Catastrophic unbalanced genome rearrangements cause somatic loss of berry color in grapevine

Pablo Carbonell-Bejerano[1*¶] Carolina Royo[1¶], Rafael Torres-Pérez[1¶], Jérôme Grimplet[1], Lucie Fernandez[2], José Manuel Franco-Zorrilla[3], Diego Lijavetzky[4], Elisa Baroja[1], Juana Martínez[1], Enrique García-Escudero[1], Javier Ibáñez[1], and José Miguel Martínez-Zapater[1]

[1] Instituto de Ciencias de la Vid y del Vino, CSIC-Universidad de La Rioja-Gobierno de La Rioja, Logroño, Spain

[2] UMR Biologie du Fruit et Pathologie, INRA, Villenave d'Ornon, France

[3] Centro Nacional de Biotecnología, CNB-CSIC, Madrid, Spain

[4] Instituto de Biología Agrícola de Mendoza, CONICET-UNCuyo-FCA, Chacras de Coria, Argentina.

[*] Corresponding author

E-mail: pablo.carbonell@icvv.es (PCB)

[¶] These authors contributed equally to this work.

Abstract (249/250 words)

Somatic mutation is a source of diversity in vegetatively propagated plants. Grape color somatic variants that can be used to develop new grapevine cultivars occasionally appear associated to genome structural variation (SV). To understand the mutational mechanisms generating somatic SV in grapevine, we compared the Tempranillo Blanco (TB) white berry somatic variant to its black berry ancestor, Tempranillo Tinto. Whole-genome sequencing uncovered a catastrophic SV in TB that caused the hemizygous deletion of 313 genes including the loss of the grape color locus functional allele. Loss of heterozygosity and decreased copy number delimited interspersed monosomic and disomic chromosome regions in the right arm of linkage groups 2 and 5 of TB. At least eleven validated clustered breakpoints involving intra- and inter-chromosomal translocations between three chromosomes flanked the deleted fragments, which are phased in a single copy of the affected chromosomes. These hallmarks along with the lack of homology between breakpoint joins, the lack of duplicated regions and the randomness of rearranged fragments order and orientation are all consistent with a chromothripsis-like pattern generated after breakage and illegitimate rejoining. This unbalanced genome reshuffling has additional consequences in reproductive development. In TB, lack of sexual transmission of rearranged chromosomes associated with low gamete viability, which compromised fruit set and decreased fruit production. Our findings show that chromothripsis-like patterns spontaneously arise and stabilize during plant somatic growth. Despite these dramatic changes may compromise sexual fitness, they also generate new interesting plant features that can be perpetuated through vegetative propagation in woody crops.

Introduction

Somatic variation is a major source of diversity in multicellular organisms and in humans it is frequently associated with the origin of acquired genetic diseases, including cancer (Malhotra et al., 2013; Poduri et al., 2013; Yang et al., 2013). Somatic variation is generated by unique mutation events affecting single dividing cells and can spread by cell division to generate mutant sectors or cell lines derived from the original mutant cell. Somatic mutations range from single nucleotide variation (SNV) and insertion-deletions (INDELs) to complex genome structural variation (SV) originated from chromosomal rearrangements. The rapidly growing number of medical genomic studies is unveiling highly complex forms of SV that are collectively known as chromoanagenesis. Among them, the most commonly reported case is chromothripsis (for chromosome shattering), a cellular catastrophe recently described in human cancer that affects a few chromosomes (1-4) and results in multiple clustered rearrangements (≥5 breakpoints) and deletions (Stephens et al., 2011; Collins et al., 2017). Chromothripsis has been reported in mammals but it has never been related with natural somatic variation in the plant kingdom.

Given their capacity for asexual propagation, plants may be better systems than animals to study somatic variation and its effects on somatic cells (Whitham and Slobodchikoff, 1981). Plant species lack a separated germ line and somatic genetic variation represents an important source of phenotypic variation that may be adaptive and transmitted to the next sexual generation when present in the meristem cell layer giving rise to gametes (D'Amato, 1997). Somatic variation is especially relevant in plants such as long living trees or vegetatively propagated species in which single genotypes are perpetuated for very long periods and can colonize vast extensions, which increase the likelihood of somatic mutation emergence. This case is applicable

to woody crops such as grapevine (*Vitis vinifera* L.), from which cultivar genotypes are asexually propagated by cuttings and perpetuated from the original sexual seedling that in many cases germinated several centuries ago (This et al., 2006). Somatic mutations generating new interesting phenotypes that are stabilized in grapevine plants as periclinal chimeras or that extend to all cell layers have been selected as new clones of wine grape cultivars or as new derivative cultivars (This et al., 2006; Pelsy et al., 2010; Torregrosa et al., 2011). Recently, genetic and genomic approaches identified single nucleotide variation (SNV) and new transposable element (TE) insertions responsible for emergent phenotypes in several grapevine somatic variants (Boss and Thomas, 2002; Fernandez et al., 2010; Battilana et al., 2011; Fernandez et al., 2013; Emanuelli et al., 2014).

Given their conspicuousness, berry color somatic variants have occasionally been selected and studied in grapevine (Kobayashi et al., 2004; Walker et al., 2006; Yakushiji et al., 2006; Furiya et al., 2009). Grape color results from the accumulation of anthocyanins generally in the berry skin (Boss et al., 1996), which is genetically controlled by a major locus on linkage group (LG) 2 (Doligez et al., 2006). This locus co-localizes with a cluster of tandemly repeated *VviMybA* genes (Wong et al., 2016). The presence of a *Gret1* retrotransposon insertion in the promoter of *VviMybA1* along with a small INDEL causing a frame-shift in *VviMybA2* are responsible for a null allele of the color locus. White berry cultivars are frequently homozygous for this null allele (Kobayashi et al., 2004; Lijavetzky et al., 2006; This et al., 2007; Walker et al., 2007). Black berry cultivars heterozygous for the null allele occasionally display grape color variants with either red/grey or white berries depending on whether only the L2 or both L1 and L2 meristem cell layers, respectively, carry mutations at the color locus (Walker et al., 2006; Furiya et al., 2009; Vezzulli et al., 2012; Migliaro et

al., 2014; Pelsy et al., 2015). Initial characterization of red/grey and white berry somatic variants of Cabernet Sauvignon and Pinot Noir cultivars using Southern blots showed that the absence of anthocyanins in the berry was related to deletion of the functional allele of the color locus (Walker et al., 2006; Yakushiji et al., 2006). Recently, the use of loss of heterozygosity (LOH) analyses along LG 2 suggested that the size of these deletions can be very variable in different berry color somatic variants (Vezzulli et al., 2012; Migliaro et al., 2014; Pelsy et al., 2015). Unfortunately, no information is so far available on the breakpoints delimiting these deletions, which is required to understand their mutational origin.

Tempranillo Blanco (TB) is a white berry somatic variant that originally appeared as a bud sport mutant of Tempranillo Tinto (TT) grapevine cultivar (Martinez et al., 2006). Like many black berry cultivars, TT is heterozygous for the color locus functional allele and TB has been proposed to appear by deletion of this allele (Ibáñez et al., 2012; Migliaro et al., 2014). Here, to understand the origin of the TB variant, SV analysis was carried out after whole-genome sequencing (WGS) of TT and TB lines. SV breakpoints were confirmed by DNA sequencing of specific amplicons and self-cross progenies were used for haplotype phasing and transmission analyses of chromosome rearrangements. Globally, the results show the relevance of complex SV as a driver of clonal variation in plants.

Results

Hemizygous deletions around the color locus in Tempranillo Blanco

TB is a grapevine cultivar generated by vegetative propagation of the white-berried mutant phenotype that appeared as a spontaneous bud sport in a TT plant (Martinez et al., 2006) (Fig 1A). In TB, the presence of a hemizygous deletion eliminating the

functional allele of the *VviMybA1* gene on the color locus at LG 2 was confirmed using DNA blot hybridization of DNA extracted from both TT and TB plants (Fig 1B). The sequence of a specific amplicon corresponding to part of the *VviMybA1* gene showed LOH in TB, further confirming the presence of a hemizygous deletion affecting at least this gene in TB (Fig 1C). Given the high heterozygosity of grapevine cultivars (Laucou et al., 2011), the extent of the deletion in TB was first assessed using primer pairs (S1 Table) intended for LOH analysis of amplified fragments along the putatively affected region in LG 2. This analysis showed that putative hemizygous deletions extended outside of the *MybA* gene cluster in LG 2 (S2 Table). Unfortunately, the efficiency of this method to identify the loss of one allele is limited to originally heterozygous genomic regions and, coincident with other studies (Migliaro et al., 2014), many tested amplicons only provided monomorphic sequences in TT original cultivar, with no polymorphic sites detected after position chr2:15,861,181 (S2 Table) in the PN40024 12X.0 reference genome assembly (https://urgi.versailles.inra.fr/Species/Vitis/Data-Sequences/Genome-sequences).

To globally predict the extension of hemizygous regions in TB irrespective of the heterozygosity level in TT, the GrapeGen GeneChip® cDNA-based microarray (Lijavetzky et al., 2012) was initially used for comparative genomic hybridization (CGH). Significant CNV was detected for probe sets corresponding to 31 annotated genes and all but one involved decreased copy number in TB compared to TT (S3 Table), 14 of them mapping on LG 2 arround the color locus and extending to regions that are homozygous in TT according to the amplicon analysis (S1 Fig). Thus, both copy number decrease and LOH supported hemizygosity around the LG 2 color locus in TB. Unexpectedly, the signal of probe sets corresponding to 14 genes mapping on LG 5 significantly decreased in TB too (S3 Table). Plotting of signal log-

148 ratios along reference genome chromosomes suggested the existence of three
149 hemizygous regions of approximately 1 Mb length each in LG 5 of TB (S1 Fig). Thus,
150 CGH microarrays suggested complex SV in the TB genome involving multiple
151 deletions in LGs 2 and 5. However, the use of a cDNA-based microarray did not
152 provide enough resolution to identify the limits of this sequence variation, which
153 required a WGS comparison.
154
155 **Whole-genome sequencing detects interspersed monosomic and disomic**
156 **regions in linkage groups 2 and 5 of Tempranillo Blanco**
157 To delimit SV occurrence in TB, WGS was performed on both Tempranillo lines using
158 90 nt paired-end reads. A total of ~$209 \cdot 10^6$ clean reads per cultivar were obtained
159 (S4 Table). Similar alignment rate was obtained in both Tempranillo lines, with >90%
160 coverage of the reference genome in an average effective depth of 30× (S5 Table).
161 **Heterozygosity loss.** According to the same logic previously followed for selected
162 amplicons, LOH analysis from WGS data was used to widely detect the loss of one
163 allele at original heterozygous sites. Coincident with microarray-based CGH results,
164 large LOH regions were detected in TB LGs 2 and 5, collectively spanning 4.3 Mb
165 (Fig 2 and S6 Table). Additionally, two small windows were detected in LGs 1 and
166 11, spanning 760 and 1027 bp, respectively (Fig 2A and S6 Table). Monomorphic
167 segments were detected for both TT and TB in LG 2 (6.3-7.7 and 15.8-18.8 Mb) and
168 LG 5 (3.8-5.2 and 12.6-12.9 Mb) (Fig 2B), precluding their LOH analysis in TB.
169 **Copy number variation.** To identify CNV in TB at the whole-genome scale
170 irrespective of the heterozygosity level, we compared the amount of reads aligned to
171 the 12X.0 reference genome between the two Tempranillo lines. Large areas with a
172 significant copy number decrease in TB were detected in LGs 2 and 5 (Fig 2A and

S7 Table). In LG 2, lower copy number in TB extended to at least three alternating segments located between 13.0-18.2 Mb, close to the 3'-end of the chromosome (Fig 2B). The extension of deletions to TT homozygous regions observed in CGH microarrays was confirmed by the CNV analysis of WGS data. Lower copy number in LG 5 of TB comprised at least three alternating segments located between 13.4-21.9 Mb, also interspersed with segments showing unaltered copy number (Fig 2B). In LGs 2 and 5, the overlap between LOH and decreased copy number strongly indicated that both events result from deletion of fragments leaving monosomic chromosome regions in TB. Similarly, the small LG 11 region with LOH in TB also showed copy number decrease (Fig 2A). A few other small regions with reduction of copy number in TB were detected in other LGs (Fig 2A). Although, microarray and LOH analyses did not confirm them as hemizygous deletions, specific positions showing LOH were detected within decreased copy number fragments that were interspersed in LG 9 between positions 14,733,335-14,920,269 (S7 Table). Altogether, these results support the presence of large hemizygous regions in LGs 2 and 5 of TB. The discontinuity of hemizygosity in these regions indicates the existence of multiple rearrangement breakpoints in the origin of the variant genome of TB. Following the same significance values (see Methods), no parallel significant increased copy number regions were detected in LGs 2 and 5 of TB, which indicates that duplications are not involved in this SV.

Structural variation junction sites reveal multiple inter- and intra-chromosomal translocations and deletions in the genome of Tempranillo Blanco

WGS reads were explored to delimit the breakpoints of fragments deleted in TB and to elucidate the nature of the SV. Discordant pair-end mapping (PEM) is useful to

198 detect SV breakpoints from short read sequencing data (Korbel et al., 2007). Thus,
199 TB-specific discordant PEM and soft-clipped reads with respect to the grapevine
200 reference genome revealed six SV junctions in the TB genome not present in TT.
201 Collectively, they comprised six breakpoints in LG 2, five in LG 5 and one in LG 9
202 (Table 1).

