The genomic diversification of 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. Email: dacantu@ucdavis.edu

Abstract

Grapevines are typically grown from cuttings rather than sexual reproduction. However, clones are not necessarily genetically identical because they accumulate somatic mutations. The purpose of this study was to better understand the consequences of clonal propagation and involved defining the nature and extent 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 the mutation occurred in individual or relatively few clones than when the site was heterozygous and shared by all or most clones. These data support three major conclusions. First, repetitive intergenic space is a major driver of clone genome diversification. Second, vegetative propagation is associated with the accumulation of putatively deleterious mutations. Third, and most interestingly, the data suggest that some degree of selection against deleterious variants in coding regions may exist such that mutations are less frequent in coding than noncoding regions of the genome.

Keywords: somatic mutations, structural variation, transposable elements, whole-genome sequencing, DNA methylation, asexual reproduction

Introduction

Cultivated grapevines are vegetatively propagated. As a result, the genome of each cultivar is preserved, except for the accumulation of mutations that accrue over time and can generate distinguishable clones (1-4). Somatic mutations tend to be heterozygous, recessive, and effectively "hidden" (2), but are responsible for several notable phenotypes. For example, a single, semi-dominant nucleotide polymorphism can affect hormone response (5) and recessive insertion of the *Gret1* retrotransposon in the promoter of the *VvmybA1* transcription factor inhibits anthocyanin accumulation in white varieties (6), as do additional mutations affecting the color locus (7-10). 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 (11,12). In citrus, undesirable mutations can be unknowingly propagated that render fruit highly acidic and inedible (13,14). Interestingly, somatic mutations in plum are associated with a switch from climacteric to non-climacteric ripening behavior (15).

49 Mutations occur in somatic cells that proliferate by mitosis. These can occur by a variety
50 of means, including single base-pair mutations (16,17) that are more prevalent in repetitive
51 regions because methylated cytosines are passively deaminated to thymines (18-20), polymerase
52 slippage that drives variable microsatellite insertions and deletions (21), and larger structural
53 rearrangements and hemizygous deletions (10,22). Transposable elements are also a major
54 source of somatic mutations in grapevines (23), though transcriptional and post-transcriptional
55 mechanisms exist to prevent transposition and maintain genome stability (24-27).
56 Distinct clones can emerge following a mutation in a shoot apical meristem that spreads
57 throughout a single cell layer, creating periclinal chimeras. This chimera is stable for Pinot
58 Meunier, a clone of Pinot Noir with distinct L1 and L2 layers in shoots (3), but cellular
59 rearrangements can result in the homogenization of the genotype in both cell layers (28) in other
60 periclinal chimeras. This is the case for green-yellow bud sports of the grey-fruited Pinot Gris,
61 wherein sub-epidermal white cells invaded and displaced epidermal pigmented cells (9).
62 Despite their significance, there is limited understanding of the extent, nature, and
63 implications of the somatic mutations that accumulate in clonally propagated crops (29).
64 Genotyping approaches based on whole genome sequencing make it possible to identify genetic
65 differences without predefined markers (23,30,31) and expedite learning the genetic basis of
66 valuable traits and developmental processes (15,22). Still, few previous studies have used
67 genomic approaches to study somatic variations among clones (22,23,30-32). The first to publish
68 a genome-wide exploration of somatic variation in grapevine was Carrier *et al.* (2012), finding
69 that transposable elements were the largest proportion of somatic mutation types affecting four
70 Pinot Noir clones. Whole genome sequencing was also used to study structural variations and
71 complex chromosomal rearrangements in Tempranillo, comparing diverse accessions of
72 phenotypically distinct Tempranillo Tinto and Tempranillo Blanco to better understand the basis

73 of somatic mutations giving rise to red versus white fruit (22). Genomic tools may also help
74 understand the consequences of clonal propagation. This study used whole genome sequencing
75 to expand the body of knowledge concerning the impact of vegetative propagation on the
76 Zinfandel winegrape.

77 Zinfandel is the third-most cultivated wine grape in California (33,34) and is particularly
78 interesting because of the intrigue surrounding its parentage, origins, and most contentiously, its
79 import to California (35-37). Zinfandel's parents remain unknown, but what is known of its
80 history was carefully reconstructed (38). Zinfandel likely arrived in California following
81 cultivation in the northeastern United States (35,39). DNA profiling produced evidence that
82 Zinfandel is synonymous with Primitivo grown in Italy (40) and Croatian Pribidrag and Crljenak
83 Kastelanski (41). In the course of its travels, Zinfandel bore many other names, probably
84 including Zinfandel and Black St. Peters, to name a couple. Historical records plus the
85 cultivation of closely related cultivars support Croatia as the likely origin of Zinfandel (36,41-43)
86 and also that Primitivo was likely brought to the Gioia del Colle region in Italy by Benedictine
87 monks in the 17th century (3,44).

88 Like other *vinifera* cultivars, there is phenotypic diversity among Zinfandel clones
89 associated with important viticultural traits (39,45,46). Comparisons of Zinfandel, Primitivo, and
90 Crljenak Kastelanski selections indicate clonal variation in yield, cluster size, width, and weight,
91 berries per cluster, berry weight, and fruit composition (46). Primitivo reportedly yields more
92 fruit that matures earlier, is less susceptible to sour rot (45) and is arranged in looser, smaller,
93 clusters with fewer berries than Zinfandel (39). The reported variability in Zinfandel and its long
94 history of cultivation make it a suitable model for studying clonal variation in grapevine,
95 specifically, and the nature of the accumulation of somatic mutations in clonally propagated
96 crops, generally. Importantly, University of California Davis and Zinfandel Advocates and

97 Producers (ZAP) established a heritage vineyard of Zinfandels, making it possible to study clonal
98 variation in a common environment.

99 The purpose of this study was to better understand the nature of the somatic variations
100 that occur during vegetative propagation. Representatives of at least a portion Zinfandel's history
101 (36,41-43) from Croatia, Italy, and California were sequenced and compared using Zin03 as
102 reference. Three conclusions were drawn from these data. First, the data support an important
103 component of Muller's ratchet (47), that asexually propagated organisms accumulate deleterious
104 mutations. Second, we show that intergenic space drives clonal diversification. As previously
105 reported, transposable element insertions varied among clones (48). This report expands that
106 understanding to implicate methylation as an indirect driver of clonal diversification; rare
107 somatic heterozygous SNPs were most observed in the repetitive intergenic regions, likely
108 because of the high levels of transposition-inhibiting methylation and associated transition
109 mutations that are prevalent there. Third, somatic mutations were relatively scarce in the coding
110 regions of genes relative to introns and intergenic space, suggesting some degree of negative
111 selection against deleterious mutations.

112

113 ## Results

114 **Phenolic diversity among Zinfandel clones**

115 University of California (UC) Davis and Zinfandel Advocates and Producers (ZAP)
116 established a Heritage Vineyard of Zinfandel clones to evaluate their viticultural and enological
117 performance when grown in the same environment. A subset of the clones used for this study are
118 grown in the Heritage Vineyard (Figure 1, Table 1). Phenolic acids, cinnamic acid, flavanones,
119 and anthocyanins are metabolites that influence fruit and wine quality (49-51); these were able to
120 distinguish some clones from others (Figure 1). Primitivo 1 and Zinfandel 8 were significantly

121 different than Zinfandels 7 and 11. Zinfandel 6, 9, and 10 were indistinguishable from one
122 another and the other clones. Interest in exploring the genetic basis of variation among
123 vegetatively propagated clones partially inspired the sequencing, assembly, and annotation of a
124 high-quality Zinfandel reference genome.
125
126 **Zinfandel genome assembly, annotation, and differences between haplotypes**
127 The clone used for the genome assembly, Zinfandel 03 (Zin03), was acquired by FPS in
128 1964 from the Reutz Vineyard near Livermore, California that was planted during Prohibition
129 (1920 - 1933, (52)). Zin03 is among some of the oldest available Zinfandel clones and had a
130 reputation for producing high quality wine at the time of its acquisition. Zin03 was sequenced
131 using Single Molecule Real-Time (SMRT; Pacific Biosciences) technology at ~98x coverage
132 and assembled using FALCON-unzip (53), a diploid-aware assembly pipeline. The genome was
133 assembled into 1,509 primary contigs (N50=1.1 Mbp) for a total assembly size of 591 Mbp,
134 similar to the genome size of Cabernet Sauvignon (590 Mbp; (53) though larger than
135 Chardonnay (490Mb; (31)) and the PN40024 genome (487 Mb, (54)). Fifty two percent of the
136 genome was phased into 2,246 additional phased sequences (haplotigs) where the homologous
137 chromosomes were distinguishable with an N50 of ~442 kbp (Table 2). A total of 53,560
138 complete protein-coding genes were annotated on the primary (33,523 genes) and haplotig
139 (20,037 genes) assemblies (Table 2).
140 Of the 20,037 genes annotated on the haplotig assembly, 18,878 aligned to the primary
141 assembly, leaving 1,159 genes that may exist hemizygously in the genome due to structural
142 variation between homologous chromosomes. These genes were annotated with a broad variety
143 of putative functions, including biosynthetic processes, secondary metabolism, and stress
144 responses. Mapping long reads to both the primary and haplotig assemblies was used for insight

145 into the differences between them. By mapping reads to the haplotig assembly, we could better
146 understand the circumstances that explain why some genes were only identified by reciprocal
147 mapping within that assembly.

148 Structural variants (SVs) between the haplotypes were examined by mapping long SMRT
149 sequencing reads onto Zin03's primary and haplotig assemblies with NGMLR and calling SVs
150 with Sniffles (55). A total of 22,399 SVs accounted for 6.94% (41.0 / 591 Mbp) of the primary
151 assembly's length and 6.02% (8.4 / 139 Mbp) of the primary assembly's gene-associated length
152 (Figure 2A; Table 3). SVs intersected 4,559 genes in the primary assembly (13.6% of primary
153 assembly genes) and 390 SVs spanned more than one gene. There was also substantial
154 hemizygosity in the genome, with long reads supporting deletions amounting to 2,521 genes and
155 4.56% of the primary assembly's length (Table 3). Manual inspection of the long reads aligned
156 to the primary assembly support that large, heterozygous deletions and inversions occurred in the
157 Zin03 genome that were either inherited from different structurally distinct parents or arose
158 during clonal propagation (Figure 2 B-D).

159 Next, we considered whether specific structural variation could account for the 1,159
160 genes uniquely found in the haplotig assembly. Three hundred eighty-two genes of the
161 previously mentioned 1,159 genes that uniquely exist within the haplotig assembly intersected
162 structural variations. Two hundred ninety of these intersected deletions, accounting for the
163 failure to identify them on the primary assembly. Some of the haplotig genes that failed to map
164 to the primary assembly intersected additional types of SVs, including duplications (80 genes),
165 insertions (89 genes), and inversions (16 genes).

