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[*†], Mélanie Massonnet[*], Amanda M. Vondras[*], Rosa Figueroa-Balderas[*], and Dario Cantu[*‡]

[†] Equal contributors
[*]Department of Viticulture and Enology, University of California Davis, Davis, California, USA
[‡] Corresponding author

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*

Keywords: 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. 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 chromosomes. The completeness and accuracy of the assembly were confirmed using a high-density linkage map of *M. rotundifolia*. Protein-coding genes were annotated using an integrated and comprehensive approach. This included using Full-length cDNA sequencing (Iso-Seq) to improve gene structure and hypothetical spliced variant predictions. Our data strongly support that *Muscadinia* chromosomes 7 and 20 are fused in *Vitis* and pinpoint the location of the fusion in Cabernet Sauvignon and PN40024 chromosome 7. Disease-related gene numbers in Trayshed and Cabernet Sauvignon were similar, but their clustering locations were different. A dramatic expansion of the Toll/Interleukin-1 Receptor-like Nucleotide-Binding Site Leucine-Rich Repeat (TIR-NBS-LRR) 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 mildews. A genome browser for Trayshed, its annotation, and an associated Blast tool are available at www.grapegenomics.com.

INTRODUCTION

Viticulture and winemaking typically make use of *Vitis vinifera* fruit. Though *V. vinifera* wines have desirable organoleptic properties, *V. vinifera* is extremely sensitive to diverse pathogens. In contrast, *Muscadinia* is a genus closely related to *Vitis* (Small 1913; Bouquet 1978; Olmo 1986; Olien 1990) that is resistant to diverse biotic and abiotic stresses. *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).

Vitis and *Muscadinia* belong to Vitaceae, a family that contains 16 other genera and approximately 950 species (Wen *et al.* 2018). These genera are estimated to have 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). *Vitis* has 19 chromosomes (2n = 38) (Olmo 1937), but like other genera in Vitaceae, including *Ampelopsis*, *Parthenocissus,* and *Ampelocissus, Muscadinia* has 20 chromosomes (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 *Vitis* and *Muscadinia* often resulted in sterile progeny; this and the graft incompatibility observed between the two genera (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 maps revealed that linkage groups LG7 and LG20 in *Muscadinia* correspond to the proximal and distal regions of chromosome 7 in *Vitis* (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 residual telomeric repeats in chromosome 7 of *Vitis* (Lewter *et al.* 2019) could provide additional support for this hypothesis, insight into the evolutionary history of Vitaceae, and understanding of the structure and function of the *Muscadinia* genome.

M. rotundifolia is a desirable partner with which to hybridize because it is resistant to many 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 resistance loci.

A draft assembly of the *M. rotundifolia* cv. Trayshed (Trayshed, hereafter) genome using Single-Molecule Real-Time (SMRT) sequencing was instrumental in resolving the genetic basis of sex determination in grapes (Massonnet *et al.*, 2020). Here, we report the phased, chromosome-scale

assembly of Trayshed, which was produced improving the previous SMRT assembly with the introduction of the optical maps in a hybrid scaffolding approach and consensus genetic map from multiple wild and cultivated grape species. 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 but also to identify genes at the *RUN1*/*RPV1* locus.

MATERIALS AND METHODS

Trayshed chromosome construction

A Saphyr Genome Imaging Instrument was used to generate a Bionano Next-Generation Mapping (NGM) of Trayshed. Ultra-high molecular weight DNA (> 500 kbp) was extracted from young leaves by Amplicon Express (Pullman, WA). DNA was then labeled with a DLE-1 non-nicking enzyme (CTTAAG) and stained according to the Bionano Prep™ Direct Label and Stain (DLS) Kit (Bionano Genomics, San Diego, CA) instructions. Labeled DNA was loaded onto the SaphyrChip nanochannel array for imaging on the Saphyr system (Bionano Genomics, San Diego, CA). Imaged molecules longer than 150 kbp were kept. These molecules were then assembled using Bionano Solve v.3.3 (Lam *et al.* 2012) with the following parameters "haplotype", "noES", and "noCut", significance cutoffs were adopted $P < 1e\text{-}10$ to generate draft consensus contigs, $P < 1e\text{-}11$ for draft consensus contig extension, $P < 1e\text{-}15$ for the final merging of the draft consensus contigs. The complete configuration file used for the assembly is available as **Supplemental File S1**. This assembly procedure produced 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*. (2020) were scaffolded with the genome maps using HybridScaffold v.04122018 (Lam *et al.* 2012) with options "-B2 -N1" for conflict resolution. The complete parameter configuration is available as **Supplemental File S2**. The ploidy state of the 2,000 rhAmpseq markers conserved across wild and domesticated grapes (Zou *et al*. 2020) was evaluated using DISPR (https://github.com/douglasgscofield/dispr) to keep the ones found in a maximum of two copies across different sequences. Hybrid scaffolds were then pre-processed to prevent marker duplication and wrong primary contigs/haplotigs co-placement. Two sets of chromosome-scale pseudomolecules were reconstructed using the pre-processed hybrid scaffolds and HaploSync v.1 tool suite (https://github.com/andreaminio/HaploSync). The hybrid scaffolds were sorted using the curated rhAmpseq markers (Zou *et al*. 2020). Sequences were treated independently from FALCON Unzip's classification as primary contigs or haplotigs but, the relationship information between the primary contigs and the corresponding alternate haplotigs was conserved during the process. Then, the two haplotypes of each chromosome were compared to identify and correct assembly errors and fill gaps.

cDNA library preparation and sequencing

Total RNA from Trayshed leaves were 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), RNA quantity with a Qubit 2.0 Fluorometer and a broad range RNA kit (Life Technologies, Carlsbad,

131 CA), and RNA integrity by electrophoresis and an Agilent 2100 Bioanalyzer (Agilent
132 Technologies, CA). Total RNA (300 ng, RNA Integrity Number > 8.0) were used for cDNA
133 synthesis and library construction. A cDNA SMRTbell library was prepared. First-strand synthesis
134 and cDNA amplification were accomplished using the NEB Next Single Cell/Low Input cDNA
135 Synthesis & Amplification Module (New England, Ipswich, MA, US). The cDNAs were
136 subsequently purified with ProNex magnetic beads (Promega, WI) following the instructions in
137 the Iso-Seq Express Template Preparation for Sequel and Sequel II Systems protocol (Pacific
138 Biosciences, Menlo Park, CA). ProNex magnetic beads (86 µL) were used to select amplified
139 cDNA >= 2 kbp. At least 80 ng of the size-selected, amplified cDNA were used to prepare the
140 cDNA SMRTbell library. This was followed by DNA damage repair and SMRTbell ligation using
141 the SMRTbell Express Template Prep Kit 2.0 (Pacific Biosciences, Menlo Park, CA) following
142 the manufacturer's protocol. One SMRT cell was sequenced on the PacBio Sequel I platform (DNA
143 Technology Core Facility, University of California, Davis). IsoSeq reads were extracted using
144 IsoSeq v3 protocol with default parameters and low-quality isoform polished wit LSC v.2.0 (Au
145 *et al*., 2012). A cDNA library was also prepared using the Illumina TruSeq RNA sample
146 preparation kit v.2 (Illumina, CA, USA) following Illumina's low-throughput protocol. This
147 library was evaluated for quantity and quality with the High Sensitivity chip in an Agilent 2100
148 Bioanalyzer (Agilent Technologies, CA) and sequenced in 100 bp paired-end reads using an
149 Illumina HiSeq4000 sequencer (DNA Technology Core Facility, University of California, Davis).
150
151 **Genome annotation**
152 The completeness of the gene space in the hybrid assembly was estimated using BUSCO v.3.0.2
153 (Waterhouse *et al*., 2018) with the embryophyta_odb9 conserved gene set and by aligning the
154 PN40024 V1 CDS against the Trayshed assembly using BLAT v.36x2 with default parameters
155 (Kent 2002). PN40024 CDS were filtered before the alignment and included only single-copy
156 genes (*i.e*., with unique mapping on its genome). The structural annotation of the *M. rotundifolia*
157 cv. Trayshed genome was performed with a modified version of the pipeline used for the *Vitis*
158 *vinifera* cv. Zinfandel genome (Vondras *et al.* 2019) and fully described here:
159 https://github.com/andreaminio/AnnotationPipeline-EVM_based-DClab. Briefly, high-quality
160 Iso-Seq Trayshed data were used to produce high-quality gene models for training gene predictors
161 in PASA v.2.3.3 (Haas 2003) along with transcript evidence obtained from RNAseq data by
162 performing transcriptome assemblies using Stringtie v.1.3.4d (Pertea *et al.* 2015) and Trinity
163 v.2.6.5 (Grabherr *et al.* 2011) and from external databases. Public databases, transcriptome
164 assemblies, and the Iso-Seq data described above were used as transcript and protein evidence.
165 They were mapped on the genome using PASA v.2.3.3 (Haas 2003), MagicBLAST v.1.4.0
166 (Boratyn *et al.* 2019), and Exonerate v.2.2.0 (Slater and Birney 2005). *Ab initio* predictions were
167 generated using BUSCO v.3.0.2 (Waterhouse *et al*., 2018), Augustus v.3.0.3 (Stanke *et al.* 2006),
168 GeneMark v.3.47 (Lomsadze 2005), and SNAP v.2006-07-28 (Korf 2004). Repeats were
169 annotated using RepeatMasker v.open-4.0.6 (Smit *et al.* 2013). Next, EvidenceModeler v.1.1.1
170 (Haas *et al.* 2008) used these predictions to generate consensus gene models. The final functional
171 annotation was produced combining blastp v.2.2.28 (Camacho *et al.* 2009) hits against the Refseq
172 plant protein database (ftp://ftp.ncbi.nlm.nih.gov/refseq, retrieved January 17th, 2019) and
173 InterProScan v.5.28-67.0 (Jones *et al.* 2014) outputs using Blast2GO v.4.1.9 (Gotz *et al.* 2008).
174 To evaluate hemizygosity, CDS from each haplotype were aligned on the opposite haplotype using
175 GMAP v. 2019-09-12 (Wu and Watanabe 2005) and alignments with at least 80% identity and
176 coverage were considered valid. To prevent over-estimation of the gene content diversity between

