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

Amanda M. Vondras¹, Andrea Minio¹, Barbara Blanco-Ulate², Rosa Figueroa-Balderas¹, Michael A. Penn¹, Yongfeng Zhou³, Danelle Seymour³, Ye Zhou¹, Dingren Liang¹, Lucero K. Espinoza¹, Michael M. Anderson¹, M. Andrew Walker¹, Brandon Gaut³, Dario Cantu¹*

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.
Mutations occur in somatic cells that proliferate by mitosis. These can occur by a variety of means, including single base-pair mutations (Ossowski et al., 2010; Hershberg & Petrov, 2010) that are more prevalent in repetitive regions because methylated cytosines are passively deaminated to thymines (Selker, 1990; Mautino & Rosa, 1998; Meunier et al., 2005), polymerase slippage that drives variable microsatellite insertions and deletions (Schlötterer & Tautz, 1992), and larger structural rearrangements and hemizygous deletions (Carbonell-Bejerano et al., 2017; Zhou et al., 2018). Transposable elements are also a major source of somatic mutations in grapevines (Carrier et al., 2012), though transcriptional and post-transcriptional mechanisms exist to prevent transposition and maintain genome stability (Chan et al., 2005; Qi et al., 2006; Cantu et al., 2010; Shen et al., 2012). Notably, methylation of transposable elements is one specific mechanism that prevents transposition, which establishes a tradeoff, then, between methylation and the transposition of mobile elements.

At the cellular level, distinct clones can emerge following a mutation in a shoot apical meristem that spreads throughout a single cell layer, creating periclinal chimeras. This chimera is stable for Pinot Meunier, a clone of Pinot Noir with distinct L1 and L2 layers in shoots (Franks et al., 2002). Each cell layer in a stratified apical meristem like that observed in grape (Thompson & Olmo, 1963) is developmentally distinct. The distinct cell layers will remain so provided cell divisions occur anticlinally. But, periclinal divisions and cellular rearrangements can result in the homogenization of a mutant genotype across cell layers (Hocquigny et al., 2004). This is the case for green-yellow bud sports of the grey-fruited Pinot Gris, wherein sub-epidermal white cells invaded and displaced epidermal pigmented cells (Pelsy et al., 2015). In contrast to replacement (L1 cells invade L2), displacement is likely more common because of the relative disorganization of the inner cell layers (Thompson & Olmo, 1963; Hocquigny et al., 2004).

Meristem architecture is related to the fate of somatic mutations, as it influences the impact of these mutations and the likelihood of competition between cell lineages, also known as diplontic selection (Klekowski et al., 1985; Tilney-Bassett, 1986; Klekowski, 2003). Provided each cellular layer is maintained by anticlinal divisions, deleterious mutations can be preserved in periclinal chimeras (Klekowski et al., 1985; Klekowski, 1998). In addition, the predominance of “hidden”, heterozygous recessive somatic mutations (Klekowski, 1998; Zhou et al., 2017) may further shield somatic mutations from selective forces. These factors are permissive of the accumulation of somatic mutations. Diplontic selection could occur if periclinal cell divisions...
result in the invasion of one cell layer by cells from another (Klekowski et al., 1985; Klekowski, 2003). This mechanism could oppose the accrual of deleterious mutations expected by Muller (Muller, 1932; Pineda-Krch & Fagerström, 1999). A recent study of the long-lived pedunculate oak, *Quercus robur*, described substantial intra-organismal genetic variation, but did not draw conclusions about the contribution of somatic variations to large-scale oak evolution (Plomion et al., 2018). Evidence of diplontic selection in plants is remarkably scarce (Klekowski, 1998), though its likelihood given different circumstances has been modeled (Klekowski et al., 1985; Orive, 2001; Klekowski, 2003). Given the prevalence of chimerism and rearrangements documented in the model (Hocquigny et al., 2004; Pelsy et al., 2015), grapevine is a suitable model for investigating the possibility of selection during vegetative propagation.

Zinfandel is the third-most cultivated wine grape in California (CDFA, 2016a,b) DNA profiling produced evidence that Zinfandel is synonymous with Primitivo grown in Italy (Bowers et al., 1993) and Croatian Pribidrag and Crljena Kastelanski (Maletic et al., 2003). Historical records plus the cultivation of closely related cultivars support Croatia as the likely origin of Zinfandel (Mirošević & Meredith, 2000; Maletic et al., 2003; 2004; Fanizza et al., 2005) and also that Primitivo was likely brought to the Gioia del Colle region in Italy by Benedictine monks in the 17th century (Franks et al., 2002; Russo et al., 2003). The reported variability in Zinfandel (Wolpert, 1996; Fidelibus et al., 2005; Zdunić et al., 2014), including variability in phenolic metabolites (Fig. S1, Methods S1), and its long history of cultivation make it a useful model for studying clonal variation in grapevine, specifically, and the nature of the accumulation of somatic mutations in clonally propagated crops, generally.

