bioRxiv preprint doi: https://doi.org/10.1101/2020.10.27.356790; this version posted October 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1 **Title page**

2 # Clonal propagation history shapes the intra-cultivar genetic diversity in 'Malbec'
3 # grapevines

4 Luciano Calderón[1*], Nuria Mauri[2], Claudio Muñoz[3], Pablo Carbonell-Bejerano[4], Laura Bree[5],
5 Cristobal Sola[5], Sebastian Gomez-Talquenca[6], Carolina Royo[2], Javier Ibáñez[2], Jose Miguel
6 Martinez-Zapater[2] & Diego Lijavetzky[1*]

7 **Authors affiliations**

8 1- Instituto de Biología Agrícola de Mendoza (IBAM, CONICET-UNCuyo). Almirante
9 Brown 500, M5528AHB. Chacras de Coria, Mendoza, Argentina
10 2- Instituto de Ciencias de la Vid y del Vino (CSIC, UR, Gobierno de La Rioja). Finca La
11 Grajera Autovía del Camino de Santiago LO-20 Salida 13, 26007. Logroño, Spain
12 3- Facultad de Ciencias Agrarias, Universidad Nacional de Cuyo. Almirante Brown 500,
13 M5528AHB. Chacras de Coria, Mendoza, Argentina
14 4- Max Planck Institute for Developmental Biology. Max-Planck-Ring 9, D-72076.
15 Tübingen, Germany
16 5- Vivero Mercier Argentina. Ruta 40 Km 3273, M5509. Perdriel, Mendoza, Argentina.
17 6- Laboratorio de Virología de Plantas, EEA Mendoza INTA, 5507. Luján de Cuyo,
18 Mendoza, Argentina

19 *** Corresponding authors contacts:**

20 Luciano Calderón:
21 lcalderon@mendoza-conicet.gob.ar ; lucianocalderon@yahoo.com.ar
22 Diego Lijavetzky:
23 dlijavetzky@conicet.gov.ar ; dlijavetzky@gmail.com

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.27.356790; this version posted October 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

24 **Abstract**

25 Grapevine (*Vitis vinifera* L.) cultivars are clonally propagated to preserve their varietal
26 attributes. However, novel genetic variation still accumulates due to somatic mutations. Aiming
27 to study the potential impact of clonal propagation history on grapevines intra-cultivar genetic
28 diversity, we have focused on 'Malbec'. This cultivar is appreciated for red wines elaboration,
29 it was originated in Southwestern France and introduced into Argentina during the 1850s. Here,
30 we generated whole-genome resequencing data for four 'Malbec' clones with different
31 historical backgrounds. A stringent variant calling procedure was established to identify reliable
32 clonal polymorphisms, additionally corroborated by Sanger sequencing. This analysis retrieved
33 941 single nucleotide variants (SNVs), occurring among the analyzed clones. Based on a set of
34 validated SNVs, a genotyping experiment was custom-designed to survey 'Malbec' genetic
35 diversity. We successfully genotyped 214 samples and identified 14 different clonal genotypes,
36 that clustered into two genetically divergent groups. Group-Ar was driven by clones with a long
37 history of clonal propagation in Argentina, while Group-Fr was driven by clones that have
38 longer remained in Europe. Findings show the ability of such approaches for clonal genotypes
39 identification in grapevines. In particular, we provide evidence on how human actions may have
40 shaped 'Malbec' extant genetic diversity pattern.