177 the two haplotypes due to pseudomolecule reconstruction partiality, unplaced sequences were also
178 searched for copies of the CDS missing in the alternative haplotype.
179
180 **Genome size quantification by flow cytometry**
181 DNA content was estimated using flow cytometry (n = 3 individual leaves).
182 *Lycopersicon esculentum* cv. Stupické polní tyčkové rané DNA was selected as the internal
183 reference standard with a genome size of 2C = 1.96 pg; 1C = 958 Mbp (Doležel *et al.* 1992). Nuclei
184 extraction was performed using the Cystain PI absolute P kit (Sysmex America Inc).
185 Approximately 5 mg (0.7 cm^2) of young healthy leaves from grapevine and tomato were finely cut
186 with a razor blade in a Petri dish containing 500 µL of extraction buffer. The nuclei suspension
187 was filtered through a 50 µm filter (CellTrics, Sysmex America Inc) and 2 mL of a propidium
188 iodide staining solution was added (Doležel and Bartoš 2005; Bertier *et al.* 2013). Measurements
189 were acquired using a Becton Dickinson FACScan (Franklin Lakes, New Jersey) equipped with a
190 488 nm laser. The data were analyzed using FlowJo v.10
191 (https://www.flowjo.com/solutions/flowjo). DNA content was inferred by linear regression using
192 the tomato DNA reference standard.
193
194 **Comparison of genome assembly and synteny analysis**
195 The Trayshed and Cabernet Sauvignon (Massonnet *et al.*, 2020) genomes were compared with
196 NUCmer from the MUMmer v.4.0.0beta5 tool suite (Marçais *et al.* 2018), and the "--mum"
197 parameter set. Descriptive statistics of the alignment were obtained using the MUMmer script
198 dnadiff. For visualization purposes, alignments with at least 90% identity and 7500 bp long were
199 kept. Blastp v.2.2.28 (Camacho *et al.* 2009) was used to align the annotated proteins of Trayshed
200 and Cabernet Sauvignon. Alignments with at least 80% reciprocal identity and coverage were
201 retained. Pairwise protein information was associated with genes and processed with McScanX
202 v.11.Nov.2013 (Wang *et al.* 2012) with default parameters to identify syntenic regions.
203
204 **Localization of *Muscadinia* SNP markers**
205 To localize the Trayshed genomic sequences the corresponding position of each marker identified
206 from *Muscadinia* populations in Lewter *et al* 2019, the sequence surrounding each of them was
207 extracted from PN40024 V0 (Lewter *et al.,* 2019) keeping five hundred base-pair of overhang
208 upstream and downstream of the SNP position. Sequences were then aligned with BLAT v.36x2
209 (Kent 2002) over the Trayshed chromosomes to recover the marker position. Alignments with at
210 least 90% identity for both query and reference, 90% of coverage on the query, and a maximum
211 of 50-bp-long gaps were considered as hits. They were then filtered for uniqueness based on the
212 number of bases matching. The previously published consensus map of *M. rotundifolia* was used
213 to assess the completeness of the chromosome reconstruction (Lewter *et al.* 2019). Sliding
214 windows (window size = 20 markers, sliding = 10 markers) were designed to move across the
215 uniquely mapped markers on the genome. The percentage of relative marker positions consistent
216 with their genetic position on the high-density genetic map was calculated per window.
217
218 **Telomeric repeats analysis**
219 Telomeric repeats of "CCCTAA" and its reverse complement were searched along chromosome 7
220 of Cabernet Sauvignon and PN40024 V2 (Canaguier *et al.* 2017) using vmatchPattern from the R
221 package Biostrings v.2.56.0 (Pagès *et al.* 2019). After peaks in the distribution of telomeric repeats
222 were identified in both genomes, genomic regions of Cabernet Sauvignon (chr7: 18,675,000-

223 18,677,000 bp) and PN40024 (chr7: 17,572,000 - 17,574,000bp) were extracted. Motif enrichment
224 analysis was performed using MEME v.5.1.0 (Bailey and Elkan 1994) and reported as "logo"
225 using the ggseqlogo R package v.0.1 (Wagih 2017).
226
227 **Identification of NBS-LRR genes**
228 Predicted Trayshed and Cabernet Sauvignon proteins were scanned using hmmsearch from
229 HMMER v.3.3 (http://hmmer.org/) with a significance threshold (sequence E-value) of 0.001 and
230 Hidden Markov Models (HMM) corresponding to the following different Pfam (El-Gebali *et al.*
231 2019) domains: NB-ARC (Pfam PF00931), TIR (PF01582) and LRR (PF00560, PF07723,
232 PF07725, and PF12799). The proteome was also scanned with NLR-annotator (Steuernagel *et al*.,
233 2020). Coiled-coil (CC) domain-containing proteins were identified by COILS (Lupas *et al.* 1991)
234 during the InterProScan annotation. The identified proteins were then divided into six protein
235 classes according to their domain composition: CC-NBS-LRR, CC-NBS, TIR-NBS-LRR, TIR-
236 NBS, NBS-LRR, and NBS. To capture the largest number of potential NBS-LRR related genes,
237 genes lacking the NB-ARC domain but with "NBS-LRR" functional annotations were also
238 selected and divided into two classes: TIR-X and CC-X. For these eight protein classes, Multiple
239 EM for Motif Elicitation (MEME) analysis was performed with the flags "-mod anr -nmotifs 20"
240 to identify conserved domains for each class. Then, FIMO v.5.1.0 (Find Individual Motif
241 Occurrences) (Grant *et al.* 2011) was run on each protein class using the corresponding MEME
242 results; proteins with at least five conserved domains were kept. NBS-LRR gene clusters were
243 defined as groups of at least two NBS-LRR genes, each separated by no more than eight non-NBS-
244 LRR genes (Richly *et al.* 2002) in a region spanning a maximum of 200 kbp (Holub 2001).
245
246 ***RUN1/RPV1* locus analysis and NBS-LRR genes phylogeny**
247 Boundaries of the *RUN1/RPV1* locus were defined by mapping two SSR markers, VMC4f3.1 and
248 VMC8g9 (Barker *et al.* 2005), on Trayshed and Cabernet Sauvignon haplotypes using GMAP v.
249 2019-09-12 (Wu and Watanabe 2005). Trayshed chromosome 12 was split into 50-kbp blocks and
250 queried with the genomic region corresponding to the locus in Cabernet Sauvignon using tblastx
251 (BLAST v.2.2.29) (Camacho *et al.* 2009). Four-way pairwise comparisons between haplotypes
252 were performed, results were filtered keeping alignments with at least 90% of identity, at least 100
253 bp long, and located within the boundaries of the resistance locus in the query and the reference.
254 Hits overlapping annotated LTR retrotransposons were discarded. Proteins corresponding to the
255 first alternative spliced variant of each NBS-LRR gene at the *RUN1/RPV1* locus were aligned
256 using MUSCLE v.3.8.31 (Edgar 2004) with default parameters and by including *GAPDH*
257 (*VIT_17s0000g10430*) ortholog sequences as an outgroup. Distances between the alignments were
258 extracted using the R package seqinr v.4.2.4 (Charif and Lobry 2007). The estimation of the
259 corresponding neighbor-joining tree, its rooting, and bootstrapping (1000 replicates) were
260 performed using the R package ape v.5.4.1 (Paradis and Schliep 2019). The tree was drawn using
261 ggtree v.2.2.4 (Yu *et al.* 2017). Species-specific clusters were identified as five or more grouped
262 proteins from the same NBS-LRR class resulting from the phylogenetic analysis with a maximum
263 of one protein coming from the other species in the same tree branch. Coding sequences of the
264 Trayshed-specific protein cluster were aligned per class using MACSE v.2.03 (Ranwez *et al.* 2018)
265 and trimmed using the option "min_percent_NT_at_ends=0.80". Synonymous substitution rates
266 (dS) were obtained using yn00 from PAML v.4.9f (Yang 2007).
267
268 **Data analysis and visualization**

