Genome diversification, deleterious mutation accumulation, and evidence of negative selection among clonally propagated grapevines

Amanda M. Vondras[1], Andrea Minio[1], Barbara Blanco-Ulate[2], Rosa Figueroa-Balderas[1], Michael A. Penn[1], Yongfeng Zhou[3], Danelle Seymour[3], Ye Zhou[1], Dingren Liang[1], Lucero K. Espinoza[1], Michael M. Anderson[1], M. Andrew Walker[1], Brandon Gaut[3], Dario Cantu[1]*

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

[2] Department of Plant Sciences, University of California, Davis, CA 95616

[3] Department of Ecology and Evolutionary Biology, University of California, Irvine, CA 92617

*Corresponding author. Telephone: +1 530-752-2929 Email: dacantu@ucdavis.edu

Abstract

Vegetatively propagated clones accumulate somatic mutations. The purpose of this study was to better understand the consequences of clonal propagation and involved defining the nature of somatic mutations throughout the genome.

Sixteen Zinfandel winegrape clone genomes were sequenced and compared to one another using a highly contiguous genome reference produced from one of the clones, Zinfandel 03.

Though most heterozygous variants were shared, somatic mutations accumulated in individual and subsets of clones. Overall, heterozygous mutations were most frequent in intergenic space and more frequent in introns than exons. A significantly larger percentage of CpG, CHG, and CHH sites in repetitive intergenic space experienced transition mutations than genic and non-repetitive intergenic spaces, likely because of higher levels of methylation in the region and the disposition of methylated cytosines to spontaneously deaminate. Of the minority of mutations that occurred in exons, larger proportions of these were putatively deleterious when they occurred in relatively few clones.

Repetitive intergenic space is a major driver of clone genome diversification. Clonal propagation is associated with the accumulation of putatively deleterious mutations. The data suggest selection against deleterious variants in coding regions such that mutations are less frequent in coding than noncoding regions of the genome.

Keywords

Clonal propagation, DNA methylation, genome diversification, somatic mutations, structural variation, transposable elements

Introduction

Cultivated grapevines are clonally propagated. As a result, the genome of each cultivar is preserved, except for the accumulation of mutations that accumulate over time and can generate distinguishable clones (Franks *et al.*, 2002; Riaz *et al.*, 2002; Ramu *et al.*, 2017; Zhou *et al.*, 2017). Somatic mutations are responsible for several notable phenotypes. For example, a single, semi-dominant nucleotide polymorphism can affect hormone response (Boss & Thomas, 2002) and recessive insertion of the *Gret1* retrotransposon in the promoter of the *VvmybA1* transcription factor inhibits anthocyanin accumulation in white varieties (Kobayashi *et al.*, 2004), as do additional mutations affecting the color locus (Walker *et al.*, 2006; Yakushiji *et al.*, 2006; Pelsy *et al.*, 2015; Zhou *et al.*, 2018). The fleshless fruit of an Ugni Blanc clone and the reiterated reproductive meristems observed in a clone of Carignan are both caused by dominant transposon insertion mutations (Fernandez *et al.*, 2010; 2013). In citrus, undesirable mutations can be unknowingly propagated that render fruit highly acidic and inedible (Soost *et al.*, 1961; Whitham & Slobodchikoff, 1981). Interestingly, somatic mutations in plum are associated with a switch from climacteric to non-climacteric ripening behavior (Farcuh *et al.*, 2017).

There is limited understanding and evidence of the extent, nature, and implications of the somatic mutations that accumulate in clonally propagated crops (McKey *et al.*, 2010). Genotyping approaches based on whole genome sequencing make it possible to identify genetic differences without predefined markers (Carrier *et al.*, 2012; Gambino *et al.*, 2017; Roach *et al.*, 2018) and expedite learning the genetic basis of valuable traits and developmental processes (Farcuh *et al.*, 2017; Carbonell-Bejerano *et al.*, 2017). Still, few previous studies have used genomic approaches to study somatic variations among clones (Carrier *et al.*, 2012; Carbonell-Bejerano *et al.*, 2017; Gambino *et al.*, 2017; Plomion *et al.*, 2018; Roach *et al.*, 2018). The first to publish a genome-wide exploration of somatic variation in grapevine was Carrier *et al.* (2012), finding that transposable elements were the largest proportion of somatic mutation types affecting four Pinot Noir clones. Whole genome sequencing was also used to study structural variations and complex chromosomal rearrangements in Tempranillo, comparing diverse accessions of phenotypically distinct Tempranillo Tinto and Tempranillo Blanco to better understand the basis of somatic mutations giving rise to red versus white fruit (Carbonell-Bejerano *et al.*, 2017). Genomic tools could be used to comprehensively describe the extent of somatic mutations and infer the processes affecting clone genomes.

75 Mutations occur in somatic cells that proliferate by mitosis. These can occur by a variety
76 of means, including single base-pair mutations (Ossowski *et al.*, 2010; Hershberg & Petrov,
77 2010) that are more prevalent in repetitive regions because methylated cytosines are passively
78 deaminated to thymines (Selker, 1990; Mautino & Rosa, 1998; Meunier *et al.*, 2005), polymerase
79 slippage that drives variable microsatellite insertions and deletions (Schlötterer & Tautz, 1992),
80 and larger structural rearrangements and hemizygous deletions (Carbonell-Bejerano *et al.*, 2017;
81 Zhou *et al.*, 2018). Transposable elements are also a major source of somatic mutations in
82 grapevines (Carrier *et al.*, 2012), though transcriptional and post-transcriptional mechanisms
83 exist to prevent transposition and maintain genome stability (Chan *et al.*, 2005; Qi *et al.*, 2006;
84 Cantu *et al.*, 2010; Shen *et al.*, 2012). Notably, methylation of transposable elements is one
85 specific mechanism that prevents transposition, which establishes a tradeoff, then, between
86 methylation and the transposition of mobile elements.
87 At the cellular level, distinct clones can emerge following a mutation in a shoot apical
88 meristem that spreads throughout a single cell layer, creating periclinal chimeras. This chimera is
89 stable for Pinot Meunier, a clone of Pinot Noir with distinct L1 and L2 layers in shoots (Franks
90 *et al.*, 2002). Each cell layer in a stratified apical meristem like that observed in grape
91 (Thompson & Olmo, 1963) is developmentally distinct. The distinct cell layers will remain so
92 provided cell divisions occur anticlinally. But, periclinal divisions and cellular rearrangements
93 can result in the homogenization of a mutant genotype across cell layers (Hocquigny *et al.*,
94 2004). This is the case for green-yellow bud sports of the grey-fruited Pinot Gris, wherein sub-
95 epidermal white cells invaded and displaced epidermal pigmented cells (Pelsy *et al.*, 2015). In
96 contrast to replacement (L1 cells invade L2), displacement is likely more common because of the
97 relative disorganization of the inner cell layers (Thompson & Olmo, 1963; Hocquigny *et al.*,
98 2004).
99 Meristem architecture is related to the fate of somatic mutations, as it influences the
100 impact of these mutations and the likelihood of competition between cell lineages, also known as
101 diplontic selection (Klekowski *et al.*, 1985; Tilney-Bassett, 1986; Klekowski, 2003). Provided
102 each cellular layer is maintained by anticlinal divisions, deleterious mutations can be preserved
103 in periclinal chimeras (Klekowski *et al.*, 1985; Klekowski, 1998). In addition, the predominance
104 of "hidden", heterozygous recessive somatic mutations (Klekowski, 1998; Zhou *et al.*, 2017)
105 may further shield somatic mutations from selective forces. These factors are permissive of the
106 accumulation of somatic mutations. Diplontic selection could occur if periclinal cell divisions

107 result in the invasion of one cell layer by cells from another (Klekowski *et al.*, 1985; Klekowski,
108 2003). This mechanism could oppose the accrual of deleterious mutations expected by Muller
109 (Muller, 1932; Pineda-Krch & Fagerström, 1999). A recent study of the long-lived pedunculate
110 oak, *Quercus robur*, described substantial intra-organismal genetic variation, but did not draw
111 conclusions about the contribution of somatic variations to large-scale oak evolution (Plomion *et*
112 *al.*, 2018). Evidence of diplontic selection in plants is remarkably scarce (Klekowski, 1998),
113 though its likelihood given different circumstances has been modeled (Klekowski *et al.*, 1985;
114 Orive, 2001; Klekowski, 2003). Given the prevalence of chimerism and rearrangements
115 documented in the model (Hocquigny *et al.*, 2004; Pelsy *et al.*, 2015), grapevine is a suitable
116 model for investigating the possibility of selection during vegetative propagation.
117 Zinfandel is the third-most cultivated wine grape in California (CDFA, 2016a,b) DNA
118 profiling produced evidence that Zinfandel is synonymous with Primitivo grown in Italy (Bowers
119 *et al.*, 1993) and Croatian Pribidrag and Crljenak Kastelanski (Maletic *et al.*, 2003). Historical
120 records plus the cultivation of closely related cultivars support Croatia as the likely origin of
121 Zinfandel (Mirošević & Meredith, 2000; Maletic *et al.*, 2003; 2004; Fanizza *et al.*, 2005) and
122 also that Primitivo was likely brought to the Gioia del Colle region in Italy by Benedictine
123 monks in the 17th century (Franks *et al.*, 2002; Russo *et al.*, 2003). The reported variability in
124 Zinfandel (Wolpert, 1996; Fidelibus *et al.*, 2005; Zdunić *et al.*, 2014), including variability in
125 phenolic metabolites (Fig. S1, Methods S1), and its long history of cultivation make it a useful
126 model for studying clonal variation in grapevine, specifically, and the nature of the accumulation
127 of somatic mutations in clonally propagated crops, generally.
128 The purpose of this study was to better understand the nature of the somatic variations
129 that occur during clonal propagation. Representatives from at least a portion of Zinfandel's
130 history (Mirošević & Meredith, 2000; Maletic *et al.*, 2003; 2004; Fanizza *et al.*, 2005) from
131 Croatia, Italy, and California were sequenced and compared using Zin03 as reference. First, we
132 show that intergenic space drives clonal diversification. As previously reported, transposable
133 element insertions varied among clones (Carrier *et al.*, 2012). This report expands that
134 understanding to implicate methylation as an indirect driver of clonal diversification; rare
135 somatic heterozygous SNPs were most observed in the repetitive intergenic regions, likely
136 because of the high levels of transposition-inhibiting methylation and associated transition
137 mutations that are prevalent there. Second, the data support an important component of Muller's
138 ratchet (Muller, 1932), that asexually propagated organisms accumulate deleterious mutations.