The purpose of this study was to better understand the nature of the somatic variations that occur during clonal propagation. Representatives from at least a portion of Zinfandel’s history (Mirošević & Meredith, 2000; Maletic et al., 2003; 2004; Fanizza et al., 2005) from Croatia, Italy, and California were sequenced and compared using Zin03 as reference. First, we show that intergenic space drives clonal diversification. As previously reported, transposable element insertions varied among clones (Carrier et al., 2012). This report expands that understanding to implicate methylation as an indirect driver of clonal diversification; rare somatic heterozygous SNPs were most observed in the repetitive intergenic regions, likely because of the high levels of transposition-inhibiting methylation and associated transition mutations that are prevalent there. Second, the data support an important component of Muller’s ratchet (Muller, 1932), that asexually propagated organisms accumulate deleterious mutations.
Third, somatic mutations were relatively scarce in the coding regions of genes relative to introns and intergenic space, suggesting some degree of negative selection against deleterious mutations.

Materials and Methods

Zinfandel plant material and additional accessions

Sixteen Zinfandel clones were used for this study. Plants were confirmed to be clones of Zinfandel using the following microsatellite markers: VVMD5, VVMD7, VVMD27, VVMD31, VVMD32, VVMS2, VRZAG62, and VRZAG79 (Thomas et al., 1994; Sefc et al., 1999; Maletic et al., 2003). Fourteen of these clones are available through Foundation Plant Services (FPS) at the University of California Davis. Nine of the sixteen clones belong to the Zinfandel Heritage Vineyard Project, a collection of rare Zinfandel vine cuttings grown in the same vineyard. The identification numbers, common names, and source of the clones used in this study are listed in Table 1. An FPS identification number suffix of “.1” indicates that the clone underwent microshoot tip tissue culture therapy, with two exceptions. Pribidrag 13 and Pribidrag 15 are directly derived from the same plants as Pribidrag 4 and Pribidrag 5, respectively, but did not undergo microshoot tip tissue culture therapy. They are labeled with identical FPS numbers to make clear that the relationship between them is known. In this manuscript, Zinfandel clones will be referred to by the clone numbers and common names listed in Table 1.

DNA extraction, library preparation, and sequencing

High quality genomic DNA was isolated from grape leaves using the method described in Chin et al. (2016). DNA purity was evaluated with a Nanodrop 2000 spectrophotometer (Thermo Scientific, Hanover Park, IL), quantity with a Qubit 2.0 Fluorometer (Life Technologies, Carlsbad, CA) and integrity by electrophoresis. For SMRT sequencing, SMRTbell libraries for the Zinfandel reference FPS clone 03 (Zin03) were prepared as described by Chin et al. (2016). For Illumina sequencing, DNA sequencing libraries for each of the sixteen Zinfandel clones were prepared using the Kapa LTP library prep kit (Kapa Biosystems) as described by Jones et al., (2014). Final libraries were evaluated for quantity and quality using a Bioanalyzer 2100 (Agilent Technologies, CA). Zin03 SMRTbell libraries were sequenced on a PacBio RS II and Illumina libraries were sequenced in 100 and 150 base-pair paired-end reads on an Illumina HiSeq3000 sequencer (DNA Technology Core Facility, University of California, Davis). Genome sequences of additional V. vinifera were used in this study, including long reads from Cabernet sauvignon...
NCBI BioProject PRJNA316730 and short reads from Cabernet franc, Chardonnay, Merlot, Pinot Noir, and Sauvignon blanc (NCBI BioProject PRJNA527006).

Zinfandel genome assembly and annotation

*De novo* assembly of Zinfandel (Zin03) was performed at DNAnexus (Mountain View, CA, USA) using PacBio RS II data and the FALCON-unzip (v. 1.7.7) pipeline (Chin et al., 2016). FALCON-unzip was used for its ability to assemble contiguous, phased diploid genomes with better resolved heterozygosity (Chin et al., 2016; Minio et al., 2017). Repetitive sequences were masked prior to error correction using TANmask and REPmask modules in Damasker (Myers, 2014). After error-correction (13,073 bp length cut-off), a total of 1.68 million error-corrected reads (N50 15Kbp, 29-fold coverage of expected genome size) were obtained and repeats were masked before overlap detection in the FALCON pipeline (v. 1.7.7). PacBio reads were assembled after testing multiple parameters to produce the least fragmented assembly. These conditions are listed in Methods S2. Haplotype reconstruction was performed with default parameters. Finally, contigs were polished with Quiver (Pacific Biosciences, bundled with FALCON-unzip v. 1.7.7). Repeats were annotated on the Zin03 assembly using RepeatMasker (v. open-4.0.6; Smit et al. 2013) and a *V. vinifera* repeat library (Minio et al., 2019).