166 These results reveal structural differences between Zinfandel's haplotypes. These
167 differences could have been inherited and/or could have occurred during vegetative propagation.
168 Overall, these structural variations affected 4,559 primary assembly genes. Importantly, these

169 data show that a notable portion of the primary assembly's length (4.56%) and genes (2,521) are
170 hemizygous.
171
172 **Differences in structure and gene content between Zinfandel and other grape genomes**
173 The Zin03 genome was compared to PN40024 and Cabernet Sauvignon to identify
174 cultivar-specific genes that may contribute to Zinfandel's characteristics. PN40024 is the inbred
175 line derived from Pinot Noir used to develop the first grape genome reference (54) and Cabernet
176 Sauvignon (CS08) was recently used to construct the first diploid, haplotype-resolved grape
177 genome for which long reads are available (53). Overall, 1,801 genes were not shared between
178 all three genotypes (Zin03, Pinot Noir, and Cabernet Sauvignon, Figure 3A). Three hundred nine
179 protein coding genes were found uniquely in Zin03 relative to PN40024 and CS08; 223 were
180 annotated on the primary assembly and 86 were annotated on the haplotigs (Figure 3A,
181 Additional file 2). These genes had a panoply of functions that included but were not limited to
182 nucleotide binding (60 genes), protein binding (58 genes), stress response (34 genes), and
183 kinases (28), and were associated with membranes (48 genes), signal transduction (23 genes),
184 carbohydrate metabolism (12 genes), and lipid metabolism (8 genes; Additional file 2).
185 Structural differences between Zin03 and CS08 were explored in more detail by mapping
186 the long SMRT reads of CS08 onto Zin03's primary and haplotig assemblies with NGMLR and
187 calling SVs with Sniffles (Figure 3B, Table 3). Overall, these SVs corresponded to 17.74% (159/
188 897 Mbp) of the Zin03 assembly's total length, 12.5% of its total protein-coding regions (28 /
189 223 Mbp), and 25.6% of all Zin03 genes. SVs spanned 9,885 genes in the primary assembly and
190 3,804 genes in the haplotigs. Manual inspection of the alignment of long CS08 reads to Zin03's
191 primary assembly support that large SVs exist between the two genotypes (Figure 3 C, D). Next,
192 we considered whether specific structural variation called by Sniffles could account for the 576

193 Zin03 genes considered absent from CS08 by the reciprocal mapping analysis (Figure 3A). Of
194 these 576 Zinfandel genes, 268 genes intersected 454 deletions supported by long CS08 reads
195 aligned to Zin03.
196 Though Zinfandel had few unique genes, high levels of structural variation between
197 Zinfandel (Zin03) and Cabernet Sauvignon (CS08) were observed and these affected
198 considerable protein-coding regions of the genome. These results justify constructing a
199 Zinfandel-specific reference to better capture genomic variability among Zinfandel clones that
200 could otherwise be missed, particularly if an alternative reference lacks sequences present in
201 Zinfandel.
202
203 **Relatedness among Zinfandel clones**
204 Sixteen Zinfandel clones, including Zin03, were sequenced using Illumina. The resulting
205 reads were aligned to the Zin03 primary assembly to characterize SNPs, small INDELs, variable
206 transposon insertions, and large structural variants. Principal Component Analysis (PCA) of
207 variants among the clones showed no clear pattern in their relationships to one another based on
208 their recorded origins prior to acquisition by FPS. The ambiguity surrounding the travels and
209 histories of these clones means that it should not be taken for granted that the Californian
210 selections, for example, ought to be more closely related to one another than to the Italian or
211 Croatian selections. Notably, Crljenak kaštelanski 3 stands notably apart from the other
212 Zinfandel clones. In addition, Pribidrags 5 and 15, which have a known and close relationship,
213 do not co-localize in the PCA (Figure 4A, B, Table 1).
214 A kinship analysis (56) was then used to quantitatively assess the relationships between
215 the Zinfandel selections. These values range from zero (unrelated) to 0.5 (self). Additional
216 cultivars were included in the analysis with known relationships to help contextualize the

217 differences between clones and the integrity of the analysis (Figure 4C). Cabernet Franc and
218 Merlot have a parent - offspring relationship, as do Pinot Noir and Chardonnay (57,58). These
219 pairs had kinship coefficients of 0.15 and 0.18, respectively (Figure 4C). As a possible
220 grandparent of Sauvignon Blanc, Pinot Noir had a kinship coefficient of 0.05 with Sauvignon
221 blanc (59,60). Most of the Zinfandel selections had kinship coefficients between 0.42 and 0.45;
222 this is likely because of the accrual of somatic mutations among clones. However, Crljenak
223 kaštelanski 3 had a noticeably low kinship coefficient (0.36 - 0.37) with every other Zinfandel
224 clone (Figure 4C). These data suggest that Crljenak kaštelanski 3 is either not a clone of
225 Zinfandel, contradicting marker analyses, or that it is a highly divergent clone.
226 Across the Zinfandel clones, the median number of homozygous and heterozygous
227 variants called relative to Zin03 were 38,092 and 717,925, respectively. Between 10-fold and
228 ~27-fold more heterozygous variants were called than homozygous variants in each clone except
229 for Crljenak kaštelanski 3, for which only ~2.5-fold more heterozygous sites were called
230 (Additional file 3: Table S1). Crljenak kaštelanski 3 had 4.3-fold more homozygous variants and
231 1.8-fold fewer heterozygous variants than the other clones (Additional file 3: Table S1).
232 Furthermore, unlike other clones, for which less than 10% of sites did not share the Zin03
233 reference allele, ~29% of variant sites were called where Crljenak kaštelanski 3 did not share the
234 Zin03 reference allele (Additional file 3: Table S1). Together, these results suggest that unlike
235 other Crljenak kaštelanski selections (41,43,46), Crljenak kaštelanski 3 is likely a close relative
236 of, but quite possibly not a clone of, Zinfandel. Because these analyses cast doubt on its identity
237 as a Zinfandel clone, Crljenak kaštelanski 3 was excluded from the clonal variation analyses
238 described in the rest of the study. Fifteen clones, including Zin03, remained in the analyses
239 (Table 1).
240

241 **Clonal versus cultivar genetic variability**

242 Overall, an average of 761,948 variant sites were identified in individual Zinfandel clones
243 when short reads were mapped on the Zin03 primary assembly. As stated previously, this
244 analysis excluded Crljenak kaštelanski 3. On average, 6,153,830 variant sites were identified in
245 other cultivars (Pinot noir, Chardonnay, Sauvignon Blanc, Merlot, Cabernet Franc) relative to
246 Zin03 (Additional file 3, Table 2). Both of these figures excluded heterozygous sites at which the
247 diploid genotype called for a given sample was identical to that called for Zin03.

248 Variants were 7.9X more frequent in other cultivars relative to Zin03 than for Zinfandel
249 clones; on average, mutations in clones occurred once every 723 bases and in once every 92
250 bases in other cultivars (Additional file 3, Table 2). However, the ratio of transitions to
251 transversion mutations and the proportions of the severities of the predicted variant effects were
252 similar for both groups (Additional file 3, Table 2). The normalized count of variants also
253 differed between cultivars and Zinfandel clones on the basis of variants' location in the genome,
254 the type of variant, and the zygosity of the variant (Figure 5).

255 Variants in non-Zinfandel cultivars and heterozygous variants among Zinfandel clones
256 were significantly more prevalent in intergenic space than introns and exons and significantly
257 more prevalent in introns than exons (Tukey HSD, $p < 0.01$). Unlike homozygous variants
258 between cultivars and as expected, homozygous variants were rare among clones (Figure 5;
259 Additional file 3, Table 1). Still, the normalized count of homozygous INDELs in intergenic
260 space, introns, and exons were significantly different among Zinfandel clones (Tukey HSD, $p <$
261 0.01), as were the normalized count of homozygous intergenic versus genic (exons and introns)
262 SNPs (Tukey HSD, $p < 0.01$). The normalized count of homozygous SNPs in exons and introns
263 were not significantly different in Zinfandel clones (Tukey HSD, $p > 0.01$). The accrual of
264 predominantly heterozygous and likely recessive variants (2) is consistent with what would be

265 expected given physically separate homologous chromosomes and the absence of sexual
266 reproduction.
267
268 **The accrual of somatic mutations in Zinfandel clones**
269 Heterozygous sites found among the 15 Zinfandel clones ought to be a mixture of sites
270 inherited from their shared ancestral plant and somatic mutations that arose during clonal
271 propagation. To better understand the nature of somatic mutations, the data were handled slightly
272 differently than they were to construct Figure 5; all 15 Zinfandel clones were included (including
273 Zin03, Crljenak kaštelanski 3 still excluded) and all heterozygous calls were considered, even if
274 all genotypes were identically heterozygous. Thirty percent of heterozygous SNPs, 24% of
275 heterozygous INDELs, and 47% of heterozygous structural positions were shared by all 15
276 Zinfandel clones (Figure 6A). Because all clones are identically heterozygous at these loci, these
277 variants are those inherited from Zinfandel's parents.
278 Individual and subsets of Zinfandel clones accumulated heterozygous mutations as clonal
279 propagation occurred (Figure 6A). Thirteen percent and 16% of heterozygous INDELs and
280 SNPs, respectively, and 1% of large (>50 bp) structural variants occurred in only one or two
281 clones (Figure 6A; Additional file 6A; Additional file 6B). The distribution of SVs called by
282 Delly is markedly different than those of SNPs and INDELs (Figure 6A). For both SNPs and
283 INDELs, there were 3 and 3.5-fold as many heterozygous variants shared by all 15 clones as
284 there were uniquely occurring variants; there were 71.5-fold more structural variants shared by
285 all clones than there were unique variants in individual clones (Figure 6A). This suggests that the
286 mechanisms that give rise to small mutations are more common among clones than the large-
287 scale changes associated with SVs.

288 The distribution of unique and shared heterozygous INDELs in exons, introns, repetitive,
289 and non-repetitive intergenic spaces were not equal (Figure 6B). The distribution of INDELs in
290 exons was significantly different than the distributions of INDELs in each other feature
291 considered (Kolmogorov-Smirnov Test, $p < 0.01$). Similarly, the distributions SNPs in genic
292 (exons, introns) and intergenic (repetitive, non-repetitive) regions were not equal (Figure 6B).
293 Shared heterozygous SNPs were most common in intergenic non-repetitive regions and introns
294 and least common in exons and repetitive intergenic regions (Figure 6B). Interestingly, unique
295 heterozygous SNPs occurred at high rates in repetitive intergenic regions (Figure 6B).
296 That shared heterozygous sites are mostly in non-repetitive intergenic space and unique
297 heterozygous sites are mostly in repetitive space may have to do with the disposition of
298 methylated cytosines to spontaneously deaminate and the prevalence of methylated repetitive
299 sequences in those regions (16,19,25,26). This is also supported by the significantly higher ratio
300 of transitions to transversions in repetitive intergenic regions than in exons, introns, and non-
301 repetitive intergenic space (Figure 6C). Furthermore, the mean percentage of CpG, CHG, and
302 CHH sites affected by transition mutations was significantly higher in repetitive intergenic space
303 than genic and non-repetitive intergenic spaces (Figure 6D, Tukey HSD, $p < 0.01$). The mean
304 percentage of CpG sites affected by transition mutations was also significantly higher in introns
305 than exons (Tukey HSD, $p < 0.01$). Compatible with this hypothesis, INDELs, which should not
306 increase in frequency due to methylation, did not occur preferentially in repeats (Figure 6B).
307 The impact of specific variants also varied with their prevalence among the clones
308 (Figure 6E). "High impact" mutations were predicted by SNPEff (61). The high impact
309 mutations identified in these data included exon losses, start and stop site gains and losses,
310 frameshifts, gene fusions, splice acceptor mutations, and splice donor mutations. These
311 mutations are predicted to be deleterious because of their disruptive effects on the coded protein.