139 Third, somatic mutations were relatively scarce in the coding regions of genes relative to introns
140 and intergenic space, suggesting some degree of negative selection against deleterious mutations.
141
142 **Materials and Methods**
143 *Zinfandel plant material and additional accessions*
144 Sixteen Zinfandel clones were used for this study. Plants were confirmed to be clones of
145 Zinfandel using the following microsatellite markers: VVMD5, VVMD7, VVMD27, VVMD31,
146 VVMD32, VVMS2, VRZAG62, and VRZAG79 (Thomas *et al.*, 1994; Sefc *et al.*, 1999; Maletic
147 *et al.*, 2003). Fourteen of these clones are available through Foundation Plant Services (FPS) at
148 the University of California Davis. Nine of the sixteen clones belong to the Zinfandel Heritage
149 Vineyard Project, a collection of rare Zinfandel vine cuttings grown in the same vineyard. The
150 identification numbers, common names, and source of the clones used in this study are listed in
151 Table 1. An FPS identification number suffix of ".1" indicates that the clone underwent
152 microshoot tip tissue culture therapy, with two exceptions. Pribidrag 13 and Pribidrag 15 are
153 directly derived from the same plants as Pribidrag 4 and Pribidrag 5, respectively, but did not
154 undergo microshoot tip tissue culture therapy. They are labeled with identical FPS numbers to
155 make clear that the relationship between them is known. In this manuscript, Zinfandel clones will
156 be referred to by the clone numbers and common names listed in Table 1.
157
158 *DNA extraction, library preparation, and sequencing*
159 High quality genomic DNA was isolated from grape leaves using the method described in
160 Chin *et al.* (2016). DNA purity was evaluated with a Nanodrop 2000 spectrophotometer (Thermo
161 Scientific, Hanover Park, IL), quantity with a Qubit 2.0 Fluorometer (Life Technologies,
162 Carlsbad, CA) and integrity by electrophoresis. For SMRT sequencing, SMRTbell libraries for
163 the Zinfandel reference FPS clone 03 (Zin03) were prepared as described by Chin *et al.* (2016).
164 For Illumina sequencing, DNA sequencing libraries for each of the sixteen Zinfandel clones were
165 prepared using the Kapa LTP library prep kit (Kapa Biosystems) as described by Jones *et al*.,
166 (2014). Final libraries were evaluated for quantity and quality using a Bioanalyzer 2100 (Agilent
167 Technologies, CA). Zin03 SMRTbell libraries were sequenced on a PacBio RS II and Illumina
168 libraries were sequenced in 100 and 150 base-pair paired-end reads on an Illumina HiSeq3000
169 sequencer (DNA Technology Core Facility, University of California, Davis). Genome sequences
170 of additional *V. vinifera* were used in this study, including long reads from Cabernet sauvignon

171 (NCBI BioProject PRJNA316730) and short reads from Cabernet franc, Chardonnay, Merlot,
172 Pinot Noir, and Sauvignon blanc (NCBI BioProject PRJNA527006).
173
174 *Zinfandel genome assembly and annotation*
175 *De novo* assembly of Zinfandel (Zin03) was performed at DNAnexus (Mountain View,
176 CA, USA) using PacBio RS II data and the FALCON-unzip (v. 1.7.7) pipeline (Chin *et al.*,
177 2016). FALCON-unzip was used for its ability to assemble contiguous, phased diploid genomes
178 with better resolved heterozygosity (Chin *et al.*, 2016; Minio *et al.*, 2017). Repetitive sequences
179 were masked prior to error correction using TANmask and REPmask modules in Damasker
180 (Myers, 2014). After error-correction (13,073 bp length cut-off), a total of 1.68 million error-
181 corrected reads (N50 15Kbp, 29-fold coverage of expected genome size) were obtained and
182 repeats were masked before overlap detection in the FALCON pipeline (v. 1.7.7). PacBio reads
183 were assembled after testing multiple parameters to produce the least fragmented assembly.
184 These conditions are listed in Methods S2. Haplotype reconstruction was performed with default
185 parameters. Finally, contigs were polished with Quiver (Pacific Biosciences, bundled with
186 FALCON-unzip v. 1.7.7). Repeats were annotated on the Zin03 assembly using RepeatMasker
187 (v. open-4.0.6; Smit *et al.* 2013) and a *V. vinifera* repeat library (Minio *et al.*, 2019).
188 The publicly available RNAseq datasets listed in Methods S2 were used as transcriptional
189 evidence for gene prediction. Each RNAseq sample was trimmed with Trimmomatic (v. 0.36;
190 Methods S2) and assembled with Stringtie (v. 1.3.3; Pertea *et al.*, 2016) to reconstruct variety-
191 specific transcripts. A detailed list of all experimental data used for the annotation procedure is in
192 Methods S2. This data was then mapped on the genome using Exonerate (v. 2.2.0, transcripts and
193 proteins; Slater & Birney, 2005) and PASA (v. 2.1.0, transcripts; Haas *et al.*, 2003). Alignments,
194 and *ab initio* predictions generated with SNAP (v. 2006-07-28; Korf, 2004), Augustus (Stanke *et*
195 *al.*, 2006), and GeneMark-ES (Lomsadze *et al.*, 2005) were used as input for EVidenceModeler
196 (v. 1.1.1; Haas *et al.*, 2008). EVidenceModeler was used to identify consensus gene structures
197 using the weight reported in Methods S2. Functional annotation was performed using the RefSeq
198 plant protein database (ftp://ftp.ncbi.nlm.nih.gov/refseq, retrieved January 17th, 2017) and
199 InteProScan (v. 5) as previously described (Minio *et al.*, 2019).
200
201 *Genetic variant calling*

202 Comparisons between Zinfandel clones and between Zin03 and other cultivars were
203 made using the Zin03 genome as reference. This pipeline is described in Methods S3. Small
204 insertions and deletions (INDELs), single nucleotide polymorphisms (SNPs), and structural
205 variations (SVs) were analyzed. The short Illumina reads belonging to the sixteen Zinfandel
206 clones and additional cultivars were trimmed using Trimmomatic (v. 0.36; Methods S2). Quality
207 filtered and trimmed paired-end reads were then randomly down-sampled to 84 million (~14X
208 coverage) in each library to mitigate the possibility of sequencing depth-dependent outcomes.
209 All libraries were aligned to Zin03 using bwa (v. 0.7.10) and the -M parameter (Li & Durbin,
210 2009). For all genotypes, the median number of reads mapping to the Zinfandel reference
211 genome was approximately 97%. Next, Picard Tools (v. 2.12.1) were used to mark optical
212 duplicates, build BAM indices, and validate SAM files (http://broadinstitute.github.io/picard).
213 Variants were called using GATK's HaplotypeCaller (v. 3.5; Van der Auwera *et al.*, 2013).
214 Then, called variants were filtered and annotated (--filterExpression "QD < 2.0 || FS > 60.0 || MQ
215 < 40.0 || MQRankSum < -12.5 || ReadPosRankSum < -8.0"). Variant call files were combined
216 using GATK's GenotypeGVCFs. Having mapped Illumina reads corresponding to the Zinfandel
217 reference onto itself, erroneous non-reference Zin03 calls (8.1%) were removed. The variants
218 called included SNPs and INDELs.
219 Next, large structural variations among clones, between Zin03 and other cultivars, and
220 between Zin03's haplotypes were studied. First, Zin03 genes were compared to PN40024 and
221 Cabernet Sauvignon (CS08) by mapping coding sequences on genome assemblies using Gmap
222 (v. 2015-09-29) and the following parameters: -K 20,000 -B 4 -f 2. Hits with at least 80%
223 identity and reciprocal coverage are reported. Genes annotated on Zin03's haplotig assembly
224 were also mapped to Zin03's primary assembly to assess differences in gene content between
225 Zin03's haplotypes. SMRT reads from Zin03 and CS08 were mapped to Zin03 using NGMLR
226 (v. 0.2.7) and structural differences were called with Sniffles (v.1.0.8; Sedlazeck *et al.*, 2018).
227 Zinfandel clones were compared to one another using Illumina short reads and Delly (v. 0.7.8)
228 with default parameters (Rausch *et al.*, 2012). The structural variations identified by Sniffles and
229 Delly in Zin03 were intersected. Several filters were applied to the results of SV analyses.
230 Transversions, non-reference Zin03 genotype calls, SVs annotated at the ends of contigs, and
231 SVs that intersected the repeat annotation were filtered from Delly output.
232
233 *Transposon insertion analysis*

234 PoPoolationTE2 (v. 1.10.04; Kofler *et al.*, 2016) was used to identify transposon
235 insertions in the Zinfandel clones; it was used following the workflow outlined in its software
236 manual (https://sourceforge.net/p/popoolation-te2/wiki/Manual/). Insertions were called relative
237 to Zin03 genome assembly and PN20024 (Jaillon *et al.*, 2007). As described in Kofler *et al.*
238 (2016), PoPoolationTE2 analyses transposable element insertions and can identify novel and
239 annotated TE insertions provided insertions fall within predefined families of TEs. The
240 annotation produced by RepeatMasker was used for the analysis. In this manuscript, the TE
241 insertions among the clones are reported using the classification system and nomenclature
242 described by Wicker *et al.* (2007). In instances where the TE order and/or superfamily was not
243 annotated, only the TE class and order, when available, are named in the associated figures and
244 text.
245
246 *Relationships between Zinfandel clones*
247 The relationships between Zinfandel clones were visualized by Principal Component
248 Analysis and their relatedness was quantified (VCFtools v. 0.1.15) based on the method
249 described by Manichaikul *et al*. (2010). This approach gives information about the relationship
250 of any pair of individuals (unrelated, 3rd degree relative, 2nd degree relative, full siblings, and
251 self) by estimating their kinship coefficient, which ranges from zero (no relationship) to 0.50
252 (self). These analyses used SNPs outside of repetitive regions.
253
254 **Results**
255 *Zinfandel genome assembly, annotation, and differences between haplotypes*
256 The clone used for the genome assembly, Zinfandel 03 (Zin03), was acquired by FPS in
257 1964 from the Reutz Vineyard near Livermore, California that was planted during Prohibition
258 (1920 – 1933; Sweet & Wolpert, 2007). Zin03 was sequenced using Single Molecule Real-Time
259 (SMRT; Pacific Biosciences) technology at ~98x coverage and assembled using FALCON-unzip
260 (Chin *et al.*, 2016), a diploid-aware assembly pipeline. The genome was assembled into 1,509
261 primary contigs (N50 = 1.1 Mbp) for a total assembly size of 591 Mbp, similar to the genome
262 size of Cabernet Sauvignon (590 Mbp; Chin *et al.*, 2016) and larger than Chardonnay (490Mb;
263 Roach *et al.*, 2018) and PN40024 (487 Mb; Jaillon *et al.*, 2007). Fifty two percent of the genome
264 was phased into 2,246 additional phased sequences (haplotigs) where the homologous
265 chromosomes were distinguishable with an N50 of ~442 kbp (Table 2). A total of 53,560

266 complete protein-coding genes were annotated on the primary (33,523 genes) and haplotig
267 (20,037 genes) assemblies (Table 2).
268 Of the 20,037 genes annotated on the haplotig assembly, 18,878 aligned to the primary
269 assembly, leaving 1,159 genes that may exist hemizygously in the genome due to structural
270 variation between homologous chromosomes or because of substantial divergence in sequence
271 between haplotypes. These genes were annotated with a broad variety of putative functions,
272 including biosynthetic processes, secondary metabolism, and stress responses. Long reads were
273 mapped to both the primary and haplotig assemblies to evaluate the circumstances that explain
274 the differences between haplotypes. Structural variants (SVs) between the haplotypes were
275 examined by mapping long SMRT sequencing reads onto Zin03's primary and haplotig
276 assemblies with NGMLR and calling SVs with Sniffles (Sedlazeck *et al.*, 2018).
277 A total of 22,399 SVs accounted for 6.94% (41.0 / 591 Mbp) of the primary assembly's
278 length and 6.02% (8.4 / 139 Mbp) of the primary assembly's gene-associated length (Fig. 1a,
279 Table 3). SVs intersected 4,559 genes in the primary assembly (13.6% of primary assembly
280 genes) and 390 SVs spanned more than one gene. Manual inspection of the long reads aligned to
281 the primary assembly support that large, heterozygous deletions and inversions occurred in the
282 Zin03 genome that were either inherited from different structurally distinct parents or arose
283 during clonal propagation (Fig. 1b,c,d). Importantly, there was substantial hemizygosity in the
284 genome, with long reads supporting deletions affecting 2,521 genes and 4.56% of the primary
285 assembly's length (Table 3).
286 Next, we considered whether specific structural variation could account for the 1,159
287 genes uniquely found in the haplotig assembly. Three hundred eighty-two genes of the
288 previously mentioned 1,159 genes that uniquely exist within the haplotig assembly intersected
289 structural variations. Two hundred ninety of these intersected deletions, accounting for the
290 failure to identify them on the primary assembly. Some of the haplotig genes that failed to map
291 to the primary assembly intersected additional types of SVs, including duplications (80 genes),
292 insertions (89 genes), and inversions (16 genes).
293 These results reveal structural differences between Zinfandel's haplotypes. These
294 differences could have been inherited and/or could have occurred during clonal propagation.
295 Overall, these structural variations affected 4,559 primary assembly genes. Importantly, these
296 data show that a notable portion of the primary assembly's length (4.56%) is hemizygous.
297

298 *Differences in structure and gene content between Zinfandel and other grape genomes*

299 The Zin03 genome was compared to PN40024 and Cabernet Sauvignon to identify
300 cultivar-specific genes that may contribute to Zinfandel's characteristics. PN40024 is the inbred
301 line derived from Pinot Noir used to develop the first grape genome reference (Jaillon *et al.*,
302 2007) and Cabernet Sauvignon (CS08) was recently used to construct the first diploid,
303 haplotype-resolved grape genome for which long reads are available (Chin *et al.*, 2016). Overall,
304 1,801 genes were not shared between all three genotypes (Zin03, Pinot Noir, and Cabernet
305 Sauvignon; Fig. 2a). Three hundred nine protein coding genes were found uniquely in Zin03
306 relative to PN40024 and CS08; 223 were annotated on the primary assembly and 86 were
307 annotated on the haplotigs (Fig. 2a, Table S1). These genes had a panoply of functions that
308 included but were not limited to nucleotide binding (60 genes), protein binding (58 genes), stress
309 response (34 genes), and kinases (28), and were associated with membranes (48 genes), signal
310 transduction (23 genes), carbohydrate metabolism (12 genes), and lipid metabolism (8 genes;
311 Table S1).