269 Data were parsed with R v.4.0.3 (R Core Team 2020) in RStudio v.1.4.869 (RStudio Team 2020)
270 using tidyverse v.1.3.0 (Wickham *et al.* 2019), GenomicFeatures v.1.40.1 for genomic ranges
271 manipulation (Lawrence *et al.* 2013), and rtracklayer for GFF files v.1.48.0 (Lawrence *et al.* 2009).
272
273 **Data availability**
274 Sequencing data are accessible at the NCBI repository under the accessions PRJNA635946 and
275 PRJNA593045. Raw optical maps are available at Zenodo under the DOI
276 10.5281/zenodo.3866087. The supplemental files are available at Figshare. The pipeline for the
277 gene annotation is available at https://github.com/andreaminio/AnnotationPipeline-EVM_based-
278 DClab. Assembly and annotation files are available at www.grapegenomics.com, which also hosts
279 a genome browser and a blast tool for Trayshed.
280
281 **RESULTS AND DISCUSSION**
282
283 **Construction of the twenty phased *Muscadinia rotundifolia* cv. Trayshed chromosomes**
284 The haploid genome size of *M. rotundifolia* cv. Trayshed evaluated by flow cytometry was
285 estimated at 483.4 ± 3.1 Mbp, a size comparable to PN40024 (487 Mbp; Jaillon et al. 2007), but
286 smaller than Cabernet Sauvignon (557.0 ± 2.4 Mbp; **Supplemental Figure S1**).
287
288 Optical maps were generated using the BioNano Genomics technology at 2,057x coverage to
289 further scaffold the Trayshed genome assembly produced previously from Single-Molecule Real-
290 Time (SMRT; Pacific Biosciences) long reads (N50 = 1,479,367 bp; Massonnet *et al.* 2020). The
291 BioNano consensus genome map was combined with the contigs to produce an 896.0 Mbp hybrid
292 assembly nearly twice the expected haploid genome size of Trayshed determined by flow
293 cytometry. Genome sizes estimated by sequencing are known to be underestimates, while the flow
294 cytometry approach tends to produce overestimates (Elliott and Gregory, 2015). In the hybrid
295 assembly, ~ 70% of the complete BUSCO genes were duplicated and, 2.16 ± 1.07 copies of each
296 PN40024 gene were present. Altogether, the size of the assembly, the amount of duplicated
297 BUSCO genes (**Supplemental Table S1**), and the doubled representation of PN20024 genes are
298 strong evidence that the draft assembly represents the diploid genome of Trayshed. Finally, we
299 reconstructed two sets of phased chromosome-scale pseudomolecules (Haplotype 1, Hap1, and
300 Haplotype 2, Hap2; **Table 1**) using a consensus grape genetic map as a guide (Zou *et al*. 2020).
301 The chromosome-scale assemblies were 400.4 Mbp for Hap1 and 369.9 Mbp for Hap2. Based on
302 the flow cytometry data, the whole assembly represents 86.2% of the estimated diploid genome
303 size (2x 483.4 Mbp). On average, chromosome size was 89.09 ± 8.87 % conserved between the
304 two haplotypes (**Supplemental Table S2**). It was not possible to recover the positional information
305 for regions lacking genetic markers, which were therefore excluded from the pseudomolecule
306 assembly (**Table 1**). The unplaced scaffolds represented only 7.5% of the genome assembly, which
307 is much less than in PN40024 V1 (12.4%; Jaillon et al., 2007) and similar to that of the V2
308 assembly (6%; Canaguier *et al*., 2017). The unplaced scaffolds were significantly more repetitive
309 than the rest of the genome (52% *vs.* 42%) and carried about 7% of the total protein-coding genes
310 (3,564 genes). The Trayshed assembly also showed a lower level of gaps overall (~ 1 Mbp per
311 haplotype) than PN40024 (~10 Mbp).
312
313 The Trayshed assembly structure was confirmed by comparing positions along each chromosome
314 with their linkage group positions within a high-density *M. rotundifolia* genetic map (Lewter *et al*.

2019) (**Supplemental Figures S2**, **S3**, and **S4**). More than 80% of markers' positions were consistent between the genetic and physical maps, with strong overall correlation ($R > 0.90$) along all chromosomes for both haplotypes.

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

Each set of Trayshed's twenty chromosomes was compared with the chromosome-scale Cabernet Sauvignon assembly (Massonnet *et al.*, 2020), which is the most complete and contiguous diploid genome assembly of grapevine currently available. The two genomes were highly syntenic (**Figure 1A**, **Supplemental Figure S5**). Based on the four haplotype comparisons, colinear regions were 2,433 ± 28 bp long and shared 92.19 ± 0.37% identity (**Supplementary Table S3**). On average, short insertions (<50bp) were 40-bp long, long insertions were 1,501-bp long, short deletions were 40-bp, and long deletions were 998-bp. The longest deletion (~50 kbp) was detected on Cabernet Sauvignon chromosome 8 and the longest insertion (~79 kbp) was observed on chromosome 2 (**Supplemental Table S4**).

Trayshed chromosomes 7 and 20 aligned at the beginning and end of Cabernet Sauvignon chromosome 7, respectively. This supports the fusion of these two chromosomes in *Vitis* reported previously using genetic maps (Blanc *et al.* 2012; Lewter *et al.* 2019) (**Figure 1A**). The marker-containing regions used for Trayshed chromosome structure validation also support this hypothesis. Markers on Trayshed LG 7 and 20 were associated with *V. vinifera* chromosome 7 (**Figure 1B**). Telomeric repeats can be used as indicators of chromosome number reduction rearrangements (Sousa and Renner, 2015). Enrichment of these sequences is expected in a genomic region where a chromosome 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 whole-genome alignments. We examined the genomic region of Cabernet Sauvignon chromosome 7 that spanned Trayshed chromosomes 7 and 20 (chr 7: 18,661,816 - 18,740,440 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 – 17,636,181 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 S6**) were generated as transcriptional evidence to support the 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 S6**). After obtaining consensus gene models, alternative splice variants were predicted if supported by Iso-Seq data, with 2.87 transcripts per gene on average (**Supplemental Figure S6**). BUSCO genes (Hap1: 92.3%, Hap2: 85.2%) were well-represented in the diploid Trayshed transcriptome. The structural annotation included 52,943 protein-coding gene loci (**Table 1**) and 83,873 proteins (including the alternative forms). There were 25,706 and 23,673 genes on Hap1 and Hap2, respectively (**Table 1**). The repetitive content composed 42.54 and 42.35% of Hap1 and Hap2, respectively, and was higher in the unplaced sequences (51.99%). The coding sequences of each haplotype were aligned on its alternative to evaluate hemizygosity. We found 2,586 and 1,034 hemizygous genes on Hap1 (10.06%) and Hap2 (4.37%), respectively. These results are consistent