312 For these reasons, we designated such mutations as putatively deleterious in this manuscript.
313 These were counted for each Zinfandel clone relative to Zin03. Relatively low percentages of
314 heterozygous variants shared by all Zinfandel clones were putatively deleterious. In contrast,
315 larger proportions of exonic SNPs and INDELs that occurred in individual or subsets of clones
316 were putatively deleterious (Figure 6E).

317 Together, these results show that mutations associated with vegetative propagation are
318 most numerous outside of coding regions of the genome, indicating that clone genomes diversify
319 most rapidly in the intergenic space, particularly in repetitive and likely methylated regions
320 (Figure 6). Though a minority of somatic mutations occurred in exons, we show that exonic
321 mutations that occur in few or individual clones are more often deleterious than exonic
322 heterozygous variants shared by all or most clones. In other words, vegetative propagation is
323 associated with the accumulation of putatively deleterious heterozygous mutations.

324

325 **Zinfandel clones incur unique transposon insertions**

326 Transposable element insertions (TEI) contribute to somatic variation in grape
327 (6,11,12,23). Relative to Zin03, 1,473 TEI were identified among the Zinfandel clones. A large
328 fraction of TEI (26.7%) occurred uniquely in individual clones (Figure 7A) and included 325
329 retrotransposons, mostly Copia and Gypsy LTRs, and 69 DNA-transposons (Figure 7B). Because
330 uniform loci are excluded, in-common TEI were not captured when clones were compared to
331 Zin03. Comparing the clones relative to PN40024, however, revealed that the majority (64.8%)
332 of TEI were shared among the 15 Zinfandel clones. Five hundred thirty TEI occurred in only
333 one, two or three clones (Figure 7A). This result supports the derivation of these selections from
334 a common ancestral plant and the accumulation of somatic variations over time.

335 In addition to being suggestive of their shared heritage, the positions of these insertions
336 and their proximity to coding genes were notable. Three-hundred forty-seven TEI occurred
337 within 314 coding genes. The remaining 938 TEIs were in intergenic regions (Figure 7C). The
338 median upstream and downstream distance of intergenic TEs from the closest feature were
339 11,811 and 11,279 base-pairs, respectively, though 25% of TEI were less than 4,345 bases
340 downstream of the closest feature or less than 3,826 bases upstream of the closest feature (Figure
341 7C).
342

343 ## Discussion

344 Somatic mutations in grapevine generate valuable phenotypes and have helped
345 understand the genetic basis of various traits (3,7,62,63). In Zinfandel, subtle but significant
346 differences in phenolic metabolites were identified between two pairs of clones of the seven
347 selected from the Heritage Vineyard (Figure 1). What manifest differences were observed could
348 reasonably be attributed to biological differences because these fruits were sampled from a
349 common, uniformly managed vineyard. However, additional studies are necessary to determine
350 the genetic basis of these differences, whether they lead to detectable sensory differences, and
351 comprehensively profile the chemical and morphological differences among clones.
352 Consideration of the genomic differences among Zinfandel clones revealed what is likely
353 a complex history not easily reconstructed, and that one selection was probably not a clone
354 despite being identified as such by SSR markers. Analyses of the relationships between clones
355 did not reveal groupings of clones per their recorded countries of origin. Somatic mutations may
356 help identify individual clones but could also blur the historical relationships between them. It is
357 also plausible that pairs of clones from any given region are not direct cuttings of one another but
358 of Zinfandels from another region; the clones now grown in California, for example, may have

359 been imported on numerous independent occasions from various other regions, meaning some
360 may indeed be more closely related to one of the Primitivo or Croatian clones than they are to
361 other Californian clones. It would be unwise to assume a single migratory path radiating from an
362 ancestral mother plant ought to be applicable to all clones.

363 Despite this ambiguity, the examination of SNPs, INDELs, transposable elements and
364 other structural variants all support the derivation of all but one of the clonal selections from a
365 common ancestral Zinfandel mother plant and show the accumulation of somatic mutations over
366 time (Figures 6 and 7).The structure of the Zinfandel genome, location of mutations among
367 clones, their frequency and prevalence, and the relationship between these factors provides some
368 insight into the nature of mutations in clonally propagated plants. Mutations among clones were
369 predominantly heterozygous (Figure 5) and uncommon heterozygous mutations shared by a
370 subset of or individual clones were increasingly deleterious when they occurred in exons (Figure
371 6E).

372 There are costs and benefits associated with clonal propagation (29). Among the benefits
373 are that the plants need not breed true-to-type; clonal propagation generally fixes heterozygous
374 loci and valuable phenotypes. However, the increase in the proportion of deleterious alleles
375 supports Muller's ratchet, which posits that sex is advantageous and that clonal propagation
376 increases mutational load (47). That mutations among Zinfandel clones were overwhelmingly
377 heterozygous reflects the absence of sexual recombination and the fact that grapevine is diploid.
378 Though these and previous data do not tell which mutations are actually recessive or dominant,
379 they could remain hidden if they are recessive or do not manifest their deleterious effects (2,64).
380 However, even after taking into consideration the total length of exons, introns, and intergenic
381 space (repetitive and non-repetitive), heterozygous mutations occurred at varying frequency in

382 these regions and were least abundant in coding regions. The rarity of mutations in exons and
383 commonality of mutations in repetitive intergenic space may have at least two components.
384 Mutations are likely more frequent in repetitive intergenic space as a result of the
385 regulation of transposition by DNA methylation. Repetitive intergenic space had the highest rate
386 of relatively unique SNPs and the ratio of transitions to transversions was significantly higher
387 there than in other regions. DNA methylation is an important epigenetic control and is one
388 mechanism that maintains genome stability and impairs the transposition of mobile elements
389 (25,65,66). Methylated cytosines, however, spontaneously deaminate faster than unmethylated
390 cytosines (18,26). Together, the expectations that intergenic regions are rich in transposable
391 elements, that these regions are typically highly methylated and as a result will experience
392 greater transition rates account for the high rates of SNPs in repetitive intergenic spaces among
393 Zinfandel clones. Also notable, these data show that some transposable elements are not entirely
394 silenced, with a substantial number inserting in genes or in close proximity to genes (Figure 7B).
395 These insertions could be effectively inconsequential or not; transposable element insertions can
396 result in novel transcripts and affect gene expression regulation (11,67).
397 The rarity of exonic mutations was surprising. After accounting for the length of these
398 spaces in the genome and their repetitiveness, we expected uniform rates of mutation in exons,
399 introns, and intergenic space. Instead, we still observed that although rare somatic mutations in
400 exons were increasingly deleterious, they were relatively scarce. This suggests the possibility of
401 an "escape hatch" from Muller's ratchet. Some degree of negative selection against deleterious
402 variants in coding regions may persist such that mutations are less frequent in coding than
403 noncoding regions of the genome. This would require somatic mutations to be exposed to
404 selection in some way, possibly as dominant and/or hemizygous alleles. Four and one half of
405 Zinfandel's genome is hemizygous; structural variations identified within the Zinfandel genome

406 and the rampant hemizygosity reported in Chardonnay (10) could expose otherwise hidden
407 somatic variations to selective pressure hostile to the accumulation of deleterious mutations. The
408 possibility of diplontic, clonal selection or competition between cell lineages that could purge
409 otherwise consequential deleterious mutations has been discussed, but evidence of its existence
410 is sparse (29,32,68).
411

412 ## Conclusions

413 This study described the nature of the mutations causing the diversification of 15 clonally
414 propagated grapevines and confirm their derivation from a single ancestral mother Zinfandel.
415 The findings indicate that repetitive intergenic space, likely because of its higher rates of
416 methylation in plants, is a significant contributor to the pool of mutations differentially observed
417 among the clones. In addition, the analyses revealed that though relatively infrequent compared
418 to intergenic mutations, mutations in exons were increasingly deleterious the less common they
419 were among Zinfandel clones; this suggests some mechanism by which deleterious alleles may
420 be purged from the genome.
421

422 ## Materials and Methods

423 **Zinfandel plant material and additional accessions**

424 Sixteen Zinfandel clones were used for this study. Plants were confirmed to be clones of
425 Zinfandel using the following microsatellite markers: VVMD5, VVMD7, VVMD27, VVMD31,
426 VVMD32, VVMS2, VRZAG62, and VRZAG79 (41,69,70). Fourteen of these clones are
427 available through Foundation Plant Services (FPS) at the University of California Davis. Nine of
428 the sixteen clones belong to the Zinfandel Heritage Vineyard Project, a collection of rare
429 Zinfandel vine cuttings grown in the same vineyard. The identification numbers, common

430 names, and source of the clones used in this study are listed in Table 1. An FPS identification
431 number suffix of ".1" indicates that the clone underwent microshoot tip tissue culture therapy,
432 with two exceptions. Pribidrag 13 and Pribidrag 15 are directly derived from the same plants as
433 Pribidrag 4 and Pribidrag 5, respectively, but did not undergo microshoot tip tissue culture
434 therapy. They are labeled with identical FPS numbers to make clear that the relationship between
435 them is known. In this manuscript, Zinfandel clones will be referred to by the clone numbers and
436 common names listed in Table 1.
437
438 **Extraction of Phenolic Compounds and HPLC-DAD**
439 Phenolic acids, cinnamic acid, flavanones, and anthocyanins in selected clones from the
440 Heritage Zinfandel Collection (Clones 1 and 6-11) were measured by HPLC-DAD. Phenolics
441 were extracted using a previously described method (71). Four biological replicates of each clone
442 were sampled at commercial harvest (~23-26 ºBrix) and technically duplicated. Berry skins were
443 blended for three minutes, mixed with 1L of solvent (1M HCl in 95% ethanol), placed in a
444 boiling water bath for 20 minutes, and allowed to cool for one hour at room temperature. The
445 solution was filtered through a 0.45-µm PTFE membrane prior to HPLC-DAD.
446 An Agilent 1100 HPLC system (Agilent Technologies, Santa Clara, CA) with a diode
447 array UV-visible detector coupled to an Agilent ChemStation (Rev. A.10.02) and 5 µm
448 ChromoSpher RP-18 column (Agilent Technologies) were used for solvent delivery and
449 detection. The flow rate was 0.5 mL/min, Solvent A was 50 mM dihydrogen ammonium
450 phosphate adjusted to pH 2.6 with orthophosphoric acid, Solvent B was 20% A in 80 %
451 acetonitrile, and Solvent C was 0.2 M orthophosphoric acid adjusted with ammonia to pH 1.5.
452 Separations were performed at 40°C and signals were recorded at 280 (Phenolic acids), 316