203 **Table 1. List of structural variation junctions predicted in the genome of**
204 **Tempranillo Blanco by discordant read mapping analysis.**

SV junction	SV Class	Breakpoint positions mapped in the reference genome	
Inv	Inversion	chr2:13,025,297	chr2:17,391,598
T1	Translocation	chr2:13,537,963	chr9:14,760,326
T2	Translocation	chr2:13,764,277	chr5:14,363,763
T3	Translocation	chr2:17,143,691	chr5:21,919,555
T4	Translocation	chr2:18,226,646	chr5:20,858,083
T5	Translocation (intra-chromosomal)	chr5:13,404,539	chr5:18,450,318

205

206 An inversion junction in LG 2 (Inv) was predicted by the identification in TB but not in
207 TT of read pairs mapping on LG 2 with mates separated one from another by more
208 than 4.36 Mb (Fig 3). Mates at both positions displayed the same alignment
209 orientation, which is compatible with an inversion junction (Korbel et al., 2007;
210 Rausch et al., 2012). TB-specific soft-clipped reads were detected at both flanks of
211 the junction (Fig 3). BLASTN identified that the unaligned extreme of these soft-
212 clipped reads matches to the 5' extreme of the mate inversion flank, unveiling the
213 putative inversion junction breakpoint at nucleotide level resolution (between
214 positions chr2:13,025,297 and chr2:17,391,598; Table 1). Similarly, discordant reads

identified four SV junctions with flanks in different chromosomes of the reference genome (T1-T4, Table 1). These rearrangements were interpreted as inter-chromosomal translocations and involved LGs 2 and 5 in all cases except T1 involving LGs 2 and 9. For the T5 event, discordant alignment identified a SV junction joining LG 5 fragments originally separated by ~5 Mb, which is compatible with an intra-chromosomal translocation (S2 Fig). Interestingly, one flank of all predicted SV breakpoints showed decreased copy number (Fig 3), suggesting that these re-combinations involved hemizygous deletion bridges. In this manner, T2 and T3 breakpoints delimit a deleted fragment in LG 2 (chr2:13,764,277 - chr2:17,143,691), which contained the allele of the grape color locus that is lost in TB (Fig 4A). Furthermore, the existence of additional breakpoints in TB is reasonable considering that discordant read analysis could not detect the breakpoints for at least one obvious copy number discontinuity in LG 5 and a likely one in LG 9 (Fig 4A), whereas missing copy number-neutral breakpoints are also conceivable.

Experimental validation of genome structural variation in Tempranillo Blanco

To validate predicted SV junctions, small contigs (450-900 bp) were re-constructed *in silico* from the sequence of the discordant reads that predicted each junction. Then, we ran PCRs using primer pairs (S1 Table) with mates mapping at each flank of the predicted SV junction breakpoint. No amplicon was expected in TT given the incompatible orientation between primer pairs in Inv, their excessive alignment distance in Inv and T5 or their alignment in different chromosomes in T1-4 (Fig 4A). As expected, amplification was obtained from TB but not from TT genomic DNA (gDNA) for all primer pairs except for the positive control that was amplified in both genotypes (Fig 4B). The actual sequence of the amplicons obtained in TB was

checked by capillary electrophoresis Sanger sequencing (S8 Table). BLAST alignment of these sequences against the 12X.0 reference genome confirmed the fragments and breakpoints/junctions predicted by WGS in the absence of point polymorphisms compared to the reference genome. Thus, these results validate all SV junctions identified in TB from the WGS comparison. Furthermore, their presence and stability was confirmed by obtaining the same amplicons in five additional TB lines that were tested, including the most direct descendant of the original bud sport that is conserved (a first round propagation plant) as well as lines from TB commercial vineyards of different age and location (S3 Fig and S9 Table). The same amplicons were obtained using either gDNA isolated from berry flesh (L2) or skin (L1 contaminated with L2) tissues of TB, suggesting that the same genome rearrangement was present in both meristem cell layers of TB bud sport. Although the original TT vineyard and TT plant in which the white-berried variant phenotype appeared are not conserved, the same PCRs were tested in six different TT accessions including four from vineyards of the same geographical location and planted at similar time as the original TT vineyard in which TB appeared and thus, likely belonging to close clonal lineages (S9 Table). No amplicon corresponding to any of the SV junctions identified in TB was obtained in any tested TT line (S3 Fig). These results suggest that all described genome rearrangements appeared together associated to the emergence of berry color variation and have been stable over successive mitotic growth and propagation cycles.

Rearrangements in Tempranillo Blanco are characterized by non-homologous breakpoints

To understand the forces driving the rearrangements originating the variant genome of TB, we searched for repetitive sequences and motifs in SV junctions. In four of them (T1-T3 and T5) one junction flank matched, with high similarity score, to fragments of Gypsy-27, VHARB-N2, VLINE1 and VIHAT3 grapevine-specific TEs, respectively (Table 2). These TE-like sequences were only found in one breakpoint donor site for each of these SV junctions (S4 Fig), which discards the existence of non-allelic homologous recombination. However, the involvement of these TE sequences in TB genome rearrangements through an unknown mechanism cannot be completely ruled out given that they appear to be enriched within the sequenced SV junctions (Table 2). On the other hand, no match with any repeat element deposited in RepBase was obtained for the Inv junction. Nevertheless, a 7 bp microhomology was observed between both breakpoints of this event (Fig 5 and S8 Table). Furthermore, 1 bp microhomology in the context of an 8 bp palindrome (CCTTAAGG) was present at the T3 junction site, while a 2 bp microhomology was found in T4.. Blunt ends with no homology were detected for T1, T2 and T5 junctions (Fig 5).

Table 2. Features of breakpoint junctions sequenced in derivative chromosomes of Tempranillo Blanco

Sequenced breakpoint junction	Sequenced length	Micro-homology in junction	Breakpoint flank	Repetitive elements	Repetitive element in total sequenced junctions	Repetitive element in grapevine reference genome
T1	730 bp	Blunt ends	T1_LG2	Gypsy-27_VV-I	8.9%	1.19%
			T1_LG9	-		
T2	621 bp	Blunt ends	T2_LG2	VHARB-N2_VV	7.1%	0.10%
			T2_LG5	-		
T3	529 bp	1 bp in 8	T3_LG2	-		

		bp palindrome	T3_LG5	VLINE1_VV	5.2%	1.91%
T4	437 bp	2 bp	T4_LG2 T4_LG5	- -		
T5	807 bp	Blunt ends	T5_Proximal T5_Proximal	VIHAT3 -	10.9%	0.12%
Inv	647 bp	7 bp	Inv_Proximal Inv_Distal	- -		

Linkage of rearrangements in a single copy of the affected chromosomes and *in silico* reconstruction of the rearranged genome of Tempranillo Blanco

Trying to reconstruct the haplotype phasing of rearrangements in the TB genome, we carried out high-throughput genotyping of a TT self-cross progeny (S_1) comprising 78 siblings. S_1 individuals were genotyped using GrapeReSeq Illumina® 18K SNP Infinium chips (Houel et al., 2015). Subsequently, the haplotype of each TT homologous chromosome in LGs 2 and 5 was reconstructed according to the data of 286 single nucleotide polymorphism (SNP) markers that were heterozygous in TT (S10 Table). The genotype of TB at these SNP positions was inferred from our WGS data and its comparison to the two TT haplotypes showed that all hemizygous deletions detected in TB were linked in a single homologous copy of LGs 2 and 5 (S10 Table). In view of that and considering detected CNV, LOH and breakpoint junctions (Fig 6A), we could *in silico* reconstruct the TB derivative chromosomes in a single homologous copy of the affected LGs (Fig 6B). In this reconstruction we hypothesized an additional translocation junction that was not detected in the PEM analysis to connect hemizygous deletion breakpoints detected by the CNV analysis at LGs 5 and 9 (Figs 2B, 4 and S2 Fig). In fact, although not passing all the significance thresholds in the SV analysis, read pairs supporting this translocation join and, delimiting a deletion of a ~16 kb fragment in LG 9 of TB according to CNV,

were detected from sequencing read alignments. Altogether, 13 breakpoints and 16 rearranged fragments were considered in the model to reconstruct the TB genome (Fig 6B).

Genetic alterations caused by the genome reshuffling in Tempranillo Blanco

The derivative genome of TB contains predicted hemizygous areas that include 165 and 148 annotated genes in LGs 2 and 5, respectively, according to the grapevine 12X V1 gene annotations (http://genomes.cribi.unipd.it/). In addition to gene dose decrease, and considering the characteristically high heterozygosity of the grapevine genome (Velasco et al., 2007), such hemizygous state might have phenotypic consequences due to the loss of functional alleles in contrasting loci as is the case for the color locus (Fig 1). In this regard, TB bears 48 and 100 predicted hemizygous genes located in LGs 2 and 5, respectively, that show heterozygous polymorphisms in TT. In LG 2, alternative alleles with respect to the PN40024 reference genome were only found for three hemizygous genes. This low rate of variation indicates that the copy present in TB regions involving hemizygous deletions in LG 2 displays a genotype very close to that existing in the white-berried PN40024 quasi-homozygous line used for the reference genome, which is homozygous for the *MybA* locus null allele. This result further confirms the loss of the color locus functional allele in TB, precluding berry color pigmentation in this variant. In contrast, 73 hemizygous genes in LG 5 displayed variant alleles in TB with respect to the reference (S11 Table). Variant alleles with possible deleterious changes remaining hemizygous in TB were specifically searched. Although no major effect variants relative to the reference genome were detected, non-synonymous changes were predicted for one gene with unknown function in LG 2 and 20 genes in LG 5 (S11 Table). Moreover, two

329 annotated genes (*VIT_09s0070g00940* and *VIT_05s0102g00090*) were truncated by
330 T1 and T3 breakpoints, respectively; while the resulting T1 junction might involve a
331 novel fusion gene substituting one allele of *VIT_09s0070g00940*. The single
332 presence of these alleles might result in specific TB phenotypes.
333
334 **Reduced sexual transmission of rearranged chromosomes associated to a**
335 **decreased fruit yield**
336 Complex and unbalanced chromosomal rearrangements found in the genome of TB
337 might lead to defects during sexual reproduction (Pellestor, 2014). This hypothesis
338 was specifically tested by studying sexual transmission of LGs 2 and 5 in selfed
339 progenies of TT and TB consisting in 38 and 46 individuals, respectively. For LG 2,
340 four heterozygous microsatellite markers were selected outside of the rearranged
341 region. The allelic distribution observed in the progenies was different for TT and TB
342 (frequencies inside the circles in Fig 7). In TT the te two alleles of TT segregated with
343 a similar frequency for every marker studied as expected for Mendelian inheritance
344 (non-significant χ^2). Nevertheless, in TB, every marker analyzed showed one allele
345 with a significantly lower frequency than expected for Mendelian inheritance. The
346 frequency of the disfavored allele decreased as the marker was closer to the
347 chromosome rearranged region. From the genotype frequencies in the TB progeny,
348 the two putative haplotypes in TB were inferred, showing that all the disfavored
349 alleles were linked in the same haplotype. The *VviMybA1* gene was also genotyped
350 by the independent amplification of the functional and non-functional alleles (Fig 7).
351 The study of genotypic frequencies for the closest microsatellite and *VviMybA1* in the
352 TT progeny allowed establishing that the disfavored haplotype in the TB progeny was
353 that carrying the functional allele for *VviMybA1* in TT and thus, the haplotype linked

to SV and deletions in TB. In the case of LG 5, two heterozygous microsatellite markers were analyzed and the results obtained were essentially in line to those described for LG 2 (Fig 7). For each marker, one allele was significantly less transmitted to the TB S_1 progeny and the disfavored alleles were linked in the same haplotype. Thus, segregation analysis further supports that SV only affected one copy of each chromosome. Rearranged chromosome segments are not transmitted to the next generation or transmitted at a very low frequency. This lack of transmission was associated with a reduction of about 59% of male gamete viability in TB with respect to TT (Fig 8), which suggests that derivative chromosomes lack genes that are essential for the viability of the haploid phase. This sterility rate would be consistent with the loss of essential genes for microgametophyte viability in both LGs 2 and 5. The effect of complex chromosome rearrangements on gamete viability might result in additional agronomical alterations present in TB considering that reduced pollen viability correlated with 44% and 29% drop in fruit set rate and fruit yield, respectively (Fig 8).

Discussion

Complex unbalanced genome rearrangements caused the loss of the berry color locus functional allele

Deletions resulting in grapevine white berry variants have been described previously (Walker et al., 2007; Vezzulli et al., 2012; Migliaro et al., 2014; Pelsy et al., 2015). However, the mutational mechanisms generating these deletions remain uncertain. So far, only gene replacement has recently been proposed as the mechanism causing LOH and originating new white berry alleles in Pinot somatic variants by recombination between chromosome 2 homologs linked to deletion of the color locus

functional allele (Pelsy et al., 2015). Nonetheless, we consider that the lack of CNV analysis of the variant lines in that study does not preclude the presence of hemizygous deletions in the origin of the described Pinot alleles. In the TB white berry variant, Southern-blotting and WGS comparisons to its black berry ancestor TT confirmed the presence of hemizygous deletions leaving at the berry color locus a single copy corresponding to the null allele (Fig 1B and S11 Table). The parallel copy number decrease observed indicates that LOH in these TB chromosome regions directly results from their hemizygous state (Fig 2), as described in other Pinot and Cabernet Sauvignon grape color variants (Walker et al., 2006; Yakushiji et al., 2006).