361 with *V. vinifera* cv. Zinfandel (4.56%; Vondras *et al*., 2019) but lower than the levels found in *V.*
362 *vinifera* cv. Chardonnay (14.6%; Zhou *et al*., 2019).
363
364 Based on blastp results (80% identity, 80% coverage), 33,822 (63.88%) Trayshed genes are
365 homologous with 31,471 (54.83%) genes of Cabernet Sauvignon. The 22,755 proteins that did not
366 have a potential ortholog in Cabernet Sauvignon were significantly enriched in functional
367 categories related to cell communication, signal transduction, signaling, regulation of biological
368 and cellular processes, and cellular responses to stimulus and biological regulation (P <0.01;
369 Fisher's exact test). On average, colinear blocks of homologous genes were ~152 genes long (block
370 extension interrupted with more than 25 unmatched genes in a row, **Supplemental Table S5**)
371 based on the synteny results with Cabernet Sauvignon. The longest block of colinear genes was
372 found on chromosome 14 and contained 613 genes (**Supplemental Table S5**).
373
374 **NBS-LRR intracellular receptors encoded in the Trayshed genome**
375 Trayshed is attractive to grape breeding programs because it is resistant to pathogens that affect *V.*
376 *vinifera*. In plants, resistance to pathogens is primarily attributed to the activity of resistance genes
377 (R genes). Nucleotide-binding site leucine-rich repeat (NBS-LRR) genes constitute the largest
378 family of R genes. NBS-LRR genes are associated with resistance to powdery and downy mildew
379 pathogens in several species, including *Arabidopsis* (Warren *et al.* 1998), wheat (He *et al.* 2018),
380 barley (Wei *et al.* 1999; Zhou *et al.* 2001), pepper (Jo *et al.* 2017), and grapes (Riaz *et al.* 2011;
381 Zini *et al.* 2019). We divided the NBS-LRR genes into eight different classes depending on the
382 domains detected in the proteins: CC-NBS-LRR, CC-NBS, CC-X, TIR-NBS-LRR, TIR-NBS,
383 TIR-X, NBS-LRR, NBS (called hereafter NBS-LRR when described as a whole). Overall, the
384 Trayshed genome shows a slightly higher number of NBS-LRR genes (1,158) than Cabernet
385 Sauvignon (1,013). Both species showed the highest number of NBS-LRR localized on
386 chromosomes 9, 12, 13, and 18 (**Figure 2A**, **Supplementary Table S6**). NBS-LRR genes can
387 occur in clusters that could be critical to their function as resistance loci (Hulbert *et al.* 2001;
388 Meyers *et al.* 2003). The average number of clusters detected in Trayshed per haplotype (122) was
389 similar to Cabernet Sauvignon (120), but their arrangement in the genome varied substantially
390 (**Supplemental Figure S7**). Among the haplotypes, in Cabernet Sauvignon, NBS-LRR clusters
391 accumulated the most on chromosome 18 Hap1 while the highest number was observed on
392 chromosome 12 in Trayshed Hap1. These differences could plausibly contribute to disparities in
393 disease resistance.
394
395 **Expansion of NBS-LRR genes at the *RUN1/RPV1* locus in Trayshed**
396 The clusters of NBS-LRR genes on chromosome 12 in Trayshed are fewer and scattered in
397 Cabernet Sauvignon (**Figures 2A** and **2B**). This chromosome contains a known powdery mildew
398 resistance locus, *RUN1/RPV1* (Pauquet *et al.* 2001), with boundaries defined by the VMC4f3.1
399 and VMC8g9 SSR markers (Barker *et al.* 2005). This region was studied in Trayshed (Hap1 -
400 chr12: 11,270,914 - 17,943,644 bp; Hap2 – chr12: 11,341,795 - 15,346,610 bp) and Cabernet
401 Sauvignon (Hap1 - chr12: 18,746,148 - 23,727,933 bp; Hap2 – chr12: 15,055,893 - 21,663,411);
402 it included 321 and 201 annotated genes in Trayshed Hap1 and Hap2, respectively, and 287 and
403 312 genes in Cabernet Sauvignon haplotypes. Within *RUN1/RPV1*, 52 and 33 NBS-LRR were
404 identified in Trayshed Hap1 and Hap2, respectively. For Cabernet Sauvignon, 33 and 40 NBS-
405 LRR were identified in Hap1 and Hap2, respectively.

406 Notably, most of the NBS-LRR classes are present for both species in this region. A phylogenetic
407 analysis of the NBS-LRR proteins at *RUN1/RPV1* identified a protein cluster specific to Trayshed
408 Hap1 (**Figure 2C**, **Supplemental Figure S8**), absent from the Hap2 (**Figure 2D**). After aligning
409 the coding sequences of those proteins, the synonymous mutation rate was calculated. We
410 estimated that TIR-NBS-LRR genes diverged at nearly 19 Ma (**Figure 2E**). This is one of the latest
411 estimates of the divergence between the *Muscadinia* and *Vitis* genera found in the literature [~18–
412 47 Ma] (Wan *et al.* 2013; Liu *et al.* 2016; Ma *et al.* 2018). Thus, a recent expansion of TIR-NBS-
413 LRR genes might have occurred in only *Muscadinia* and other classes may have expanded prior.
414 To characterize gene duplication in the *RUN1/RPV1* alleles, Cabernet Sauvignon haplotypes were
415 aligned on Trayshed haplotypes to evaluate how many homologous regions are detected using
416 tblastx. Within the resistance locus boundaries of both genomes (and excluding LTR
417 retrotransposons), the highest number of alignments was observed when Trayshed Hap1 was used
418 as the target (**Figure 3**, **Supplemental Figure S9**). There were 10,270 regions aligned with
419 Cabernet Sauvignon Hap1 (**Figure 3A**) and 8,372 alignments between Hap2 (**Figure 3B**). On
420 average, one region in Cabernet Sauvignon Hap1 had 1.75 hits in Trayshed Hap1 and 1.18 hits in
421 Hap2. For the most duplicated region, a single sequence from Cabernet Sauvignon Hap1 matched
422 up to 16 different regions in Trayshed Hap1 (**Figure 3A**). The increased number of hits per window
423 that was observed in Trayshed Hap1 evidenced the expansion of the locus through an increase of
424 the copy number of homologous regions. The regions in Trayshed Hap1 with the highest number
425 of homologs (with at least 10 hits matching a single Cabernet Sauvignon region) were positioned
426 between ~11.83 Mbp and 14.24 Mbp on chromosome 12 (**Figure 3A**) and is similar to the genomic
427 region identified by Feechan *et al*. (2013). It includes UDP-glucosyltransferase-coding genes
428 followed by R gene analogs. In Cabernet Sauvignon, highly duplicated regions were mostly
429 composed of kinetochore NDC80 complex genes; these matched genes with similar or NBS-LRR
430 functions in Trayshed. It is important to note that the assembly of the two *RUN1/RPV1* loci of
431 Trayshed was complex. Despite the currently available long-read sequencing technologies, the
432 highly duplicated content at *RUN1/RPV1* still presents several technical limitations.
433
434 Between Trayshed and Cabernet Sauvignon, functionally annotated disease-related genes were
435 differently clustered along the chromosomes. An expansion of TIR-NBS-LRR was identified,
436 which likely occurred after the divergence of their genera. Similar research could be undertaken
437 to characterize the genes or features at other known resistance loci in *M. rotundifolia*, like *RUN2*
438 and *REN5* (Riaz *et al.* 2011; Blanc *et al.* 2012). Acquiring this understanding could be expedited
439 with the availability of high-quality reference sequences for resistant selections, like Trayshed, and
440 susceptible cultivars, like Cabernet Sauvignon and others (Massonnet *et al.*, 2020).
441

442 **ACKNOWLEDGMENTS**

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

445 **TABLES AND FIGURES**
446
447 **Table 1. Genome assembly statistics**

	Haplotype 1	Haplotype 2	Unplaced
Assembly length (bp)	400,450,509	369,985,741	62,961,436
Number of scaffolds	20	20	1,690
Average length (bp)	20,022,525	18,499,287	37,255
Maximum length (bp)	34,098,201	26,289,366	1,487,392
N50 length (bp)	20,338,664	18,578,327	46,805
Total gap length (bp)	645,752	1,182,032	125,041
Repetitive content (%)	42.54	42.35	51.99
Number of protein-coding genes	25,706	23,673	3,564
Complete BUSCO copies (%)	90.6	86.2	9.6

448
449 **Figure 1: Chromosome fusion localization in *Vitis vinifera***
450 **A**. Whole-genome alignment of Trayshed Hap1 (y-axis) on Cabernet Sauvignon Hap1 (x-axis).
451 The putative fusion of Trayshed chromosomes 7 and 20 (green) and Cabernet Sauvignon
452 chromosome 7 (purple) is inset. The percentage identity (% id) between the alignments is displayed
453 as a color gradient. **B**. The positions of markers on Trayshed chromosomes 7 and 20 (green) and
454 their corresponding location on PN40024 V0 chromosome 7. **C**. The distribution of telomeric
455 repeats in 1 Mbp containing the expected chromosome fusion in Cabernet Sauvignon (left panel)
456 and PN40024 (right panel). The frequency of telomere repeats is represented as a color gradient.
457 The most enriched motif in the 2-kbp region surrounding the peak of telomeric repeats is inset for
458 each genotype.
459
460 **Figure 2: NBS-LRR genes in Trayshed and Cabernet Sauvignon *RUN1/RPV1* loci**
461 Location of the different classes of NBS-LRR along the different chromosomes of the two
462 haplotypes (Hap1: dark grey, Hap2: light grey) in Trayshed (**A**) and Cabernet Sauvignon (**B**). The
463 locations of *RUN1/RPV1* are indicated with an orange box. Phylogenetic trees of NBS-LRR
464 proteins at *RUN1/RPV1* locus in Trayshed and Cabernet Sauvignon, Hap1 (**C**) and Hap2 (**D**).
465 GAPDH was used to root the tree and is colored in grey. **E.** The boxplot represents an
466 approximation of the expansion time of the protein cluster highlighted in orange in panel **C**. The
467 approximation of the divergence time between *Vitis* and *Muscadinia* genera is highlighted in grey.
468 The same color coding is used throughout the figure to distinguish the NBS-LRR classes.
469
470 **Figure 3: Synteny between the *RUN1/RPV1* alleles in Trayshed and Cabernet Sauvignon**
471 Chromosome 12 of Trayshed (left, target) and Cabernet Sauvignon (right, query) Hap1 (**A**) or
472 Hap2 (**B**) are compared with tblastx of 50Kbp blocks. The *RUN1/RPV1* resistance locus is
473 highlighted with an orange box. Syntenic gene content within the locus is represented as lines in
474 the center of each panel. Duplications of Cabernet Sauvignon sequences are shown using a color
475 gradient. Highly duplicated sequences in Trayshed are dark purple. The functional annotations
476 depicted include categories represented by at least 5 genes. Categories represented by less than 10
477 genes and not related to disease resistance genes are included in the category, "other".