312 Structural differences between Zin03 and CS08 were explored in more detail by mapping
313 the long SMRT reads of CS08 onto Zin03's primary and haplotig assemblies with NGMLR and
314 calling SVs with Sniffles (Fig. 2b, Table 3). Overall, these SVs corresponded to 17.74% (159/
315 897 Mbp) of the Zin03 assembly's total length, 12.5% of its total protein-coding regions (28 /
316 223 Mbp), and 25.6% of all Zin03 genes. SVs affected 9,885 genes in the primary assembly and
317 3,804 genes in the haplotigs. Manual inspection of the alignment of long CS08 reads to Zin03's
318 primary assembly support that large SVs exist between the two genotypes (Fig. 2c,d). Next, we
319 considered whether specific structural variation called by Sniffles could account for the 576
320 Zin03 genes absent from CS08 according to the reciprocal mapping analysis (Fig. 2a). Of these
321 576 Zinfandel genes, 268 genes intersected 454 deletions supported by long CS08 reads aligned
322 to Zin03.

323 Though Zinfandel had few unique genes, high levels of structural variation between
324 Zinfandel (Zin03) and Cabernet Sauvignon (CS08) were observed and these affected
325 considerable protein-coding regions of the genome. These results justify constructing a
326 Zinfandel-specific reference to better capture genomic variability among Zinfandel clones that
327 could otherwise be missed, particularly if an alternative reference lacks sequences present in
328 Zinfandel.

329

330 *Relatedness among Zinfandel clones*

331 Sixteen Zinfandel clones, including Zin03, were sequenced using Illumina. The resulting
332 reads were aligned to the Zin03 primary assembly to characterize SNPs, small INDELs, variable
333 transposon insertions, and large structural variants. Principal Component Analysis (PCA) of
334 variants among the clones showed no clear pattern in their relationships to one another based on
335 their recorded origins prior to acquisition by FPS (Fig. 3a,b). The ambiguity surrounding the
336 travels and histories of these clones means that it should not be taken for granted that the
337 Californian selections, for example, ought to be more closely related to one another than to the
338 Italian or Croatian selections. Notably, Crljenak kaštelanski 3 stands notably apart from the other
339 Zinfandel clones. In addition, Pribidrags 5 and 15, which have a known and close relationship,
340 do not co-localize in the PCA (Fig. 3a,b, Table 1).

341 A kinship analysis (Manichaikul *et al.*, 2010) was then used to quantitatively assess the
342 relationships between the Zinfandel selections. These values range from zero (unrelated) to 0.5
343 (self). Additional cultivars were included in the analysis with known relationships to help
344 contextualize the differences between clones and the integrity of the analysis (Fig. 3c). Cabernet
345 Franc and Merlot have a parent - offspring relationship, as do Pinot Noir and Chardonnay
346 (Bowers *et al.*, 1999; Boursiquot *et al.*, 2009). These pairs had kinship coefficients of 0.15 and
347 0.18, respectively (Fig. 3c). As a possible grandparent of Sauvignon Blanc, Pinot Noir had a
348 kinship coefficient of 0.05 with Sauvignon blanc (Regner *et al.*, 2000; Imazio *et al.*, 2002). Most
349 of the Zinfandel selections had kinship coefficients between 0.42 and 0.45; this is likely because
350 of the accrual of somatic mutations among clones. However, Crljenak kaštelanski 3 had a
351 noticeably low kinship coefficient (0.36 - 0.37) with every other Zinfandel clone (Fig. 3c). These
352 data suggest that Crljenak kaštelanski 3 is either not a clone of Zinfandel, contradicting marker
353 analyses, or that it is a highly divergent clone.

354 Across the Zinfandel clones, the median number of homozygous and heterozygous
355 variants called relative to Zin03 were 38,092 and 717,925, respectively. Between 10-fold and
356 ~27-fold more heterozygous variants were called than homozygous variants in each clone except
357 for Crljenak kaštelanski 3, for which only ~2.5-fold more heterozygous sites were called (Table
358 S2). Crljenak kaštelanski 3 had 4.3-fold more homozygous variants and 1.8-fold fewer
359 heterozygous variants than the other clones (Table S2). Furthermore, unlike other clones, for
360 which less than 10% of sites did not share the Zin03 reference allele, ~29% of variant sites were
361 called where Crljenak kaštelanski 3 did not share the Zin03 reference allele (Table S2). Together,

362 these results suggest that unlike other Crljenak kaštelanski selections (Maletic *et al.*, 2003;
363 Fanizza *et al.*, 2005; Zdunić *et al.*, 2014), Crljenak kaštelanski 3 is likely a close relative of, but
364 quite possibly not a clone of, Zinfandel. Because these analyses cast doubt on its identity as a
365 Zinfandel clone, Crljenak kaštelanski 3 was excluded from the clonal variation analyses
366 described in the rest of the study. Fifteen clones, including Zin03, remained in the analyses
367 (Table 1).
368
369 *Clonal versus cultivar genetic variability*
370 Overall, an average of 761,948 variant sites were identified in individual Zinfandel clones
371 when short reads were mapped on the Zin03 primary assembly. As stated previously, this
372 analysis excluded Crljenak kaštelanski 3. On average, 6,153,830 variant sites were identified in
373 other cultivars (Pinot noir, Chardonnay, Sauvignon Blanc, Merlot, Cabernet Franc) relative to
374 Zin03 (Table S2). Both of these figures excluded heterozygous sites at which the diploid
375 genotype called for a given sample was identical to that called for Zin03.
376 Variants were 7.9X more frequent in other cultivars relative to Zin03 than for Zinfandel
377 clones; on average, mutations in clones occurred once every 723 bases and once every 92 bases
378 in other cultivars (Table S2). However, the ratio of transitions to transversion mutations and the
379 proportions of the severities of the predicted variant effects were similar for both groups (Table
380 S2). The normalized count of variants differed between cultivars and Zinfandel clones on the
381 basis of variants' location in the genome, the type of variant, and the zygosity of the variant (Fig.
382 4).
383 Variants in non-Zinfandel cultivars and heterozygous variants among Zinfandel clones
384 were significantly more prevalent in intergenic space than introns and exons and significantly
385 more prevalent in introns than exons (Tukey HSD, $p < 0.01$). Unlike homozygous variants
386 between cultivars and as expected, homozygous variants were rare among clones (Fig. 4, Table
387 S2). Still, the normalized count of homozygous INDELs in intergenic space, introns, and exons
388 were significantly different among Zinfandel clones (Tukey HSD, $p < 0.01$), as were the
389 normalized count of homozygous intergenic versus genic (exons and introns) SNPs (Tukey HSD,
390 $p < 0.01$). The normalized count of homozygous SNPs in exons and introns were not
391 significantly different in Zinfandel clones (Tukey HSD, $p > 0.01$). The accrual of predominantly
392 heterozygous and likely recessive variants (Zhou *et al.*, 2017) is consistent with what would be
393 expected given physically separate homologous chromosomes and the absence of sexual

394 reproduction. The differences in mutation abundances observed were initially surprising; if
395 somatic mutations occurred randomly and absent mechanisms that make certain sites more or
396 less susceptible to mutation, then different regions of the genome should have had equal levels of
397 mutations. This was not the case (Figure 5).
398
399 *The accrual of somatic mutations in Zinfandel clones*
400 Heterozygous sites found among the 15 Zinfandel clones ought to be a mixture of sites
401 inherited from their shared ancestral plant and somatic mutations that arose during clonal
402 propagation. To better understand the nature of somatic mutations, the data were handled slightly
403 differently than they were to construct Figure 5; all 15 Zinfandel clones were included (including
404 Zin03, Crljenak kaštelanski 3 still excluded) and all heterozygous calls were considered, even if
405 all genotypes were identically heterozygous. Thirty percent of heterozygous SNPs, 24% of
406 heterozygous INDELs, and 47% of heterozygous structural positions were shared by all 15
407 Zinfandel clones (Fig. 5a). Because all clones are identically heterozygous at these loci, these
408 variants are those inherited from Zinfandel's parents.
409 Individual and subsets of Zinfandel clones accumulated heterozygous mutations as clonal
410 propagation occurred (Fig. 5a). Thirteen percent and 16% of heterozygous INDELs and SNPs,
411 respectively, and 1% of large (>50 bp) structural variants occurred in only one or two clones
412 (Fig. 5a). The distribution of SVs called by Delly is markedly different than those of SNPs and
413 INDELs (Fig. 5a). For both SNPs and INDELs, there were 3 and 3.5-fold as many heterozygous
414 variants shared by all 15 clones as there were uniquely occurring variants; there were 71.5-fold
415 more structural variants shared by all clones than there were unique variants in individual clones
416 (Fig. 5a). This might imply that the mechanisms that give rise to small mutations are more
417 common among clones than the large-scale changes associated with SVs.
418 The distribution of unique and shared heterozygous INDELs in exons, introns, repetitive,
419 and non-repetitive intergenic spaces were not equal (Fig. 5b). The distribution of INDELs in
420 exons was significantly different than the distributions of INDELs in each other feature
421 considered (Kolmogorov-Smirnov Test, $p < 0.01$). Similarly, the distributions SNPs in genic
422 (exons, introns) and intergenic (repetitive, non-repetitive) regions were not equal (Fig. 5b).
423 Shared heterozygous SNPs were most common in intergenic non-repetitive regions and introns
424 and least common in exons and repetitive intergenic regions (Fig. 5b). Interestingly, unique
425 heterozygous SNPs occurred at high rates in repetitive intergenic regions (Fig. 5b).