453 (Cinnamic acid), 365 (Flavanones), and 520 (Anthocyanins) nm. ChemStation was used to
454 identify and quantify metabolite classes.
455
456 **DNA extraction, library preparation, and sequencing**
457 High quality genomic DNA was isolated from grape leaves using the method described in
458 Chin *et al.* (2016) (53). DNA purity was evaluated with a Nanodrop 2000 spectrophotometer
459 (Thermo Scientific, Hanover Park, IL), quantity with a Qubit 2.0 Fluorometer (Life
460 Technologies, Carlsbad, CA) and integrity by electrophoresis. For SMRT sequencing, SMRTbell
461 libraries for the Zinfandel reference FPS clone 03 (Zin03) were prepared as described by Chin *et*
462 *al.* (2016) (53). For Illumina sequencing, DNA sequencing libraries for each of the sixteen
463 Zinfandel clones were prepared using the Kapa LTP library prep kit (Kapa Biosystems) as
464 described by Jones *et al*., 2014 (72). Final libraries were evaluated for quantity and quality using
465 a Bioanalyzer 2100 (Agilent Technologies, CA). Zin03 SMRTbell libraries were sequenced on a
466 PacBio RS II and Illumina libraries were sequenced in 100 and 150 base-pair paired-end reads
467 on an Illumina HiSeq3000 sequencer (DNA Technology Core Facility, University of California,
468 Davis). Genome sequences of additional *V. vinifera* were used in this study, including long reads
469 from Cabernet sauvignon (NCBI BioProject PRJNA316730) and short reads from Cabernet
470 franc, Chardonnay, Merlot, Pinot Noir, and Sauvignon blanc (NCBI BioProject PRJNA527006).
471
472 **Zinfandel genome assembly and annotation**
473 *De novo* assembly of Zinfandel (Zin03) was performed at DNAnexus (Mountain View,
474 CA, USA) using PacBio RS II data and the FALCON-unzip (v. 1.7.7) pipeline (53). FALCON-
475 unzip was used for its ability to assemble contiguous, phased diploid genomes with better
476 resolved heterozygosity (53,73). Repetitive sequences were masked prior to error correction

477 using TANmask and REPmask modules in the Damasker (74). After error-correction (13,073 bp
478 length cut-off), a total of 1.68 million error-corrected reads (N50 15Kbp, 29-fold coverage of
479 expected genome size) were obtained and repeats were masked before overlap detection in the
480 FALCON pipeline (v. 1.7.7). PacBio reads were assembled after testing multiple parameters to
481 produce the least fragmented assembly. These conditions are listed in Additional file 1.
482 Haplotype reconstruction was performed with default parameters. Finally, contigs were polished
483 with Quiver (Pacific Biosciences, bundled with FALCON-unzip v. 1.7.7). Repeats were
484 annotated on the Zin03 assembly using RepeatMasker (v. open-4.0.6) (75) and a *V. vinifera*
485 repeat library (76).
486 The publicly available RNAseq datasets listed in Additional file 1 were used as
487 transcriptional evidence for gene prediction. Each RNAseq sample was trimmed with
488 Trimmomatic (v. 0.36; settings are listed in Additional file 1) and assembled with Stringtie (v.
489 1.3.3) (77) to reconstruct variety-specific transcripts. A detailed list of all experimental data used
490 for the annotation procedure is listed in Additional file 1. This data was then mapped on the
491 genome using Exonerate (v. 2.2.0, transcripts and proteins) (78) and PASA (v. 2.1.0, transcripts)
492 (79). Alignments, and *ab initio* predictions generated with SNAP (80)(v. 2006-07-28), Augustus
493 (81) (v. 3.0.3), and GeneMark-ES (82) (v. 4.32), were used as input for EVidenceModeler (v.
494 1.1.1) (83). EVidenceModeler was used to identify consensus gene structures using the weight
495 reported in Additional file 1 Functional annotation was performed using the RefSeq plant protein
496 database (ftp://ftp.ncbi.nlm.nih.gov/refseq, retrieved January 17th, 2017) and InteProScan (v. 5)
497 as previously described (76).
498
499 **Genetic variant calling**

500 Comparisons between Zinfandel clones and between Zin03 and other cultivars were
501 made using the Zin03 genome as reference. Small insertions and deletions (INDELs), single
502 nucleotide polymorphisms (SNPs), and structural variations (SVs) were analyzed. The short
503 Illumina reads belonging to the sixteen Zinfandel clones and additional cultivars were trimmed
504 using Trimmomatic (v. 0.36; settings are listed in Additional file 1). Quality filtered and trimmed
505 paired-end reads were then randomly down-sampled to 84 million (~14X coverage) in each
506 library to mitigate the possibility of sequencing depth-dependent outcomes. All libraries were
507 aligned to Zin03 using bwa (v. 0.7.10) and the -M parameter (84). For all genotypes, the median
508 number of reads mapping to the Zinfandel reference genome was approximately 97%. Next,
509 Picard Tools (v. 2.12.1) were used to mark optical duplicates, build BAM indices, and validate
510 SAM files (http://broadinstitute.github.io/picard). Variants were called using GATK's
511 HaplotypeCaller (v. 3.5) (85). Then, called variants were filtered and annotated (--
512 filterExpression "QD < 2.0 || FS > 60.0 || MQ < 40.0 || MQRankSum < -12.5 || ReadPosRankSum
513 < -8.0"). Variant call files were combined using GATK's GenotypeGVCFs. Having mapped
514 Illumina reads corresponding to the Zinfandel reference onto itself, erroneous non-reference
515 Zin03 calls (8.1%) were removed. The variants called included SNPs and INDELs.
516 Next, large structural variations among clones, between Zin03 and other cultivars, and
517 between Zin03's haplotypes were studied. First, Zin03 genes were compared to PN40024 and
518 Cabernet Sauvignon (CS08) by mapping coding sequences on genome assemblies using Gmap
519 (v. 2015-09-29) and the following parameters: -K 20,000 -B 4 -f 2. Hits with at least 80%
520 identity and reciprocal coverage are reported. Genes annotated on Zin03's haplotig assembly
521 were also mapped to Zin03's primary assembly to assess differences in gene content between
522 Zin03's haplotypes. SMRT reads from Zin03 and CS08 were mapped to Zin03 using NGMLR
523 (v. 0.2.7) and structural differences were called with Sniffles v.1.0.8 (55). Zinfandel clones were

524 compared to one another using Illumina short reads and Delly (v. 0.7.8) with default parameters
525 (86). The structural variations identified by Sniffles and Delly in Zin03 were intersected. Several
526 filters were applied to the results of SV analyses. Transversions, non-reference Zin03 genotype
527 calls, SVs that affect the ends of contigs, and SVs that intersected the repeat annotation were
528 filtered from Delly output.
529
530 **Transposon insertion analysis**
531 PoPoolationTE2 (v. 1.10.04) (87) was used to identify transposon insertions in the
532 Zinfandel clones; it was used following the workflow outlined in its software manual
533 (https://sourceforge.net/p/popoolation-te2/wiki/Manual/). Insertions were called relative to Zin03
534 genome assembly and PN20024 (54). As described in Kofler *et al.* 2016, PoPoolationTE2
535 analyses transposable element insertions and can identify novel and annotated TE insertions
536 provided insertions fall within predefined families of TEs. The annotation produced by
537 RepeatMasker was used for the analysis. In this manuscript, the TE insertions among the clones
538 are reported using the classification system and nomenclature described by Wicker *et al.* (2007)
539 (88). In instances where the TE order and/or superfamily was not annotated, only the TE class
540 and, when available, order are named in the associated figures and text.
541
542 **Relationships between Zinfandel clones**
543 The relationships between Zinfandel clones were visualized by Principal Component
544 Analysis and their relatedness was quantified (VCFtools v. 0.1.15) based on the method
545 described by Manichaikul *et al*. (2010) (56). This approach gives information about the
546 relationship of any pair of individuals (unrelated, 3rd degree relative, 2nd degree relative, full

547 siblings, and self) by estimating their kinship coefficient, which ranges from zero (no
548 relationship) to 0.50 (self). These analyses used SNPs outside of repetitive regions.
549

550 **Availability of data and materials**

551 The datasets supporting the conclusions of this article are available in two locations. Raw
552 sequences are available at NCBI (Bioproject PRJNA527006). Other relevant data, such as
553 genome sequence, gene and protein sequences, gene and repeat coordinates and annotation,
554 along with a genome browser and a blast tool, are available at http://cantulab.github.io/data.html.
555

556 **Funding**

557 This work was partially supported by start-up funds from the College of Agricultural and
558 Environmental Sciences (UC Davis) to DC, the Louis P. Martini Endowment in Viticulture to
559 DC and the NSF PGRP grant #1741627 to DC, MAW, and BG.
560

561 **Acknowledgements**

562 We are grateful for the vision of the late James A. Wolpert, who established the original
563 Zinfandel clone trials with the support of the Zinfandel Advocates and Producers (ZAP).
564

565 ## References

566 1. Riaz S, Garrison KE, Dangl GS, Boursiquot JM, Meredith CP. Genetic divergence
567 and chimerism within ancient asexually propagated winegrape cultivars. Journal of
568 the American Society for Horticultural Science. 2002;127(4):508–14.

569 2. Zhou Y, Massonnet M, Sanjak JS, Cantu D, Gaut BS. Evolutionary genomics of
570 grape (*Vitis vinifera* ssp. *vinifera*) domestication. PNAS. 2017; 48:201709257–6.

571 3. Franks T, Botta R, Thomas MR, Franks J. Chimerism in grapevines: Implications
572 for cultivar identity, ancestry and genetic improvement. Theor Appl Genet.
573 2002;104(2-3):192–9.

574 4. Ramu P, Esuma W, Kawuki R, Rabbi IY, Egesi C, Bredeson JV, *et al.* Cassava
575 haplotype map highlights fixation of deleterious mutations during clonal
576 propagation. Nat Genet. 2017;49(6):959–63.

577 5. Boss PK, Thomas MR. Association of dwarfism and floral induction with a grape
578 "green revolution" mutation. Nature. 2002;416(6883):847–50.

579 6. Kobayashi S, Goto-Yamamoto N, Hirochika H. Retrotransposon-induced mutations
580 in grape skin color. Science. 2004;304(5673):982–2.

581 7. Walker AR, Lee E, Robinson SP. Two new grape cultivars, bud sports of Cabernet
582 Sauvignon bearing pale-coloured berries, are the result of deletion of two regulatory
583 genes of the berry colour locus. Plant Mol Biol. 2006;62(4-5):623–35.

584 8. Yakushiji H, Kobayashi S, Goto-Yamamoto N, Tae Jeong S, Sueta T, Mitani N, *et*
585 *al.* A skin color mutation of grapevine, from black-skinned Pinot Noir to white-
586 skinned Pinot Blanc, is caused by deletion of the functional VvmybA1 allele. Biosci
587 Biotechnol Biochem. 2006;70(6):1506–8.

588 9. Pelsy F, Dumas V, Bévilacqua L, Hocquigny S, Merdinoglu D. Chromosome
589 Replacement and Deletion Lead to Clonal Polymorphism of Berry Color in
590 Grapevine. Richardson PM, editor. PLoS Genet. 2015;11(4):e1005081–18.

591 10. Zhou YS, Minio A, Massonnet M, Solares EA, Lyu Y, Beridze T, *et al.* Structural
592 variants, clonal propagation, and genome evolution in grapevine (*Vitis vinifera*):
593 Supplemental Figures and Tables. bioRxiv. 2018;1–48.

594 11. Fernandez L, Torregrosa L, Segura V, Bouquet A, Martínez-Zapater JM.
595 Transposon-induced gene activation as a mechanism generating cluster shape
596 somatic variation in grapevine. The Plant Journal. 2010;61(4):545–57.

597 12. Fernandez L, Chaïb J, Zapater JMM, Thomas MR, Torregrosa L. Mis-expression of
598 a PISTILLATA-like MADS box gene prevents fruit development in grapevine. The
599 Plant Journal. 2013;73(6):918–28.

600 13. Whitham TG, Slobodchikoff CN. Evolution by Individuals, Plant-Herbivore
601 Interactions, and Mosaics of Genetic Variability: The Adaptive Significance of
602 Somatic Mutations in Plants. Oecologia. 1981;49(3):287–92.