478 **LITERATURE CITED**

479 Au, K. F., J. G. Underwood, L. Lee, and W. H. Wong, 2012 Improving PacBio Long Read
480 Accuracy by Short Read Alignment (Y. Xing, Ed.). PLoS ONE 7: e46679.
481 Bailey, T. L., and C. Elkan, 1994 Fitting a mixture model by expectation maximization to
482 discover motifs in biopolymers. Proc Int Conf Intell Syst Mol Biol 2: 28–36.
483 Barker, C. L., T. Donald, J. Pauquet, M. B. Ratnaparkhe, A. Bouquet *et al.*, 2005 Genetic and
484 physical mapping of the grapevine powdery mildew resistance gene, Run1, using a
485 bacterial artificial chromosome library. Theor Appl Genet 111: 370–377.
486 Bertier, L., L. Leus, L. D'hondt, A. W. A. M. de Cock, and M. Höfte, 2013 Host adaptation and
487 speciation through hybridization and polyploidy in *Phytophthora*. PLOS ONE 8: e85385.
488 Blanc, S., S. Wiedemann-Merdinoglu, V. Dumas, P. Mestre, and D. Merdinoglu, 2012 A
489 reference genetic map of *Muscadinia rotundifolia* and identification of *Ren5*, a new
490 major locus for resistance to grapevine powdery mildew. Theor Appl Genet 125: 1663–
491 1675.
492 Blanco-Ulate, B., E. Vincenti, A. L. T. Powell, and D. Cantu, 2013 Tomato transcriptome and
493 mutant analyses suggest a role for plant stress hormones in the interaction between fruit
494 and *Botrytis cinerea*. Front. Plant Sci. 4: 142.
495 Boratyn, G. M., J. Thierry-Mieg, D. Thierry-Mieg, B. Busby, and T. L. Madden, 2019 Magic-
496 BLAST, an accurate RNA-seq aligner for long and short reads. BMC Bioinformatics 20:
497 405.
498 Bouquet, A., 1978 La muscadine (Vitis rotundifolia michx) et sa culture aux Etats-Unis. OENO
499 One 12: 1–20.
500 Bouquet, A., 1980 *Vitis* x *Muscadinia* hybridization: A new way in grape breeding for disease
501 resistance in France, pp. 42–61 in *Proceedings of the 3rd International Symposium on*
502 *Grape Breeding*, University of California, Davis, Department of Viticulture and Enology.
503 Branas, M. M., 1932 Sur la caryologie des Ampélidées. C R Acad Sci Paris 194: 121–123.
504 Camacho, C., G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos *et al.*, 2009 BLAST+:
505 architecture and applications. BMC Bioinformatics 10: 421.
506 Canaguier, A., J. Grimplet, G. Di Gaspero, S. Scalabrin, E. Duchene *et al.*, 2017 A new version
507 of the grapevine reference genome assembly (12X.v2) and of its annotation (VCost.v3).
508 Genom Data 14: 56–62.
509 Charif, D., and J. R. Lobry, 2007 SeqinR 1.0-2: A Contributed Package to the R Project for
510 Statistical Computing Devoted to Biological Sequences Retrieval and Analysis, pp. 207–
511 232 in *Structural Approaches to Sequence Evolution: Molecules, Networks, Populations*,
512 edited by U. Bastolla, M. Porto, H. E. Roman, and M. Vendruscolo. Springer Berlin
513 Heidelberg, Berlin, Heidelberg.
514 Chu, Z.-F., J. Wen, Y.-P. Yang, Z.-L. Nie, and Y. Meng, 2018 Genome size variation and
515 evolution in the grape family Vitaceae: Genome size variation in Vitaceae. Jnl of
516 Sytematics Evolution 56: 273–282.
517 Davidis, U. X., and H. P. Olmo, 1964 The Vitis vinifera x V. rotundifolia hybrids as phylloxera
518 resistant rootstocks. Vitis 4: 129–143.
519 Dearing, C., 1917 Muscadine grape breeding. J. Hered. 8: 409–424.
520 Delame, M., E. Prado, S. Blanc, G. Robert-Siegwald, C. Schneider *et al.*, 2019 Introgression
521 reshapes recombination distribution in grapevine interspecific hybrids. Theor Appl Genet
522 132: 1073–1087.

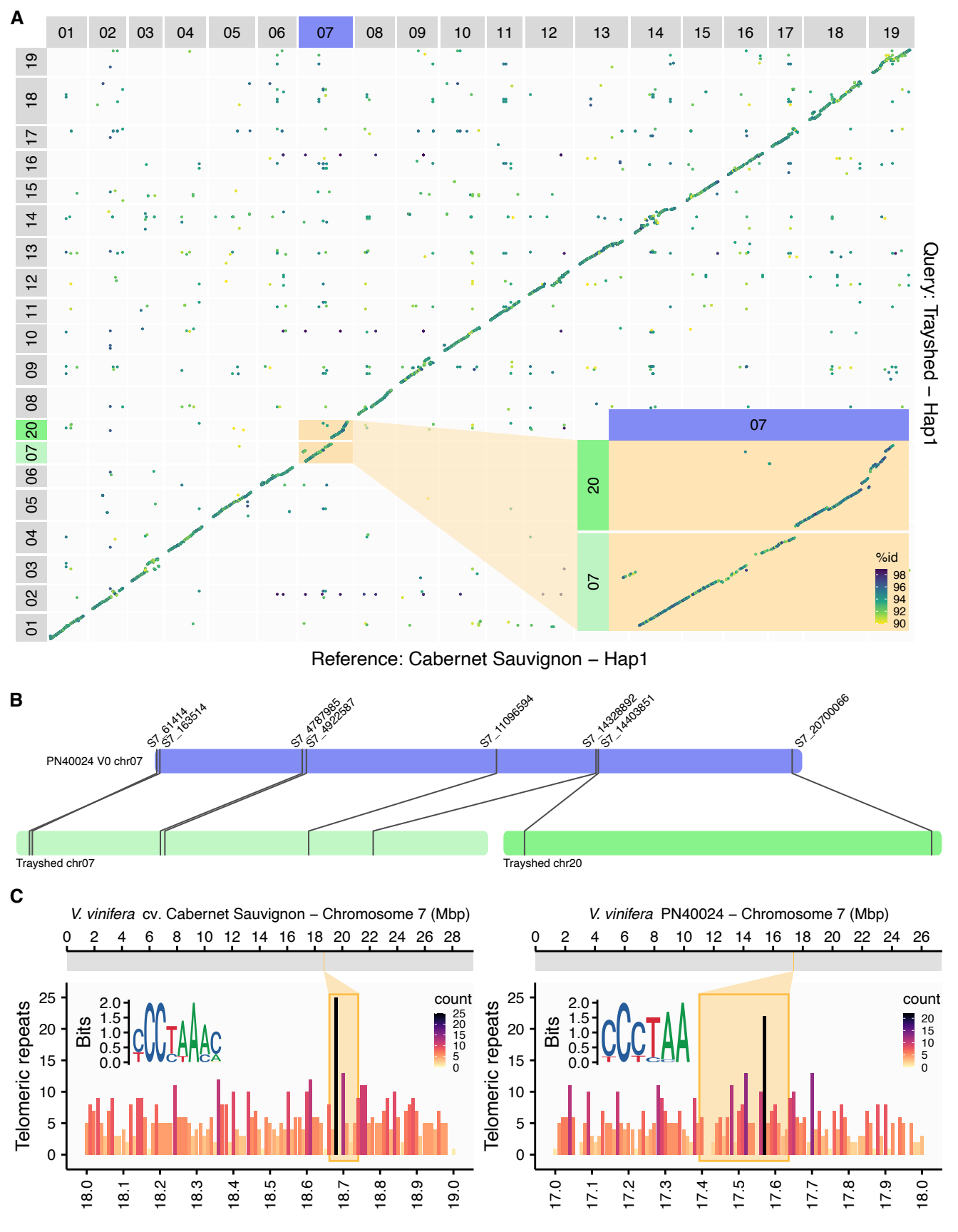
523 Doležel, J., and J. Bartoš, 2005 Plant DNA Flow Cytometry and Estimation of Nuclear Genome
524 Size. Ann. Bot. 95: 99–110.
525 Doležel, J., S. Sgorbati, and S. Lucretti, 1992 Comparison of three DNA fluorochromes for flow
526 cytometric estimation of nuclear DNA content in plants. Physiol. Plant. 85: 625–631.
527 Edgar, R. C., 2004 MUSCLE: multiple sequence alignment with high accuracy and high
528 throughput. Nucleic Acids Res. 32: 1792–1797.
529 Elliott, T. A., and T. R. Gregory, 2015 What's in a genome? The C-value enigma and the
530 evolution of eukaryotic genome content. Philos. Trans. R. Soc. B Biol. Sci. 370:
531 20140331.
532 El-Gebali, S., J. Mistry, A. Bateman, S. R. Eddy, A. Luciani *et al.*, 2019 The Pfam protein
533 families database in 2019. Nucleic Acids Res. 47: D427–D432.
534 Feechan, A., C. Anderson, L. Torregrosa, A. Jermakow, P. Mestre *et al.*, 2013 Genetic dissection
535 of a TIR-NB-LRR locus from the wild North American grapevine species *Muscadinia*
536 *rotundifolia* identifies paralogous genes conferring resistance to major fungal and
537 oomycete pathogens in cultivated grapevine. Plant J 76: 661–674.
538 Firoozabady, E., and H. P. Olmo, 1982 Resistance to grape phylloxera in Vitis vinifera x V.
539 rotundifolia grape hybrids. Vitis 21: 1–4.
540 Gotz, S., J. M. Garcia-Gomez, J. Terol, T. D. Williams, S. H. Nagaraj *et al.*, 2008 High-
541 throughput functional annotation and data mining with the Blast2GO suite. Nucleic Acids
542 Res. 36: 3420–3435.
543 Grabherr, M. G., B. J. Haas, M. Yassour, J. Z. Levin, D. A. Thompson *et al.*, 2011 Full-length
544 transcriptome assembly from RNA-Seq data without a reference genome. Nat.
545 Biotechnol. 29: 644–652.
546 Grant, C. E., T. L. Bailey, and W. S. Noble, 2011 FIMO: scanning for occurrences of a given
547 motif. Bioinformatics 27: 1017–1018.
548 Haas, B. J., 2003 Improving the Arabidopsis genome annotation using maximal transcript
549 alignment assemblies. Nucleic Acids Res. 31: 5654–5666.
550 Haas, B. J., S. L. Salzberg, W. Zhu, M. Pertea, J. E. Allen *et al.*, 2008 Automated eukaryotic
551 gene structure annotation using EVidenceModeler and the Program to Assemble Spliced
552 Alignments. Genome Biol 9: R7.
553 He, H., S. Zhu, R. Zhao, Z. Jiang, Y. Ji *et al.*, 2018 Pm21, Encoding a Typical CC-NBS-LRR
554 Protein, Confers Broad-Spectrum Resistance to Wheat Powdery Mildew Disease. Mol.
555 Plant 11: 879–882.
556 Heinitz, C. C., J. Uretsky, J. C. Dodson Peterson, K. G. Huerta-Acosta, and M. A. Walker, 2019
557 Crop Wild Relatives of Grape (Vitis vinifera L.) Throughout North America, pp. 329–
558 351 in *North American Crop Wild Relatives, Volume 2: Important Species*, edited by S.
559 L. Greene, K. A. Williams, C. K. Khoury, M. B. Kantar, and L. F. Marek. Springer
560 International Publishing, Cham.
561 Holub, E. B., 2001 The arms race is ancient history in Arabidopsis, the wildflower. Nat Rev
562 Genet 2: 516–527.
563 Hulbert, S. H., C. A. Webb, S. M. Smith, and Q. Sun, 2001 RESISTANCE GENE
564 COMPLEXES: Evolution and Utilization. Annu. Rev. Phytopathol. 39: 285–312.
565 Jaillon, O., J. M. Aury, B. Noel, A. Policriti, C. Clepet *et al.*, 2007 The grapevine genome
566 sequence suggests ancestral hexaploidization in major angiosperm phyla. Nature 449:
567 463–468.