Genome-wide deep sequencing analyses uncovered a much more complex SV in the origin of TB berry color variant than previously identified in any other grapevine somatic variant. Hemizygous segments detected in LGs 2 and 5 of TB are flanked by breakpoints corresponding to several translocations and inversions (Table 1, Fig 6A). These events collectively resulted in the hemizygous deletion of more than 8 Mb (Fig 2), *ca* 1.7% of the grapevine haploid genome (Jaillon et al., 2007). The same hemizygous deletions were repeatedly detected using different approaches (Figs 1, 2 and 4, S1 Fig, S3 Fig and S2 Table) and the corresponding breakpoints were experimentally confirmed irrespective of the clonal lineage proximity of TB and TT accessions to the original TB bud sport (S3 Fig). Thus, all identified SV events are characteristic of the TB bud sport and stable in subsequent multiplication cycles and therefore, do not seem to correspond to the accumulation of secondary mutations. Remarkably, the longest deleted fragment in TB was that containing the color locus functional allele (~3.4 Mb between T2 and T3 breakpoints in chromosome 2, Table 1). This coincidence might reflect the fact that grapevine cells are more tolerant to the hemizygous state of this chromosome region that has been submitted to intensive

selective sweep (Di Gaspero and Foria, 2015) and consequently, might carry less heterozygous deleterious alleles.

For WGS, we used whole leaves to obtain DNA assuming that both L1 and L2 meristem layers were mutated in TB as reported before for this and other grapevine white berry variants (Walker et al., 2006; Vezzulli et al., 2012; Migliaro et al., 2014; Pelsy et al., 2015). The existence of independent chimeric mutations in L1 and L2 involving the loss of the color locus functional allele was observed in a white berry variant of Pinot (Pelsy et al., 2015). However, in TB the frequency of WGS discordant reads detected in SV junctions appears close to 50% (Fig 3), while no read for a second allele was identified in SNP positions showing LOH (S11 Table). These results indicate that the meristem cell lineage in which the complex genome rearrangement that caused the loss of the grape color functional allele appeared had colonized L1 and L2 cell layers in the meristem that generated the TB bud sport.

Genome reshuffling in Tempranillo Blanco shows hallmarks of chromosome breakage and illegitimate fusion

The existence of SV has been reported among the few grapevine cultivars so far sequenced at the whole genome level when compared with the PN40024 reference sequence (Jaillon et al., 2007; Velasco et al., 2007; Da Silva et al., 2013; Di Genova et al., 2014; Cardone et al., 2016; Xu et al., 2016). Moreover, apart from TE insertions, major SV were not considered in the previous comparison of whole-genome sequences performed between four clones of Pinot Noir (Carrier et al., 2012). Our study provides the opportunity to identify possible mutational causes of somatic SV in grapevine by directly comparing the derivative genome of TB to the original genome in TT. In this manner, an analysis of the rearrangement pattern

429 detected in TB indicates that it fulfills all the criteria proposed for the inference of a
430 chromothripsis phenomenon (Stephens et al., 2011; Korbel and Campbell, 2013): i)
431 CNV, LOH and PEM analyses revealed intercalated regions with monosomic and
432 disomic copy number states flanked by clustered breakpoints in the right arm of LGs
433 2 and 5 in the absence of significant copy number gain in the affected chromosome
434 regions (Fig 2). A goodness-of-fit test confirmed (95% confidence in Kolmogorov-
435 Smirnov test) that PCR-validated breakpoints are not exponentially distributed along
436 LGs 2 and 5, which is in agreement with clustering rather than random distribution of
437 breakpoints along the rearranged chromosomes as expected in chromothripsis; ii)
438 Different segregation analyses indicate that all rearrangements are linked in a single
439 homologous copy of LGs 2 and 5 (Fig 7 and S10 Table), which, along with CNV and
440 breakpoint sequence data, allowed us reconstructing and walk the tentative single
441 copy of derivative chromosomes following head-to-tail sequence of fragments (Fig 6)
442 as expected in a chromothripsis scenario (Korbel and Campbell, 2013); iii) Simulation
443 analyses confirmed the randomness of the predicted fragment order in TB derivative
444 chromosomes. A rearrangement order value of 81 was estimated for TB considering
445 the fragment order coordinates. This value is within the normal distribution or
446 randomness interval (68.9 ±15.7) estimated by three independent rounds of 10^8
447 Monte Carlo simulations. iv) The proportion of join types in validated SV junctions is a
448 probable result (P = 0.703) for a multinomial distribution with equal probability for
449 each of the four possible join types as reported in the structural rearrangement graph
450 (Fig 6A), suggesting that fragments were randomly joined as expected in a single
451 catastrophic event (Stephens et al., 2011; Korbel and Campbell, 2013); and v) The
452 DNA sequences of the breakpoint joins validated in TB show lack of homology (Fig
453 5), which is a hallmark of breakage-repair mechanisms involved in chromothripsis

454 (Kloosterman et al., 2012; Malhotra et al., 2013). Altogether, the rearrangement
455 pattern of the TB derivative genome is characteristic of chromothripsis events.

456 Breakage-fusion-bridge (BFB) cycles and degradation of lagged chromosomes
457 during mitosis inside of micronuclei are mechanisms that have been proposed as
458 triggers of chromothripsis (Li et al., 2014; Zhang et al., 2015). Proof for the
459 micronuclei origin of chromothripsis arose from the observation that mis-segregated
460 anaphase chromosomes are enclosed in extranuclear bodies (micronuclei) where
461 they can undergo fragmentation and illegitimate repair before their restitution into the
462 nucleus in 2-3 consecutive cell division cycles (Crasta et al., 2012; Zhang et al.,
463 2015; Ly et al., 2017). Similar complex rearrangements associated with micronuclei
464 formation and dependent on the DNA repair machinery were observed during the
465 process of genome elimination taking place after hybridization of transgenic
466 Arabidopsis plants expressing mutant forms of H3 histones (Tan et al., 2015), which
467 indicate that cellular mechanisms involved in chromothripsis are present in plants. In
468 TB, , the lack of homology or the microhomology of 1-7 bp that characterize SV
469 junctions (Fig 5 and S8 Table) are compatible with non-homologous end joining
470 (NHEJ) at blunt ends and microhomology-mediated end joining (MMEJ) repair
471 mechanisms of double-strand breaks generating these junctions (Jones and
472 Jallepalli, 2012; Sinha et al., 2016).

473 Nevertheless, the fact that the left arm of chromosomes 2 and 5 is protected from
474 rearrangement in TB might contradict the idea of a simultaneous breakage and repair
475 of the affected chromosomes inside of a micronucleus. Rather, the type of
476 chromothripsis pattern found in TB has been proposed to be generated by
477 successive BFB cycles after chromosome fusions generating dicentric chromosomes
478 that are broken at random positions between the two centromeres by the action of

the mitotic spindle and rejoined in successive cell divisions that produce the deletion of fragments segregated in different daughter cells (Sorzano et al., 2013). The putative pseudodicentric nature of the grapevine chromosome 9 (Di Gaspero and Foria, 2015), might have triggered this event as the two breakpoints detected in LG 9 of TB fall within the small region between the two putative centromeres (Fig 6B). Irrespectively of the cellular origin for this chromothripsis-like conformation, the causal genome instability appears stabilized in TB derivative cultivar (S3 Fig), showing that this type of chromoanagenesis event leads to genetic and phenotype diversity that can be exploited as a source of novel traits in vegetatively propagated plants.

Consequences of complex unbalanced breakage and rejoin genome rearrangements as a source of somatic variation in plants

SV is a common driver of genome evolution in sexually reproduced plants (Cao et al., 2011; Chia et al., 2012; Qi et al., 2013), and a source of genetic diversity and new traits in vegetatively propagated plant crops (Terol et al., 2015). Resembling TB derivative cultivar, chromoanagenesis events can rapidly produce dramatic genome reshuffle resulting in multiple molecular and phenotype alterations (Leibowitz et al., 2015). At the gene level, the partial monosomy of large chromosome segments yields more than 300 hemizygous genes in TB. In highly heterozygous species such as grapevine, monosomy may unmask the expression of deleterious alleles originally present in a heterozygous state. In TB this is the case for the berry color locus at which rearrangements caused deletions leaving only the null allele (Fig 1B) (Migliaro et al., 2014). Other non-synonymous SNV originally heterozygous in TT are present at hemizygous state in TB (S11 Table), which might account for additional phenotype

variation in this cultivar. Furthermore, given that chimeric fusion proteins might acquire novel dominant functions (Borrow et al., 1996), further variant TB phenotypes might arise from the putative fusion protein generated by the T1 junction at the original *VIT_09s0070g00940* locus.

At the reproductive level, the sexual fitness of TB seems to be compromised by the genome reshuffling. Segregation distortion in a TB S_1 progeny indicates lack of transmission of rearranged chromosome copy arms (Fig 7). Similar lack of transmission linked to meiotic recombination might have been considered as a proof of gene replacement in Pinot grape color variants (Pelsy et al., 2015). However, considering the proportion of hemizygous genome and the reduction of pollen viability observed in TB (Fig 2 and 8), it is more likely to assume that chromothriptic-like chromosomes as those present in TB (Fig 6B) are deleterious for the resulting unbalanced haploid gametes (Pellestor, 2014). Although sexual reproduction is not necessary for the maintenance of vegetatively propagated crops, gametophyte viability is required for proper fruit set (Iyer, 1966) and thus, complex SV may have agronomic consequences decreasing fruit yield as observed in TB (Fig 8). Different levels of sterility were also observed in Arabidopsis lines with shattered chromosomes (Tan et al., 2015).

Conclusions

A whole-genome scale approach was helpful to show that in TB white grape variant, the deletion of the grape color locus functional allele is linked to a much more complex genome rearrangement than previously observed in any other grapevine somatic variant. This is the first study delimiting the deletion of this allele at a base resolution level. In TB this deletion occurred associated to an unbalanced chromoanagenesis event entailing an illegitimate rejoin of fragments after multiple

chromosome breakage. Taken together, our results show that complex chromosome rearrangements with chromothriptic features naturally emerge and stabilize during plant vegetative growth. Although possibly reducing sexual fitness and compromising fruit and seed production, these complex unbalanced rearrangements might be relevant for the genetic improvement of woody crops as they rapidly generate new phenotypes that can be selected and vegetatively propagated.

Materials and Methods

Plant material

Molecular biology and viticultural experiments, respectively, were carried out using materials collected from the two ancestral TB multiplication vineyards (TB-ICVV2 and TB-ICVV3 accessions, second and third propagation cycles from the original TB bud sport, respectively). These vineyards are located in “Finca Valdegón” (in Agoncillo, La Rioja, Spain) and “Finca La Grajera” (in Logroño, La Rioja, Spain), respectively. As a representative accession, TT clone RJ51 (the most cultivated TT clone in the Rioja DOC region) from both sites was used as respective control. Both plots belong to the Grapevine Germplasm Collection of the Instituto de Ciencias de la Vid y del Vino (ICVV) (ESP-217) and are maintained under the same agronomical conditions. All plants are grafted in Richter-110 rootstock, trellised in double cordon Royat system and cultivated in a similar way. The genotype of nine microsatellite loci located in different chromosomes was obtained as described elsewhere (Ibanez et al., 2009), confirming that TB matches the genotype of TT (S12 Table). Additionally, another five TT accessions from different clonal lineages conserved in the clone bank of the ICVV at Finca La Grajera and TB mother plants (first round propagation plants

553 from the TB bud sport) as well as TB plants from four vineyards of different age were
554 used to assess the presence of SV breakpoints (S9 Table).

555 Selfed progenies of TT-RJ51 and TB-ICVV2 accessions were generated in 2008,
556 consisting in 78 and 46 individuals, respectively. These progenies were maintained
557 ungrafted in pots at Finca Valdegón.

558

559 **DNA extraction**

560 In all experiments, total gDNA was isolated from young leaves using the DNeasy
561 Plant Mini Kit (Qiagen, Germany) according to the protocol described by the
562 manufacturer. Additionally, mature fruits of TT-RJ51 and TB-ICVV1 were harvested
563 in 2016 and gDNA was obtained separately from berry skin and flesh. Independent
564 DNA samples were used for each technique: microsatellite genotyping, Southern
565 blot, GrapeGen GeneChip® hybridization, Illumina sequencing and PCR
566 amplification.

567

568 **Southern blot**

569 DNA blot hybridization analysis was performed as described by Sambrook et al.
570 (Sambrook J., 1989). Briefly, gDNA from TT and TB was digested with ApaI,
571 transferred to a nylon membrane and hybridized to a *VviMybA1* probe as previously
572 described (Lijavetzky et al., 2006). The *VviMybA1* probe was generated by PCR
573 amplification with primers Ps (5'-TCACGGGGTTTAGAAAGTGG-3') and Pas (5'-
574 ATCAATTGGGGAATTGGTGA-3') using TT gDNA as template. The hybridized
575 membrane was scanned with a STORM PhosphorImager (Molecular Dynamics).

576

577 **Loss of heterozygosity analysis in specific amplicons.**

To analyze LOH in TB, primer pairs for 30 amplicons were designed along LG 2 (S1 Table). PCR amplifications were carried out using Taq DNA Polymerase (Qiagen) as recommended by the manufacturer. PCR products were purified with ExoSAP-IT (USB Products Affymetrix, Inc., Cleveland, OH, USA) according to manufacturer's instructions and then used for Sanger sequencing at the Genomic Unit of 'Parque Científico de Madrid' with the same primers used for amplification.