426 That shared heterozygous sites are mostly in non-repetitive intergenic space and unique
427 heterozygous sites are mostly in repetitive space may have to do with the disposition of
428 methylated cytosines to spontaneously deaminate and the prevalence of methylated repetitive
429 sequences in those regions (Meunier *et al.*, 2005; Cantu *et al.*, 2010; Ossowski *et al.*, 2010; Shen
430 *et al.*, 2012). This is also supported by the significantly higher ratio of transitions to transversions
431 in repetitive intergenic regions than in exons, introns, and non-repetitive intergenic space (Fig.
432 5c). Furthermore, the mean percentage of CpG, CHG, and CHH sites affected by transition
433 mutations was significantly higher in repetitive intergenic space than genic and non-repetitive
434 intergenic spaces (Fig. 5d; Tukey HSD, $p < 0.01$). The mean percentage of CpG sites affected by
435 transition mutations was also significantly higher in introns than exons (Tukey HSD, $p < 0.01$).
436 Compatible with this hypothesis, INDELs, which should not increase in frequency due to
437 methylation, did not occur preferentially in repeats (Fig. 5b).
438 The impact of specific variants also varied with their prevalence among the clones (Fig.
439 5e). "High impact" mutations were predicted by SNPEff (Cingolani *et al.*, 2012). The high
440 impact mutations identified in these data included exon losses, start and stop site gains and
441 losses, frameshifts, gene fusions, splice acceptor mutations, and splice donor mutations. These
442 mutations are predicted to be deleterious because of their disruptive effects on the coded protein.
443 For these reasons, we designated such mutations as putatively deleterious in this manuscript.
444 These were counted for each Zinfandel clone relative to Zin03. Relatively low percentages of
445 heterozygous variants shared by all Zinfandel clones were putatively deleterious. In contrast,
446 larger proportions of exonic SNPs and INDELs that occurred in individual or subsets of clones
447 were putatively deleterious (Fig. 5e).
448 Together, these results show that mutations associated with clonal propagation are most
449 numerous outside of coding regions of the genome, indicating that clone genomes diversify most
450 rapidly in the intergenic space, particularly in repetitive and likely methylated regions (Fig. 5).
451 Though a minority of somatic mutations occurred in exons, we show that exonic mutations that
452 occur in few or individual clones are more often deleterious than exonic heterozygous variants
453 shared by all or most clones. In other words, clonal propagation is associated with the
454 accumulation of putatively deleterious heterozygous mutations.
455
456 *Zinfandel clones incur unique transposon insertions*

457 Transposable element insertions (TEI) contribute to somatic variation in grape
458 (Kobayashi *et al.*, 2004; Fernandez *et al.*, 2010; Carrier *et al.*, 2012; Fernandez *et al.*, 2013).
459 Relative to Zin03, 1,473 TEI were identified among the Zinfandel clones. A large fraction of TEI
460 (26.7%) occurred uniquely in individual clones (Fig. 6a) and included 325 retrotransposons,
461 mostly Copia and Gypsy LTRs, and 69 DNA-transposons (Fig. 6b). Because uniform loci are
462 excluded, in-common TEI were not captured when clones were compared to Zin03. Comparing
463 the clones relative to PN40024, however, revealed that the majority (64.8%) of TEI were shared
464 among the 15 Zinfandel clones. Five hundred thirty TEI occurred in only one, two or three
465 clones (Fig. 6a). This result supports the derivation of these selections from a common ancestral
466 plant and the accumulation of somatic variations over time.
467 In addition to being suggestive of their shared heritage, the positions of these insertions
468 and their proximity to coding genes were notable. Three-hundred forty-seven TEI occurred
469 within 314 coding genes. The remaining 938 TEIs were in intergenic regions (Fig. 6c). The
470 median upstream and downstream distance of intergenic TEs from the closest feature were
471 11,811 and 11,279 base-pairs, respectively, and 25% of TEI were less than 4,345 bases
472 downstream of the closest feature and/or less than 3,826 bases upstream of the closest feature
473 (Fig. 6c).
474
475 **Discussion**
476 Consideration of the genomic differences among Zinfandel clones revealed what is likely
477 a complex history not easily reconstructed, and that one selection was probably not a clone
478 despite being identified as such by SSR markers. Analyses of the relationships between clones
479 did not reveal groupings of clones per their recorded countries of origin. Somatic mutations may
480 help identify individual clones but could also blur the historical relationships between them. It is
481 also plausible that pairs of clones from any given region are not direct cuttings of one another but
482 of Zinfandels from another region; the clones now grown in California, for example, may have
483 been imported on numerous independent occasions from various other regions, meaning some
484 may indeed be more closely related to one of the Primitivo or Croatian clones than they are to
485 other Californian clones. It would be unwise to assume a single migratory path radiating from an
486 ancestral mother plant ought to be applicable to the clones.
487 Despite this ambiguity, the examination of SNPs, INDELs, transposable elements and
488 other structural variants all support the derivation of all but one of the clonal selections from a

489 common ancestral Zinfandel mother plant and show the accumulation of somatic mutations over
490 time (Figs. 6 and 7). The structure of the Zinfandel genome, location of mutations among clones,
491 their frequency and prevalence, and the relationship between these factors provides some insight
492 into the nature of mutations in clonally propagated plants. Mutations among clones were
493 predominantly heterozygous (Fig. 4) and uncommon heterozygous mutations shared by a subset
494 of or individual clones were increasingly deleterious when they occurred in exons (Fig. 5e).
495 There are costs and benefits associated with clonal propagation (McKey *et al.*, 2010).
496 Among the benefits are that the plants need not breed true-to-type; clonal propagation generally
497 fixes heterozygous loci and valuable phenotypes. However, the increase in the proportion of
498 deleterious alleles supports Muller's ratchet, which posits that sex is advantageous and that
499 clonal propagation increases mutational load (Muller, 1932). Though these and previous data do
500 not tell which mutations are actually recessive or dominant, they could remain hidden if they are
501 recessive or do not manifest their deleterious effects (Zhou *et al.*, 2017; Gaut *et al.*, 2018).
502 However, even after taking into consideration the total length of exons, introns, and intergenic
503 space (repetitive and non-repetitive), heterozygous mutations occurred at varying frequency in
504 these regions and were least abundant in coding regions. The rarity of mutations in exons and
505 commonality of mutations in repetitive intergenic space may have at least two components.
506 Mutations are likely more frequent in repetitive intergenic space as a result of the
507 regulation of transposition by DNA methylation. Repetitive intergenic space had the highest rate
508 of relatively unique SNPs and the ratio of transitions to transversions was significantly higher
509 there than in other regions. DNA methylation is an important epigenetic control and is one
510 mechanism that maintains genome stability and impairs the transposition of mobile elements
511 (Chen *et al.*, 1998; Hirochika *et al.*, 2000; Shen *et al.*, 2012). Methylated cytosines, however,
512 spontaneously deaminate faster than unmethylated cytosines (Selker, 1990; Cantu *et al.*, 2010).
513 Together, the expectations that intergenic regions are rich in transposable elements, that these
514 regions are typically highly methylated and as a result will experience greater transition rates
515 account for the high rates of SNPs in repetitive intergenic spaces among Zinfandel clones. Also
516 notable, these data show that some transposable elements are not entirely silenced, with a
517 substantial number inserting in genes or in close proximity to genes (Fig. 6b). These insertions
518 could be effectively inconsequential or not; transposable element insertions can result in novel
519 transcripts and affect gene expression regulation (Fernandez *et al.*, 2010; Hirsch & Springer,
520 2017). Gene body methylation is appreciated as a mutagenic "double-edged sword" (Zemach *et*

521 *al.*, 2010), with benefits coming at the price. Recent work observed region-specific methylation
522 in vegetatively propagated Sardinian white poplar that may serve an advantageous function
523 (Guarino *et al.*, 2015) and others have suggested that the epigenome contributes to the success of
524 vegetatively propagated plants (Douhovnikoff & Dodd, 2014). Future work might also consider
525 the long-term price associated with intergenic mutagenesis and the potential loss of methylation
526 in vegetatively propagated plants.
527 The rarity of exonic mutations was surprising. After accounting for the length of these
528 spaces in the genome and their repetitiveness, we expected uniform rates of mutation in exons,
529 introns, and intergenic space. Instead, we observed that although rare somatic mutations in exons
530 were increasingly deleterious, they were relatively scarce. Some degree of negative selection
531 against deleterious variants in coding regions could explain why mutations were less frequent in
532 coding than noncoding regions of the genome. The possibility of diplontic, clonal selection or
533 competition between cell lineages that could purge otherwise consequential deleterious
534 mutations has been modeled, but evidence of its occurrence is sparse (Pineda-Krch &
535 Fagerström, 1999; Klekowski, 2003; McKey *et al.*, 2010). The structures of apical meristems
536 (Klekowski & Kazarinova-Fukshansky, 1984; Klekowski *et al.*, 1985) and the tendency of
537 somatic mutations to be heterozygous and recessive (Zhou *et al.*, 2017) place constraints on the
538 likelihood that deleterious mutations would be subjected to negative selection. Periclinal
539 divisions across cell layers could enhance diplontic selection (Klekowski, 2003) against
540 dominant and/or hemizygous recessive alleles. Four and one half of Zinfandel's genome is
541 hemizygous; structural variations identified within the Zinfandel genome and the rampant
542 hemizygosity reported in Chardonnay (Zhou *et al.*, 2018) could also expose otherwise hidden
543 somatic variations to selective pressure hostile to the accumulation of deleterious mutations.
544 Additional work should explore to what degree each of these factors, or others not considered
545 here, explain why somatic mutations in exons were relatively infrequent and characterize the
546 realized long-term consequences of mutation accumulation versus selection for grapevine and
547 other clonally propagated plants.
548
549 **Acknowledgements**
550 We are grateful for the vision of the late James A. Wolpert, who established the original
551 Zinfandel clone trials with the support of the Zinfandel Advocates and Producers (ZAP).
552

553 **Authors' contributions**

554 AMV, MAW, BG, and DC designed the experiments. BBU, YZ, MMA, and RFB
555 collected the biological material and generated the data. MAP carried out the chemical analysis
556 of the clones. AMV, AM, YZ, DS, DL, and LKE analyzed the data. AMV and DC prepared the
557 figures and wrote the manuscript. All authors contributed to the final version of the manuscript.
558 All authors read and approved the final manuscript.

559

560 **References**

561

562 **Boss PK, Thomas MR**. **2002**. Association of dwarfism and floral induction with a grape 'green
563 revolution' mutation. *Nature* **416**: 847–850.

564 **Boursiquot J-M, Lacombe T, Laucou V, Julliard S, Perrin FX, Lanier N, Legrand D,**
565 **Meredith C, This P**. **2009**. Parentage of Merlot and related winegrape cultivars of southwestern
566 France: discovery of the missing link. *Australian Journal of Grape and Wine Research* **15**: 144–
567 155.

568 **Bowers J, Boursiquot J-M, This P, Chu K, Johansson K, Meredith C**. **1999**. Historical
569 genetics: The parentage of chardonnay, gamay, and other wine grapes of northeastern France.
570 *Science* **285**: 1562–1565.

571 **Bowers JE, Bandman EB, Meredith CP**. **1993**. DNA Fingerprint Characterization of Some
572 Wine Grape Cultivars. *American Journal of Enology and Viticulture* **44**: 266–274.

573 **Cantu D, Vanzetti LS, Sumner A, Dubcovsky M, Matvienko M, Distelfeld A, Michelmore**
574 **RW, Dubcovsky J**. **2010**. Small RNAs, DNA methylation and transposable elements in wheat.
575 *BMC genomics* doi: 10.1186/1471-2164-11-408

576 **Carbonell-Bejerano P, Royo C, Torres-Pérez R, Grimplet J, Fernandez L, Franco-Zorrilla**
577 **JM, Lijavetzky D, Baroja E, Martínez J, García-Escudero E, *et al.* 2017**. Catastrophic
578 unbalanced genome rearrangements cause somatic loss of berry color in grapevine. *Plant*
579 *Physiology* **175**: 786–801.

580 **Carrier G, Le Cunff L, Dereeper A, Legrand D, Sabot F, Bouchez O, Audeguin L,**
581 **Boursiquot J-M, This P**. **2012**. Transposable Elements Are a Major Cause of Somatic
582 Polymorphism in *Vitis vinifera* L. *PLOS ONE* doi: 10.1371/journal.pone.0032973

583 **CDFA**. **2016a**. *Grape Crush Report, Final 2016 Crop*.

584 **CDFA**. **2016b**. *California Grape Crush Report Preliminary 2015*.

585 **Chan SW-L, Henderson IR, Jacobsen SE**. **2005**. Gardening the genome: DNA methylation in
586 Arabidopsis thaliana. *Nature Reviews Genetics* **6**: 351–360.

587 **Chen RZ, Pettersson U, Beard C, Jackson-Grusby L, Jaenisch R**. **1998**. DNA
588 hypomethylation leads to elevated mutation rates. *Nature* **395**: 89–93.

589 **Chin C-S, Peluso P, Sedlazeck FJ, Nattestad M, Concepcion GT, Clum A, Dunn C,**
590 **O'Malley R, Figueroa-Balderas R, Morales-Cruz A, *et al.* 2016**. Phased diploid genome
591 assembly with single-molecule real-time sequencing. *Nature Methods* **13**: 1050–1054.

592 **Cingolani P, Platts A, Le Lily Wang, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden**
593 **DM**. **2012**. A program for annotating and predicting the effects of single nucleotide
594 polymorphisms, SnpEff. *Fly* **6**: 80–92.

595 **Douhovnikoff V, Dodd RS**. **2014**. Epigenetics: a potential mechanism for clonal plant success.
596 *Plant Ecology* **216**: 227–233.