41 **Introduction**

42 Clonal propagation is a common practice in perennial crops. In this kind of growing system, a
43 scarce genetic variability could be expected among clones within a given cultivar. However,
44 intrinsic genetic mechanisms such as somatic mutations keep occurring and accumulating along
45 cultivars' history [[1]]. Grapevine (*Vitis vinifera* L.) cultivars are perennial crops that consist on
46 highly heterozygous genotypes, originated from a sexual cross and clonally propagated to
47 preserve their productive traits [[2]]. Grapevine is among the top five fruit crops in terms of tons
48 produced worldwide [[3]] and it possesses a rather relatively small genome size (~480 Mb) [[4]].
49 The described features, turn this species into an attractive model for studying the impact of
50 somatic mutations on the genetic diversity of clonal crops [[5–7]]. In this regard, there are many
51 well-documented cases of somatic mutations affecting traits of productive interest in
52 grapevines, mainly involved in berry color determination [[8–10]], berry aroma [[11]], cluster shape
53 [[12,13]] and reproductive development [[14,15]]. However, somatic mutations do not always have
54 qualitative consequences, and quantitative effects have also been reported among clones [[16,17]],
55 even with the responsible mutations identified at the nucleotide resolution level [[18]]. But most
56 of the occurring somatic mutations might not have phenotypic consequences, nonetheless these

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.27.356790; this version posted October 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

57 'silent' variants still constitute a valuable resource of genetic diversity [5,6]. For example, they
58 may be used in marker assisted selection programs [19,20] or to provide insight on the historical
59 processes shaping current genetic diversity patterns [1,21,22].
60 When analyzing genetic diversity among different grapevine cultivars genetic variation
61 turns very clear [23,24]. However, studying the genetic diversity at the intra-cultivar level is more
62 challenging. This limitation is based on the expected low variability and because traditional
63 markers such as SSRs and SNPs selected from inter-cultivar polymorphisms, have shown low
64 efficiency in such approaches [25–28]. The increased accessibility to genome-wide scale
65 sequencing has made possible to more accurately address this issue [5,6,29,30].
66 Here we focused on 'Malbec' cultivar, which prime name is 'Cot' [31]. This cultivar has
67 for long been appreciated for the elaboration of high-quality red wines [32]. According to the
68 genetic evidence [33] and historical records [32,34], 'Malbec' was originated from the outcrossing
69 of cultivars 'Prunelard' and 'Magdeleine Noir des Charentes', in Southwestern France (Cahors
70 region). 'Malbec' was then introduced into Argentina (Mendoza province) during the 1850s
71 [32,34]. In fact, in this South American region is where the largest volumes of 'Malbec' wine has
72 been produced for the past two decades [35]. 'Malbec' shows a notorious clonal phenotypic
73 diversity [16,36] and a great adaptation capacity, being successfully introduced into a wide range
74 of agroecological conditions across Argentina [37]. However, little is known about 'Malbec'
75 inter-clonal genetic diversity. If we track back its clonal propagation history, we can spotlight
76 milestones that could have shaped the current pattern of genetic diversity. Starting from the
77 single seedling that became cultivated after the mentioned outcross, followed by a "bottleneck
78 effect" when it was initially introduced into South America. To the independent accumulation
79 of somatic mutations, as consequence of clonal propagation under different environmental
80 pressures and selection criteria.
81 In this work, we surveyed 'Malbec' intra-cultivar genetic diversity with focus on the
82 impact of its particular clonal propagation history. We implemented a whole genome
83 resequencing (WGR) approach to discover single nucleotide variants (SNVs), occurring among
84 four clones with different historical backgrounds. Then, after a validation process, a reduced
85 set of the identified SNVs was employed to perform a genotyping analysis to survey the genetic
86 diversity across an extensive sampling.