603 14. Soost RK, Cameron JW, Bitters WP, Platt RG. Citrus bud variation, old and new.
604 Calif Citrograph. 1961;46(176):188–93.

605 15. Farcuh M, Li B, Rivero RM, Shlizerman L, Sadka A, Blumwald E. Sugar
606 metabolism reprogramming in a non-climacteric bud mutant of a climacteric plum

607 fruit during development on the tree. Journal of Experimental Botany. 2017;68(21-
608 22):5813–28.

609 16. Ossowski S, Schneeberger K, Lucas-Lledó JI, Warthmann N, Clark RM, Shaw RG,
610 *et al.* The Rate and Molecular Spectrum of Spontaneous Mutations in *Arabidopsis*
611 *thaliana*. Science. 2010;327(5961):92–4.

612 17. Hershberg R, Petrov DA. Evidence That Mutation Is Universally Biased towards AT
613 in Bacteria. Nachman MW, editor. PLoS Genet. 2010;6(9).

614 18. Selker EU. Premeiotic instability of repeated sequences in Neurospora crassa. Annu
615 Rev Genet. Annual Reviews 1990;24(1):579–613.

616 19. Meunier J, Khelifi A, Navratil V, Duret L, Lipman DJ. Homology-Dependent
617 Methylation in Primate Repetitive DNA. Proc Natl Acad Sci. 2005;102(15):5471–6.

618 20. Mautino MR, Rosa AL. Analysis of Models Involving Enzymatic Activities for the
619 Occurrence of C-T Transition Mutations During Repeat-Induced Point Mutation
620 (RIP) in *Neurospora crassa.* J. theor. Biol. 1998;192:61–71

621 21. Schlötterer C, Tautz D. Slippage synthesis of simple sequence DNA. Nucleic Acids
622 Research. 1992;20(2):211–5.

623 22. Carbonell-Bejerano P, Royo C, Torres-Pérez R, Grimplet J, Fernandez L, Franco-
624 Zorrilla JM, *et al.* Catastrophic unbalanced genome rearrangements cause somatic
625 loss of berry color in grapevine. Plant Physiology. 2017;175:786-801

626 23. Carrier G, Le Cunff L, Dereeper A, Legrand D, Sabot F, Bouchez O, *et al.*
627 Transposable Elements Are a Major Cause of Somatic Polymorphism in *Vitis*
628 *vinifera*. PLoS ONE. 2012;7(3):e32973–10.

629 24. Qi Y, He X, Wang X-J, Kohany O, Jurka J, Hannon GJ. Distinct catalytic and non-
630 catalytic roles of ARGONAUTE4 in RNA-directed DNA methylation. Nature.
631 2006;443(7114):1008–12.

632 25. Shen H, He H, Li J, Chen W, Wang X, Guo L, *et al.* Genome-Wide Analysis of
633 DNA Methylation and Gene Expression Changes in Two Arabidopsis Ecotypes and
634 Their Reciprocal Hybrids. The Plant Cell Online. 2012;24(3):875–92.

635 26. Cantu D, Vanzetti LS, Sumner A, Dubcovsky M, Matvienko M, Distelfeld A, *et al.*
636 Small RNAs, DNA methylation and transposable elements in wheat. BMC
637 Genomics. 2010;11(1).

638 27. Chan SW-L, Henderson IR, Jacobsen SE. Gardening the genome: DNA methylation
639 in *Arabidopsis thaliana*. Nat Rev Genet. 2005;6(5):351–60.

640 28. Hocquigny S, Pelsy F, Dumas V, Kindt S, Heloir M-C, Merdinoglu D.
641 Diversification within grapevine cultivars goes through chimeric states. Genome.
642 2004;47(3):579–89.

643 29. McKey D, Elias M, Pujol B, Duputié A. The evolutionary ecology of clonally
644 propagated domesticated plants. New Phytologist. 2010;186(2):318–32.

645 30. Gambino G, Molin AD, Boccacci P, Minio A, Chitarra W, Avanzato CG, *et al.*
646 Whole-genome sequencing and SNV genotyping of "Nebbiolo" (*Vitis vinifera* L.)
647 clones. Sci Rep. 2017;7(1):17294.

648 31. Roach MJ, Johnson DL, Bohlmann J, van Vuuren HJJ, Jones SJM, Pretorius IS, *et*
649 *al.* Population sequencing reveals clonal diversity and ancestral inbreeding in the
650 grapevine cultivar Chardonnay. Richardson PM, editor. PLoS Genet. Public Library
651 of Science; 2018;14(11)

652 32. Plomion C, Aury J-M, Amselem J, Leroy T, Murat F, Duplessis S, *et al.* Oak
653 genome reveals facets of long lifespan. Nature Plants. 2018;4:440–52.

654 33. California Grape Crush Report Preliminary 2015. CDFA

655 34. California Grape Crush Report 2016. CDFA

656 35. Sullivan CL. A Viticultural Mystery Solved: The Historical Origins of Zinfandel in
657 California. California History. 1978;57(2):114–29.

658 36. Mirošević N, Meredith CP. A review of research and literature related to the origin
659 and identity of the cultivars Plavac mali, Zinfandel and Primitivo (*Vitis vinifera* L.).
660 Agriculturae Conspectus Scientificus. 2000;65(1):45–9.

661 37. Lévai C. Ágoston Haraszthy: "Father of California Viticulture?" Debates in the
662 Mirror of Recent Revisionist Literature. Hungarian Studies Review. 2013

663 38. Sullivan C. Zinfandel: a history of a grape and its wine. 2003

664 39. Wolpert JA. Performance of Zinfandel and Primitivo Clones in a Warm Climate.
665 AJEV. 1996;47(2):124.

666 40. Bowers JE, Bandman EB, Meredith CP. DNA Fingerprint Characterization of Some
667 Wine Grape Cultivars. AJEV. 1993;44(3):266–74.

668 41. Maletic E, Pejic I, Karoglan Kontic J, Piljac J, Dangl G, Vokurka A, *et al.* The
669 Identification of Zinfandel on the Dalmatian Cost of Croatia. Agriculturae
670 Conspectus Scientificus. 2003;(603):251–4.

671 42. Maletic E, Pejic I, Kontic JK, Piljac J, Dangl GS, Vokurka A, *et al.* Zinfandel,
672 Dobricic, and Plavac mali: The genetic relationship among three cultivars of the
673 Dalmatian Coast of Croatia. AJEV. 2004;55(2):174–80.

674 43. Fanizza G, Lamaj F, Ricciardi L, Resta P, Savino V. Grapevine cvs Primitivo,
675 Zinfandel and Crljenak kastelanski: Molecular analysis by AFLP. Vitis.
676 2005;44(3):147–8.

677 44. Russo G, Liuzzi V, D'Andrea L, Alviti G. Comparison among Five Clones of
678 "Primitivo" Vine in Southern Italy. Agriculturae Conspectus Scientificus.
679 2003;(603):779–86.

680 45. Fidelibus MW, Christensen LP, Katayama DG, Verdenal PT. Performance of
681 Zinfandel and Primitivo grapevine selections in the central San Joaquin Valley,
682 California. AJEV. 2005;56(3):284–6.

683 46. Zdunić G, Simon S, Malenica N, Budić-Leto I, Maletic E, Karoglan Kontić J, *et al.*
684 Intravarietal variability of Crljenak Kastelanski' and Its Relationship with 'Zinfandel'
685 and 'Primitivo' Selections. Agriculturae Conspectus Scientificus. 2014;(1046):573–
686 80.

687 47. Muller HJ. Some genetic aspects of sex. The American Naturalist. 1932;66:118–38.

688 48. Carrier G, Le Cunff L, Dereeper A, Legrand D, Sabot F, Bouchez O, *et al.*
689 Transposable Elements Are a Major Cause of Somatic Polymorphism in *Vitis*
690 *vinifera* L. PLoS ONE. 2012;7(3)

691 49. Liu Y, Zhang B, He F, Duan C-Q, Shi Y. The Influence of Prefermentative Addition
692 of Gallic Acid on the Phenolic Composition and Chromatic Characteristics of
693 Cabernet Sauvignon Wines. Journal of Food Science. 2016;81(7):C1669–78.

694 50. Price SF, Breen PJ, Valladao M, Watson BT. Cluster Sun Exposure and Quercetin in
695 Pinot noir Grapes and Wine. AJEV. 1995;46(2):187–94.

696 51. Kontoudakis N, Esteruelas M, Fort F, Canals JM, De Freitas V, Zamora F. Influence
697 of the heterogeneity of grape phenolic maturity on wine composition and quality.
698 Food Chemistry. 2011;124(3):767–74.

699 52. Sweet NL, The Zinfandels of FPS. FPS Grape Program Newsletter. 2007

700 53. Chin C-S, Peluso P, Sedlazeck FJ, Nattestad M, Concepcion GT, Clum A, *et al.*
701 Phased diploid genome assembly with single-molecule real-time sequencing. Nat
702 Meth. 2016;13(12):1050–4.

703 54. Jaillon O, Aury J-M, Noel B, Policriti A, Clepet C, Casagrande A, *et al.* The
704 grapevine genome sequence suggests ancestral hexaploidization in major
705 angiosperm phyla. Nature. 2007;449(7161):463–7.

706 55. Sedlazeck FJ, Rescheneder P, Smolka M, Fang H, Nattestad M, Haeseler A, *et al.*
707 Accurate detection of complex structural variations using single-molecule
708 sequencing. Nat Meth. 2018;15:461–8.

709 56. Manichaikul A, Mychaleckyj JC, Rich SS, Daly K, Sale M, Chen W-M. Robust
710 relationship inference in genome-wide association studies. Bioinformatics.
711 2010;26(22):2867–73.

712 57. Boursiquot JM, Lacombe T, Laucou V, Julliard S, Perrin Fx, Lanier N, *et al.*
713 Parentage of Merlot and related winegrape cultivars of southwestern France:

714 discovery of the missing link. Australian Journal of Grape and Wine Research.
715 2009;15(2):144–55.

716 58. Bowers J, Boursiquot JM, This P, Chu K, Johansson H, Meredith C. Historical
717 genetics: The parentage of chardonnay, gamay, and other wine grapes of
718 northeastern France. Science. 1999;285(5433):1562–5.

719 59. Regner F, Stadlbauer A, Eisenheld C, Kaserer H. Genetic Relationships Among
720 Pinots and Related Cultivars. AJEV. 2000;51(1):7–14.

721 60. Imazio S, Labra M, Grassi F, Winfield M, Bardini M, Scienza A. Molecular tools
722 for clone identification: the case of the grapevine cultivar "Traminer." Plant
723 Breeding. 2nd ed. 2002;121(6):531–5.

724 61. Cingolani P, Platts A, Le Lily Wang, Coon M, Nguyen T, Wang L, *et al.* A program
725 for annotating and predicting the effects of single nucleotide polymorphisms,
726 SnpEff. Fly. 2012;6(2):80–92.

727 62. Vezzulli S, Leonardelli L, Malossini U, Stefanini M, Velasco R, Moser C. Pinot
728 blanc and Pinot gris arose as independent somatic mutations of Pinot noir. Journal of
729 Experimental Botany. 2012;63(18):6359–69.