The publicly available RNAseq datasets listed in Methods S2 were used as transcriptional evidence for gene prediction. Each RNAseq sample was trimmed with Trimmomatic (v. 0.36; Methods S2) and assembled with Stringtie (v. 1.3.3; Pertea et al., 2016) to reconstruct variety-specific transcripts. A detailed list of all experimental data used for the annotation procedure is in Methods S2. This data was then mapped on the genome using Exonerate (v. 2.2.0, transcripts and proteins; Slater & Birney, 2005) and PASA (v. 2.1.0, transcripts; Haas et al., 2003). Alignments, and *ab initio* predictions generated with SNAP (v. 2006-07-28; Korf, 2004), Augustus (Stanke et al., 2006), and GeneMark-ES (Lomsadze et al., 2005) were used as input for EVidenceModeler (v. 1.1.1; Haas et al., 2008). EVidenceModeler was used to identify consensus gene structures using the weight reported in Methods S2. Functional annotation was performed using the RefSeq plant protein database (ftp://ftp.ncbi.nlm.nih.gov/refseq, retrieved January 17th, 2017) and InteProScan (v. 5) as previously described (Minio et al., 2019).

Genetic variant calling
Comparisons between Zinfandel clones and between Zin03 and other cultivars were made using the Zin03 genome as reference. This pipeline is described in Methods S3. Small insertions and deletions (INDELs), single nucleotide polymorphisms (SNPs), and structural variations (SVs) were analyzed. The short Illumina reads belonging to the sixteen Zinfandel clones and additional cultivars were trimmed using Trimmomatic (v. 0.36; Methods S2). Quality filtered and trimmed paired-end reads were then randomly down-sampled to 84 million (~14X coverage) in each library to mitigate the possibility of sequencing depth-dependent outcomes.

All libraries were aligned to Zin03 using bwa (v. 0.7.10) and the -M parameter (Li & Durbin, 2009). For all genotypes, the median number of reads mapping to the Zinfandel reference genome was approximately 97%. Next, Picard Tools (v. 2.12.1) were used to mark optical duplicates, build BAM indices, and validate SAM files (http://broadinstitute.github.io/picard). Variants were called using GATK’s HaplotypeCaller (v. 3.5; Van der Auwera et al., 2013). Then, called variants were filtered and annotated (--filterExpression "QD < 2.0 || FS > 60.0 || MQ < 40.0 || MQRankSum < -12.5 || ReadPosRankSum < -8.0"). Variant call files were combined using GATK’s GenotypeGVCFs. Having mapped Illumina reads corresponding to the Zinfandel reference onto itself, erroneous non-reference Zin03 calls (8.1%) were removed. The variants called included SNPs and INDELs.

Next, large structural variations among clones, between Zin03 and other cultivars, and between Zin03’s haplotypes were studied. First, Zin03 genes were compared to PN40024 and Cabernet Sauvignon (CS08) by mapping coding sequences on genome assemblies using Gmap (v. 2015-09-29) and the following parameters: -K 20,000 -B 4 -f 2. Hits with at least 80% identity and reciprocal coverage are reported. Genes annotated on Zin03’s haplotig assembly were also mapped to Zin03’s primary assembly to assess differences in gene content between Zin03’s haplotypes. SMRT reads from Zin03 and CS08 were mapped to Zin03 using NGMLR (v. 0.2.7) and structural differences were called with Sniffles (v.1.0.8; Sedlazeck et al., 2018).

Zinfandel clones were compared to one another using Illumina short reads and Delly (v. 0.7.8) with default parameters (Rausch et al., 2012). The structural variations identified by Sniffles and Delly in Zin03 were intersected. Several filters were applied to the results of SV analyses. Transversions, non-reference Zin03 genotype calls, SVs annotated at the ends of contigs, and SVs that intersected the repeat annotation were filtered from Delly output.