Copy number variation analysis by GeneChip® hybridization

Two biological replicates of TT and TB were hybridized to the GrapeGen GeneChip® (Lijavetzky et al., 2012). For each replicate, gDNA samples (10 µg each) were fragmented to an average size of 0.5 kb with a sonicator Labsonic U set at 50% intensity and repeating duty cycle 0.5, 4 pulses of 10 seconds each. Labeling of DNA was performed as recommended in GeneChip® Whole Transcript Double-Stranded Target Assay (Affymetrix). Briefly, fragmented DNA was amplified, purified and biotin terminal labeled with Double-Stranded DNA Terminal Labeling Kit (Affymetrix), following recommendations. Finally, 7.5 µg of labeled DNA was used for hybridization of microarrays according to the Affymetrix GeneChip® Expression Analysis Technical Manual. CEL files were RMA normalized and differential probe set hybridization between TB and TT assessed using the RankProd package from Bioconductor (Breitling et al., 2004). A cut-off value of 1.8 fold-change absolute value and P <0.01 was established to detect genes with genomic CNV. To this end the annotation of the GrapeGen GeneChip® according to 12X V1 gene predictions from Lijavetzky et al. (Lijavetzky et al., 2012) was used.

Genome re-sequencing and computational comparative analyses

603 **Sequencing and pre-processing.** Two gDNA libraries were built from a sample of
604 young leaves from one individual of TT-RJ51 and another from TB-ICVV2 (second
605 round multiplication plant), respectively. The gDNA of each individual was
606 fragmented randomly. After electrophoresis, DNA fragments of approximately 470 bp
607 were gel-purified. Adapter ligation and DNA cluster preparation were then performed
608 and subjected to sequencing. Each Tempranillo line was sequenced in a different
609 lane at Beijing Genomics Institute (BGI) facilities (Shenzen, China) using the Illumina
610 HiSeq 2000 sequencing system. On each sample, a total of ~$209 \cdot 10^6$ paired-end
611 reads of 90 nt length were obtained (~18.8 Gb of total sequence per sample) (S4
612 Table).

613 Sequence data were processed by removing the adapter sequence from reads and
614 subsequently taking out the reads with low quality (Phred quality ≤5 in ≥50% of the
615 positions of the read) to obtain clean data described in S4 Table. The final Q20 of
616 each lane was above 95%. Two files containing the clean reads in FastQ format,
617 corresponding to TT and TB samples, were submitted to an in-house pipeline
618 composed of several bash shell scripts (available upon request to the authors).
619 These scripts used shell commands, perl scripts and open source programs to
620 process the reads, produce the alignments, select particular subsets of alignments
621 for the PEM study and carry out the CNV study and the variant calling. The pipeline
622 also included a prediction of effects for the retrieved variants and a scanning for LOH
623 regions in TB. Each specific procedure is described below.

624 **Alignment.** Each FastQ file was aligned to the grapevine PN40024 12X.0 reference
625 genome assembly using BWA version 0.5.9-r16 with the option “sampe” (for paired-
626 ends) and default parameters (Li and Durbin, 2009). Approximately 76% of the reads
627 aligned with the reference genome (S5 Table). Redundant reads comprising the

same span were then removed using the 'rmdup' option of SAMtools (Li et al., 2009). To improve the alignments by minimizing the number of artefactual mismatching bases due to the close presence of INDELs with respect to the reference genome, local realignment was then done using GATK version 1.0.5777 (tools "RealignerTargetCreator" and "IndelRealigner") (McKenna et al., 2010). To increase the ability of SV breakpoint detection, another script was used to split each of the 90-nt clean reads into two fragments of 45 nt. The new sets of FastQ files containing the 45 nt split reads were also aligned as described above and used for SV junction detection.

Variant calling and loss of heterozygosity genome scanning. Files containing read alignments of TT and TB in BAM format were used to search for Tempranillo line-specific polymorphisms (SNPs and INDELs). Aligned reads were pre-processed as described above and moreover, reads with mapping Quality (MAPQ) <40 after local re-alignment in IndelRealigner were discarded too. Variant calling on each line was carried out using the BCFtools utility from SAMtools for the whole genome and HaplotypeCaller tool from the GATK package for LGs 2 and 5. Initially, for both tools, VCF files independently comparing each Tempranillo line to the grapevine 12X.0 reference genome were generated. In order to remove false positives, the following filters and cut-off thresholds were used: polymorphisms were considered when ≥15 reads covered the position, the variant sequence was observed in ≥35% for heterozygous sites and ≥90% with ≤2 reads of the non-variant allele for homozygous sites. Finally, resulting TT and TB sets of heterozygous polymorphisms identified in SAMtools were binned in 100-Kb bins and these bins were reported and plotted for each LG. To delimit segments of LOH along the TB genome, intervals of ≥3 consecutive heterozygous sites in TT (according to the filters described above) that

fulfilled the criteria of homozygous sites in TB were considered. To estimate the genotype in TB irrespective of the variation relative to the reference genome, the sequence at these positions was obtained from the compilation in the “pileup” file of the alignments. In these cases, a Perl script was constructed to detect homozygous sites in TB according to ≥8 reads coverage and ≥90% frequency of the most frequent allele, and ≤2 for alleles other than the most frequent one.

GATK HaplotypeCaller was also used for variant calling in LGs 2 and 5 and results were cross-checked with these obtained from SAMtools+BCFtools. Although results were highly equivalent, HaplotypeCaller was used to study the effect of polymorphisms involving LOH in hemizygous regions of TB by assuming that this application could perform better on INDEL detection (Hwang et al., 2015). The same filters described for the output of SAMtools were used to detect TT polymorphic sites with LOH in TB (heterozygous in TT and homozygous in TB). In this case, to estimate the genotype of TB for non-variant LOH sites, the variant calling was repeated after merging the reads from the two lines. The genotype in the non-variant line was inferred by deducting the reads of the variant line from the total of both. Finally, the effect of detected polymorphisms was predicted using SnpEff version 2.0.3 (Cingolani et al., 2012) according to the grapevine 12X V1 gene predictions from CRIBI (http://www.cribi.unipd.it/) and the functional annotation and classification of genes described by Grimplet *et al*. (2012) (Grimplet et al., 2012).

Genome re-sequencing copy number variation. Alignment files of TT and TB were compared to detect chromosome regions involving CNV between the two Tempranillo lines. To this end, CNV-seq (Xie and Tammi, 2009) was applied using TB and TT as “test” and “reference”, respectively, on the corresponding files containing the initial position of each aligned read (called “hits” files by the CNV-seq

678 authors). Significance values were set in CNV-seq to Log_2 (copy number TB/TT) ≤-
679 0.5 and *P* ≤0.00001 for each chromosomal window, and ≥4 consecutive sliding
680 windows for annotating a CNV zone. The CNV zones detected by CNV-seq having a
681 *P* ≤0.05 and Log_2 (copy number TB/TT) ≤-0.5 or ≥0.37 were considered as
682 significantly decreased or increased, respectively, in copy number in TB relative to
683 TT.

684 **Structural variation breakpoint search.** PEM and soft-clipped reads indicative of
685 the presence of SV junctions were specifically searched in the two SAM files
686 containing the TB and TT reads, respectively, aligned to the PN40024 12X.0
687 reference genome. The two alignment sets of whole 90 nt reads and reads split into
688 two 45 nt fragments were equally processed in parallel. Firstly, read pairs with MAPQ
689 equal to 0 in BWA were removed as it indicates mapping to multiple locations. From
690 the remaining set of aligned reads and using bash shell commands (grep and awk)
691 and the *samtools merge* utility, putative chromosome rearrangement breakpoint sites
692 were searched by extracting two subsets of paired-end reads with discordant mate
693 alignment: 1) Read pairs with increased insert size indicative of intra-chromosomal
694 translocations or large deletions: absolute value of insert size (TLEN field in the SAM
695 file) >4 times the median TLEN value of all aligned pairs in the sample. 2) Read pairs
696 with mates in different chromosomes indicative of inter-chromosome translocations
697 (RNEXT field different than '=' or RNAME value in the SAM file). To determine the
698 orientation of the rearranged chromosome fragments comprising putative
699 breakpoints, paired-end reads with unexpected mate orientation alignment were
700 searched from the previous two subsets. Two different discordant orientations were
701 considered: A, both mates with unexpected orientation (FLAG field of the SAM file
702 with values = 81, 161, 97 or 145); B, only one mate with unexpected orientation

(FLAG with values 65, 129, 113 or 177). An additional subset comprising soft-clipped reads (only partially mapping to the reference in one read extreme) was searched for from genomic regions with discordant mate alignment 1 or 2. Reads containing CIGAR alignment operation S was the criterion used to select soft-clipped reads.

To select breakpoints distinguishing TB and TT genomes, BEDtools (Quinlan and Hall, 2010) and a Perl script were used to compare TT and TB discordant read alignments detected as described above. Candidate breakpoint regions were selected when the depth coverage of TB-specific discordant reads (discordant PEM + soft-clipped reads) was ≥2. From these regions, the subsets of TB-specific discordant reads were extracted to delimit SV candidate breakpoints, which were visually inspected using Integrative Genomics Viewer (IGV, version 2.2) software (Thorvaldsdottir et al., 2013).

The non-clipped part of TB-specific soft-clipped reads mapping on candidate breakpoint areas was aligned to the PN40024 12X.0 genome assembly using BLASTN suite (https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE_TYPE=BlastSearch).

Analysis of repetitive sequences at breakpoints

For each SV junction detected in TB, a consensus breakpoint junction sequence was built from the sequence of discordant reads and submitted to RepeatMasker version 4.0.3. (http://www.repeatmasker.org/cgi-bin/WEBRepeatMasker). *V. vinifera* was the "DNA source" and ABBlast/WUBlast the search engine to query for matches with repeat elements deposited in RepBase database update 2013/04/22 (http://www.girinst.org/repbase), which includes all described repeat elements specific of *V. vinifera*. TE-like sequences were searched by BLAST of RepBase sequences against 1) the 12X.0 reference genome in regions where TB SV

breakpoints mapped and 2) TB consensus sequences at SV junctions. The structure and domains of detected TE sequences were studied with NCBI CDD tools (http://www.ncbi.nlm.nih.gov/cdd/).

Validation of structural variation junctions

PCR primers were designed at each flank of genomically detected SV junctions (S1 Table). The consensus breakpoint junction sequence reconstructed from discordant reads was used as template for primer pair design. Primer specificity was checked by BLAST against the PN40024 12X grapevine reference genome assembly in the Genoscope website (http://www.genoscope.cns.fr/externe/GenomeBrowser/Vitis/). These primer pairs were tested for PCR amplification using as template gDNA from different TB and TT accessions (S3 Fig, S9 Table). PCRs were carried out using MyTaq™ DNA Polymerase (Bioline, Meridian Life Science, Memphis, USA) following the manufacturer's protocol. *VviDXS* (*VIT_05s0020g02130*) gene, mapping on a presumably disomic region of LG 5 in both TT and TB, was included in the experiment as positive control using primers described elsewhere (Emanuelli et al., 2010). An aliquot of the amplification product was subjected to electrophoresis through 1% agarose gels to check for the presence and size of PCR products. PCR products were sequenced as described above.

SNP chip genotyping and haplotype phasing.

Total gDNA from 78 TT S_1 progeny individuals was genotyped using the GrapeReSeq Illumina® 18 K SNP Infinium as described elsewhere (Houel et al., 2015). SNP genotype data was processed as described in GrapeReSeq project (https://urgi.versailles.inra.fr/Species/Vitis/GrapeReSeq_Illumina_20K). To

753 reconstruct the original phasing of the two haplotypes in TT, segregation for each
754 pair of adjacent SNP loci represented in the chip was studied, assuming that the
755 most frequent genotype combination in the progeny correspond to the parental
756 haplotypes. The genotype of TB at SNP positions represented in the chip mapping
757 on LGs 2 and 5 was obtained from our WGS dataset. The haplotype remaining in TB
758 at positions involving LOH in this variant was compared to the two haplotypes
759 reconstructed in TT to infer the homologous copy affected by each deletion.
760

761 **Sexual transmission analysis.**

762 A series of markers were genotyped in a set of TT and TB self-cross S_1 progenies to
763 study sexual transmission of each homologous chromosome at LGs 2 and 5. Four
764 heterozygous microsatellite markers in both Tempranillo lines mapping on LG 2 were
765 selected: Chr2b, Chr2a, C6F1, and C5G7. Genotyping of *VviMybA1* alleles in TT, TB
766 and TT self-cross S_1 progeny was carried out as described previously (Lijavetzky et
767 al., 2006). Primers pairs 'a'–'d3' and 'b'-'d' were used to amplify null and functional
768 alleles, respectively. Allele segregation was determined for each locus and progeny,
769 and their deviation from expected values (0.5) was evaluated through χ^2 tests. This
770 statistical test was performed using Microsoft Excel software. Genotypic segregation
771 for each pair of adjacent loci was studied to determine microsatellite haplotypes,
772 assuming that the most frequent double homozygous genotypes in the progeny
773 correspond to the parental haplotypes. *VviMybA1* genotyping was used to associate
774 the microsatellite haplotypes with the alleles of the berry color locus. Genotypes were
775 assigned by assuming that non-amplification of one of the alleles corresponds to a
776 homozygous genotype for the other allele. Joint genotypic segregation of C5G7 and
777 *VviMybA1* was used as described above to establish the haplotypes at LG 2. For LG

5 two microsatellite markers were selected, VVIT68 and VMC5E11, and the procedure was similar to that described for LG 2.

Fruit production and pollen viability.