597 **Fanizza G, Lamaj F, Ricciardi L, Resta P, Savino V**. **2005**. Grapevine cvs Primitivo,
598 Zinfandel and Crljenak kastelanski: Molecular analysis by AFLP. *Vitis* **44**: 147–148.

599 **Farcuh M, Li B, Rivero RM, Shlizerman L, Sadka A, Blumwald E**. **2017**. Sugar metabolism
600 reprogramming in a non-climacteric bud mutant of a climacteric plum fruit during development
601 on the tree. *Journal of Experimental Botany* **68**: 5813–5828.

602 **Fernandez L, Chaïb J, Zapater JMM, Thomas MR, Torregrosa L**. **2013**. Mis-expression of a
603 PISTILLATA-like MADS box gene prevents fruit development in grapevine. *The Plant Journal*
604 **73**: 918–928.

605 **Fernandez L, Torregrosa L, Segura V, Bouquet A, Martínez-Zapater JM**. **2010**.
606 Transposon-induced gene activation as a mechanism generating cluster shape somatic variation
607 in grapevine. *The Plant Journal* **61**: 545–557.

608 **Fidelibus MW, Christensen LP, Katayama DG, Verdenal P-T**. **2005**. Performance of
609 Zinfandel and Primitivo grapevine selections in the central San Joaquin Valley, California.
610 *American Journal of Enology and Viticulture* **56**: 284–286.

611 **Franks T, Botta R, Thomas MR**. **2002**. Chimerism in grapevines: implications for cultivar
612 identity, ancestry and genetic improvement. *Theoretical and Applied Genetics* **104**: 192–199.

613 **Gambino G, Molin AD, Boccacci P, Minio A, Chitarra W, Avanzato CG, Tononi P,**
614 **Perrone I, Raimondi S, Schneider A, *et al.* 2017**. Whole-genome sequencing and SNV
615 genotyping of 'Nebbiolo' (*Vitis vinifera* L.) clones. *Scientific Reports* **7**: 1–15.

616 **Gaut BS, Seymour DK, Liu Q, Zhou Y**. **2018**. Demography and its effects on genomic
617 variation in crop domestication. *Nature Plants* doi: 0.1038/s41477-018-0210-1

618 **Guarino F, Cicatelli A, Brundu G, Heinze B, Castiglione S**. **2015**. Epigenetic Diversity of
619 Clonal White Poplar (Populus alba L.) Populations: Could Methylation Support the Success of
620 Vegetative Reproduction Strategy? (M Labra, Ed.). *PLOS ONE* **10**: e0131480–20.

621 **Haas BJ, Delcher AL, Mount SM, Wortman JR, Smith RK Jr, Hannick LI, Maiti R,**
622 **Ronning CM, Rusch DB, Town CD, *et al.* 2003**. Improving the Arabidopsis genome annotation
623 using maximal transcript alignment assemblies. *Nucleic Acids Research* **31**: 5654–5666.

624 **Haas BJ, Salzberg SL, Zhu W, Pertea M, Allen JE, Orvis J, White O, Buell CR, Wortman**
625 **JR**. **2008**. Automated eukaryotic gene structure annotation using EVidenceModeler and the
626 Program to Assemble Spliced Alignments. *Genome biology* **9**.

627 **Hershberg R, Petrov DA**. **2010**. Evidence That Mutation Is Universally Biased towards AT in
628 Bacteria (MW Nachman, Ed.). *PLOS Genetics* doi: 10.1371/journal.pgen.1001115

629 **Hirochika H, Okamoto H, Kakutani T**. **2000**. Silencing of Retrotransposons in Arabidopsis
630 and Reactivation by the ddm1 Mutation. *The Plant Cell* **12**: 357–368.

631 **Hirsch CD, Springer NM**. **2017**. Transposable element influences on gene expression in plants.
632 *Biochimica et Biophysica Acta* **1860**: 157–165.

633 **Hocquigny S, Pelsy F, Dumas V, Kindt S, Heloir M-C, Merdinoglu D**. **2004**. Diversification
634 within grapevine cultivars goes through chimeric states. *Genome* **47**: 579–589.

635 **Imazio S, Labra M, Grassi F, Winfield M, Bardini M, Scienza A**. **2002**. Molecular tools for
636 clone identification: the case of the grapevine cultivar 'Traminer'. *Plant Breeding* **121**: 531–535.

637 **Jaillon O, Aury J-M, Noel B, Policriti A, Clepet C, Casagrande A, Choisne N, Aubourg S,**
638 **Vitulo N, Jubin C, *et al.* 2007**. The grapevine genome sequence suggests ancestral
639 hexaploidization in major angiosperm phyla. *Nature* **449**: 463–467.

640 **Jones L, Riaz S, Morales-Cruz A, Amrine KCH, McGuire B, Gubler WD, Walker MA,**
641 **Cantu D**. **2014**. Adaptive genomic structural variation in the grape powdery mildew pathogen,
642 Erysiphe necator. *BMC genomics* doi: 10.1186/1471-2164-15-1081

643 **Klekowski EJ**. **1998**. Mutation rates in mangroves and other plants. *Genetica* **102/103**: 325-331

644 **Klekowski EJ**. **2003**. Plant clonality, mutation, diplontic selection and mutational meltdown.
645 *Biological Journal of the Linnean Society* **79**: 61–67.

646 **Klekowski EJ Jr., Kazarinova-Fukshansky N**. **1984**. Shoot Apical Meristems and Mutation:
647 Selective Loss of Disadvantageous Cell Genotypes. *American Journal of Botany* **71**: 28–34.

648 **Klekowski EJ, Kazarinova-Fukshansky N, Mohr H**. **1985**. Shoot Apical Meristems and
649 Mutation - Stratified Meristems and Angiosperm Evolution. *American Journal of Botany* **72**:
650 1788–1800.

651 **Kobayashi S, Goto-Yamamoto N, Hirochika H**. **2004**. Retrotransposon-induced mutations in
652 grape skin color. *Science* **304**: 982.

653 **Kofler R, Gómez-Sánchez D, Schlötterer C**. **2016**. PoPoolationTE2: Comparative Population
654 Genomics of Transposable Elements Using Pool-Seq. *Molecular Biology and Evolution* **33**:
655 2759–2764.

656 **Korf I**. **2004**. Gene finding in novel genomes. *BMC Bioinformatics* doi: 10.1186/1471-2105-5-
657 59

Pineda-Krch M, Fagerström T. **1999**. On the potential for evolutionary change in meristematic cell lineages through intraorganismal selection. *Journal of Evolutionary Biology* **12**: 681–688.

Li H, Durbin R. **2009**. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**: 1754–1760.

Lomsadze A, Ter-Hovhannisyan V, Chernoff YO, Borodovsky M. **2005**. Gene identification in novel eukaryotic genomes by self-training algorithm. *Biological Journal of the Linnean Society* **33**: 6494–6506.

Maletic E, Pejic I, Karoglan Kontic J, Piljac J, Dangl G, Vokurka A, Lacombe T, MiroÅ evic N, Meredith C. **2003**. The Identification of Zinfandel on the Dalmatian Cost of Croatia. *Acta Hort* **603:** 251–254.

Maletic E, Pejic I, Kontic JK, Piljac J, Dangl GS, Vokurka A, Lacombe T, Mirošević N, Meredith CP. **2004**. Zinfandel, Dobricic, and Plavac mali: The genetic relationship among three cultivars of the Dalmatian Coast of Croatia. *American Journal of Enology and Viticulture* **55**: 174–180.

Manichaikul A, Mychaleckyj JC, Rich SS, Daly K, Sale M, Chen W-M. **2010**. Robust relationship inference in genome-wide association studies. *Bioinformatics* **26**: 2867–2873.

Mautino MR, Rosa AL. **1998**. Analysis of Models Involving Enzymatic Activities for the Occurrence of C-T Transition Mutations During Repeat-Induced Point Mutation (RIP) in *Neurospora crassa*. *Journal of Theoretical Biology* **192**: 61–71.

McKey D, Elias M, Pujol B, Duputié A. **2010**. The evolutionary ecology of clonally propagated domesticated plants. *New Phytologist* **186**: 318–332.

Meunier J, Khelifi A, Navratil V, Duret L. **2005**. Homology-Dependent Methylation in Primate Repetitive DNA. *Proceedings of the National Academy of Sciences* **102**: 5471–5476.

Minio A, Lin J, Gaut BS, Cantu D. **2017**. How Single Molecule Real-Time Sequencing and Haplotype Phasing Have Enabled Reference-Grade Diploid Genome Assembly of Wine Grapes. *Frontiers in plant science* **8**: 481–486.

Minio A, Massonnet M, Figueroa-Balderas R, Vondras AM, Blanco-Ulate B, Cantu D. **2019**. Iso-Seq Allows Genome-Independent Transcriptome Profiling of Grape Berry Development. *G3* **9**: 755–767.

Mirošević N, Meredith CP. **2000**. A review of research and literature related to the origin and identity of the cultivars Plavac mali, Zinfandel and Primitivo (*Vitis vinifera* L.). *Acta Hort* **65**: 45–49.

Muller HJ. **1932**. Some genetic aspects of sex. *The American Naturalist* **66**: 118–138.

Myers G. **2014**. Efficient Local Alignment Discovery amongst Noisy Long Reads. Wroclaw, Poland: Springer, Berlin, Heidelberg, 52–67.

Orive ME. **2001**. Somatic Mutations in Organisms with Complex Life Histories. *Theoretical Population Biology* **59:** 235-249.

Ossowski S, Schneeberger K, Lucas-Lledó JI, Warthmann N, Clark RM, Shaw RG, Weigel D, Lynch M. **2010**. The Rate and Molecular Spectrum of Spontaneous Mutations in *Arabidopsis thaliana*. *Science* **327**: 92–94.

Pelsy F, Dumas V, Bévilacqua L, Hocquigny S, Merdinoglu D. **2015**. Chromosome Replacement and Deletion Lead to Clonal Polymorphism of Berry Color in Grapevine. *PLOS Genetics* **11**.

Pertea M, Kim D, Pertea GM, Leek JT, Salzberg SL. **2016**. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nature Protocols* **11**: 1650–1667.

Plomion C, Aury J-M, Amselem J, Leroy T, Murat F, Duplessis S, Faye S, Francillonne N, Labadie K, Provost G, *et al.* 2018. Oak genome reveals facets of long lifespan. *Nature Plants* **4**: 440–452.

Qi Y, He X, Wang X-J, Kohany O, Jurka J, Hannon GJ. **2006**. Distinct catalytic and non-catalytic roles of ARGONAUTE4 in RNA-directed DNA methylation. *Nature* **443**: 1008–1012.

Ramu P, Esuma W, Kawuki R, Rabbi IY, Egesi C, Bredeson JV, Bart RS, Verma J, Buckler ES, Lu F. **2017**. Cassava haplotype map highlights fixation of deleterious mutations during clonal propagation. *Nature Genetics* **49**: 959–963.

Rausch T, Zichner T, Schlattl A, Stutz AM, Benes V, Korbel JO. **2012**. DELLY: structural variant discovery by integrated paired-end and split-read analysis. *Bioinformatics* **28**: 333–339.

Regner F, Stadlbauer A, Eisenheld C, Kaserer H. **2000**. Genetic Relationships Among Pinots and Related Cultivars. *American Journal of Enology and Viticulture* **51**: 7–14.

Riaz S, Garrison KE, Dangl GS, Boursiquot J-M, Meredith CP. **2002**. Genetic divergence and chimerism within ancient asexually propagated winegrape cultivars. *J Amer Soc Hort Sci* **127**: 508–514.

Roach MJ, Johnson DL, Bohlmann J, van Vuuren HJJ, Jones SJM, Pretorius IS, Schmidt SA, Borneman AR. **2018**. Population sequencing reveals clonal diversity and ancestral inbreeding in the grapevine cultivar Chardonnay. *PLOS Genetics* **14**. doi: 10.1371/journal.pgen.1007807

Russo G, Liuzzi V, D'Andrea L, Alviti G. **2003**. Comparison among Five Clones of 'Primitivo' Vine in Southern Italy. *Acta Hort* **603:** 779–786.