Transposon insertion analysis
PoPoolationTE2 (v. 1.10.04; Kofler et al., 2016) was used to identify transposon insertions in the Zinfandel clones; it was used following the workflow outlined in its software manual (https://sourceforge.net/p/popoolation-te2/wiki/Manual/). Insertions were called relative to Zin03 genome assembly and PN20024 (Jaillon et al., 2007). As described in Kofler et al. (2016), PoPoolationTE2 analyses transposable element insertions and can identify novel and annotated TE insertions provided insertions fall within predefined families of TEs. The annotation produced by RepeatMasker was used for the analysis. In this manuscript, the TE insertions among the clones are reported using the classification system and nomenclature described by Wicker et al. (2007). In instances where the TE order and/or superfamily was not annotated, only the TE class and order, when available, are named in the associated figures and text.

Relationships between Zinfandel clones

The relationships between Zinfandel clones were visualized by Principal Component Analysis and their relatedness was quantified (VCFtools v. 0.1.15) based on the method described by Manichaikul et al. (2010). This approach gives information about the relationship of any pair of individuals (unrelated, 3rd degree relative, 2nd degree relative, full siblings, and self) by estimating their kinship coefficient, which ranges from zero (no relationship) to 0.50 (self). These analyses used SNPs outside of repetitive regions.

Results

Zinfandel genome assembly, annotation, and differences between haplotypes

The clone used for the genome assembly, Zinfandel 03 (Zin03), was acquired by FPS in 1964 from the Reutz Vineyard near Livermore, California that was planted during Prohibition (1920 – 1933; Sweet & Wolpert, 2007). Zin03 was sequenced using Single Molecule Real-Time (SMRT; Pacific Biosciences) technology at ~98x coverage and assembled using FALCON-unzip (Chin et al., 2016), a diploid-aware assembly pipeline. The genome was assembled into 1,509 primary contigs (N50 = 1.1 Mbp) for a total assembly size of 591 Mbp, similar to the genome size of Cabernet Sauvignon (590 Mbp; Chin et al., 2016) and larger than Chardonnay (490Mb; Roach et al., 2018) and PN40024 (487 Mb; Jaillon et al., 2007). Fifty two percent of the genome was phased into 2,246 additional phased sequences (haplotigs) where the homologous chromosomes were distinguishable with an N50 of ~442 kbp (Table 2). A total of 53,560
complete protein-coding genes were annotated on the primary (33,523 genes) and haplotig (20,037 genes) assemblies (Table 2).

Of the 20,037 genes annotated on the haplotig assembly, 18,878 aligned to the primary assembly, leaving 1,159 genes that may exist hemizygosly in the genome due to structural variation between homologous chromosomes or because of substantial divergence in sequence between haplotypes. These genes were annotated with a broad variety of putative functions, including biosynthetic processes, secondary metabolism, and stress responses. Long reads were mapped to both the primary and haplotig assemblies to evaluate the circumstances that explain the differences between haplotypes. Structural variants (SVs) between the haplotypes were examined by mapping long SMRT sequencing reads onto Zin03’s primary and haplotig assemblies with NGMLR and calling SVs with Sniffles (Sedlazeck et al., 2018).

A total of 22,399 SVs accounted for 6.94% (41.0 / 591 Mbp) of the primary assembly’s length and 6.02% (8.4 / 139 Mbp) of the primary assembly’s gene-associated length (Fig. 1a, Table 3). SVs intersected 4,559 genes in the primary assembly (13.6% of primary assembly genes) and 390 SVs spanned more than one gene. Manual inspection of the long reads aligned to the primary assembly support that large, heterozygous deletions and inversions occurred in the Zin03 genome that were either inherited from different structurally distinct parents or arose during clonal propagation (Fig. 1b,c,d). Importantly, there was substantial hemizygosity in the genome, with long reads supporting deletions affecting 2,521 genes and 4.56% of the primary assembly’s length (Table 3).

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

These results reveal structural differences between Zinfandel’s haplotypes. These differences could have been inherited and/or could have occurred during clonal propagation. Overall, these structural variations affected 4,559 primary assembly genes. Importantly, these data show that a notable portion of the primary assembly’s length (4.56%) is hemizygous.
Differences in structure and gene content between Zinfandel and other grape genomes

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

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

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

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

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

Across the Zinfandel clones, the median number of homozygous and heterozygous variants called relative to Zin03 were 38,092 and 717,925, respectively. Between 10-fold and ~27-fold more heterozygous variants were called than homozygous variants in each clone except for Crljenak kaštelanski 3, for which only ~2.5-fold more heterozygous sites were called (Table S2). Crljenak kaštelanski 3 had 4.3-fold more homozygous variants and 1.8-fold fewer heterozygous variants than the other clones (Table S2). Furthermore, unlike other clones, for which less than 10% of sites did not share the Zin03 reference allele, ~29% of variant sites were called where Crljenak kaštelanski 3 did not share the Zin03 reference allele (Table S2). Together,
these results suggest that unlike other Crljenak kaštelanski selections (Maletic et al., 2003; Fanizza et al., 2005; Zdunić et al., 2014), Crljenak kaštelanski 3 is likely a close relative of, but quite possibly not a clone of, Zinfandel. Because these analyses cast doubt on its identity as a Zinfandel clone, Crljenak kaštelanski 3 was excluded from the clonal variation analyses described in the rest of the study. Fifteen clones, including Zin03, remained in the analyses (Table 1).