87 Results

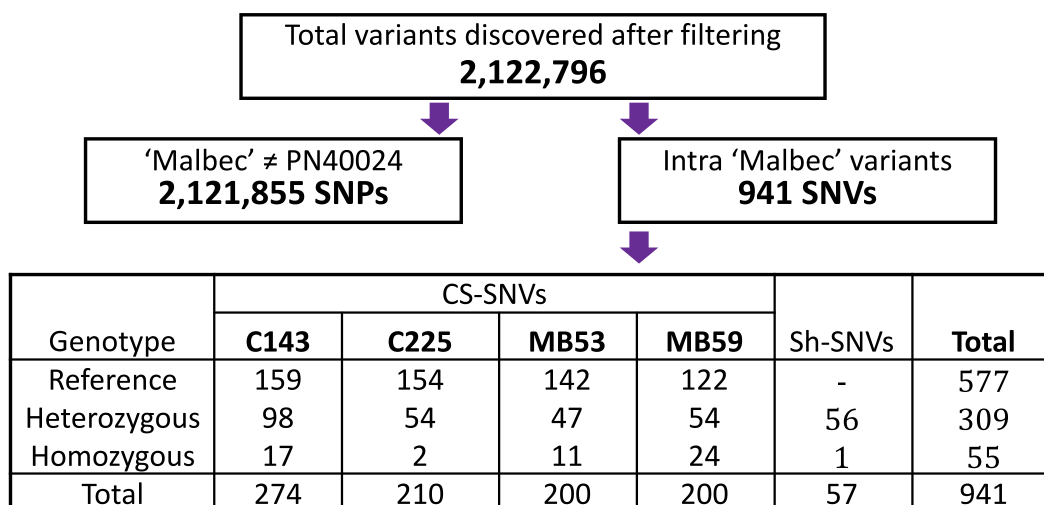
88 **a. Genetic diversity among 'Malbec' clones is 2000-fold lower than compared to**
89 **grapevine's reference genome.** We performed WGR of four Malbec clones: MB53, MB59,

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.27.356790; this version posted October 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

C225 and C143, that differed in their time span of clonal propagation in Argentina. In total, ~90 million paired-end reads per clone were produced, adding more than 45 Gb of sequence (Supplementary Table S1). Filtered reads were aligned to the *Vitis vinifera* L. reference genome PN40024 [4] (hereafter: PN40024), covering ~78% of its length, with a read depth of ~30x (Supplementary Table S1).

After variant calling and filtering processes, we discovered 2,122,796 variants in total (Figure 1). More precisely, we detected 2,121,855 single nucleotide polymorphisms

Figure 1



Genotype	CS-SNVs C143	CS-SNVs C225	CS-SNVs MB53	CS-SNVs MB59	Sh-SNVs	Total
Reference	159	154	142	122	-	577
Heterozygous	98	54	47	54	56	309
Homozygous	17	2	11	24	1	55
Total	274	210	200	200	57	941

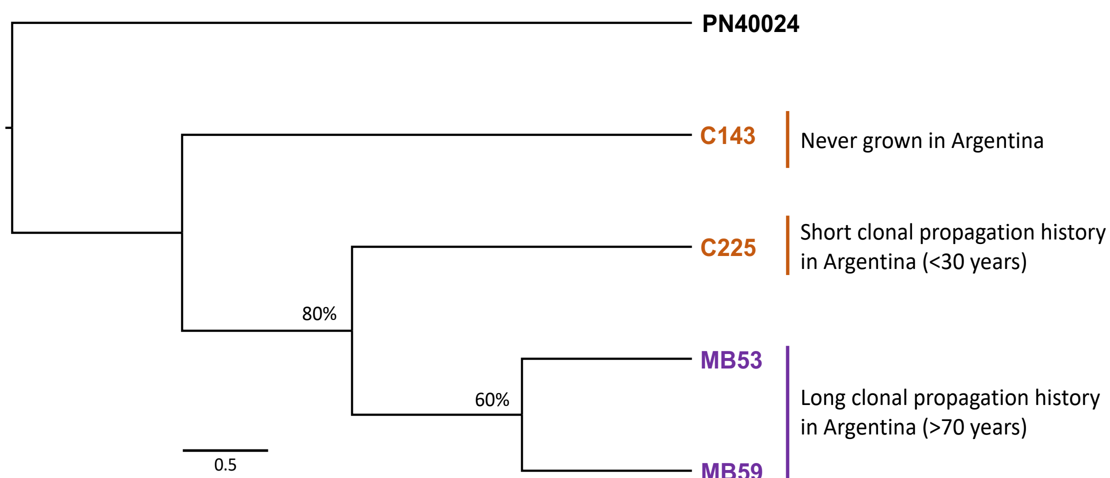
(SNPs), defined here as common variants to the four clones differentiating 'Malbec' from PN40024. We also identified 941 single nucleotide variants (SNVs), defined as variants distinguishing 'Malbec' clones among each other. From which, 884 were clone-specific (hereafter: CS-SNVs), meaning that one clone had a genotype different from the other three clones. While 57 were shared SNVs (hereafter: Sh-SNVs), meaning that two clones presented the same genotype, different from the other two. Genotypes for CS-SNVs were classified as: Heterozygous (*Het*) = one clone with a heterozygous alternative allele not observed in the other three (253 CS-SNVs); Reference (*Ref*) = one clone showed the reference allele in homozygosis and the other three shared an alternative allele (577 CS-SNVs); and Homozygous (*Hom*): one clone with an homozygous alternative allele and the other three clones were either *Het* or *Ref* (54 CS-SNVs) (Figure 1). Even though CS-SNVs were rather evenly distributed among the four analyzed clones, C143 still showed the highest number (Figure 1). Sh-SNVs genotypes were defined as *Het* and *Hom* when two clones shared the same alternative allele in a heterozygous or homozygous state, respectively. Only a single Sh-SNVs was *Hom*, shared by