730 63. Duchêne E, Legras Jl, Karst F, Merdinoglu D, Claudel P, Jaegli N, *et al.* Variation
731 of linalool and geraniol content within two pairs of aromatic and non-aromatic
732 grapevine clones. Australian Journal of Grape and Wine Research. 2009;15(2):120–
733 30.

734 64. Gaut BS, Seymour DK, Liu Q, Zhou Y. Demography and its effects on genomic
735 variation in crop domestication. Nature Plants. 2018;4:512-20

736 65. Chen RZ, Pettersson U, Beard C, Jackson-Grusby L, Jaenisch R. DNA
737 hypomethylation leads to elevated mutation rates. Nature. 1998;395(6697):89–93.

738 66. Hirochika H, Okamoto H, Kakutani T. Silencing of Retrotransposons in Arabidopsis
739 and Reactivation by the ddm1 Mutation. Plant Cell. 2000;12(3):357.

740 67. Hirsch CD, Springer NM. Transposable element influences on gene expression in
741 plants. BBA - Gene Regulatory Mechanisms. 2017;1860(1):157–65.

742 68. Klekowski EJ. Plant clonality, mutation, diplontic selection and mutational
743 meltdown. Biological Journal of the Linnean Society. 2003;79(1):61–7.

744 69. Thomas MR, Cain P, Scott NS. Dna Typing of Grapevines - a Universal
745 Methodology and Database for Describing Cultivars and Evaluating Genetic
746 Relatedness. Plant Mol Biol. 1994;25(6):939–49.

747 70. Sefc KM, Regner F, Turetschek E, Glössl J, Steinkellner H. Identification of
748 microsatellite sequences in Vitis riparia and their applicability for genotyping of
749 different Vitis species. Genome. 1999;42(3):367–73.

750 71. Erickson AC. Evolution of pigments and cofactors in the skin of Pinot noir and
751 Cabernet sauvignon grapes during ripening, and their effect on wine color.
752 University of California Davis; 2000.

753 72. Jones L, Riaz S, Morales-Cruz A, Amrine KCH, McGuire B, Gubler WD, *et al.*
754 Adaptive genomic structural variation in the grape powdery mildew pathogen,
755 Erysiphe necator. BMC Genomics. 2014;15(1):1081.

756 73. Minio A, Lin J, Gaut BS, Cantu D. How Single Molecule Real-Time Sequencing
757 and Haplotype Phasing Have Enabled Reference-Grade Diploid Genome Assembly
758 of Wine Grapes. Front Plant Sci. 2017;8:481–6.

759 74. Myers G. Efficient Local Alignment Discovery amongst Noisy Long Reads. In:
760 Algorithms in Bioinformatics. 2014; 52–67.

761 75. Smit A, Hubley R, Green P. *RepeatMasker Open-4.0* [Internet]. Available from:
762 http://www.repeatmasker.org

763 76. Minio A, Massonnet M, Figueroa-Balderas R, Vondras AM, Blanco-Ulate B, Cantu
764 D. Iso-Seq Allows Genome-Independent Transcriptome Profiling of Grape Berry
765 Development. G3 (Bethesda). G3: Genes, Genomes, Genetics;
766 2019;g3.201008.2018.

767 77. Pertea M, Kim D, Pertea GM, Leek JT, Salzberg SL. Transcript-level expression
768 analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. Nat Protoc.
769 2016;11(9):1650–67.

770 78. Slater GSC, Birney E. Automated generation of heuristics for biological sequence
771 comparison. BMC Bioinformatics. BioMed Central; 2005;6(1):31.

772 79. Haas BJ, Delcher AL. Improving the Arabidopsis genome annotation using maximal
773 transcript alignment assemblies. Nucleic Acids Research. 2003;31(19):5654–5666,

774 80. Korf I. Gene finding in novel genomes. BMC Bioinformatics. BioMed Central;
775 2004;5(1):59.

776 81. Stanke M, Keller O, Gunduz I, Hayes A, Waack S, Morgenstern B. AUGUSTUS: ab
777 initio prediction of alternative transcripts. Nucleic Acids Research. 2006;34:W435–
778 9.

779 82. Lomsadze A, Ter-Hovhannisyan V, Chernoff YO, Borodovsky M. Gene
780 identification in novel eukaryotic genomes by self-training algorithm. Nucleic Acids
781 Research. 2005;33:6494–506.

782 83. Haas BJ, Salzberg SL, Genome WZ, 2008. Automated eukaryotic gene structure
783 annotation using EVidenceModeler and the Program to Assemble Spliced
784 Alignments. Genome Biology. 2008;9:R7

785 84. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler
786 transform. Bioinformatics. 2009;25(14):1754–60.

85. Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, del Angel G, Moonshine AL, *et al.* From FastQ Data to High-Confidence Variant Calls: The Genome Analysis Toolkit Best Practices Pipeline. Current Protocols in Bioinformatics. 2013;43(1):11.10.1–11.10.33.

86. Rausch T, Zichner T, Schlattl A, Stutz AM, Benes V, Korbel JO. DELLY: structural variant discovery by integrated paired-end and split-read analysis. Bioinformatics. 2012;28(18):i333–9.

87. Kofler R, Gómez-Sánchez D, Schlötterer C. PoPoolationTE2: Comparative Population Genomics of Transposable Elements Using Pool-Seq. Mol Biol Evol. 2016 Sep 16;33(10):2759–64.

88. Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, Chalhoub B, *et al.* A unified classification system for eukaryotic transposable elements. Nat Rev Genet. 2007;8(12):973–82.

Tables and Figures

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

802

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

803

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

804

805

806 ## Figure legends

807 **Figure 1.** Linear discriminant analysis of seven Zinfandel selections using their phenolic content
808 (anthocyanins, catechins, gallic acid, and quercetin) measured by HPLC. 95% Confidence ellipses
809 around replicates are shown. There were statistically significant differences between clones per
810 their phenolic profiles. MANOVA, $p < 0.0005$.

811

812 **Figure 2.** Structural variation between Zin03 haplotypes. **A**. Distribution of structural variation
813 sizes. Boxplots show the 25th quartile, median, and 75th quartile for each type of SV. Whiskers are
814 1.5$^{\text{Inter-Quartile Range}}$. Diamonds indicate the mean $\log_{10}$(length) of each type of SV; **B,C,D.** Examples
815 of heteryzygous structural variants between haplotypes that intersect genes. For each reported
816 deletion, (from top to bottom) the coverage, haplotype-resolved alignment of reads, and the genes

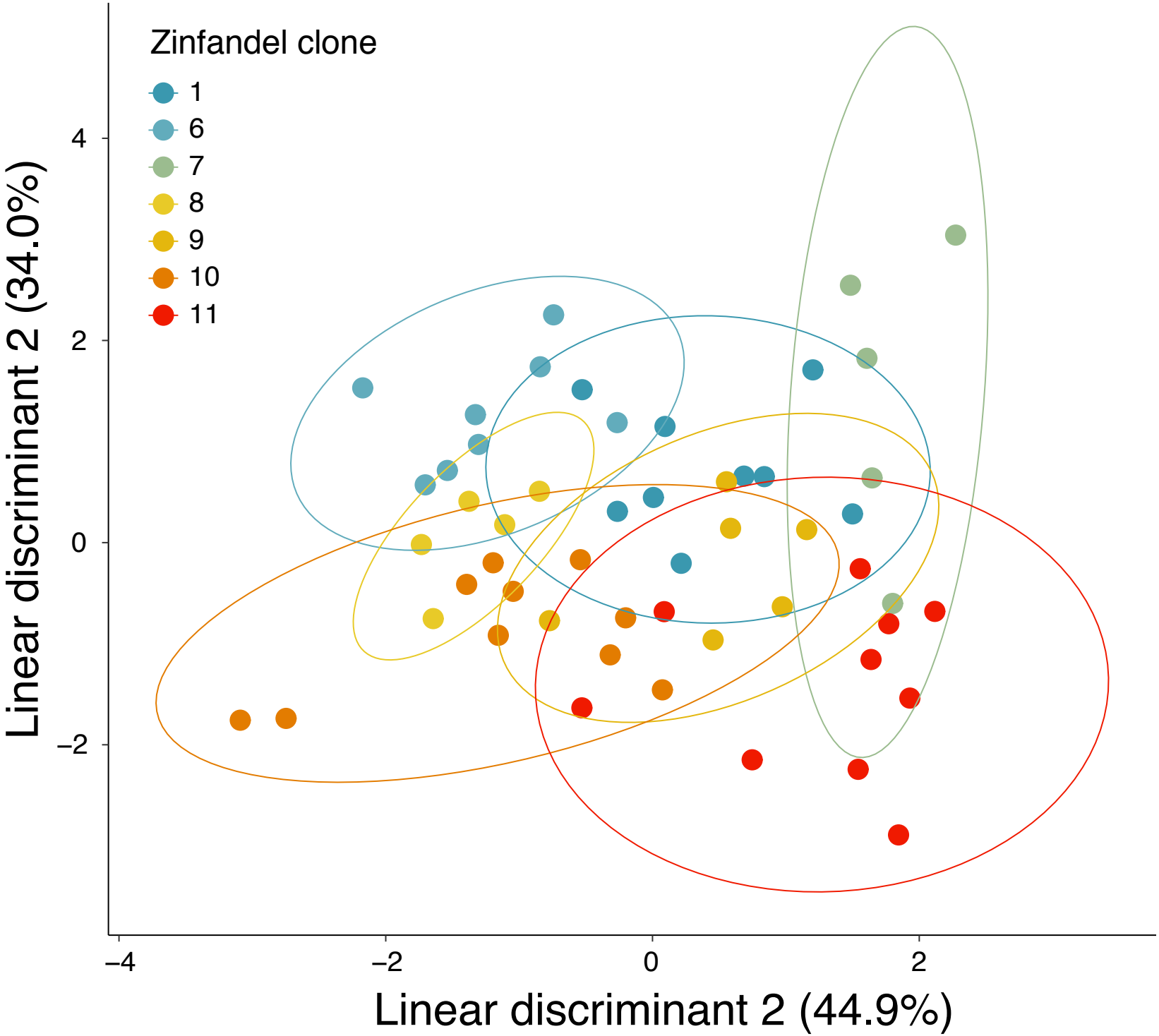
817 annotated in the region are shown; **B.** 4 kbp heterozygous deletion of two genes; **C.** 11 kbp
818 heterozygous deletion of two genes; **D.** 22 kbp inversion that intersects a single gene. Triangles
819 indicate boundaries of the inversion. A gap is shown rather than the center of the inverted region.
820
821 **Figure 3.** Gene content and structural variability between Zin03 and other *V. vinifera* genomes.
822 **A**. Uniquely occurring Zinfandel genes and the number of Zinfandel genes that align well to
823 other cultivars with >=80% identity and reciprocal coverage. The total number of hits (or total
824 gene content for Zin03) is indicated by the "Set Size" and the exclusive hits for each intersection
825 is indicated as the "Intersection Size"; **B,C.** Selected deletions in Cabernet sauvignon relative to
826 Zin03 that intersect genes. For each reported deletion, (from top to bottom) the coverage of reads
827 over the region by long Zinfandel and Cabernet Sauvignon reads, haplotype-resolved alignment
828 of the reads, and the genes annotated in the region are shown; **B.** Two genes are completely
829 deleted in Cabernet Sauvignon relative to Zinfandel and are deleted in one Zinfandel haplotype;
830 **C.** One gene contains a homozygous partial deletion in Cabernet Sauvignon.
831
832 **Figure 4.** The relationships between Zinfandel selections. **A.** Principal component analysis of
833 Zinfandel selections based on SNP data. Zin03 was not included in the analysis; **B.** Zoomed-in
834 view of **A.**, excluding Crljenak kaštelanski 3 and Pribidrag 15; **C.** Kinship analysis of Zinfandel
835 selections and other cultivars with known relationships based on SNP data and outside of annotated
836 repeats. The Kinship coefficient, PHI, is shown, as well as a dendogram constructed by
837 hierarchically clustering genotypes using their kinship coefficients.
838
839 **Figure 5**. Characterization of variants and their frequency among Zinfandel selections and other
840 *vinifera* cultivars (Pinot Noir, Chardonnay, Merlot, Cabernet Franc, and Sauvignon Blanc). The