Schlötterer C, Tautz D. **1992**. Slippage synthesis of simple sequence DNA. *Nucleic Acids Research* **20**: 211–215.

Sedlazeck FJ, Rescheneder P, Smolka M, Fang H, Nattestad M, Haeseler A, Schatz MC. **2018**. Accurate detection of complex structural variations using single-molecule sequencing. *Nature Methods* **15**: 461–468.

730 **Sefc KM, Regner F, Turetschek E, Glössl J, Steinkellner H**. **1999**. Identification of
731 microsatellite sequences in Vitis riparia and their applicability for genotyping of different Vitis
732 species. *Genome* **42**: 367–373.

733 **Selker EU**. **1990**. Premeiotic instability of repeated sequences in Neurospora crassa. *Annual*
734 *Review of Genetics* **24**: 579–613.

735 **Shen H, He H, Li J, Chen W, Wang X, Guo L, Peng Z, He G, Zhong S, Qi Y, *et al.* 2012**.
736 Genome-Wide Analysis of DNA Methylation and Gene Expression Changes in Two Arabidopsis
737 Ecotypes and Their Reciprocal Hybrids. *The Plant Cell* **24**: 875–892.

738 **Slater GSC, Birney E**. **2005**. Automated generation of heuristics for biological sequence
739 comparison. *BMC Bioinformatics.* doi:10.1186/1471-2105-6-31

740 **Smit A, Hubley R, Green P**. **2013-2015** *RepeatMasker Open-4.0*. http://www.repeatmasker.org

741 **Soost RK, Cameron JW, Bitters WP, Platt RG**. **1961**. Citrus bud variation, old and new. *Calif*
742 *Citrograph* **46**: 188–193.

743 **Stanke M, Keller O, Gunduz I, Hayes A, Waack S, Morgenstern B**. **2006**. AUGUSTUS: ab
744 initio prediction of alternative transcripts. *Nucleic Acids Research* **34**: W435–W439.

745 **Sweet NL, Wolpert JA**. **2007**. The Zinfandels of FPS. *FPS Grape Program Newsletter*: 10–19.

746 **Thomas MR, Cain P, Scott NS**. **1994**. DNA typing of grapevines: a universal methodology and
747 database for describing cultivars and evaluating genetic relatedness. *Plant Molecular Biology* **25**:
748 939–949.

749 **Thompson MM, Olmo HP**. **1963**. Cytohistological Studies of Cytochimeric and Tetraploid
750 Grapes. *American Journal of Botany* **50**: 901–906.

751 **Tilney-Bassett RAE**. **1986**. *Plant chimeras.* Edward Arnold (Publishers) Ltd.

752 **Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, del Angel G, Levy-Moonshine A,**
753 **Jordan T, Shakir K, Roazen D, Thibault J, *et al.* 2013**. From FastQ data to high confidence
754 variant calls: the Genome Analysis Toolkit best practices pipeline. *Current Protocols in*
755 *Bioinformatics*.

756 **Walker AR, Lee E, Robinson SP**. **2006**. Two new grape cultivars, bud sports of Cabernet
757 Sauvignon bearing pale-coloured berries, are the result of deletion of two regulatory genes of the
758 berry colour locus. *Plant Molecular Biology* **62**: 623–635.

759 **Whitham TG, Slobodchikoff CN**. **1981**. Evolution by Individuals, Plant-Herbivore Interactions,
760 and Mosaics of Genetic Variability: The Adaptive Significance of Somatic Mutations in Plants.
761 *Oecologia* **49**: 287–292.

762 **Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, Chalhoub B, Flavell A, Leroy P,**
763 **Morgante M, Panaud O, *et al.* 2007**. A unified classification system for eukaryotic transposable
764 elements. *Nature Reviews Genetics* **8**: 973–982.

765 **Wolpert JA**. **1996**. Performance of Zinfandel and Primitivo Clones in a Warm Climate.
766 *American Journal of Enology and Viticulture* **47**: 124–126.

767 **Yakushiji H, Kobayashi S, Goto-Yamamoto N, Tae Jeong S, Sueta T, Mitani N, Azuma A**.
768 **2006**. A skin color mutation of grapevine, from black-skinned Pinot Noir to white-skinned Pinot
769 Blanc, is caused by deletion of the functional VvmybA1 allele. *Bioscience, biotechnology, and*
770 *biochemistry* **70**: 1506–1508.

771 **Zdunić G, Simon S, Malenica N, Budić-Leto I, Maletic E, Karoglan Kontić J, Pejic I**. **2014**.
772 Intravarietal variability of Crljenak Kastelanski' and Its Relationship with 'Zinfandel' and
773 "Primitivo" Selections. *Acta Hort* **1046:** 573–580.

774 **Zemach A, McDaniel IE, Silva P, Zilberman D**. **2010**. Genome-Wide Evolutionary Analysis
775 of Eukaryotic DNA Methylation. *Science* **328**: 916–919.

776 **Zhou Y, Massonnet M, Sanjak JS, Cantu D, Gaut BS**. **2017**. Evolutionary genomics of grape
777 (*Vitis vinifera* ssp. *vinifera*) domestication. *PNAS* **114**:11715-11720.

778 **Zhou Y, Minio A, Massonnet M, Solares E, Lyu Y, Beridze T, Cantu D, Gaut BS**. **2018**.
779 Structural variants, clonal propagation, and genome evolution in grapevine (*Vitis vinifera*).
780 *bioRxiv*: 1–48.

781

782 **Figure legends**

783

784 **Figure 1.** Structural variation between Zin03 haplotypes. **a**. Distribution of structural variation
785 sizes. Boxplots show the 25th quartile, median, and 75th quartile for each type of SV. Whiskers
786 are 1.5$^{\text{Inter-Quartile Range}}$. Diamonds indicate the mean $\log_{10}$(length) of each type of SV; **b,c,d.**
787 Examples of heterozygous structural variants between haplotypes that intersect genes. For each
788 reported deletion, (from top to bottom) the coverage, haplotype-resolved alignment of reads, and
789 the genes annotated in the region are shown; **b.** 4 kbp heterozygous deletion of two genes; **c.** 11
790 kbp heterozygous deletion of two genes; **d.** 22 kbp inversion that intersects a single gene.
791 Triangles indicate boundaries of the inversion. A gap is shown rather than the center of the
792 inverted region.

793

794 **Figure 2.** Gene content and structural variability between Zin03 and other *V. vinifera* genomes.
795 **(a)** Uniquely occurring Zinfandel genes and the number of Zinfandel genes that align well to
796 other cultivars with >=80% identity and reciprocal coverage. The total number of hits (or total
797 gene content for Zin03) is indicated by the "Set Size" and the exclusive hits for each intersection

is indicated as the "Intersection Size"; (**b,c)** Selected deletions in Cabernet sauvignon relative to Zin03 that intersect genes. For each reported deletion, (from top to bottom) the coverage of reads over the region by long Zinfandel and Cabernet Sauvignon reads, haplotype-resolved alignment of the reads, and the genes annotated in the region are shown; **(b)** Two genes are completely deleted in Cabernet Sauvignon relative to Zinfandel and are deleted in one Zinfandel haplotype; **(c)** One gene contains a homozygous partial deletion in Cabernet Sauvignon.

Figure 3. The relationships between Zinfandel selections. (**a)** Principal component analysis of Zinfandel selections based on SNP data. Zin03 was not included in the analysis; **(b)** Zoomed-in view of **(a)**, excluding Crljenak kaštelanski 3 and Pribidrag 15; **(c)** Kinship analysis of Zinfandel selections and other cultivars with known relationships based on SNP data and outside of annotated repeats. The Kinship coefficient, PHI, is shown, as well as a dendrogram constructed by hierarchically clustering genotypes using their kinship coefficients.

Figure 4. Characterization of variants and their frequency among Zinfandel selections and other *vinifera* cultivars (Pinot Noir, Chardonnay, Merlot, Cabernet Franc, and Sauvignon Blanc). The normalized rate of variants (number of variants divided by the total feature length in the genome * 1k) by type (SNP, INDEL), feature (Intergenic, Intron, Exon), and genotype (Non-Zinfandel Cultivars, Zinfandel selections). Boxplots show the 25th quartile, median, and 75th quartile.

Figure 5. The abundance and impact of shared and unique heterozygous mutations among Zinfandel clones. **(a)** The number of heterozygous SNPs, INDELs, and SVs are shared by N Zinfandel clones; **(b)** The number of SNPs and INDELs shared by N clones in exons, introns, intergenic repeats ("Repeats"), and non-repetitive intergenic space; **(c)** The ratio of transitions (Tr) to transversions (Tv) for heterozygous SNPs that uniquely occur in single Zinfandel clones and in different genome features. Different letters correspond to significant differences in Tr/Tv rates between features (ANOVA, Tukey HSD, $p < 0.01$); **(d)** The percentage of CpG, CHG, and CHH in exons, introns, intergenic repeats ("Repeats"), and non-repetitive intergenic space that experiences transition mutations. Comparisons were made between features for each type of C-repeat separately. Different letters correspond to significant differences (Tukey HSD, $p < 0.01$); **(e)** Proportion of exonic SNPs and INDELs that are deleterious and shared by N Zinfandel clones

830

831 **Figure 6**. Transposable element insertions among Zinfandel selections. **(a)** Transposable element
832 insertions shared among N Zinfandel selections relative to Zin03 and PN40024; **(b)** The
833 proximity of intergenic transposable element insertions to genes; **(c)** Types of transposable
834 element insertions shared by N Zinfandel selections.

835

836 **Supporting information**

837

838 **Methods S1.** Extraction of phenolic metabolites from Heritage Vineyard Zinfandel clones.

839

840 **Figure S1.** Discriminant analysis of Zinfandel clones based on their phenolic profiles.

841

842 **Table S1.** Unique genes identified in Zinfandel, not identified in Pinot Noir and Cabernet
843 Sauvignon (309), with associated Gene Ontology categories.

844

845 **Table S2.** The first tab of this excel file is a summary of variants relative to the Zinfandel
846 reference genome and the second is a summary of the SnpEff analysis of variants, with mean
847 values ± SEM shown, and excluding sites where samples and Zin03 have identical heterozygous
848 genotypes at the locus.

849

850 **Methods S2.** Settings and data used for Zin03 genome assembly, annotation, and variant calling.

851

852 **Methods S3.** Bioinformatic pipeline for SNP, INDEL, and SV calling.

853

854 **Tables**

855

Table 1. Clone identifying information

Clone #	Common name	Origin	Foundation Plant Services
1	Primitivo	Bari, Italy	Primitivo FPS 03
2	Primitivo	Conegliano, Italy	Primitivo FPS 06
3	Crljenak kaštelanski	Kaštel Novi, Croatia	Zinfandel FPS 42.1
4	Pribidrag	Svinšće, Croatia	Zinfandel FPS 43.1
5	Pribidrag	Svinšće, Croatia	Zinfandel FPS 44.1
6	Zinfandel	California, USA	Zinfandel FPS 10
7	Zinfandel	California, USA	Zinfandel FPS 24
8	Zinfandel	California, USA	Zinfandel FPS 37
9	Zinfandel	California, USA	Zinfandel FPS 39
10	Zinfandel	California, USA	Zinfandel FPS 56.1
11	Zinfandel	California, USA	Zinfandel FPS 40
12	Pribidrag	Marušići, Croatia	In testing at FPS
13	Pribidrag	Svinšće, Croatia	Zinfandel FPS 43.1
14	Crljenak kaštelanski	University of Zagreb, Croatia	-
15	Pribidrag	Svinšće, Croatia	Zinfandel FPS 44.1
Zin03	Zinfandel	California, USA	Zinfandel FPS 03

856

857

858

Table 2. Summary statistics of the Zinfandel genome assembly and annotation.