Clonal versus cultivar genetic variability

Overall, an average of 761,948 variant sites were identified in individual Zinfandel clones when short reads were mapped on the Zin03 primary assembly. As stated previously, this analysis excluded Crljenak kaštelanski 3. On average, 6,153,830 variant sites were identified in other cultivars (Pinot noir, Chardonnay, Sauvignon Blanc, Merlot, Cabernet Franc) relative to Zin03 (Table S2). Both of these figures excluded heterozygous sites at which the diploid genotype called for a given sample was identical to that called for Zin03.

Variants were 7.9X more frequent in other cultivars relative to Zin03 than for Zinfandel clones; on average, mutations in clones occurred once every 723 bases and once every 92 bases in other cultivars (Table S2). However, the ratio of transitions to transversion mutations and the proportions of the severities of the predicted variant effects were similar for both groups (Table S2). The normalized count of variants differed between cultivars and Zinfandel clones on the basis of variants’ location in the genome, the type of variant, and the zygosity of the variant (Fig. 4).

Variants in non-Zinfandel cultivars and heterozygous variants among Zinfandel clones were significantly more prevalent in intergenic space than introns and exons and significantly more prevalent in introns than exons (Tukey HSD, p < 0.01). Unlike homozygous variants between cultivars and as expected, homozygous variants were rare among clones (Fig. 4, Table S2). Still, the normalized count of homozygous INDELs in intergenic space, introns, and exons were significantly different among Zinfandel clones (Tukey HSD, p < 0.01), as were the normalized count of homozygous intergenic versus genic (exons and introns) SNPs (Tukey HSD, p < 0.01). The normalized count of homozygous SNPs in exons and introns were not significantly different in Zinfandel clones (Tukey HSD, p > 0.01). The accrual of predominantly heterozygous and likely recessive variants (Zhou et al., 2017) is consistent with what would be expected given physically separate homologous chromosomes and the absence of sexual...
reproduction. The differences in mutation abundances observed were initially surprising; if somatic mutations occurred randomly and absent mechanisms that make certain sites more or less susceptible to mutation, then different regions of the genome should have had equal levels of mutations. This was not the case (Figure 5).

The accrual of somatic mutations in Zinfandel clones

Heterozygous sites found among the 15 Zinfandel clones ought to be a mixture of sites inherited from their shared ancestral plant and somatic mutations that arose during clonal propagation. To better understand the nature of somatic mutations, the data were handled slightly differently than they were to construct Figure 5; all 15 Zinfandel clones were included (including Zin03, Crljenak kaštelanski 3 still excluded) and all heterozygous calls were considered, even if all genotypes were identically heterozygous. Thirty percent of heterozygous SNPs, 24% of heterozygous INDELs, and 47% of heterozygous structural positions were shared by all 15 Zinfandel clones (Fig. 5a). Because all clones are identically heterozygous at these loci, these variants are those inherited from Zinfandel’s parents.

Individual and subsets of Zinfandel clones accumulated heterozygous mutations as clonal propagation occurred (Fig. 5a). Thirteen percent and 16% of heterozygous INDELs and SNPs, respectively, and 1% of large (>50 bp) structural variants occurred in only one or two clones (Fig. 5a). The distribution of SVs called by Delly is markedly different than those of SNPs and INDELs (Fig. 5a). For both SNPs and INDELs, there were 3 and 3.5-fold as many heterozygous variants shared by all 15 clones as there were uniquely occurring variants; there were 71.5-fold more structural variants shared by all clones than there were unique variants in individual clones (Fig. 5a). This might imply that the mechanisms that give rise to small mutations are more common among clones than the large-scale changes associated with SVs.