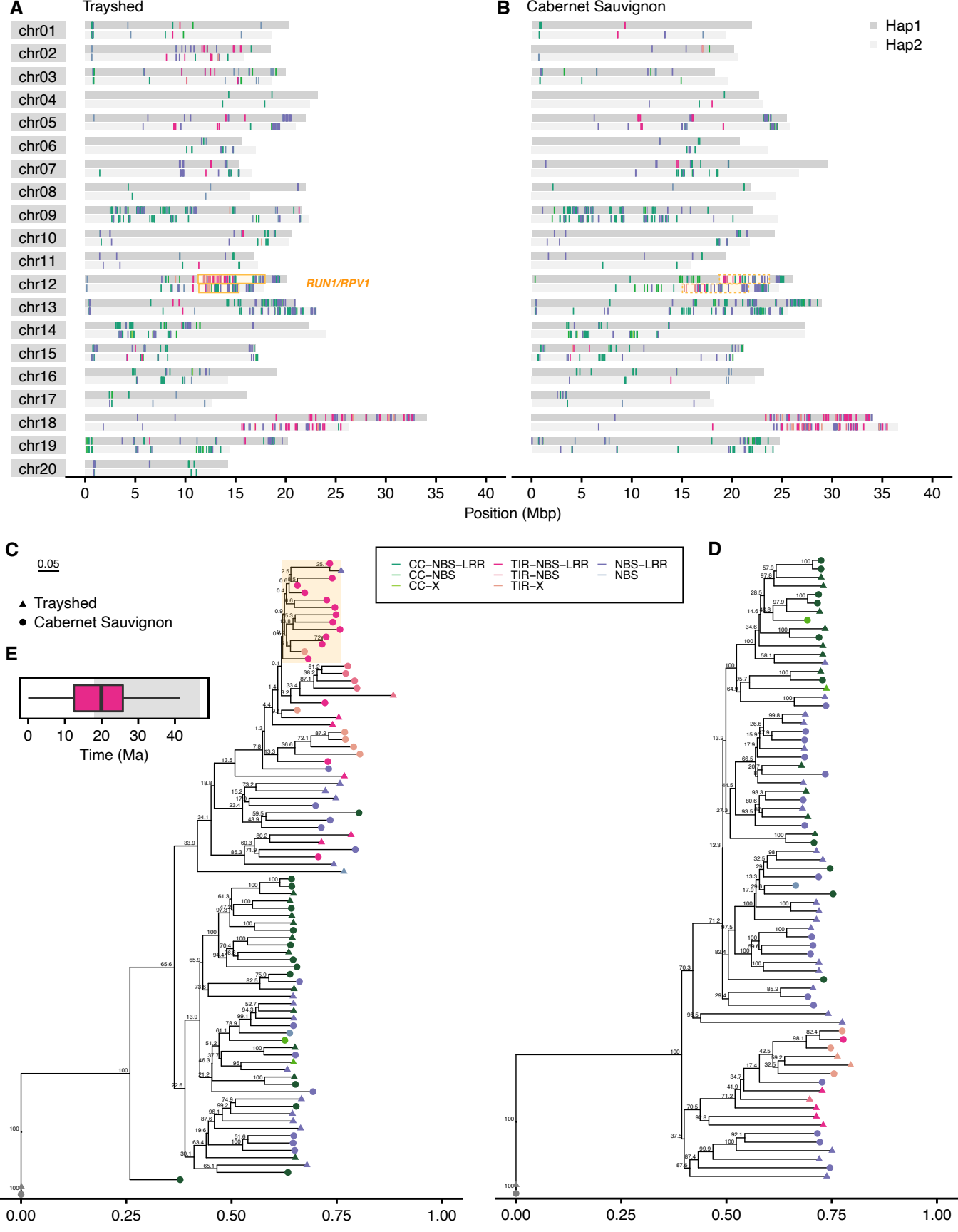
568 Jo, J., J. Venkatesh, K. Han, H.-Y. Lee, G. J. Choi *et al.*, 2017 Molecular Mapping of PMR1, a
569 Novel Locus Conferring Resistance to Powdery Mildew in Pepper (Capsicum annuum).
570 Front. Plant Sci. 8: 2090.
571 Jones, P., D. Binns, H.-Y. Chang, M. Fraser, W. Li *et al.*, 2014 InterProScan 5: genome-scale
572 protein function classification. Bioinformatics 30: 1236–1240.
573 Karkamkar, S. P., S. G. Patil, and S. C. Misra, 2010 Cyto–morphological studies and their
574 significance in evolution of family Vitaceae. Nucleus 53: 37–43.
575 Kent, W. J., 2002 BLAT : The Blast-Like Alignment Tool. Genome Res. 12: 656–664.
576 Korf, I., 2004 Gene finding in novel genomes. BMC Bioinformatics 5: 59.
577 Lam, E. T., A. Hastie, C. Lin, D. Ehrlich, S. K. Das *et al.*, 2012 Genome mapping on
578 nanochannel arrays for structural variation analysis and sequence assembly. Nat
579 Biotechnol 30: 771–776.
580 Lawrence, M., R. Gentleman, and V. Carey, 2009 rtracklayer: an R package for interfacing with
581 genome browsers. Bioinformatics 25: 1841–1842.
582 Lawrence, M., W. Huber, H. Pagès, P. Aboyoun, M. Carlson *et al.*, 2013 Software for
583 Computing and Annotating Genomic Ranges (A. Prlic, Ed.). PLoS Comput Biol 9:
584 e1003118.
585 Lewter, J., M. L. Worthington, J. R. Clark, A. V. Varanasi, L. Nelson *et al.*, 2019 High-density
586 linkage maps and loci for berry color and flower sex in muscadine grape (Vitis
587 rotundifolia). Theor Appl Genet 132: 1571–1585.
588 Liu, X.-Q., S. M. Ickert-Bond, Z.-L. Nie, Z. Zhou, L.-Q. Chen *et al.*, 2016 Phylogeny of the
589 Ampelocissus–Vitis clade in Vitaceae supports the New World origin of the grape genus.
590 Mol Phylogenet Evol 95: 217–228.
591 Lomsadze, A., 2005 Gene identification in novel eukaryotic genomes by self-training algorithm.
592 Nucleic Acids Res. 33: 6494–6506.
593 Lupas, A., M. Van Dyke, and J. Stock, 1991 Predicting coiled coils from protein sequences.
594 Science 252: 1162–1164.
595 Ma, Z.-Y., J. Wen, S. M. Ickert-Bond, Z.-L. Nie, L.-Q. Chen *et al.*, 2018 Phylogenomics,
596 biogeography, and adaptive radiation of grapes. Mol Phylogenet Evol 129: 258–267.
597 Marçais, G., A. L. Delcher, A. M. Phillippy, R. Coston, S. L. Salzberg *et al.*, 2018 MUMmer4: A
598 fast and versatile genome alignment system (A. E. Darling, Ed.). PLoS Comput Biol 14:
599 e1005944.
600 Massonnet, M., N. Cochetel, A. Minio, A. M. Vondras, J. Lin *et al*., 2020 The genetic basis of
601 sex determination in grapes. Nat. Commun. 11: 2902.
602 Merdinoglu, D., C. Schneider, E. Prado, S. Wiedemann-Merdinoglu, and P. Mestre, 2018
603 Breeding for durable resistance to downy and powdery mildew in grapevine. OENO One
604 52: 203–209.
605 Merdinoglu, D., S. Wiedeman-Merdinoglu, P. Coste, V. Dumas, S. Haetty *et al.*, 2003 Genetic
606 analysis of downy mildew resistance derived from Muscadinia rotundifolia. Acta Hortic.
607 603: 451–456.
608 Meyers, B. C., A. Kozik, A. Griego, H. Kuang, and R. W. Michelmore, 2003 Genome-Wide
609 Analysis of NBS-LRR–Encoding Genes in Arabidopsis. Plant Cell 15: 809–834.
610 Olien, W. C., 1990 The Muscadine Grape: Botany, Viticulture, History, and Current Industry.
611 HortSci 25: 732–739.
612 Olmo, H. P., 1937 Chromosome Numbers in the European Grape (*Vitis vinifera*). CYTOLOGIA
613 Fujii Jubilae: 606–613.