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.27.356790; this version posted October 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

113 C143 and C225, that position remained *Het* in MB53 and MB59. The remaining Sh-SNVs were
114 *Het* and distributed as follows: 17 were shared by MB53-MB59 and 17 by C143-C225, while
115 the remaining 22 Sh-SNVs were shared in different combinations: C143-MB53 = 3, C225-
116 MB59 = 4, C225-MB53 = 6 and C143-MB59 = 9.

117 We performed a phylogenetic analysis based on the 941 SNVs using PN40024 genotype
118 as an outgroup, and observed that the genetic relations among the four resequenced clones were
119 associated to their clonal propagation history (Figure 2).

120 **Figure 2.**



121

122 More precisely, clone C143 -never grown in Argentina- turned out to be the most genetically
123 divergent from the other three. While C225, with a short history of clonal propagation in
124 Argentina (<30 years), differentiated (80% bootstrap support) from MB53 and MB59. Finally,
125 MB53 and MB59, the two clones that have been longer propagated in Argentina (>70 years)
126 appeared also divergent (60% bootstrap support), but more closely related between each other
127 than to the other two clones.

128 Out of the 941 described SNVs, 34 were chosen for validation through Sanger
129 sequencing (Supplementary Table S2). All the sequenced SNVs showed the expected allelic
130 states for the corresponding clone, demonstrating the reliability of the employed bioinformatic
131 procedures. As an example, we show the electropherogram alignments of four validated CS-
132 SNVs (one for each of the resequenced clones) (Supplementary Fig. S1).

133 **b. Genotyping analysis shows that genetic diversity pattern in 'Malbec' is related to**
134 **clones' propagation history.** Genetic diversity was surveyed using a custom-designed
135 genotyping chip. We selected 48 SNVs (including the mentioned 34 validated ones), with 42
136 CS-SNVs and 6 Sh-SNVs (Supplementary Table S3). Only heterozygous alternative variants