The distribution of unique and shared heterozygous INDELs in exons, introns, repetitive, and non-repetitive intergenic spaces were not equal (Fig. 5b). The distribution of INDELs in exons was significantly different than the distributions of INDELs in each other feature considered (Kolmogorov-Smirnov Test, \( p < 0.01 \)). Similarly, the distributions SNPs in genic (exons, introns) and intergenic (repetitive, non-repetitive) regions were not equal (Fig. 5b). Shared heterozygous SNPs were most common in intergenic non-repetitive regions and introns and least common in exons and repetitive intergenic regions (Fig. 5b). Interestingly, unique heterozygous SNPs occurred at high rates in repetitive intergenic regions (Fig. 5b).
That shared heterozygous sites are mostly in non-repetitive intergenic space and unique heterozygous sites are mostly in repetitive space may have to do with the disposition of methylated cytosines to spontaneously deaminate and the prevalence of methylated repetitive sequences in those regions (Meunier et al., 2005; Cantu et al., 2010; Ossowski et al., 2010; Shen et al., 2012). This is also supported by the significantly higher ratio of transitions to transversions in repetitive intergenic regions than in exons, introns, and non-repetitive intergenic space (Fig. 5c). Furthermore, the mean percentage of CpG, CHG, and CHH sites affected by transition mutations was significantly higher in repetitive intergenic space than genic and non-repetitive intergenic spaces (Fig. 5d; Tukey HSD, $p < 0.01$). The mean percentage of CpG sites affected by transition mutations was also significantly higher in introns than exons (Tukey HSD, $p < 0.01$).

Compatible with this hypothesis, INDELs, which should not increase in frequency due to methylation, did not occur preferentially in repeats (Fig. 5b).

The impact of specific variants also varied with their prevalence among the clones (Fig. 5e). “High impact” mutations were predicted by SNPEff (Cingolani et al., 2012). The high impact mutations identified in these data included exon losses, start and stop site gains and losses, frameshifts, gene fusions, splice acceptor mutations, and splice donor mutations. These mutations are predicted to be deleterious because of their disruptive effects on the coded protein. For these reasons, we designated such mutations as putatively deleterious in this manuscript. These were counted for each Zinfandel clone relative to Zin03. Relatively low percentages of heterozygous variants shared by all Zinfandel clones were putatively deleterious. In contrast, larger proportions of exonic SNPs and INDELs that occurred in individual or subsets of clones were putatively deleterious (Fig. 5e).

Together, these results show that mutations associated with clonal propagation are most numerous outside of coding regions of the genome, indicating that clone genomes diversify most rapidly in the intergenic space, particularly in repetitive and likely methylated regions (Fig. 5). Though a minority of somatic mutations occurred in exons, we show that exonic mutations that occur in few or individual clones are more often deleterious than exonic heterozygous variants shared by all or most clones. In other words, clonal propagation is associated with the accumulation of putatively deleterious heterozygous mutations.

Zinfandel clones incur unique transposon insertions
Transposable element insertions (TEI) contribute to somatic variation in grape (Kobayashi et al., 2004; Fernandez et al., 2010; Carrier et al., 2012; Fernandez et al., 2013).

Relative to Zin03, 1,473 TEI were identified among the Zinfandel clones. A large fraction of TEI (26.7%) occurred uniquely in individual clones (Fig. 6a) and included 325 retrotransposons, mostly Copia and Gypsy LTRs, and 69 DNA-transposons (Fig. 6b). Because uniform loci are excluded, in-common TEI were not captured when clones were compared to Zin03. Comparing the clones relative to PN40024, however, revealed that the majority (64.8%) of TEI were shared among the 15 Zinfandel clones. Five hundred thirty TEI occurred in only one, two or three clones (Fig. 6a). This result supports the derivation of these selections from a common ancestral plant and the accumulation of somatic variations over time.

In addition to being suggestive of their shared heritage, the positions of these insertions and their proximity to coding genes were notable. Three-hundred forty-seven TEI occurred within 314 coding genes. The remaining 938 TEIs were in intergenic regions (Fig. 6c). The median upstream and downstream distance of intergenic TEs from the closest feature were 11,811 and 11,279 base-pairs, respectively, and 25% of TEI were less than 4,345 bases downstream of the closest feature and/or less than 3,826 bases upstream of the closest feature (Fig. 6c).

Discussion

Consideration of the genomic differences among Zinfandel clones revealed what is likely a complex history not easily reconstructed, and that one selection was probably not a clone despite being identified as such by SSR markers. Analyses of the relationships between clones did not reveal groupings of clones per their recorded countries of origin. Somatic mutations may help identify individual clones but could also blur the historical relationships between them. It is also plausible that pairs of clones from any given region are not direct cuttings of one another but of Zinfandels from another region; the clones now grown in California, for example, may have been imported on numerous independent occasions from various other regions, meaning some may indeed be more closely related to one of the Primitivo or Croatian clones than they are to other Californian clones. It would be unwise to assume a single migratory path radiating from an ancestral mother plant ought to be applicable to the clones.