Experiments were carried out at “Finca La Grajera” during 2015. In this vineyard plot, three rows (~50 plants each) of TB plants are interspersed by three rows of TT plants. Each row was used as independent biological replicate and four plants per replicate were analyzed. Pollen was simultaneously collected from TT and TB inflorescences (one per plant) at 50% bloom. Pollen viability was analyzed using Alexander’s modified staining as described previously (Royo et al., 2016). For each Tempranillo line, more than 1000 pollen grains per biological replicate were analyzed. Fruit set percentage was calculated as the rate between the number of ripen fruits per cluster at maturity and the number of flowers at flowering time in the same cluster (one cluster per plant was analyzed). The number of flowers per inflorescence was estimated by bagging the inflorescence before the onset of flowering and counting flower caps inside the bag after fruit set. Fruit yield was estimated as the total mass of grape bunches per plant.

Data Reporting

GrapeGen GeneChip® and WGS data are deposited in NCBI under Gene Expression Omnibus (GEO) GSE80801 and Sequence Read Archive SRP065756 (BioProject PRJNA301084) accession numbers, respectively.

Abbreviations

CGH = Comparative genomic hybridization

BFB = Breakage-fusion-bridge

CNV = Copy number variation

gDNA: genomic DNA

Inv = inversion

LG = linkage group

MAPQ = mapping quality

PEM = paired-end mapping

SNP = single nucleotide polymorphism

SNV = single nucleotide variation

SV = genome structural variation

T = translocation

TB = Tempranillo Blanco

TE = transposable element

TT = Tempranillo Tinto

WGS = whole-genome sequencing

Acknowledgements

The authors are very grateful to Drs. Karel H. M. van Wely and Daniel Trujillano for thoughtful comments and feedback, and Guillermo Juárez and Silvia Hernáiz for technical assistance. We also thank the Genomics service of the CNB-CSIC for running the CGH hybridizations.

References

Battilana J, Emanuelli F, Gambino G, Gribaudo I, Gasperi F, Boss PK, Grando MS (2011) Functional effect of grapevine 1-deoxy-D-xylulose 5-phosphate synthase substitution K284N on Muscat flavour formation. J Exp Bot **62:** 5497-5508

Borrow J, Stanton VP, Jr., Andresen JM, Becher R, Behm FG, Chaganti RS, Civin CI, Disteche C, Dube I, Frischauf AM, Horsman D, Mitelman F, Volinia S, Watmore AE, Housman DE (1996) The translocation t(8;16)(p11;p13) of acute myeloid leukaemia fuses a putative acetyltransferase to the CREB-binding protein. Nat Genet **14:** 33-41

Boss PK, Davies C, Robinson SP (1996) Expression of anthocyanin biosynthesis pathway genes in red and white grapes. Plant Mol Biol **32:** 565-569

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

Breitling R, Armengaud P, Amtmann A, Herzyk P (2004) Rank products: a simple, yet powerful, new method to detect differentially regulated genes in replicated microarray experiments. FEBS Lett **573:** 83-92

Cao J, Schneeberger K, Ossowski S, Gunther T, Bender S, Fitz J, Koenig D, Lanz C, Stegle O, Lippert C, Wang X, Ott F, Muller J, Alonso-Blanco C, Borgwardt K, Schmid KJ, Weigel D (2011) Whole-genome sequencing of multiple Arabidopsis thaliana populations. Nat Genet **43:** 956-963

Cardone MF, D'Addabbo P, Alkan C, Bergamini C, Catacchio CR, Anaclerio F, Chiatante G, Marra A, Giannuzzi G, Perniola R, Ventura M, Antonacci D (2016) Inter-varietal structural variation in grapevine genomes. Plant J **88:** 648-661

Carrier G, Le Cunff L, Dereeper A, Legrand D, Sabot F, Bouchez O, Audeguin L, Boursiquot JM, This P (2012) Transposable elements are a major cause of somatic polymorphism in Vitis vinifera L. PLoS One **7:** e32973

Cingolani P, Platts A, Wang le L, 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: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly (Austin) **6:** 80-92

Collins RL, Brand H, Redin CE, Hanscom C, Antolik C, Stone MR, Glessner JT, Mason T, Pregno G, Dorrani N, Mandrile G, Giachino D, Perrin D, Walsh C, Cipicchio M, Costello M, Stortchevoi A, An JY, Currall BB, Seabra CM, Ragavendran A, Margolin L, Martinez-Agosto JA, Lucente D, Levy B, Sanders SJ, Wapner RJ, Quintero-Rivera F, Kloosterman W, Talkowski ME (2017) Defining the diverse spectrum of inversions, complex structural variation, and chromothripsis in the morbid human genome. Genome Biol **18:** 36

Crasta K, Ganem NJ, Dagher R, Lantermann AB, Ivanova EV, Pan Y, Nezi L, Protopopov A, Chowdhury D, Pellman D (2012) DNA breaks and chromosome pulverization from errors in mitosis. Nature **482:** 53-58

Chia JM, Song C, Bradbury PJ, Costich D, de Leon N, Doebley J, Elshire RJ, Gaut B, Geller L, Glaubitz JC, Gore M, Guill KE, Holland J, Hufford MB, Lai J, Li M, Liu X, Lu Y, McCombie R, Nelson R, Poland J, Prasanna BM, Pyhajarvi T, Rong T, Sekhon RS, Sun Q, Tenaillon MI, Tian F, Wang J, Xu X, Zhang Z, Kaeppler SM, Ross-Ibarra J, McMullen MD, Buckler ES, Zhang G, Xu Y, Ware D (2012) Maize HapMap2 identifies extant variation from a genome in flux. Nat Genet **44:** 803-807

D'Amato F (1997) Role of somatic mutations in the evolution of higher plants. Caryologia **50:** 1-15

Da Silva C, Zamperin G, Ferrarini A, Minio A, Dal Molin A, Venturini L, Buson G, Tononi P, Avanzato C, Zago E, Boido E, Dellacassa E, Gaggero C, Pezzotti M, Carrau F, Delledonne M (2013) The high polyphenol content of grapevine cultivar tannat berries is conferred primarily by genes that are not shared with the reference genome. Plant Cell **25:** 4777-4788

Di Gaspero G, Foria S (2015) Molecular grapevine breeding techniques. *In* AG Reynolds, ed, Grapevine Breeding Programs for the Wine Industry. Woodhead Publishing, Oxford, pp 23-37

Di Genova A, Almeida AM, Munoz-Espinoza C, Vizoso P, Travisany D, Moraga C, Pinto M, Hinrichsen P, Orellana A, Maass A (2014) Whole genome comparison between table and wine grapes reveals a comprehensive catalog of structural variants. BMC Plant Biol **14:** 7

Doligez A, Adam-Blondon AF, Cipriani G, Di Gaspero G, Laucou V, Merdinoglu D, Meredith CP, Riaz S, Roux C, This P (2006) An integrated SSR map of grapevine based on five mapping populations. Theor Appl Genet **113:** 369-382

Emanuelli F, Battilana J, Costantini L, Le Cunff L, Boursiquot JM, This P, Grando MS (2010) A candidate gene association study on muscat flavor in grapevine (Vitis vinifera L.). BMC Plant Biol **10:** 241

Emanuelli F, Sordo M, Lorenzi S, Battilana J, Grando MS (2014) Development of user-friendly functional molecular markers for VvDXS gene conferring muscat flavor in grapevine. Mol Breed **33:** 235-241

Fernandez L, Chaib J, Martinez-Zapater JM, Thomas MR, Torregrosa L (2013) Mis-expression of a PISTILLATA-like MADS box gene prevents fruit development in grapevine. Plant J **73:** 918-928

Fernandez L, Torregrosa L, Segura V, Bouquet A, Martinez-Zapater JM (2010) Transposon-induced gene activation as a mechanism generating cluster shape somatic variation in grapevine. Plant J **61:** 545-557

Furiya T, Suzuki S, Sueta T, Takayanagi T (2009) Molecular Characterization of a Bud Sport of Pinot gris Bearing White Berries. Am J Enol Viticult **60:** 66-73

Grimplet J, Van Hemert J, Carbonell-Bejerano P, Diaz-Riquelme J, Dickerson J, Fennell A, Pezzotti M, Martinez-Zapater JM (2012) Comparative analysis of grapevine whole-genome gene predictions, functional annotation, categorization and integration of the predicted gene sequences. BMC Res Notes **5:** 213

Houel C, Chatbanyong R, Doligez A, Rienth M, Foria S, Luchaire N, Roux C, Adiveze A, Lopez G, Farnos M, Pellegrino A, This P, Romieu C, Torregrosa L (2015) Identification of stable QTLs for vegetative and reproductive traits in the microvine (Vitis vinifera L.) using the 18 K Infinium chip. BMC Plant Biol **15:** 205

Hwang S, Kim E, Lee I, Marcotte EM (2015) Systematic comparison of variant calling pipelines using gold standard personal exome variants. Sci Rep **5:** 17875

Ibanez J, Velez MD, de Andres MT, Borrego J (2009) Molecular markers for establishing distinctness in vegetatively propagated crops: a case study in grapevine. Theor Appl Genet **119:** 1213-1222

Ibáñez J, Muñoz-Organero G, Zinelabidine LH, de Andrés MT, Cabello F, Martínez-Zapater JM (2012) Genetic Origin of the Grapevine Cultivar Tempranillo. Am J Enol Viticult **63:** 549-553

Iyer CPAR, G S (1966) Induction of pollen sterility in grapes (*Vitis vinifera*). Vitis **5:** 433-445

Jaillon O, Aury JM, Noel B, Policriti A, Clepet C, Casagrande A, Choisne N, Aubourg S, Vitulo N, Jubin C, Vezzi A, Legeai F, Hugueney P, Dasilva C, Horner D, Mica E, Jublot D, Poulain J, Bruyere C, Billault A, Segurens B, Gouyvenoux M, Ugarte E, Cattonaro F, Anthouard V, Vico V, Del Fabbro C, Alaux M, Di Gaspero G, Dumas V, Felice N, Paillard S, Juman I, Moroldo M, Scalabrin S, Canaguier A, Le Clainche I, Malacrida G, Durand E, Pesole G, Laucou V, Chatelet P, Merdinoglu D, Delledonne M, Pezzotti M, Lecharny A, Scarpelli C, Artiguenave F, Pe ME, Valle G, Morgante M, Caboche M, Adam-Blondon AF, Weissenbach J, Quetier F, Wincker P, French-Italian Public Consortium for Grapevine Genome C (2007) The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. Nature **449:** 463-467

Jones MJ, Jallepalli PV (2012) Chromothripsis: chromosomes in crisis. Dev Cell **23:** 908-917

Kloosterman WP, Tavakoli-Yaraki M, van Roosmalen MJ, van Binsbergen E, Renkens I, Duran K, Ballarati L, Vergult S, Giardino D, Hansson K, Ruivenkamp CA, Jager M, van Haeringen A, Ippel EF, Haaf T, Passarge E, Hochstenbach R, Menten B, Larizza L, Guryev V, Poot M, Cuppen E (2012) Constitutional chromothripsis rearrangements involve clustered double-stranded DNA breaks and nonhomologous repair mechanisms. Cell Rep **1:** 648-655

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

Korbel JO, Campbell PJ (2013) Criteria for inference of chromothripsis in cancer genomes. Cell **152:** 1226-1236

Korbel JO, Urban AE, Affourtit JP, Godwin B, Grubert F, Simons JF, Kim PM, Palejev D, Carriero NJ, Du L, Taillon BE, Chen Z, Tanzer A, Saunders AC, Chi J, Yang F, Carter NP, Hurles ME,

Weissman SM, Harkins TT, Gerstein MB, Egholm M, Snyder M (2007) Paired-end mapping reveals extensive structural variation in the human genome. Science **318:** 420-426

Laucou V, Lacombe T, Dechesne F, Siret R, Bruno JP, Dessup M, Dessup T, Ortigosa P, Parra P, Roux C, Santoni S, Vares D, Peros JP, Boursiquot JM, This P (2011) High throughput analysis of grape genetic diversity as a tool for germplasm collection management. Theor Appl Genet **122:** 1233-1245

Leibowitz ML, Zhang CZ, Pellman D (2015) Chromothripsis: A New Mechanism for Rapid Karyotype Evolution. Annu Rev Genet **49:** 183-211

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

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Genome Project Data Processing S (2009) The Sequence Alignment/Map format and SAMtools. Bioinformatics **25:** 2078-2079

Li Y, Schwab C, Ryan SL, Papaemmanuil E, Robinson HM, Jacobs P, Moorman AV, Dyer S, Borrow J, Griffiths M, Heerema NA, Carroll AJ, Talley P, Bown N, Telford N, Ross FM, Gaunt L, McNally RJ, Young BD, Sinclair P, Rand V, Teixeira MR, Joseph O, Robinson B, Maddison M, Dastugue N, Vandenberghe P, Haferlach C, Stephens PJ, Cheng J, Van Loo P, Stratton MR, Campbell PJ, Harrison CJ (2014) Constitutional and somatic rearrangement of chromosome 21 in acute lymphoblastic leukaemia. Nature **508:** 98-102

Lijavetzky D, Carbonell-Bejerano P, Grimplet J, Bravo G, Flores P, Fenoll J, Hellin P, Oliveros JC, Martinez-Zapater JM (2012) Berry flesh and skin ripening features in Vitis vinifera as assessed by transcriptional profiling. PLoS One **7:** e39547

Lijavetzky D, Ruiz-Garcia L, Cabezas JA, De Andres MT, Bravo G, Ibanez A, Carreno J, Cabello F, Ibanez J, Martinez-Zapater JM (2006) Molecular genetics of berry colour variation in table grape. Mol Genet Genomics **276:** 427-435

Ly P, Teitz LS, Kim DH, Shoshani O, Skaletsky H, Fachinetti D, Page DC, Cleveland DW (2017) Selective Y centromere inactivation triggers chromosome shattering in micronuclei and repair by non-homologous end joining. Nat Cell Biol **19:** 68–75

Malhotra A, Lindberg M, Faust GG, Leibowitz ML, Clark RA, Layer RM, Quinlan AR, Hall IM (2013) Breakpoint profiling of 64 cancer genomes reveals numerous complex rearrangements spawned by homology-independent mechanisms. Genome Res **23:** 762-776

Martinez J, Vicente T, Martinez T, Chavarri JB, Garcia-Escudero E (2006) Una nueva variedad blanca para la D.O.Ca. Rioja: el Tempranillo Blanco[A new white variety for the DOC Rioja: 'Tempranillo Blanco']. *In* 29th OIV World Congress of Vine and Wine. 4th General Assembly of the OIV. Organisation Internationale de la Vigne et du Vin, Logrono; Spain, p 9 pp.