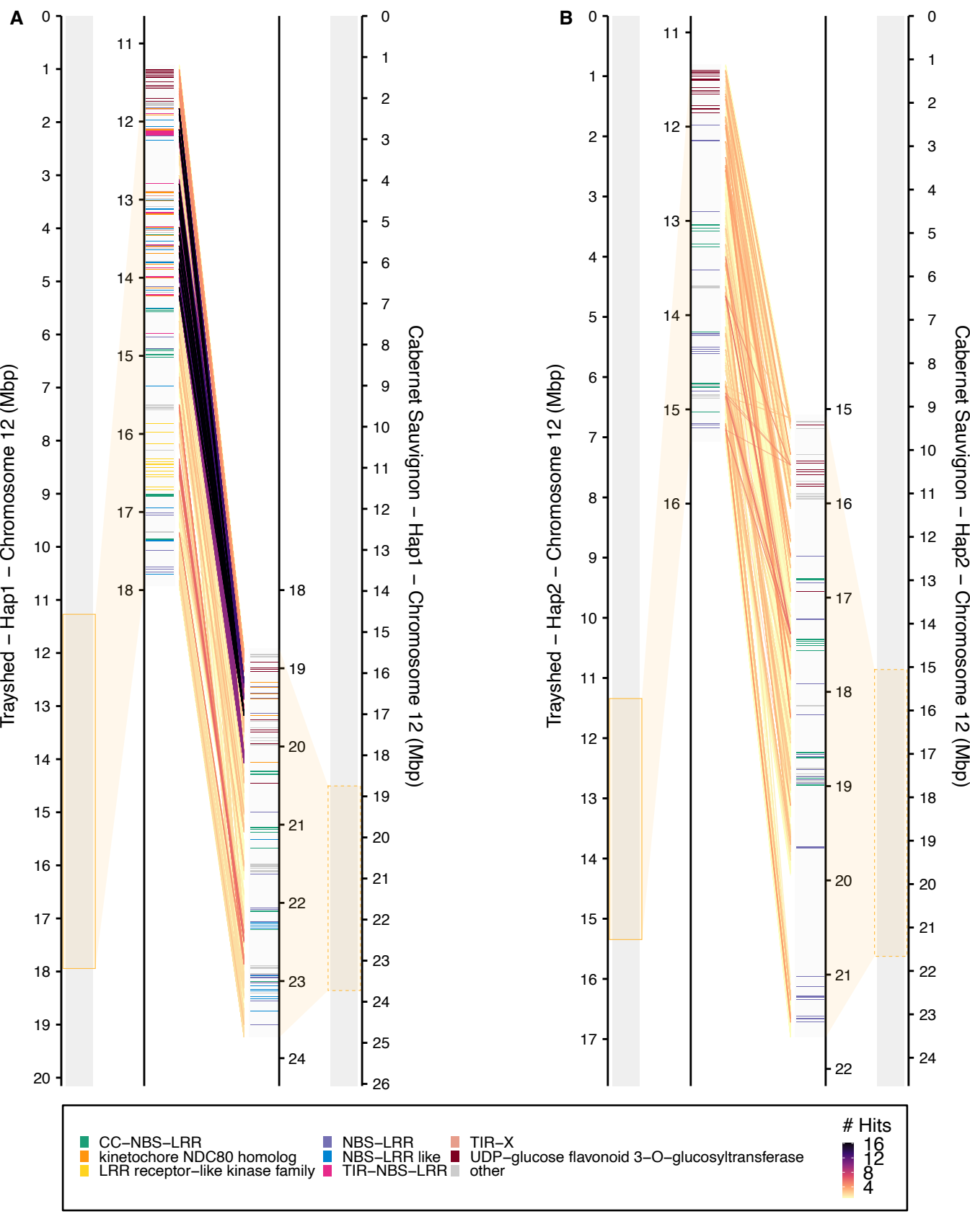
614 Olmo, H. P., 1986 The potential role of (vinifera x rotundifolia) hybrids in grape variety
615 improvement. Experientia 42: 921–926.
616 Olmo, H. P., 1971 Vinifera Rotundifolia Hybrids as Wine Grapes. Am. J. Enol. Vitic. 22: 87–91.
617 Pagès, H., P. Aboyoun, R. Gentleman, and S. DebRoy, 2019 Biostrings: Efficient manipulation
618 of biological strings.
619 Paradis, E., and K. Schliep, 2019 ape 5.0: an environment for modern phylogenetics and
620 evolutionary analyses in R (R. Schwartz, Ed.). Bioinformatics 35: 526–528.
621 Patel, G. I., and H. P. Olmo, 1955 Cytogenetics of Vitis: I. The hybrid V. vinifera x V.
622 rotundifolia. Am. J. Bot. 42: 141–159.
623 Patil, S. G., and V. P. Patil, 1992 Karyomorphology of Vitis vinifera, V. rotundifolia and their
624 hybrid. Cytologia 57: 91–95.
625 Pauquet, J., A. Bouquet, P. This, and A.-F. Adam-Blondon, 2001 Establishment of a local map
626 of AFLP markers around the powdery mildew resistance gene Run1 in grapevine and
627 assessment of their usefulness for marker assisted selection. Theor Appl Genet 103:
628 1201–1210.
629 Pertea, M., G. M. Pertea, C. M. Antonescu, T.-C. Chang, J. T. Mendell *et al.*, 2015 StringTie
630 enables improved reconstruction of a transcriptome from RNA-seq reads. Nat Biotechnol
631 33: 290–295.
632 R Core Team, 2020 *R: A language and environment for statistical computing*. R Foundation for
633 Statistical Computing, Vienna, Austria.
634 Ranwez, V., E. J. P. Douzery, C. Cambon, N. Chantret, and F. Delsuc, 2018 MACSE v2: Toolkit
635 for the Alignment of Coding Sequences Accounting for Frameshifts and Stop Codons (C.
636 Wilke, Ed.). Mol Biol Evol 35: 2582–2584.
637 Ravaz, L., 1902 *Les Vignes américaines. Porte greffes et producteurs directs, caractères,*
638 *aptitudes*.
639 Riaz, S., A. C. Tenscher, D. W. Ramming, and M. A. Walker, 2011 Using a limited mapping
640 strategy to identify major QTLs for resistance to grapevine powdery mildew (Erysiphe
641 necator) and their use in marker-assisted breeding. Theor Appl Genet 122: 1059–1073.
642 Richly, E., J. Kurth, and D. Leister, 2002 Mode of Amplification and Reorganization of
643 Resistance Genes During Recent Arabidopsis thaliana Evolution. Mol Biol Evol 19: 76–
644 84.
645 RStudio Team, 2020 *RStudio: Integrated Development Environment for R*. RStudio, Inc.,
646 Boston, MA.
647 Ruel, J. J., and M. A. Walker, 2006 Resistance to Pierce's Disease in Muscadinia rotundifolia
648 and Other Native Grape Species. Am. J. Enol. Vitic. 57: 158–165.
649 Slater, G., and E. Birney, 2005 Automated generation of heuristics for biological sequence
650 comparison. BMC Bioinformatics 6: 31.
651 Small, J. K., 1913 *Flora of the southeastern United States: being descriptions of the seed-plants,*
652 *ferns and fern-allies growing naturally in North Carolina, South Carolina, Georgia,*
653 *Florida, Tennessee, Alabama, Mississippi, Arkansas, Louisiana and in Oklahoma and*
654 *Texas east of the one hundredth meridian*. J.K. Small, New York.
655 Smit, A., R. Hubley, and P. Green, 2013 *RepeatMasker Open-4.0*.
656 Sousa, A., and S. S. Renner, 2015 Interstitial telomere-like repeats in the monocot family
657 Araceae: Chromosome Evolution in the Araceae. Bot J Linn Soc 177: 15–26.
658 Stanke, M., O. Keller, I. Gunduz, A. Hayes, S. Waack *et al.*, 2006 AUGUSTUS: ab initio
659 prediction of alternative transcripts. Nucleic Acids Res. 34: W435–W439.