Despite this ambiguity, the examination of SNPs, INDELs, transposable elements and other structural variants all support the derivation of all but one of the clonal selections from a
common ancestral Zinfandel mother plant and show the accumulation of somatic mutations over time (Figs. 6 and 7). The structure of the Zinfandel genome, location of mutations among clones, their frequency and prevalence, and the relationship between these factors provides some insight into the nature of mutations in clonally propagated plants. Mutations among clones were predominantly heterozygous (Fig. 4) and uncommon heterozygous mutations shared by a subset of or individual clones were increasingly deleterious when they occurred in exons (Fig. 5e).

There are costs and benefits associated with clonal propagation (McKey et al., 2010). Among the benefits are that the plants need not breed true-to-type; clonal propagation generally fixes heterozygous loci and valuable phenotypes. However, the increase in the proportion of deleterious alleles supports Muller’s ratchet, which posits that sex is advantageous and that clonal propagation increases mutational load (Muller, 1932). Though these and previous data do not tell which mutations are actually recessive or dominant, they could remain hidden if they are recessive or do not manifest their deleterious effects (Zhou et al., 2017; Gaut et al., 2018). However, even after taking into consideration the total length of exons, introns, and intergenic space (repetitive and non-repetitive), heterozygous mutations occurred at varying frequency in these regions and were least abundant in coding regions. The rarity of mutations in exons and commonality of mutations in repetitive intergenic space may have at least two components.

Mutations are likely more frequent in repetitive intergenic space as a result of the regulation of transposition by DNA methylation. Repetitive intergenic space had the highest rate of relatively unique SNPs and the ratio of transitions to transversions was significantly higher there than in other regions. DNA methylation is an important epigenetic control and is one mechanism that maintains genome stability and impairs the transposition of mobile elements (Chen et al., 1998; Hirochika et al., 2000; Shen et al., 2012). Methylated cytosines, however, spontaneously deaminate faster than unmethylated cytosines (Selker, 1990; Cantu et al., 2010). Together, the expectations that intergenic regions are rich in transposable elements, that these regions are typically highly methylated and as a result will experience greater transition rates account for the high rates of SNPs in repetitive intergenic spaces among Zinfandel clones. Also notable, these data show that some transposable elements are not entirely silenced, with a substantial number inserting in genes or in close proximity to genes (Fig. 6b). These insertions could be effectively inconsequential or not; transposable element insertions can result in novel transcripts and affect gene expression regulation (Fernandez et al., 2010; Hirsch & Springer, 2017). Gene body methylation is appreciated as a mutagenic “double-edged sword” (Zemach et
(al., 2010), with benefits coming at the price. Recent work observed region-specific methylation in vegetatively propagated Sardinian white poplar that may serve an advantageous function (Guarino et al., 2015) and others have suggested that the epigenome contributes to the success of vegetatively propagated plants (Douhovnikoff & Dodd, 2014). Future work might also consider the long-term price associated with intergenic mutagenesis and the potential loss of methylation in vegetatively propagated plants.

The rarity of exonic mutations was surprising. After accounting for the length of these spaces in the genome and their repetitiveness, we expected uniform rates of mutation in exons, introns, and intergenic space. Instead, we observed that although rare somatic mutations in exons were increasingly deleterious, they were relatively scarce. Some degree of negative selection against deleterious variants in coding regions could explain why mutations were less frequent in coding than noncoding regions of the genome. The possibility of diplontic, clonal selection or competition between cell lineages that could purge otherwise consequential deleterious mutations has been modeled, but evidence of its occurrence is sparse (Pineda-Krch & Fagerström, 1999; Klekowski, 2003; McKey et al., 2010). The structures of apical meristems (Klekowski & Kazarinova-Fukshansky, 1984; Klekowski et al., 1985) and the tendency of somatic mutations to be heterozygous and recessive (Zhou et al., 2017) place constraints on the likelihood that deleterious mutations would be subjected to negative selection. Periclinal divisions across cell layers could enhance diplontic selection (Klekowski, 2003) against dominant and/or hemizygous recessive alleles. Four and one half of Zinfandel’s genome is hemizygous; structural variations identified within the Zinfandel genome and the rampant hemizygosity reported in Chardonnay (Zhou et al., 2018) could also expose otherwise hidden somatic variations to selective pressure hostile to the accumulation of deleterious mutations. Additional work should explore to what degree each of these factors, or others not considered here, explain why somatic mutations in exons were relatively infrequent and characterize the realized long-term consequences of mutation accumulation versus selection for grapevine and other clonally propagated plants.