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA (2010) The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res **20:** 1297-1303

Migliaro D, Crespan M, Muñoz-Organero G, Velasco R, Moser C, Vezzulli S (2014) Structural dynamics at the berry colour locus inVitis vinifera L. somatic variants. Aust J Grape Wine R **20:** 485-495

Pelsy F, Dumas V, Bevilacqua L, Hocquigny S, Merdinoglu D (2015) Chromosome replacement and deletion lead to clonal polymorphism of berry color in grapevine. PLoS Genet **11:** e1005081

Pelsy F, Hocquigny S, Moncada X, Barbeau G, Forget D, Hinrichsen P, Merdinoglu D (2010) An extensive study of the genetic diversity within seven French wine grape variety collections. Theor Appl Genet **120:** 1219-1231

Pellestor F (2014) Chromothripsis: how does such a catastrophic event impact human reproduction? Hum Reprod **29:** 388-393

Poduri A, Evrony GD, Cai X, Walsh CA (2013) Somatic mutation, genomic variation, and neurological disease. Science **341:** 1237758

Qi J, Liu X, Shen D, Miao H, Xie B, Li X, Zeng P, Wang S, Shang Y, Gu X, Du Y, Li Y, Lin T, Yuan J, Yang X, Chen J, Chen H, Xiong X, Huang K, Fei Z, Mao L, Tian L, Stadler T, Renner SS, Kamoun S,

Lucas WJ, Zhang Z, Huang S (2013) A genomic variation map provides insights into the genetic basis of cucumber domestication and diversity. Nat Genet **45:** 1510-1515

Quinlan AR, Hall IM (2010) BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics **26:** 841-842

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:** i333-i339

Royo C, Carbonell-Bejerano P, Torres-Perez R, Nebish A, Martinez O, Rey M, Aroutiounian R, Ibanez J, Martinez-Zapater JM (2016) Developmental, transcriptome, and genetic alterations associated with parthenocarpy in the grapevine seedless somatic variant Corinto bianco. J Exp Bot **67:** 259-273

Sambrook J. FEF, Maniatis T. (1989) Molecular cloning : a laboratory manual, Ed 2. New York: Cold Spring Harbor Laboratory Press

Sinha S, Villarreal D, Shim EY, Lee SE (2016) Risky business: Microhomology-mediated end joining. Mutat Res **788:** 17-24

Sorzano CO, Pascual-Montano A, Sánchez de Diego A, Martínez AC, van Wely KH (2013) Chromothripsis: breakage-fusion-bridge over and over again. Cell Cycle **12:** 2016-2023

Stephens PJ, Greenman CD, Fu B, Yang F, Bignell GR, Mudie LJ, Pleasance ED, Lau KW, Beare D, Stebbings LA, McLaren S, Lin ML, McBride DJ, Varela I, Nik-Zainal S, Leroy C, Jia M, Menzies A, Butler AP, Teague JW, Quail MA, Burton J, Swerdlow H, Carter NP, Morsberger LA, Iacobuzio-Donahue C, Follows GA, Green AR, Flanagan AM, Stratton MR, Futreal PA, Campbell PJ (2011) Massive genomic rearrangement acquired in a single catastrophic event during cancer development. Cell **144:** 27-40

Tan EH, Henry IM, Ravi M, Bradnam KR, Mandakova T, Marimuthu MP, Korf I, Lysak MA, Comai L, Chan SW (2015) Catastrophic chromosomal restructuring during genome elimination in plants. Elife **4**

Terol J, Ibanez V, Carbonell J, Alonso R, Estornell LH, Licciardello C, Gut IG, Dopazo J, Talon M (2015) Involvement of a citrus meiotic recombination TTC-repeat motif in the formation of gross deletions generated by ionizing radiation and MULE activation. BMC Genomics **16:** 69

This P, Lacombe T, Cadle-Davidson M, Owens CL (2007) Wine grape (Vitis vinifera L.) color associates with allelic variation in the domestication gene VvmybA1. Theor Appl Genet **114:** 723-730

This P, Lacombe T, Thomas MR (2006) Historical origins and genetic diversity of wine grapes. Trends Genet **22:** 511-519

Thorvaldsdottir H, Robinson JT, Mesirov JP (2013) Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief Bioinform **14:** 178-192

Torregrosa L, Fernandez L, Bouquet A, Boursiquot JM, Pelsy F, Martínez-Zapater JM (2011) Origins and Consequences of Somatic Variation in Grapevine. *In* A-F Adam-Blondon, J-M Martinez-Zapater, C Kole, eds, Genetics, Genomics, and Breeding of Grapes, Ed 1st. Science Publishers, Jersey, British Isles, pp 68-92

Velasco R, Zharkikh A, Troggio M, Cartwright DA, Cestaro A, Pruss D, Pindo M, Fitzgerald LM, Vezzulli S, Reid J, Malacarne G, Iliev D, Coppola G, Wardell B, Micheletti D, Macalma T, Facci M, Mitchell JT, Perazzolli M, Eldredge G, Gatto P, Oyzerski R, Moretto M, Gutin N, Stefanini M, Chen Y, Segala C, Davenport C, Dematte L, Mraz A, Battilana J, Stormo K, Costa F, Tao Q, Si-Ammour A, Harkins T, Lackey A, Perbost C, Taillon B, Stella A, Solovyev V, Fawcett JA, Sterck L, Vandepoele K, Grando SM, Toppo S, Moser C, Lanchbury J, Bogden R, Skolnick M, Sgaramella V, Bhatnagar SK, Fontana P, Gutin A, Van de Peer Y, Salamini F, Viola R (2007) A high quality draft consensus sequence of the genome of a heterozygous grapevine variety. PLoS One **2:** e1326

Vezzulli S, Leonardelli L, Malossini U, Stefanini M, Velasco R, Moser C (2012) Pinot blanc and Pinot gris arose as independent somatic mutations of Pinot noir. J Exp Bot **63:** 6359-6369

Walker AR, Lee E, Bogs J, McDavid DA, Thomas MR, Robinson SP (2007) White grapes arose through the mutation of two similar and adjacent regulatory genes. Plant J **49:** 772-785

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 Mol Biol **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

Wong DC, Schlechter R, Vannozzi A, Holl J, Hmmam I, Bogs J, Tornielli GB, Castellarin SD, Matus JT (2016) A systems-oriented analysis of the grapevine R2R3-MYB transcription factor family uncovers new insights into the regulation of stilbene accumulation. DNA Res **23:** 451-466

Xie C, Tammi MT (2009) CNV-seq, a new method to detect copy number variation using high-throughput sequencing. BMC Bioinformatics **10:** 80

Xu Y, Gao Z, Tao J, Jiang W, Zhang S, Wang Q, Qu S (2016) Genome-Wide Detection of SNP and SV Variations to Reveal Early Ripening-Related Genes in Grape. PLoS One **11:** e0147749

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. Biosci Biotechnol Biochem **70:** 1506-1508

Yang L, Luquette LJ, Gehlenborg N, Xi R, Haseley PS, Hsieh CH, Zhang C, Ren X, Protopopov A, Chin L, Kucherlapati R, Lee C, Park PJ (2013) Diverse mechanisms of somatic structural variations in human cancer genomes. Cell **153:** 919-929

Zhang CZ, Spektor A, Cornils H, Francis JM, Jackson EK, Liu S, Meyerson M, Pellman D (2015) Chromothripsis from DNA damage in micronuclei. Nature **522:** 179-184

Supporting Information

S1 Fig. Microarray-based copy number variation in Tempranillo Blanco white berry somatic variant. Log_2 hybridization signal (TB/TT) ratio for GapeGen GeneChip® probe sets mapping on LGs 2 and 5 is depicted. Values for *VviMybA* probe sets are highlighted in different colors.

S2 Fig. Intra-chromosome translocation in Tempranillo Blanco linkage group 5. The upper panel shows a sharp decay in copy number at the 3' and 5' flanks of the upstream and downstream breakpoints of T5, respectively. The lower panel depicts aligned reads in TT and TB around the positions of the 5' (left) and 3' (right) flanks of the T5 junction (images obtained from IGV viewer). Reads are represented as arrows showing the orientation between paired-end mates. From the PEM analysis, read pairs with abnormally increased insert size are colored and almost only appeared in

TB. Soft-clipped reads spanning the breakpoint show an unaligned extreme (in colors), which matches to the other flank of the translocation junction. All aligned reads are displayed in the IGV first (TT) and second (TB) upper windows, while only discordant reads from TB are displayed in the lower window. The subsets of burgundy read mates align to the reference genome separated by more than 5 Mb.

S3 Fig. Validation of Tempranillo Blanco structural variation junctions in additional Tempranillo accessions. Agarose gel electrophoresis picture showing the PCR amplification product obtained using primer pairs designed to amplify Inv and T1-5 SV junctions. Primer pairs designed for breakpoint junctions predicted in TB genome from WGS data (T1-5 and Inv) as well as the positive control (C) were tested in different TB and TT lines: 1, TT-121; 2, TT-989; 3, TT-1048; 4, TT-1071; 5, TT-1381; 6-8 TT-RJ51; 9, TB-Albelda; 10, TB-Cenicero; 11, TB-Nalda; 12, TB-ICVV3; 13 TB-ICVV2; 14-16, TB-ICVV1 (the origin of each Tempranillo line is described in S9 Table). gDNA obtained from young leaves was used in all PCRs, except for wells 7 and 15 obtained from mature berry skin and wells 8 and 16 from mature berry flesh. Primer pair combinations were used as indicated in Fig 4.

S4 Fig. Transposon and repetitive sequences in breakpoints of Tempranillo Blanco. TE-like sequences matching breakpoint sites are highlighted according to colors in original TEs on the lower part. Breakpoints in the context of the PN40024 12X.0 reference genome assembly chromosomes are represented in the upper part. Original TEs mapping in proximity with breakpoints are represented in the lower part (numbers indicate the start and end positions of the TE).

S1 Table. Oligonucleotides used as primers for PCR amplification.

S2 Table. Heterozygosity analysis in PCR amplified regions around the color locus in linkage group 2.

S3 Table. GrapeGen GeneChip® probe sets showing differential CGH between TB and TT.

S4 Table. Summary of re-sequencing data production (Clean data).

S5 Table. Summary of re-sequencing read alignments.

S6 Table. Genome regions showing LOH in TB.

S7 Table. Genome windows showing copy number decrease in TB.

S8 Table. Sanger capillary electrophoresis-obtained sequences of PCR amplicons confirming breakpoint junctions predicted by genome re-sequencing in TB.

S9 Table. Tempranillo sample origins.

S10 Table. Haplotype phasing of fragments deleted in LGs 2 and 5 of TB.

S11 Table. List of predicted hemizygous variant sites in TB genome and functional effect.

S12 Table. Microsatellite genotyping identification.

Figures

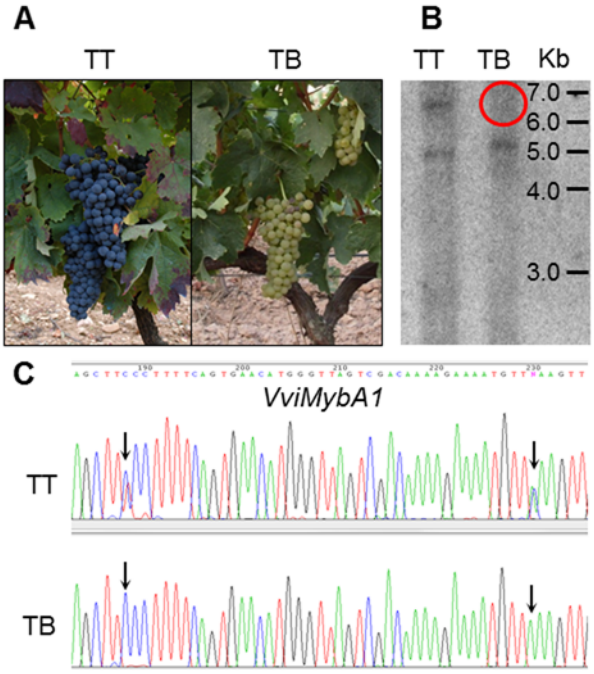


Fig 1. Initial characterization of Tempranillo Blanco white grape variant line. (**A)** TT and TB plant phenotype. Mature plants showing the absence of anthocyanin pigments in TB bunches. (**B)** Southern blot for a *VviMybA1* probe showing the loss of one copy in TB. Only the smallest positive fragment, corresponding to the non-functional allele with a predicted ApaI target site within the *Gret1* retroelement, is kept in TB. (**C)** Loss of *VviMybA1* heterozygosity in TB. Only one *VviMybA1* haplotype is kept in TB according to the loss of heterozygosity at original heterozygous SNPs in TT.