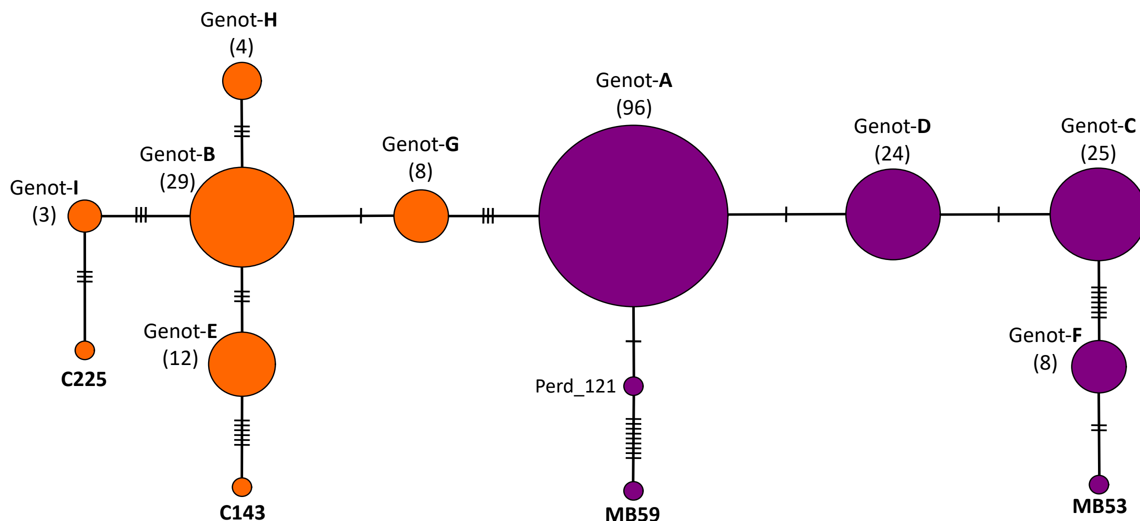
bioRxiv preprint doi: https://doi.org/10.1101/2020.10.27.356790; this version posted October 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

137 were selected, based on their ability to distinguish among the four resequenced clones. Final
138 analyses were performed on 214 successfully genotyped 'Malbec' accessions based on 41
139 properly working SNVs (37 CS-SNVs and 4 Sh-SNVs). We discarded seven out of the 48
140 starting SNVs and five out of the 219 starting samples, due to technical problems related to
141 missing data. Based on the resequenced clone for which they were originally identified, the 37
142 properly working CS-SNVs distributed as follows: nine for C143, seven for C225, eleven for
143 MB53 and ten for MB59; while the four Sh-SNVs corresponded to variants shared by MB53
144 and MB59. Regarding the 41 SNVs variability, as expected for *de novo* mutations, most of them
145 consisted in transitional mutations and only eight were transversion
146 (transitions/transversions=4.1). A total of 22 SNVs markers in the chip were widely informative
147 across the surveyed clonal population, as they ranged from 2 to 164 samples showing the
148 alternative heterozygous allele. Only one of the latter (C225-snv4) showed the three possible
149 genotypes, including the alternative allele in homozygosis. Finally, 19 CS-SNVs showed the
150 alternative heterozygous allele only for one of the four resequenced clones (Supplementary
151 Table S4), which were analyzed with the genotyping chip as a proof of concept of its precision.
152 In fact, the four resequenced clones showed in the chip the expected alternative allele for the
153 respective CS-SNVs, in agreement with the WGR data (Supplementary Table S5).

154 The genotypes of the 214 samples based on 41 SNVs (Supplementary Table S5),
155 constituted the genotypic dataset used in the subsequent genetic diversity analyses. We built a
156 Median-Joining network, which identified 14 different clonal genotypes: five singletons (i.e.
157 genotypes observed uniquely for one sample) and nine genotypes that were represented by more
158 than one sample (named A to I) (Figure 3). Most genotypes

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.27.356790; this version posted October 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