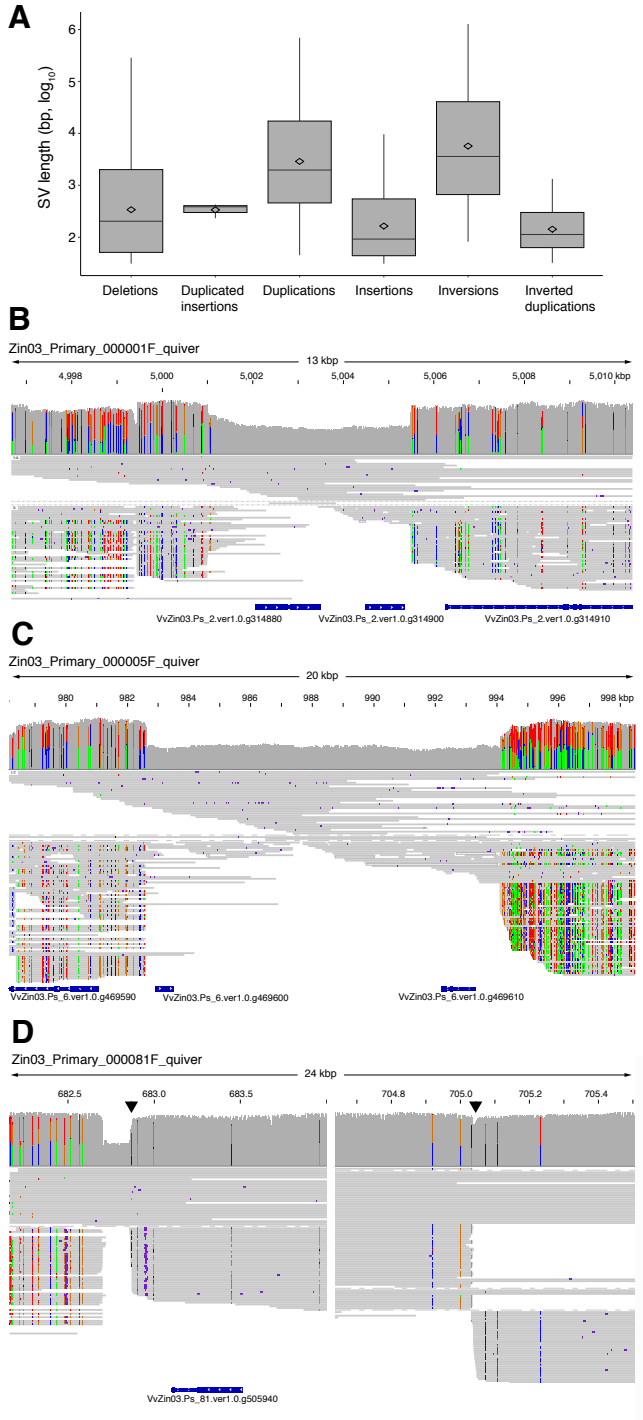
841 normalized rate of variants (number of variants divided by the total feature length in the genome
842 * 1k) by type (SNP, INDEL), feature (Intergenic, Intron, Exon), and genotype (Non-Zinfandel
843 Cultivars, Zinfandel selections). Boxplots show the 25th quartile, median, and 75th quartile.
844
845 **Figure 6**. The abundance and impact of shared and unique heterozygous mutations among
846 Zinfandel clones. **A.** The number of heterozygous SNPs, INDELs, and SVs are shared by N
847 Zinfandel clones; **B.** The number of SNPs and INDELs shared by N clones in exons, introns,
848 intergenic repeats ("Repeats"), and non-repetitive intergenic space; **C.** The ratio of transitions (Tr)
849 to transversions (Tv) for heterozygous SNPs that uniquely occur in single Zinfandel clones and in
850 different genome features. Different letters correspond to significant differences in Tr/Tv rates
851 between features (ANOVA, Tukey HSD, $p < 0.01$); **D.** The percentage of CpG, CHG, and CHH
852 in exons, introns, intergenic repeats ("Repeats"), and non-repetitive intergenic space that
853 experiences transition mutations. Comparisons were made between features for each type of C-
854 repeat separately. Different letters correspond to significant differences (Tukey HSD, $p < 0.01$);
855 **E.** Proportion of exonic SNPs and INDELs that are deleterious and shared by N Zinfandel clones
856
857 **Figure 7**. Transposable element insertions among Zinfandel selections. **A.** Transposable element
858 insertions shared among N Zinfandel selections relative to Zin03 and PN40024; **B.** The
859 proximity of intergenic transposable element insertions to genes; **C.** Types of transposable
860 element insertions shared by N Zinfandel selections.
861
862 ## Additional files
863

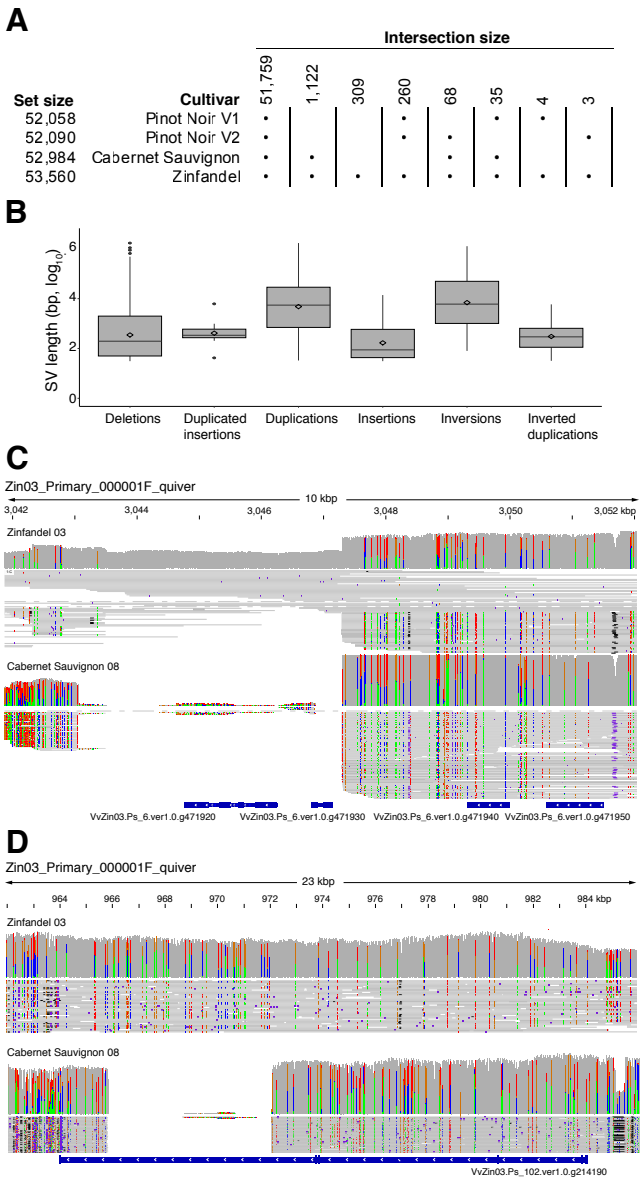
Additional File 1. Text file, .txt ; Settings and data used for Zin03 genome assembly, annotation, and variant calling. Related to FALCON pipeline, RNAseq datasets and experimental data for annotation procedure, Trimmomatic settings, and EVidenceModeler.

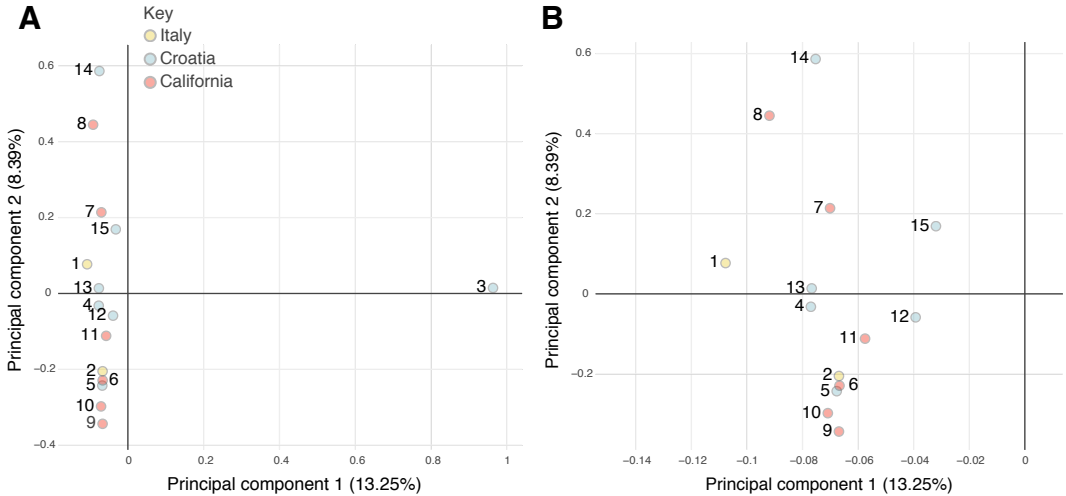
Additional File 2. Excel file, .xlsx ; Unique genes identified in Zinfandel, not identified in Pinot Noir and Cabernet Sauvignon (309), with associated Gene Ontology categories

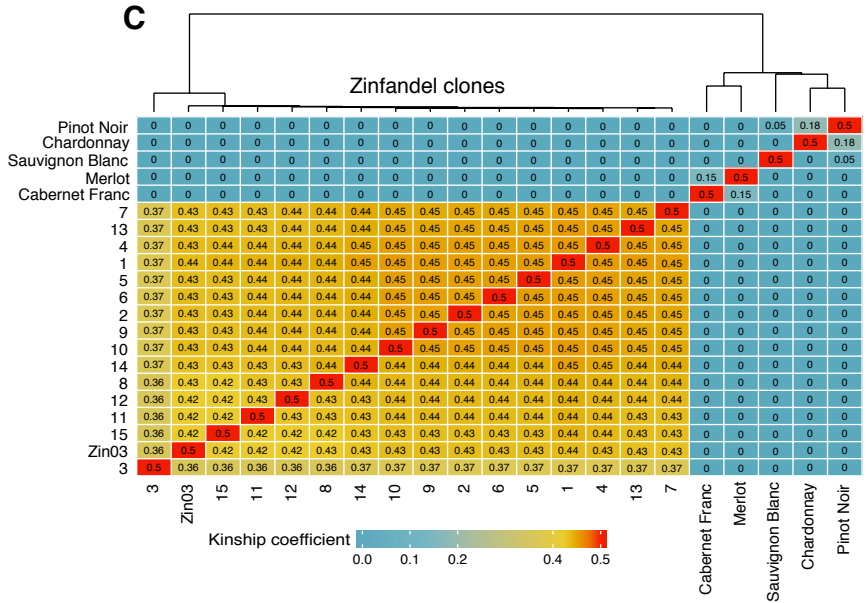
Additional File 3, Excel file, .xlsx ; Table S1, Summary of variants relative to the Zinfandel reference genome; Table S2, SnpEff analysis of variants. Mean values ± SEM are shown. Excludes sites and variants where samples and Zin03 have identical heterozygous genotypes at the locus.

