate change. BMC Evol Biol 13: 141.

Wang, Y., H. Tang, J. D. DeBarry, X. Tan, J. Li *et al.*, 2012 MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. Nucleic Acids Res. 40: e49.

Warren, R. F., A. Henk, P. Mowery, E. Holub, and R. W. Innes, 1998 A Mutation within the Leucine-Rich Repeat Domain of the Arabidopsis Disease Resistance Gene RPS5 Partially Suppresses Multiple Bacterial and Downy Mildew Resistance Genes. Plant Cell 10: 1439–1452.

Wei, F., K. Gobelman-Werner, S. M. Morroll, J. Kurth, and L. Mao, 1999 The Mla (Powdery Mildew) Resistance Cluster Is Associated With Three NBS-LRR Gene Families and Suppressed Recombination Within a 240-kb DNA Interval on Chromosome 5S (1HS) of 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.