159 **Figure 3.**



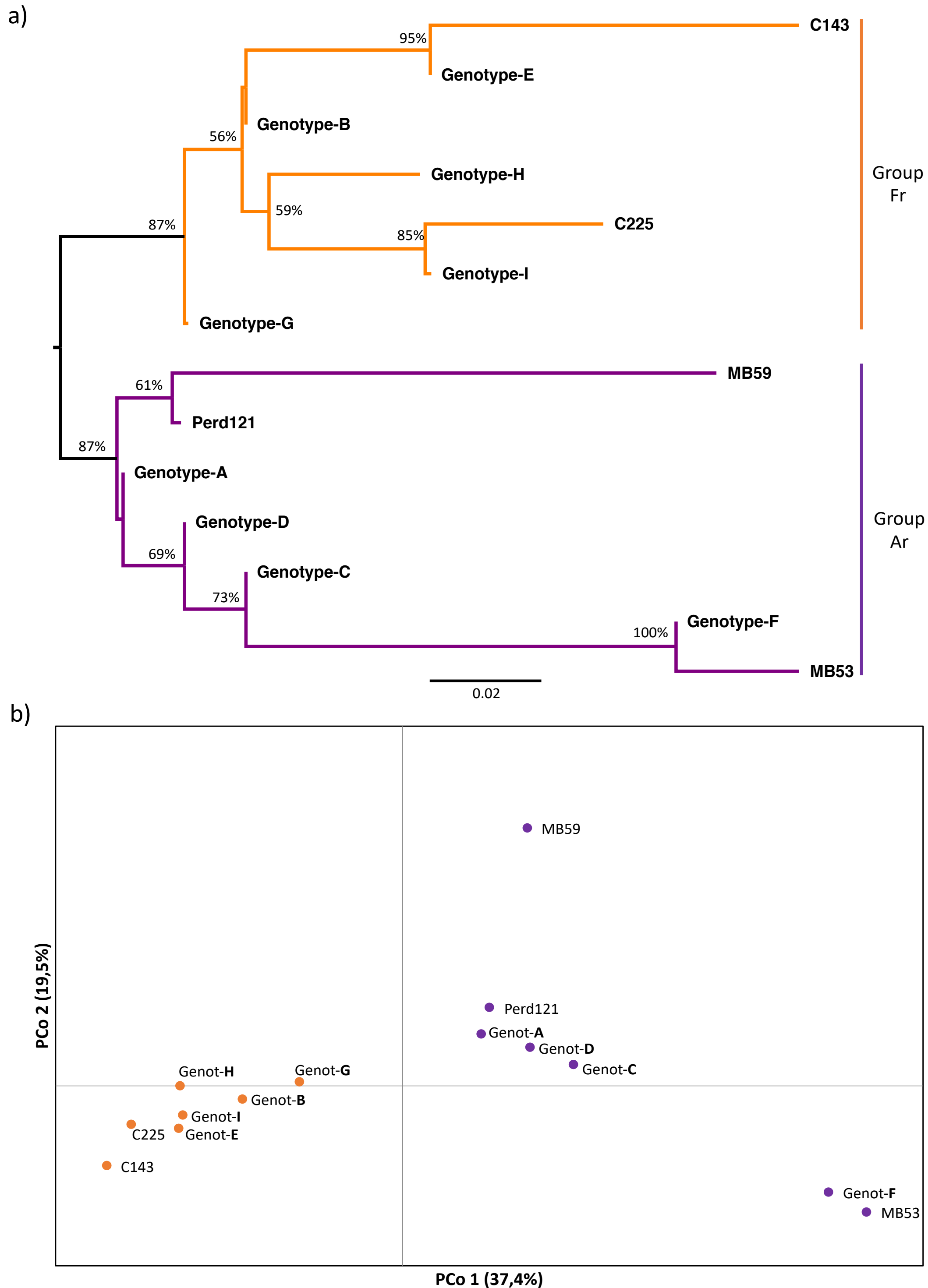
160

161 differentiated each other by one, two or three SNVs; except for Genotype-F that accumulated
162 seven, C143 six and MB59 nine SNVs, that differentiated them from their respective closest
163 genotype. The number of samples represented by each genotype ranged from 96 (Genotype-A),
164 comprising 45% of the analyzed accessions, to three (Genotype-I). After inspecting the origin
165 of the samples, no association was observed between the mass selections and the genotypes
166 assignment. Meaning that most genotypes had representatives of samples coming from different
167 mass selections (Supplementary Table S6). The five singleton genotypes corresponded to a
168 sample from *Perdriel* mass selection (Perd_121) and to the four resequenced clones. As
169 expected from the WGR origin of markers in the chip, MB53, MB59, C143 and C225 were the
170 most differentiated samples (Figure 3), due to the effect of CS-SNVs that were variable only
171 for each of them (Supplementary Table S4). Nonetheless, after a more stringent analysis based
172 only on the 22 SNVs showing at least two samples with the alternative allele, the main nine
173 genotypes were recovered (Supplementary Fig. S2). The difference was that C225 and MB53
174 were the only two samples still differentiating as singletons, while C143, MB59 and Perd_121
175 were included in Genotypes E, F and A respectively.