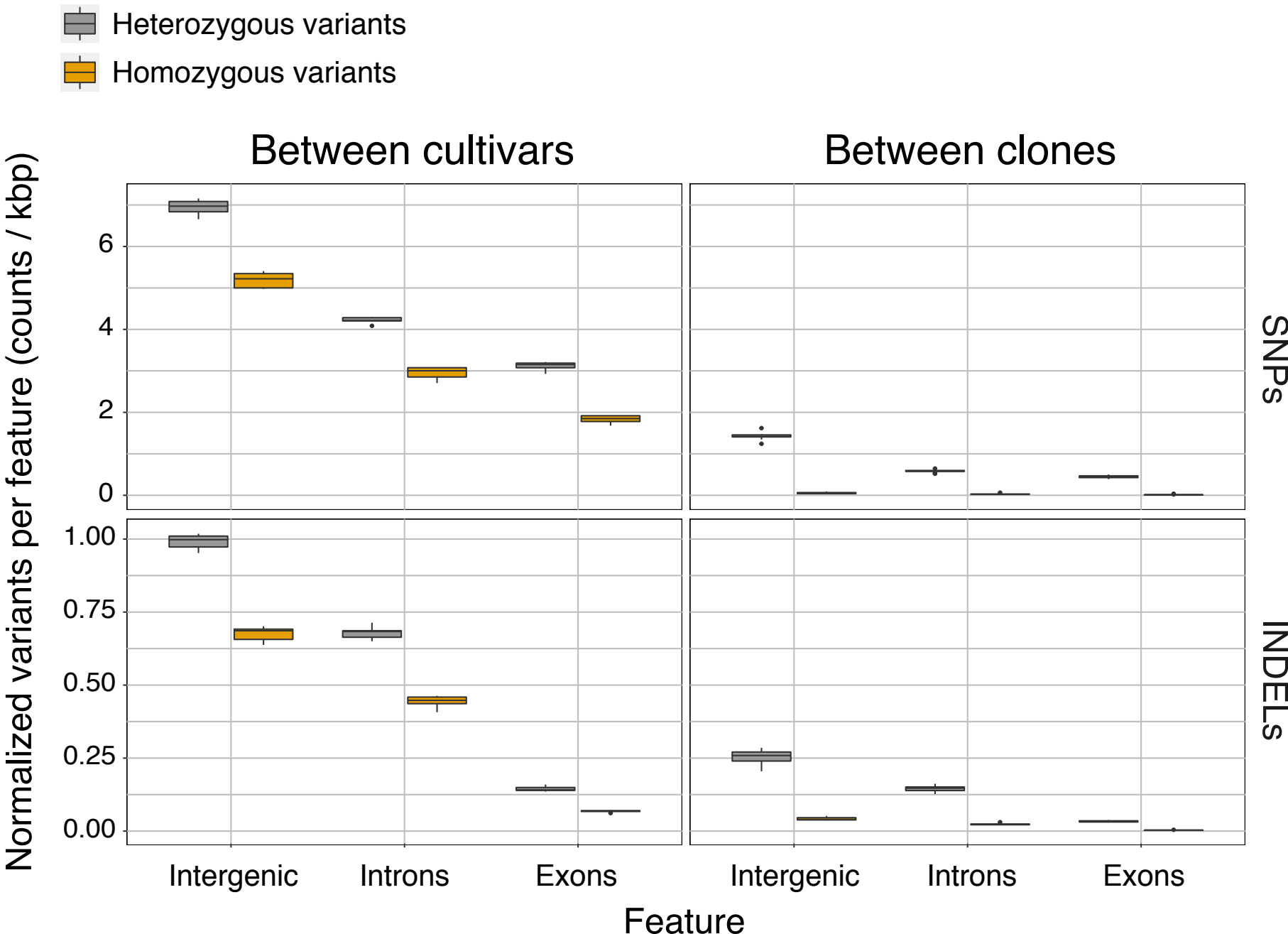

Zinfandel clone
1
6
7
8
9
10
11
4
2
0
−2
−4
−2
0
2
Linear discriminant 2 (34.0%)
Linear discriminant 2 (44.9%)

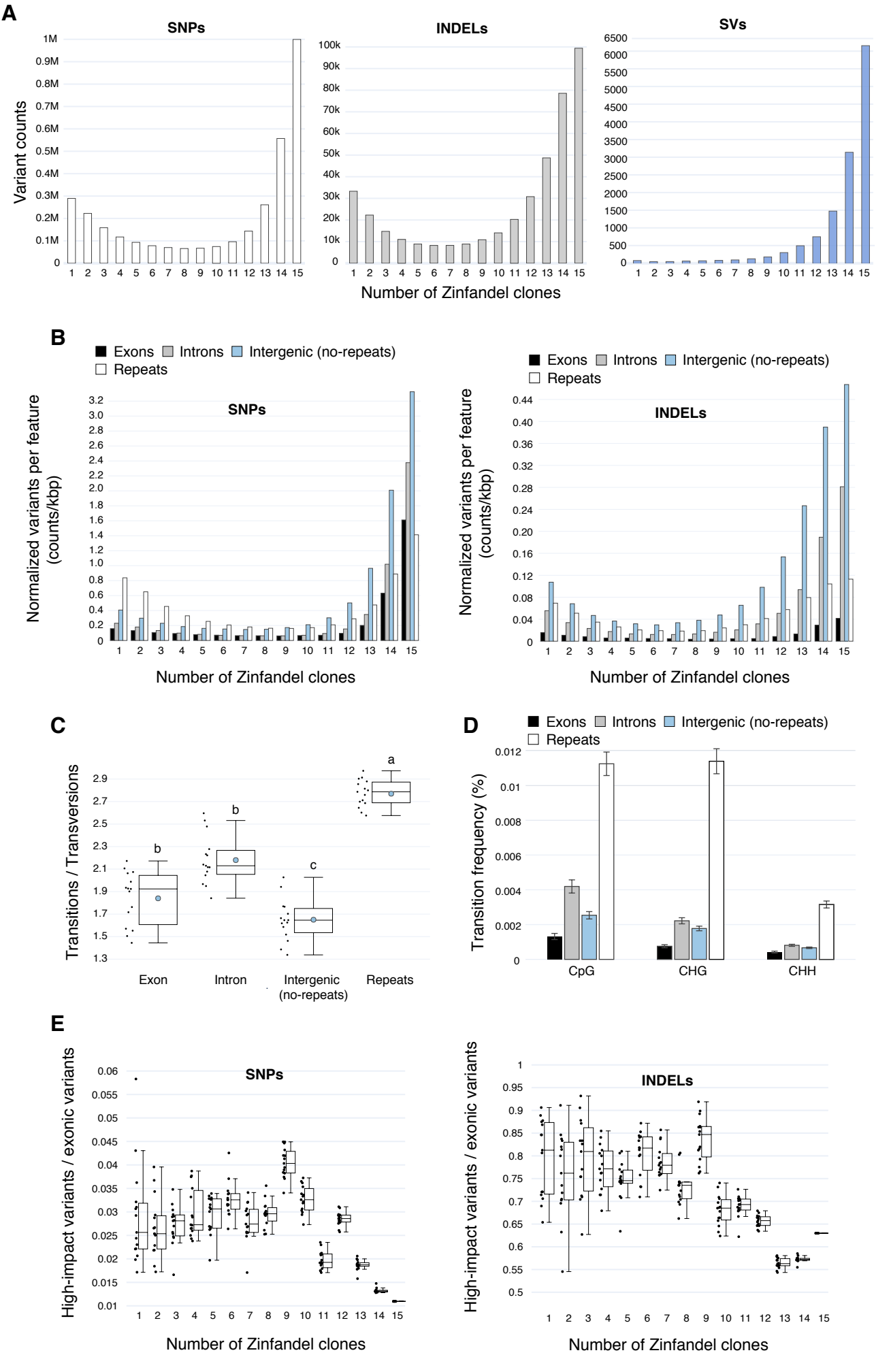

A
SV length (bp, log10)
6
5
4
3
2
Deletions
Duplicated insertions
Duplications
Insertions
Inversions
Inverted duplications
B
Zin03_Primary_000001F_quiver
13 kbp
4,998
5,000
5,002
5,004
5,006
5,008
5,010 kbp
VvZin03.Ps_2.ver1.0.g314880
VvZin03.Ps_2.ver1.0.g314900
VvZin03.Ps_2.ver1.0.g314910
C
Zin03_Primary_000005F_quiver
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

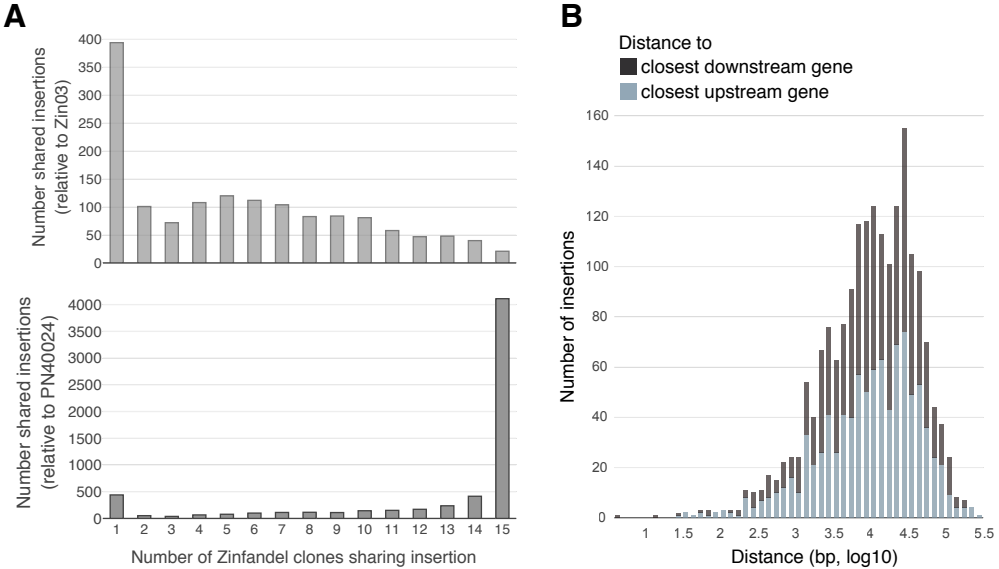

A
Intersection size
Set size
Cultivar
51,759
1,122
309
260
68
35
4
3
52,058 Pinot Noir V1
52,090 Pinot Noir V2
52,984 Cabernet Sauvignon
53,560 Zinfandel
B
SV length (bp, log10)
Deletions
Duplicated insertions
Duplications
Insertions
Inversions
Inverted duplications
C
Zin03_Primary_000001F_quiver
10 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
7
13
4
1
5
6
2
9
10
14
8
12
11
15
Zin03
3
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