660 Staudt, G., and H. H. Kassemeyer, 1995 Evaluation of downy mildew resistance in various
661 accessions of wild Vitis species. Vitis 34: 225–228.
662 Steuernagel, B., K. Witek, S. G. Krattinger, R. H. Ramirez-Gonzalez, H. Schoonbeek *et al*., 2020
663 The NLR-Annotator Tool Enables Annotation of the Intracellular Immune Receptor
664 Repertoire. Plant Physiol. 183: 468–482.
665 Vondras, A. M., A. Minio, B. Blanco-Ulate, R. Figueroa-Balderas, M. A. Penn *et al.*, 2019 The
666 genomic diversification of grapevine clones. BMC Genomics 20: 972.
667 Wagih, O., 2017 *ggseqlogo: A "ggplot2" Extension for Drawing Publication-Ready Sequence*.
668 Walker, M. A., L. A. Lider, A. C. Goheen, and H. P. Olmo, 1991 VR 039-16 Grape Rootstock.
669 HortSci 26: 1224–1225.
670 Walker, M. A., K. Lund, C. Agüero, S. Riaz, K. Fort *et al.*, 2014 Breeding grape rootstocks for
671 resistance to phylloxera and nematodes - it's not always easy. Acta Hortic. 1045: 89–97.
672 Wan, Y., H. R. Schwaninger, A. M. Baldo, J. A. Labate, G.-Y. Zhong *et al.*, 2013 A
673 phylogenetic analysis of the grape genus (Vitis L.) reveals broad reticulation and
674 concurrent diversification during neogene and quaternary climate change. BMC Evol
675 Biol 13: 141.
676 Wang, Y., H. Tang, J. D. DeBarry, X. Tan, J. Li *et al.*, 2012 MCScanX: a toolkit for detection
677 and evolutionary analysis of gene synteny and collinearity. Nucleic Acids Res. 40: e49.
678 Warren, R. F., A. Henk, P. Mowery, E. Holub, and R. W. Innes, 1998 A Mutation within the
679 Leucine-Rich Repeat Domain of the Arabidopsis Disease Resistance Gene RPS5 Partially
680 Suppresses Multiple Bacterial and Downy Mildew Resistance Genes. Plant Cell 10:
681 1439–1452.
682 Waterhouse, R. M., M. Seppey, F. A. Simão, M. Manni, P. Ioannidis et al., 2018 BUSCO
683 Applications from Quality Assessments to Gene Prediction and Phylogenomics. Mol.
684 Biol. Evol. 35: 543–548.
685 Wei, F., K. Gobelman-Werner, S. M. Morroll, J. Kurth, and L. Mao, 1999 The Mla (Powdery
686 Mildew) Resistance Cluster Is Associated With Three NBS-LRR Gene Families and
687 Suppressed Recombination Within a 240-kb DNA Interval on Chromosome 5S (1HS) of
688 Barley. Genetics 153: 1929–1948.
689 Wen, J., L.-M. Lu, Z.-L. Nie, X.-Q. Liu, N. Zhang *et al.*, 2018 A new phylogenetic tribal
690 classification of the grape family (Vitaceae): Tribal classification of Vitaceae. Jnl of
691 Sytematics Evolution 56: 262–272.
692 Wickham, H., M. Averick, J. Bryan, W. Chang, L. McGowan *et al.*, 2019 Welcome to the
693 Tidyverse. Journal of Open Source Software 4: 1686.
694 Wu, T. D., and C. K. Watanabe, 2005 GMAP: a genomic mapping and alignment program for
695 mRNA and EST sequences. Bioinformatics 21: 1859–1875.
696 Yang, Z., 2007 PAML 4: Phylogenetic Analysis by Maximum Likelihood. Mol Biol Evol 24:
697 1586–1591.
698 Yu, G., D. K. Smith, H. Zhu, Y. Guan, and T. T.-Y. Lam, 2017 ggtree: an r package for
699 visualization and annotation of phylogenetic trees with their covariates and other
700 associated data. Methods Ecol. Evol. 8: 28–36.
701 Zhou, F., J. Kurth, F. Wei, C. Elliott, G. Valè *et al.*, 2001 Cell-Autonomous Expression of Barley
702 Mla1 Confers Race-Specific Resistance to the Powdery Mildew Fungus via a Rar1-
703 Independent Signaling Pathway. Plant Cell 13: 337–350.
704 Zhou, Y., A. Minio, M. Massonnet, E. Solares, Y. Lv et al., 2019 The population genetics of
705 structural variants in grapevine domestication. Nat. Plants 5: 965–979.

706 Zini, E., C. Dolzani, M. Stefanini, V. Gratl, P. Bettinelli *et al.*, 2019 R-Loci Arrangement Versus
707 Downy and Powdery Mildew Resistance Level: A Vitis Hybrid Survey. IJMS 20: 3526.
708 Zou, C., A. Karn, B. Reisch, A. Nguyen, Y. Sun et al., 2020 Haplotyping the Vitis collinear core
709 genome with rhAmpSeq improves marker transferability in a diverse genus. Nat.
710 Commun. 11: 413.
711


A
Query: Trayshed – Hap1
Reference: Cabernet Sauvignon – Hap1
%id
98
96
94
92
90
B
PN40024 V0 chr07
S7_61414
S7_163514
S7_4787985
S7_4922587
S7_11096594
S7_14328892
S7_14403851
S7_20700066
Trayshed chr07
Trayshed chr20
C
V. vinifera cv. Cabernet Sauvignon – Chromosome 7 (Mbp)
V. vinifera PN40024 – Chromosome 7 (Mbp)
Telomeric repeats
Bits
count


A
Trayshed
B
Cabernet Sauvignon
Hap1
Hap2
chr01
chr02
chr03
chr04
chr05
chr06
chr07
chr08
chr09
chr10
chr11
chr12
chr13
chr14
chr15
chr16
chr17
chr18
chr19
chr20
RUN1/RPV1
0 5 10 15 20 25 30 35 40
Position (Mbp)
C
0.05
Trayshed
Cabernet Sauvignon
CC–NBS–LRR
CC–NBS
CC–X
TIR–NBS–LRR
TIR–NBS
TIR–X
NBS–LRR
NBS
D
E
0 10 20 30 40
Time (Ma)
0.00 0.25 0.50 0.75 1.00


A
B
Trayshed – Hap1 – Chromosome 12 (Mbp)
Cabernet Sauvignon – Hap1 – Chromosome 12 (Mbp)
Trayshed – Hap2 – Chromosome 12 (Mbp)
Cabernet Sauvignon – Hap2 – Chromosome 12 (Mbp)
CC–NBS–LRR
kinetochore NDC80 homolog
LRR receptor–like kinase family
NBS–LRR
NBS–LRR like
TIR–NBS–LRR
TIR–X
UDP–glucose flavonoid 3–O–glucosyltransferase
other
Hits
16
12
8
4