176 We tested for the phylogenic relations among the 14 identified clonal genotypes based
177 on 41 SNVs. The analysis included a unique sequence representing each of the nine Genotypes
178 from A to I and the five singletons. The resulted tree displayed the existence of two divergent
179 clades, named Group-Ar (Argentina) and Group-Fr (France) (Figure 4a).

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.27.356790; this version posted October 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

180 **Figure 4.**



181

182 Group-Ar was driven by the resequenced clones with >70 years of clonal propagation in
183 Argentina (MB53 and MB59) and clustered the closely related genotypes A, C, D and F. Jointly,
184 these genotypes represented the great majority of the analyzed samples (155), including also
185 the singleton genotype Perd_121. While Group-Fr was driven by the resequenced clones that
186 longer remained close to the origin of Malbec in France, never grown or with less than 30 years

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.27.356790; this version posted October 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

187 of clonal propagation in Argentina (C143 and C225), clustering all the genotypes closer related
188 to them: E, B, G, H, I. In total, 64 samples clustered in Group-Fr, including all the other
189 analyzed samples with less than 30 years of clonal propagation in Argentina (Cot42, Cot46,
190 Cot595, Cot596, Cot598, Inta19) or never grown outside Europe (Esp217). Even though
191 Genotype-G was clearly differentiated from other genotypes of Group-Fr (Figure 4a), all
192 performed analyses consistently placed it closer to genotypes from this group (Figure 4b and
193 AMOVA). The distinction between Group-Ar and Group-Fr was also observed in the Principal
194 Coordinate Analysis (PCoA), where the PCo1 and PCo2 explained almost 55% of the genotypic
195 variance (Figure 4b). The separation between the two Groups was mainly depicted by PCo1
196 (37.4%), all genotypes with a closer genetic distance to C143 and C225 clustered together
197 (including Genotype-G) and the same occurred for genotypes closer related to MB53 and
198 MB59, although with a larger dispersion (Figure 4b). PCoA based only on the four Sh-SNVs,
199 also recovered the distinction between Groups Ar and Fr. Again Genotype-G clearly
200 differentiated from the two Groups, but it was still closer to Group-Fr (Supplementary Fig. S3).
201 Finally, the AMOVA results indicated that a significant proportion of the total molecular
202 variance, was explained after grouping and contrasting the genotypes included in Group-Ar and
203 Group-Fr. The highest AMOVA value was reached when Genotype-G was included in Group-
204 Fr, Phi_{PT} = 0,39 (p = 0,001).

205 **Discussion**

206 Extant cultivated grapevines (*V. vinifera* ssp. *sativa*) have retained most of the genetic diversity
207 present in their wild counterpart, ssp. *sylvestris* [21,38]. This genetic diversity is evidenced
208 through the great variability observed among cultivars [7,24]. However, genetic variation is
209 strongly reduced at the intra-cultivar level. Here, we surveyed the genetic variation in *V.*
210 *vinifera* L. cv. 'Malbec' and found evidence on how clonal propagation history has shaped the
211 diversity pattern of this cultivar.

212 Somatic mutations mostly accumulate as heterozygous variants, which are more prone
213 to generate false positives in variant calling analyses. Therefore, a major challenge when
214 processing high-throughput genomic data, for clonal genetic diversity studies, consists on
215 avoiding variants overestimation [39,40]. Being stringent with the bioinformatic procedures, as
216 well as experimental corroboration of the called variants might provide more certainty on this
217 regard. Here, we worked with a set of variants that were consistently called by three different
218 software: GATK [41], BCFTOOLS [42] and VARSCAN2 [43]. Afterwards, stringent
219 bioinformatic filters, particularly related to the read edit distance and the variant allele

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.27.356790; this version posted October 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

220 frequency (VAF), were applied to discard spurious variants. Moreover, experimental
221 corroboration was successfully performed by means of two alternative technologies, Sanger
222 sequencing (Supplementary Fig. S1) and a Fluidigm genotyping-chip (Supplementary Table
223 S5). All tested SNVs showed the expected alternative allele for the expected sample, so we
224 could assume for a non-significant proportion of false positives among the identified variants.
225 This stringent workflow allowed us to obtain a reliable set of SNVs and set-up a genotyping
226 experiment, to analyze the clonal genetic diversity in 'Malbec'.