Acknowledgements

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

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

References

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

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

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

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

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

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

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

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

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

Chan SW-L, Henderson IR, Jacobsen SE. 2005. Gardening the genome: DNA methylation in Arabidopsis thaliana. *Nature Reviews Genetics* 6: 351–360.
Chen RZ, Pettersson U, Beard C, Jackson-Grusby L, Jaenisch R. 1998. DNA hypomethylation leads to elevated mutation rates. Nature 395: 89–93.

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

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

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

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

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

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

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

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

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

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

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

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

Haas BJ, Delcher AL, Mount SM, Wortman JR, Smith RK Jr, Hannick LI, Maiti R, Ronning CM, Rusch DB, Town CD, et al. 2003. Improving the Arabidopsis genome annotation using maximal transcript alignment assemblies. Nucleic Acids Research 31: 5654–5666.
Haas BJ, Salzberg SL, Zhu W, Pertea M, Allen JE, Orvis J, White O, Buell CR, Wortman JR. 2008. Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. *Genome biology* 9.

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

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

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

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

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

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

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

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

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

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

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

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

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

Korf I. 2004. Gene finding in novel genomes. *BMC Bioinformatics* doi: 10.1186/1471-2105-5-9
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, Mirošević 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.
Oribe 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.
Sefc KM, Regner F, Turetschek E, Glössl J, Steinkellner H. 1999. Identification of microsatellite sequences in Vitis riparia and their applicability for genotyping of different Vitis species. *Genome* 42: 367–373.

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

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

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

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

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

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

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

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

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

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

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

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

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

Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, Chalhoub B, Flavell A, Leroy P, Morgante M, Panaud O, et al. 2007. A unified classification system for eukaryotic transposable elements. *Nature Reviews Genetics* 8: 973–982.
Wolpert JA. 1996. Performance of Zinfandel and Primitivo Clones in a Warm Climate. *American Journal of Enology and Viticulture* 47: 124–126.

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

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

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

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

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

Figure legends

**Figure 1.** Structural variation between Zin03 haplotypes. (a) Distribution of structural variation sizes. Boxplots show the 25th quartile, median, and 75th quartile for each type of SV. Whiskers are 1.5*Inter-Quartile Range*. Diamonds indicate the mean log10(length) of each type of SV; b,c,d. Examples of heterozygous structural variants between haplotypes that intersect genes. For each reported deletion, (from top to bottom) the coverage, haplotype-resolved alignment of reads, and the genes annotated in the region are shown; b. 4 kbp heterozygous deletion of two genes; c. 11 kbp heterozygous deletion of two genes; d. 22 kbp inversion that intersects a single gene. Triangles indicate boundaries of the inversion. A gap is shown rather than the center of the inverted region.

**Figure 2.** Gene content and structural variability between Zin03 and other *V. vinifera* genomes. (a) Uniquely occurring Zinfandel genes and the number of Zinfandel genes that align well to other cultivars with >=80% identity and reciprocal coverage. The total number of hits (or total 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 Crljnak 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.
Figure 6. Transposable element insertions among Zinfandel selections. (a) Transposable element insertions shared among N Zinfandel selections relative to Zin03 and PN40024; (b) The proximity of intergenic transposable element insertions to genes; (c) Types of transposable element insertions shared by N Zinfandel selections.

Supporting information

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

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

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

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

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

Methods S3. Bioinformatic pipeline for SNP, INDEL, and SV calling.
| 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 | Zinfandel FPS 10 |
| 15      | Pribidrag        | Svinšće, Croatia      | Zinfandel FPS 44.1        |
| Zin03   | Zinfandel        | California, USA       | Zinfandel FPS 03          |
|                          | 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       |

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

|                          | 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      |
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    |
**Intersection size**

| Set size | Cultivar       | Intersection size |
|----------|----------------|-------------------|
| 52,058   | Pinot Noir V1  | 51,759            |
| 52,090   | Pinot Noir V2  | 1,122             |
| 52,984   | Cabernet Sauvignon | 309           |
| 53,560   | Zinfandel      | 260               |

**SV length (bp, log10)**

- Deletions
- Duplicated insertions
- Duplications
- Insertions
- Inversions
- Inverted duplications

**C**

**Zin03_Primary_000001F_quiver**

- Zinfandel 03
- Cabernet Sauvignon 08

**d**

**Zin03_Primary_000001F_quiver**

- Zinfandel 03
- Cabernet Sauvignon 08
The copyright holder for this preprint. http://dx.doi.org/10.1101/585869

doi: bioRxiv preprint first posted online Mar. 22, 2019;