	Primary	Haplotig
Total length	591,171,721	306,029,957
Number of contigs	1,509	2,246
N50	1,062,797	442,393
N75	366,308	185,785
L50	154	200
L75	395	463
Median contig length (bp)	161,249	37,307
Longest contig (bp)	7,901,503	2,609,171
Shortest contig (bp)	17,787	1,970
Average GC content (%)	34.45%	34.37%
Number of genes	33,523	20,037
	Total	*Average per gene*
Number of exons	244,880	4.57
Number of introns	191,320	3.57
	Average (bp)	*Maximum (bp)*
mRNA lengths	4,166	94,143
Exon lengths	245.79	7,992
Intron lengths	191,320	41,647
Intergenic distances	10,309	302,473

859

860

861

862

863

864

865

866

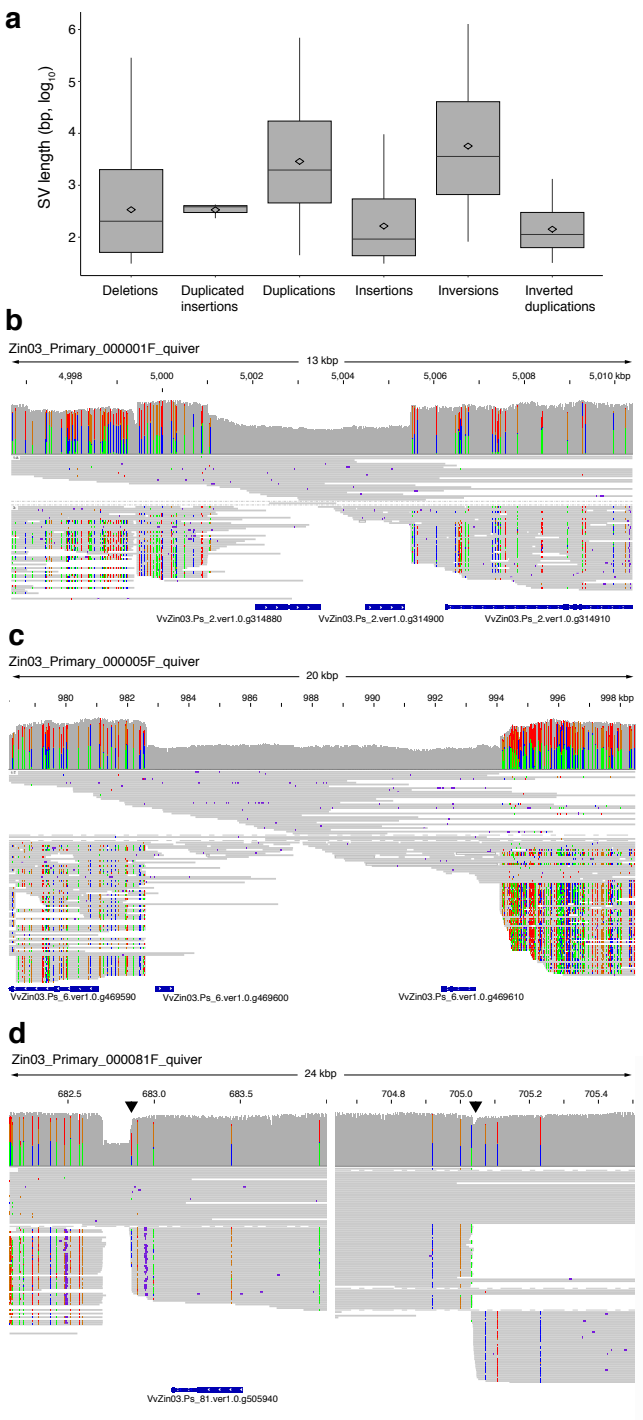
867

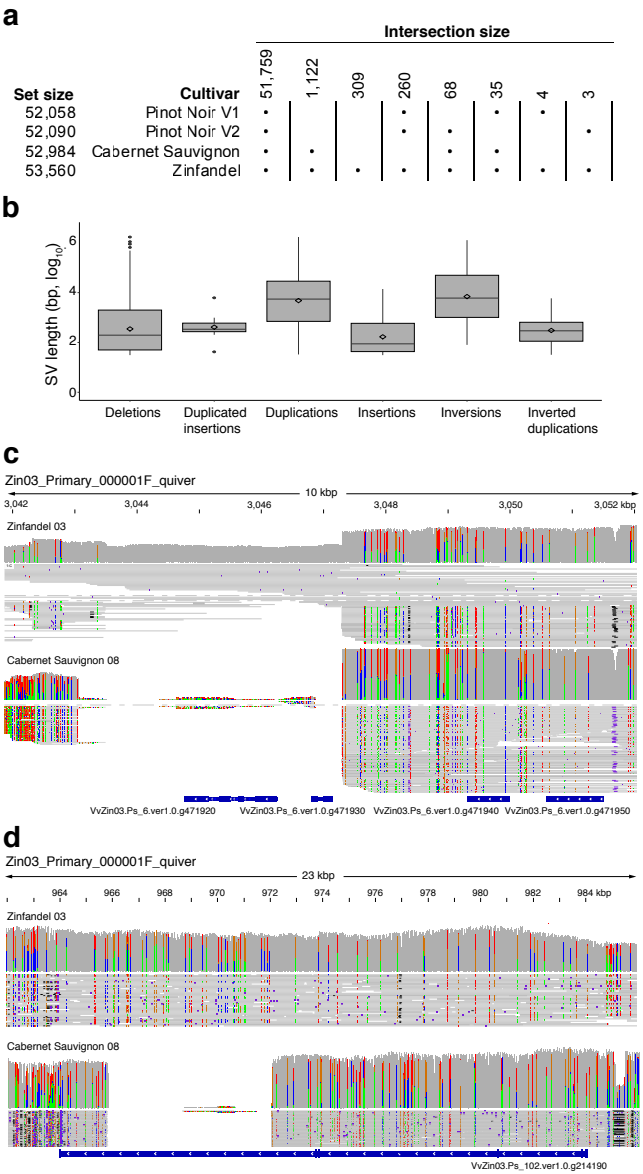
Table 3. Sniffles analysis of structural variation between cultivars and between Zinfandel parental haplotypes

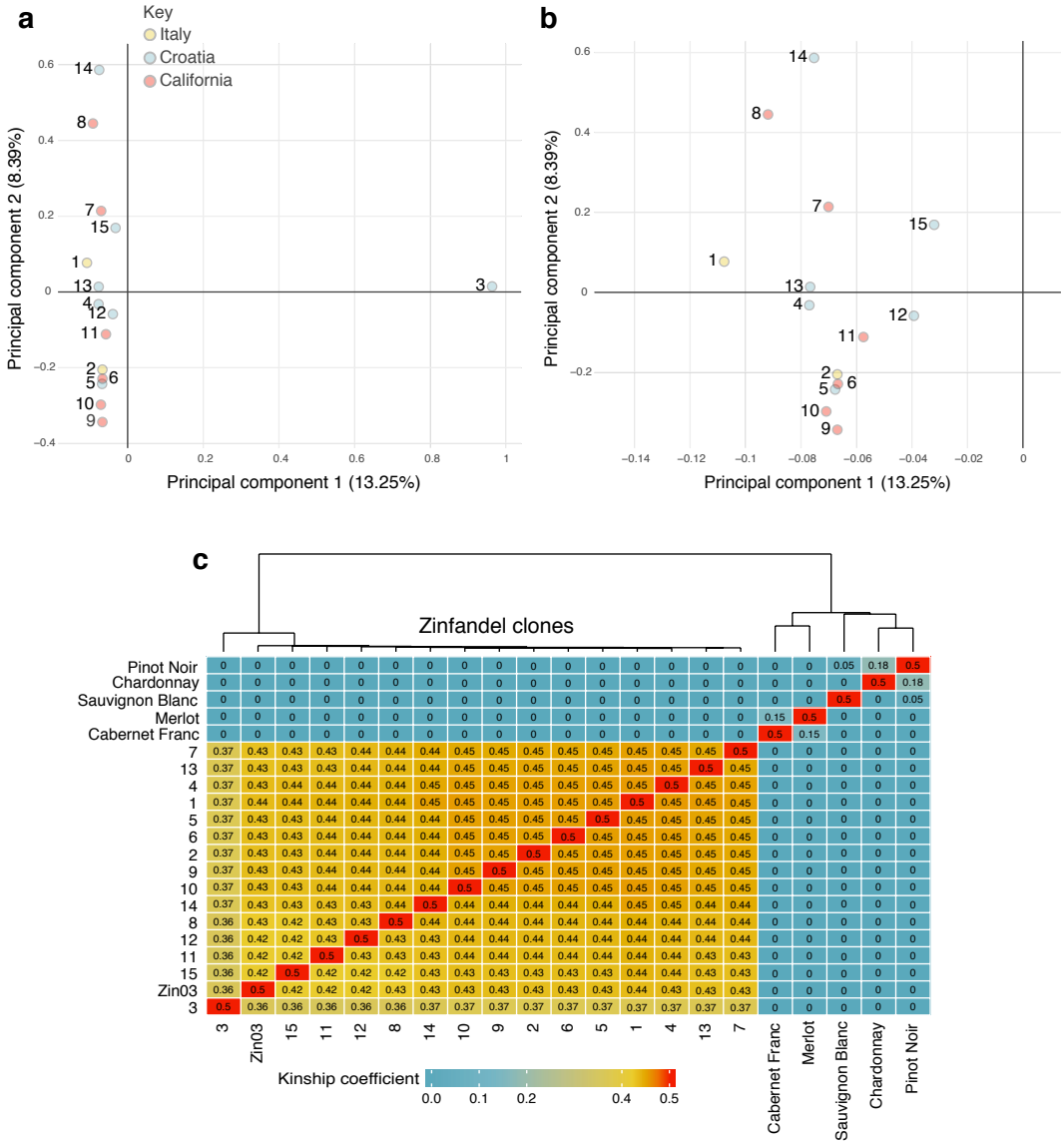
	Cabernet Sauvignon vs. Zinfandel					Zinfandel haplotig vs. Zinfandel primary				
	Median Size (bp)	*Count*	*Genes*	*Total SV size (Mb)*	*% genome*	*Median Size (bp)*	*Count*	*Genes*	*Total SV size (Mb)*	*% genome*
Deletions	196	46,363	9,219	115.0	12.82	203	12,031	2,521	26,953,558	4.56
Duplications	5,518	2,884	3,286	48.7	5.43	1,966	553	535	7,604,041	1.29
Insertions	88	37,407	5,225	23.9	2.66	92	9,647	2,081	5,594,259	0.95
Inversions	6,037	607	1,440	20.6	2.30	3,592	111	391	5,521,214	0.93
Duplicated Insertions	339	9	2	0.0439	0.0049	385	3	2	6,861	0.0012
Inverted Duplications	293	65	12	0.0418	0.0047	113	54	11	12,930	0.0022