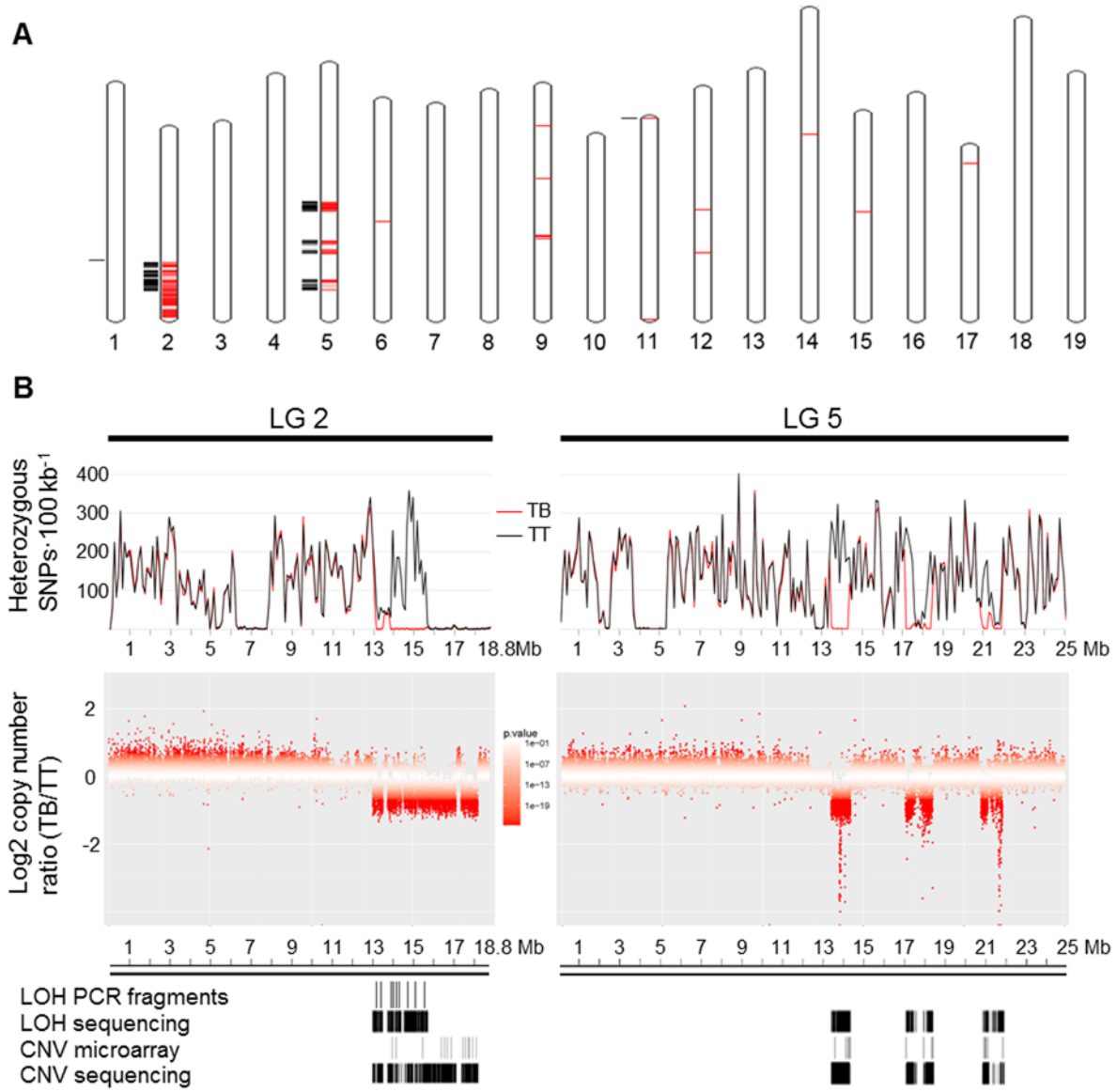


Fig 2. Genome-wide scanning of loss of heterozygosity and copy number decrease in Tempranillo Blanco. A) Genome-wide representation of regions with significant LOH or copy number decrease in TB. Significant intervals were mapped to chromosomes of the PN40024 12X.0 reference genome using the *V. vinifera* whole-genome tool of EnsemblPlants (http://plants.ensembl.org/Vitis_vinifera/Location/Genome). Black outer line, chromosome segment with significant LOH in TB. Red inner line, chromosome segment with significantly decreased copy number in TB. **B)** Close-up of heterozygosity and CNV comparison between TT and TB re-sequencing data

mapping on LGs 2 and 5. Upper panel, the heterozygosity level in TT (black) and TB (red) is depicted. Central panel, copy number ratio between TB and TT is represented. Windows are represented in increasing red color tones corresponding with CNV significance p-values. Lower panel, comparison of LOH and CNV segments in TB detected by WGS and other approaches.

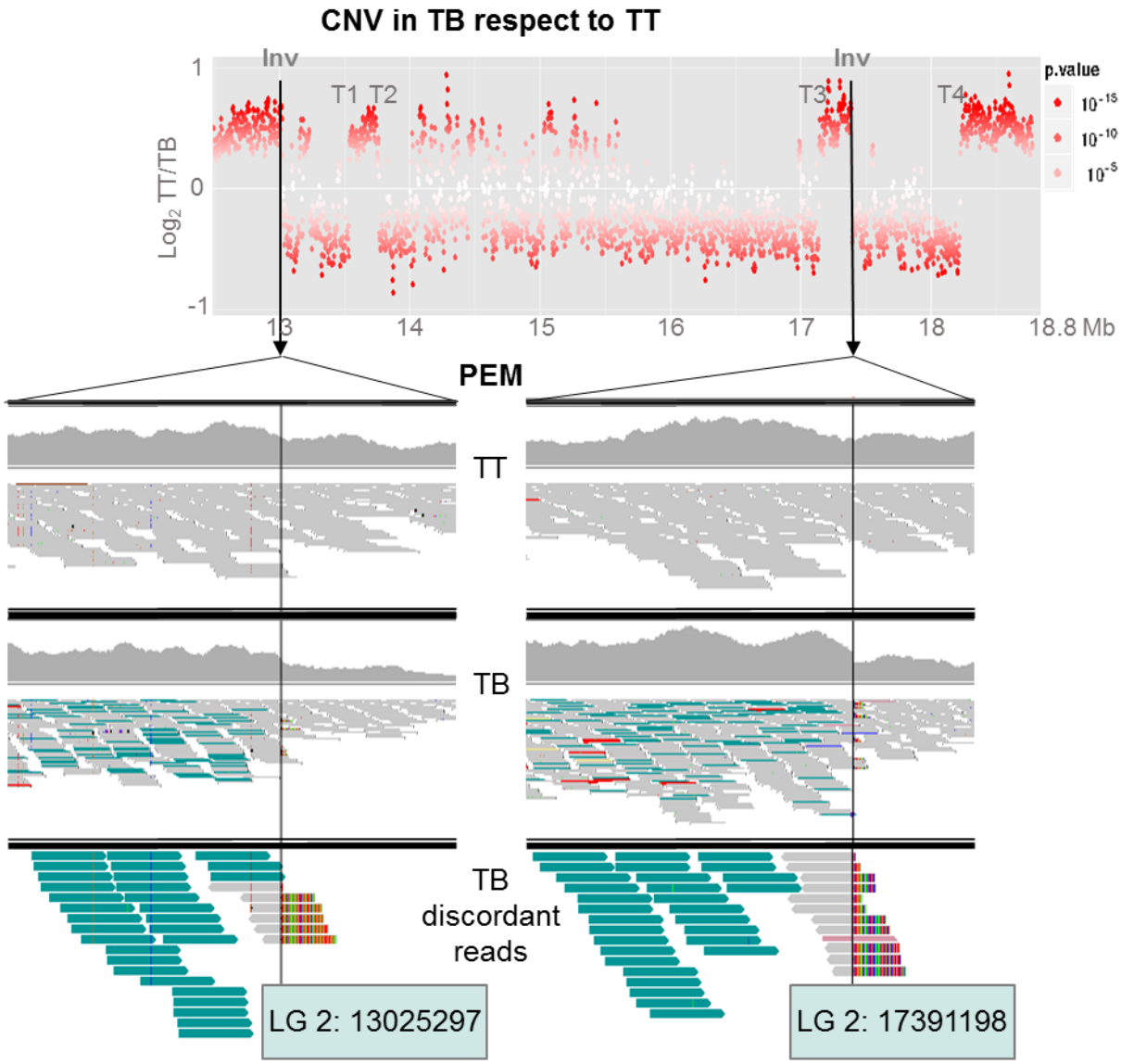


Fig 3. Inversion breakpoints in Tempranillo Blanco linkage group 2. The upper panel shows a sharp decay in copy number at the 3' flank of the two Inv breakpoints. The lower panel shows reads aligned at both Inv breakpoint sites in TT and TB (images obtained from IGV viewer). All aligned reads are displayed in the IGV first (TT) and second (TB) upper windows, while only discordant reads from TB are displayed in the lower window. Reads are represented as arrows showing the orientation of paired-end mates. From the PEM analysis, read pairs with abnormally increased insert size are colored in turquoise and were only detected in TB. These

read pairs are sense oriented as expected in an inversion junction. Soft-clipped reads spanning the breakpoint shows an unaligned extreme in colors.

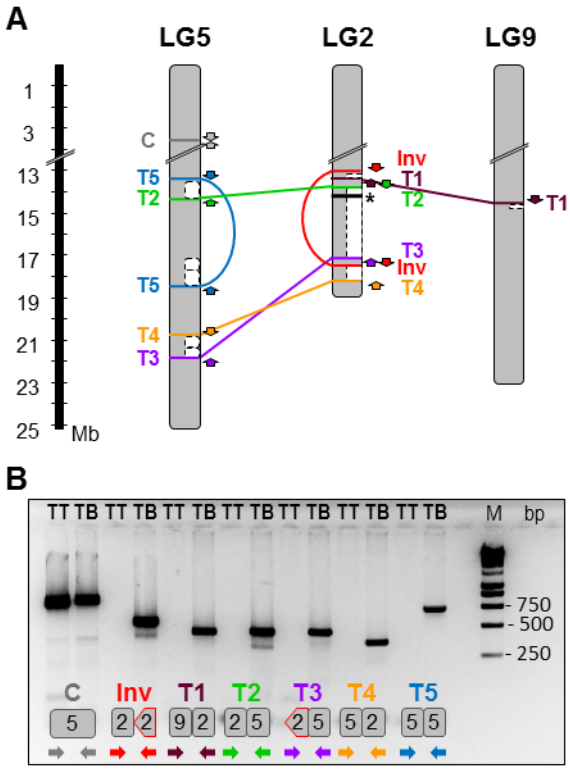


Fig 4. **Mapping and validation of structural variation junctions detected in the genome of Tempranillo Blanco. A)** Scheme depicting SV breakpoints detected in the TB genome. Breakpoints detected by discordant read analysis from WGS data are mapped on the affected chromosomes according to the PN40024 12X.0 reference genome assembly. One inversion (Inv), four inter-chromosomal translocations (T1-4) and one intra-chromosomal translocation (T5) junctions were predicted. Hemizygous chromosome regions detected by LOH and/or CNV from WGS data are indicated with empty boxes. PCR primers used for validation are represented by arrows according to their mapping position and orientation with respect to the reference genome (Inv, red; T1, brown; T2, green; T3, orange; T4, purple; T5, blue; arrow head indicates the 3' end of the primer). An asterisk indicates the position of the cluster of *MybA* genes at the grape color locus on LG 2. **B)** Agarose gel electrophoresis picture showing the PCR amplification product obtained using primer pairs designed to amplify Inv and T1-5 SV junctions. Primer pairs

designed for the TB junctions as well as the positive control (C) were tested for TB and TT. In the gel image, below the two electrophoresis wells corresponding to the same primer pair combination, a scheme illustrating the expected amplicon in TB is shown. Chromosome fragments involved in the rearrangement junction included in the amplicon are represented in the scheme. Inverted fragments are represented as arrow-ended blocks. The alignment of PCR primers colored as in panel A is depicted by arrows below the amplicon. M, DNA size marker; TB, TB gDNA was used as template; TT, TT gDNA was used as template.