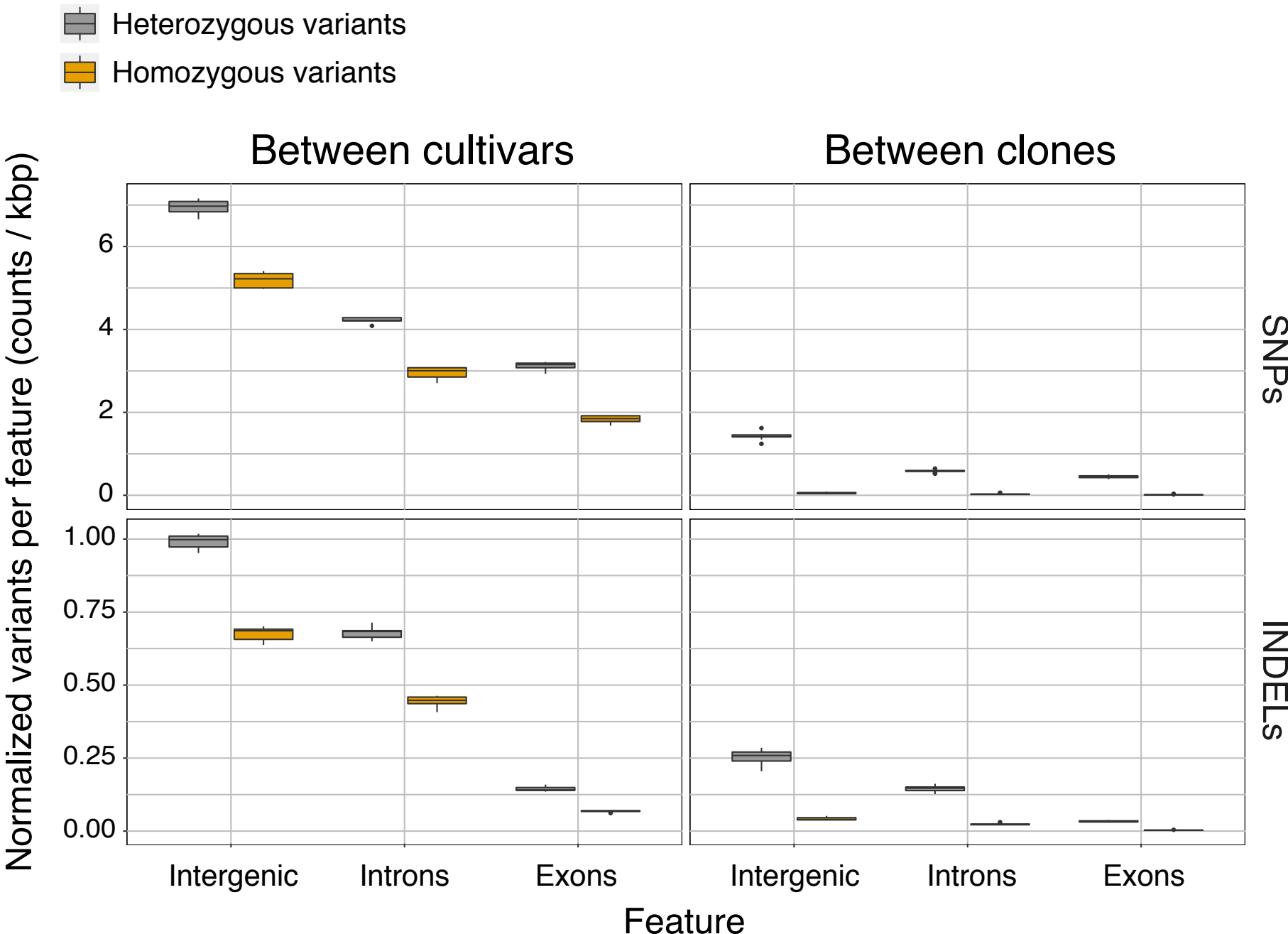
868

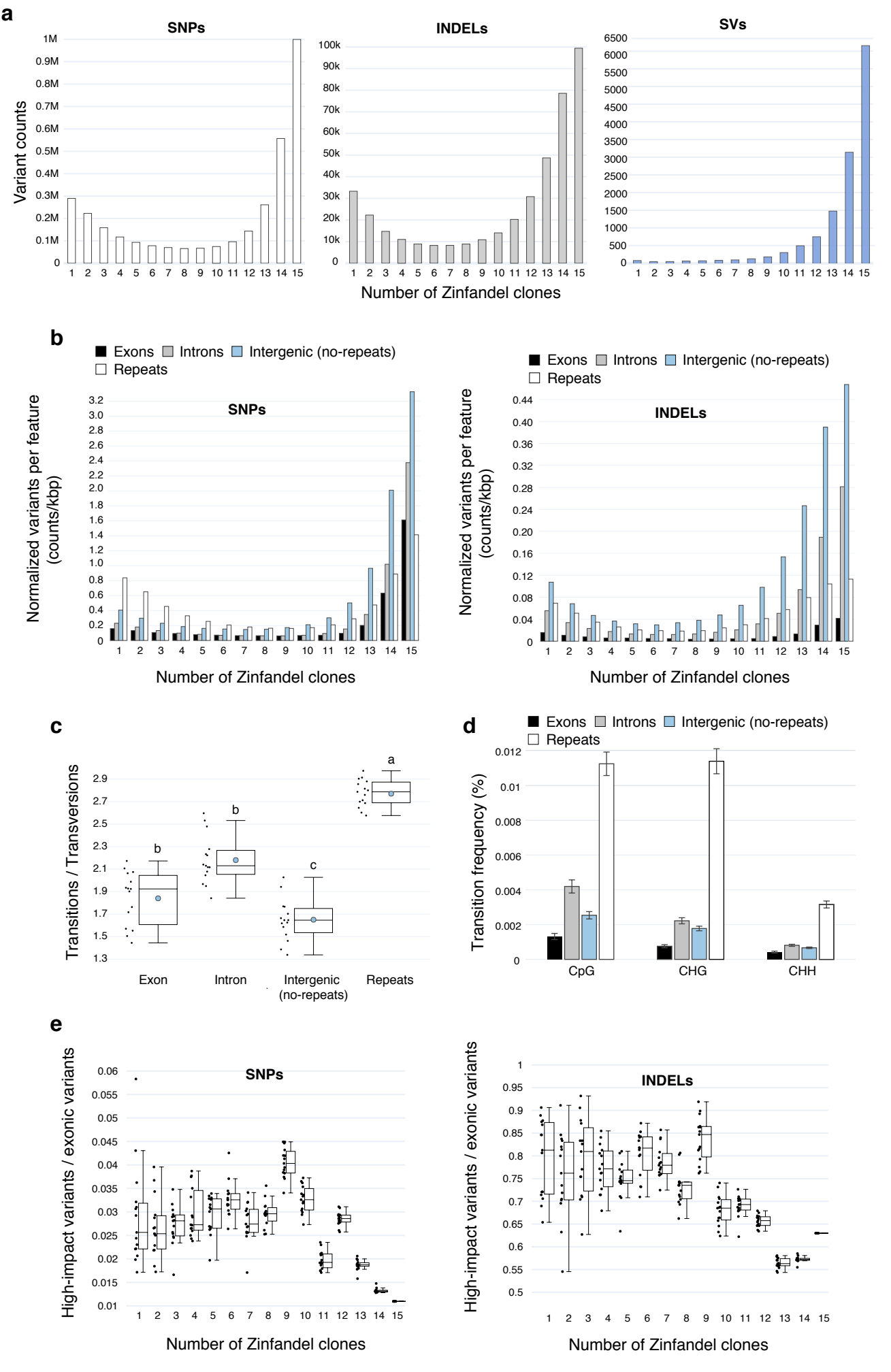
869

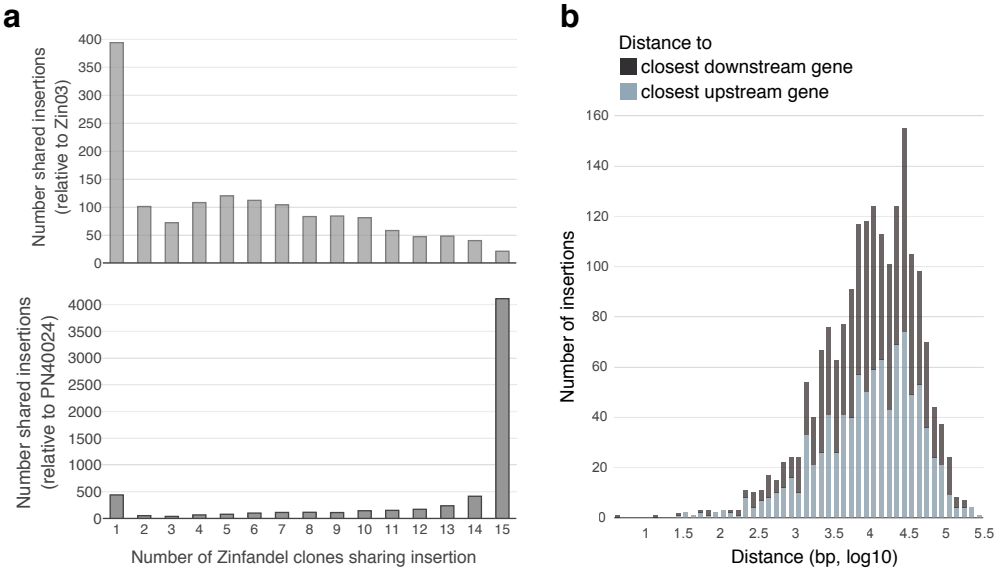

a
SV length (bp, log10)
6
5
4
3
2
Deletions
Duplicated insertions
Duplications
Insertions
Inversions
Inverted duplications
b
Zin03_Primary_000001F_quiver
13 kbp
4,998
5,000
5,002
5,004
5,006
5,008
5,010 kbp
VvZin03.Ps_2.ver1.0.g314880
VvZin03.Ps_2.ver1.0.g314900
VvZin03.Ps_2.ver1.0.g314910
c
Zin03_Primary_000005F_quiver
20 kbp
980
982
984
986
988
990
992
994
996
998 kbp
VvZin03.Ps_6.ver1.0.g469590
VvZin03.Ps_6.ver1.0.g469600
VvZin03.Ps_6.ver1.0.g469610
d
Zin03_Primary_000081F_quiver
24 kbp
682.5
683.0
683.5
704.8
705.0
705.2
705.4
VvZin03.Ps_81.ver1.0.g505940


a
Intersection size
Set size
Cultivar
51,759
1,122
309
260
68
35
4
3
52,058 Pinot Noir V1
52,090 Pinot Noir V2
52,984 Cabernet Sauvignon
53,560 Zinfandel
b
SV length (bp, log10)
Deletions
Duplicated insertions
Duplications
Insertions
Inversions
Inverted duplications
c
Zin03_Primary_000001F_quiver
10 kbp
Zinfandel 03
Cabernet Sauvignon 08
VvZin03.Ps_6.ver1.0.g471920
VvZin03.Ps_6.ver1.0.g471930
VvZin03.Ps_6.ver1.0.g471940
VvZin03.Ps_6.ver1.0.g471950
d
Zin03_Primary_000001F_quiver
23 kbp
Zinfandel 03
Cabernet Sauvignon 08
VvZin03.Ps_102.ver1.0.g214190


a
Key
Italy
Croatia
California
Principal component 2 (8.39%)
Principal component 1 (13.25%)
b
Principal component 2 (8.39%)
Principal component 1 (13.25%)
c
Zinfandel clones
Pinot Noir
Chardonnay
Sauvignon Blanc
Merlot
Cabernet Franc
Zin03
Kinship coefficient
0.0
0.1
0.2
0.4
0.5


Heterozygous variants
Homozygous variants
Between cultivars
Between clones
SNPs
INDELs
Normalized variants per feature (counts / kbp)
0
2
4
6
0.00
0.25
0.50
0.75
1.00
Intergenic
Introns
Exons
Intergenic
Introns
Exons
Feature


a
SNPs
INDELs
SVs
Variant counts
Number of Zinfandel clones
b
Exons
Introns
Intergenic (no-repeats)
Repeats
SNPs
INDELs
Normalized variants per feature (counts/kbp)
Number of Zinfandel clones
c
Transitions / Transversions
Exon
Intron
Intergenic (no-repeats)
Repeats
d
Transition frequency (%)
CpG
CHG
CHH
e
High-impact variants / exonic variants
SNPs
INDELs
Number of Zinfandel clones



c

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Retrotransposon / PLE : Penelope	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
Retrotransposon / LTR : Gypsy	194	60	36	60	62	62	44	41	38	33	21	19	21	13	10
Retrotransposon / LTR : Copia	84	11	14	18	23	13	17	10	12	17	10	6	6	7	6
Retrotransposon / LTR	40	14	6	6	13	13	20	10	10	8	7	8	2	6	1
Retrotransposon / LINE : L1	7	1	2	3	4	3	0	2	2	0	1	0	0	0	0
Retrotransposon / LINE	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0
Retrotransposon / DIRS : Ngaro	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
DNA-transposon / TIR : PIF-Harbinger	6	4	5	6	3	6	7	7	9	8	6	4	3	4	0
DNA-transposon / TIR : P	1	0	0	1	0	0	1	0	0	0	0	0	0	0	0
DNA-transposon / TIR : Mutator	6	0	1	1	1	5	0	0	0	1	2	1	0	1	1
DNA-transposon / TIR : hAT	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0
DNA-transposon / TIR	8	1	4	3	4	0	3	2	3	1	2	4	5	1	0
DNA-transposon	46	10	3	9	10	10	10	11	10	13	9	5	11	8	3

Transposable elements class / order : superfamily

Number of Zinfandel clones sharing insertion