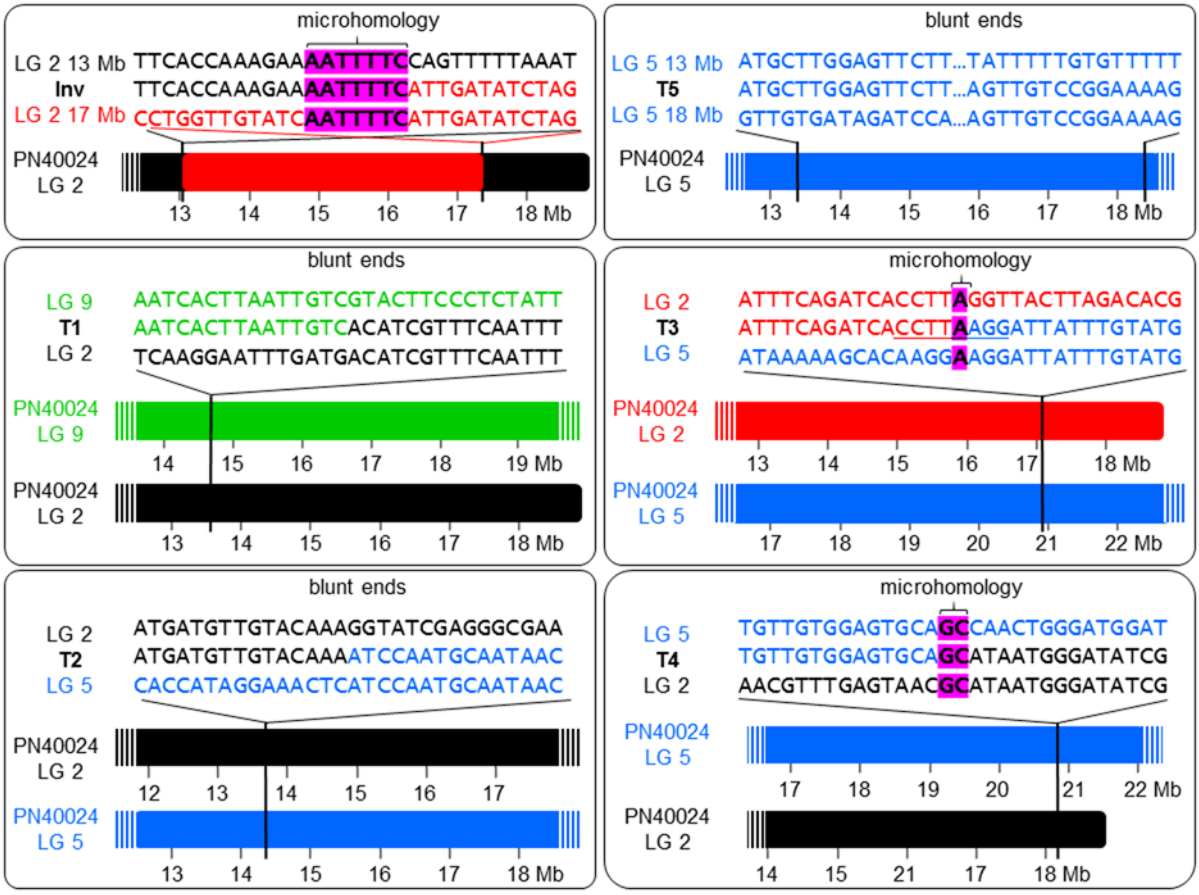
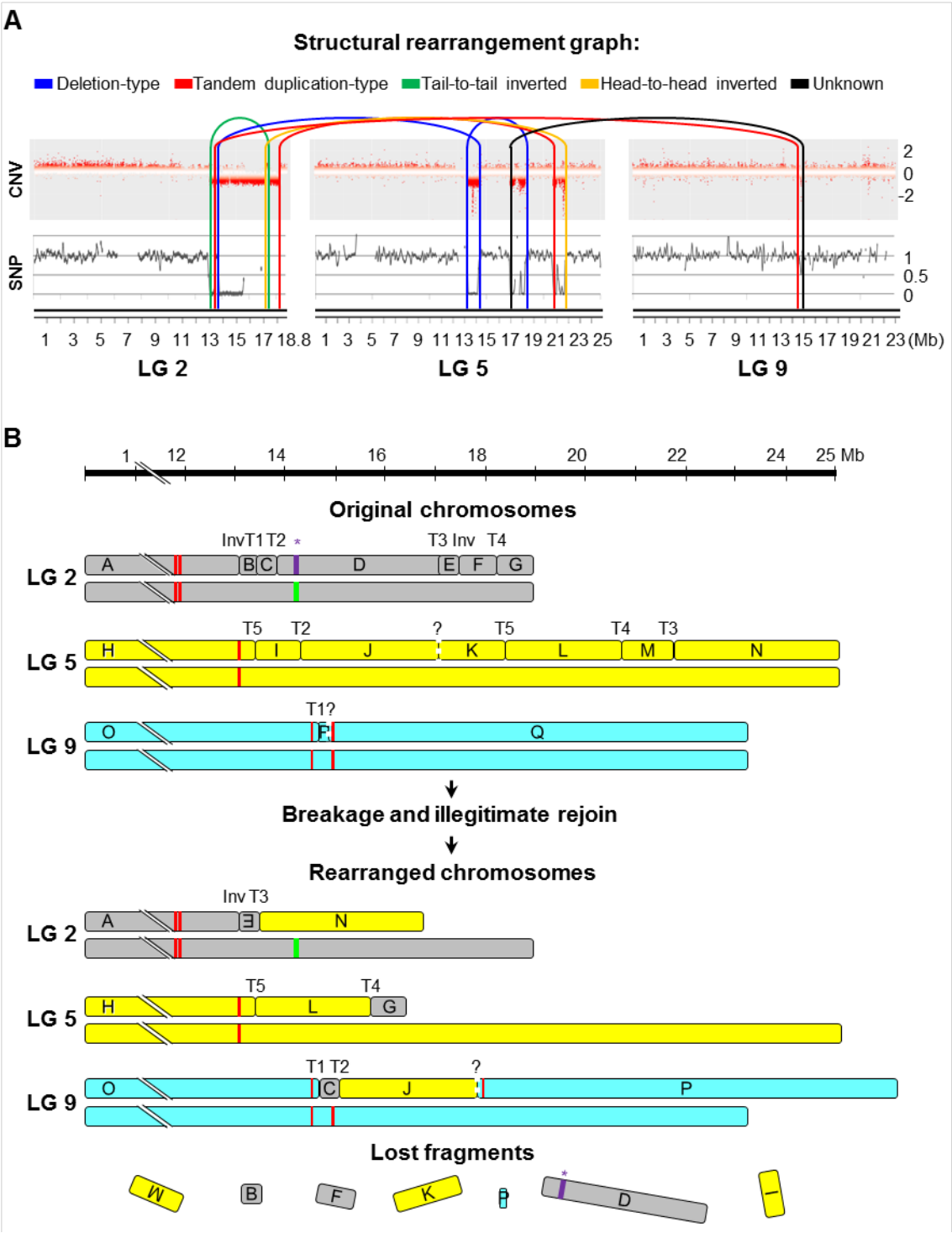


Fig 5. Nature of breakpoint junction sequences detected in Tempranillo Blanco. For each validated SV junction site, close-ups of breakpoint and junction site sequences are shown in the upper panel (LG 2 with regular orientation in black, LG 2 with inverted orientation in red, LG 5 in blue, LG 9 in green). Breakpoints in the context of the PN40024 12X.0 reference genome assembly chromosomes are represented in the lower part.



1186

1187 **Fig 6. DNA structural rearrangement graph and predicted rearranged**
1188 **chromosomes in Tempranillo Blanco. A)** Structural rearrangement graph depicting
1189 CNV, LOH and rearrangement joins and types in the affected LGs. CNV is shown as
1190 Log_2 (copy number ratio TB/TT) calculated for each window of the affected
1191 chromosomes by CNV-seq. SNP heterozygosity level is depicted as Log_2

(heterozygous SNPs per 100 kb ratio TB/TT). Connections between rearranged fragments are depicted by curved lines colored as indicated at the top of the image according to join types observed in the paired-end alignment analysis. **B)** Diagram depicting the original and the predicted rearranged order of chromosome fragments affected by the chromothripsis-like restructuring that originated the variant genome of TB. Deleted fragments in TB were predicted by LOH and CNV from WGS data and their extremes were confirmed by PCR of breakpoint junctions in most cases. Fragment order in TB derivative chromosomes was predicted considering the six deleted fragments and the six breakpoint junction sequences detected in the discordant read analysis that were validated by PCR (Inv, inversion; T, translocation). Question marks (?) represent an additional breakpoint junction that was hypothesized from WGS read pairs that collectively did not pass all the SV analysis significance thresholds. Additional undetected rearranged fragments are also conceivable as explained in the text. Rearrangements are represented in a single copy of the affected chromosomes according to the linkage of deleted fragments in the same chromosome haplotype, which was inferred from a TT self-cross progeny segregation analysis and WGS data of TB. In purple, an asterisk indicates the segment comprising the cluster of *MybA* (*) genes belonging to the grape color locus functional allele, which is present in heterozygous state in TT and is lost after genome rearrangements in TB. The white allele is represented by a green line. Red lines denote centromeric repeats according to predictions on the grapevine reference genome (Di Gaspero and Foria 2015 and Genoscope website).

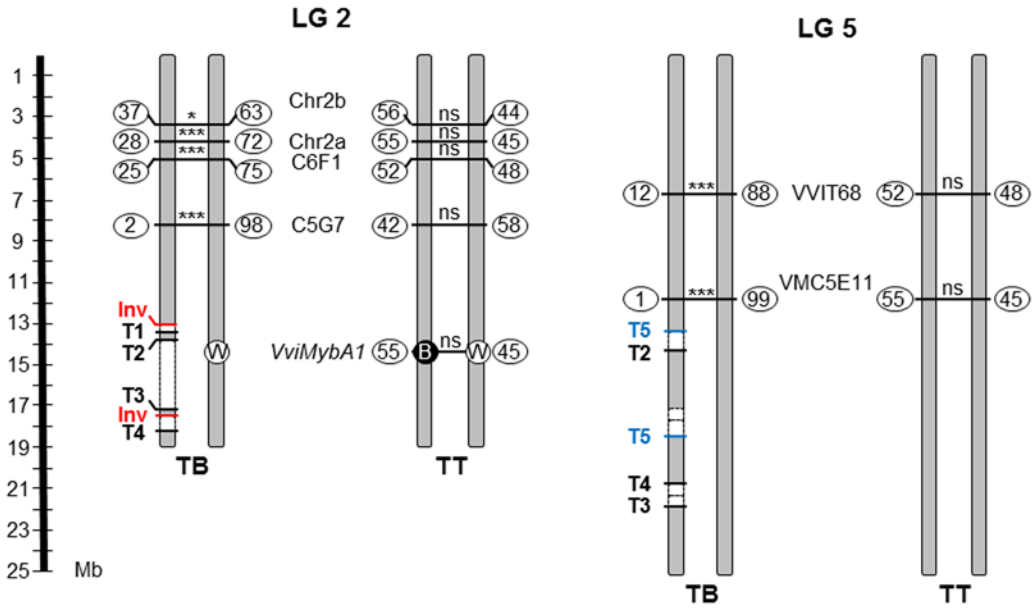


Fig 7. Segregation of molecular markers at LGs 2 (left) and 5 (right) in TB and TT selfed progenies. Markers are mapped to the PN40024 reference genome chromosomes. Absolute frequencies (%) for each allele are shown inside circles, and significance value of χ^2 tests for the difference between the two allele frequencies is indicated for every marker (*: P <0.05, and ***: P <0.0001). W and B, white and black berry color alleles in the *VviMybA1* locus, respectively. Validated inversion (Inv) and translocation (T1-5) breakpoints are depicted as well. Deleted chromosome segments detected by CNV analysis from WGS data are depicted as white boxes and linked in the same chromosome copy as determined by haplotype phasing using a TT self-cross progeny.

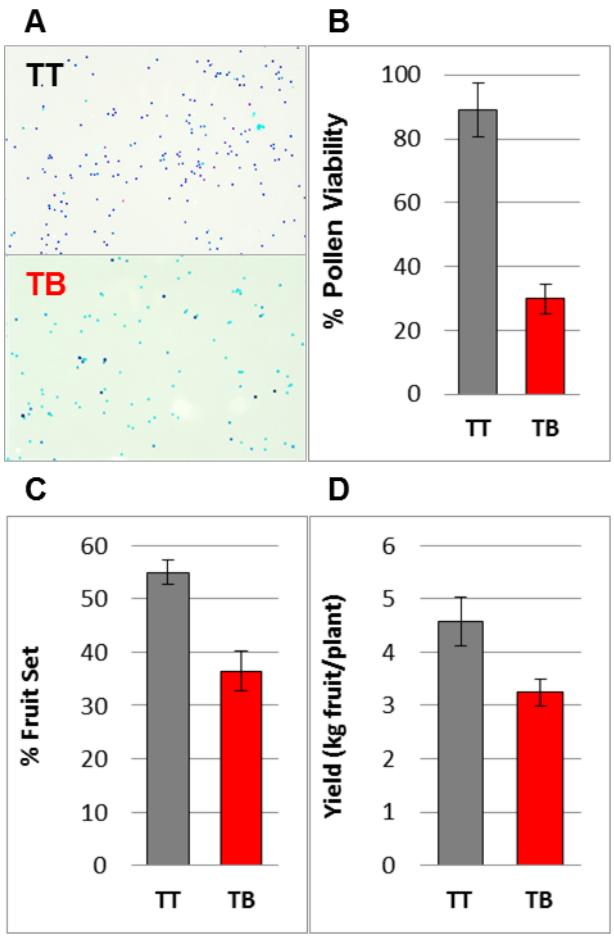


Fig 8. Pollen viability, fruit set and yield in Tempranillo Blanco and Tempranillo Tinto. **A)** TT and TB pollen after Alexander's staining. Dark pollen is viable, while empty (pale) pollen is sterile. **B)** Percentage of pollen viability, **C)** fruit set rate and **D)** fruit yield in both Tempranillo lines. In B-D panels, depicted data correspond to mean±SD values. All data was recorded in the experimental vineyard of "Finca La Grajera" during 2015 season. Similar results were obtained in other two seasons.