227 We observed that the number of SNPs distinguishing 'Malbec' from the PN40024
228 grapevine reference genome, exceed the intra-cultivar SNVs by three orders of magnitude
229 (Figure 1). The identified number of SNPs in the present study is within the range of those
230 reported in other works, that have also compared the genetic diversity between grapevine
231 cultivars [[44–46]]. While previous works that studied the intra-cultivar genetic diversity (using
232 WGR data) identified total numbers of SNVs ranging from the few thousand in 'Chardonnay'
233 [[30]] and 'Nebbiolo' [[5]], to the several thousand in 'Zinfandel' [[6]]. Here, we present the lowest
234 total number of SNVs reported so far (Figure 1). A reason for this might be that assuming the
235 presence of putative false negatives in the variant calling procedures, we aimed for a stringent
236 filtering to yield reliable markers for clonal lineages identification. However, final results
237 reported in each genetic diversity analysis might be differentially influenced by other technical
238 (e.g. sequencing methods) and biological aspects (e.g. genetic distance among the analyzed
239 clones), as well as by the aim of the analysis [[47,48]]. Regardless of the differences in the absolute
240 numbers of SNVs identified, it is clear that in grapevines the intra-cultivar genetic diversity
241 drops drastically when compared to the inter-cultivar. This observation corroborates the role of
242 vegetative propagation in preserving the desired phenotypes of cultivars, by stabilizing the
243 accumulation rate of novel genetic variation [[2]].

244 Despite the scarce intra-cultivar genetic diversity in grapevines, the identified variants
245 successfully distinguished 14 different clonal genotypes of 'Malbec' (Figure 3). We found no
246 association between the genotype assignment and the mass selection origin of our samples.
247 Plants from the same mass selection share particular phenotypic traits of productive interest
248 (Supplementary Table S7). However, the sought phenotypic homogeneity contrasts with the
249 observed genetic diversity, suggesting that SNVs analyzed here are not associated to genes
250 responsible for the selected traits. On the other hand, the number of samples represented by
251 each genotype was highly variable. Genotype-A was the most abundant, including almost half
252 of the studied accessions (Figure 3). Genotype-A abundancy could indicate that this has been
253 the most propagated lineage in Argentina. Either as consequence of a "bottleneck effect" caused

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.27.356790; this version posted October 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

254 by ancestral introductions of 'Malbec' in South America, and/or posterior selections favored by
255 its productive performance. However, we cannot rule out that Genotype-A abundancy could be
256 consequence of a sampling bias. It is expectable that including more samples with diverse
257 origins, as well as employing additional SNVs markers, would turn into a greater number of
258 genotypes represented by fewer samples. Despite of these caveats, even with a reduced set of
259 the identified genetic markers here, it was possible to recover the main identified clonal
260 genotypes (Supplementary Fig. S2).

261 The identified clonal genotypes clustered in two genetically divergent groups, Groups
262 Ar and Fr (Figures 4a and 4b). The observed pattern of genetic diversity in 'Malbec' is likely
263 resembling the combination of natural and human directed processes. The only natural source
264 of genetic variation in grapevine cultivars are somatic mutations and epimutations, which arise
265 during the vine growth and might be pass to daughter vines through vegetative propagation
266 [49,50]. Therefore, shared mutated positions turn into fingerprints that provide information on
267 the history of a given clonal lineage [51]. Here, by using only four Sh-SNVs was enough to
268 recover the distinction between the two main identified groups (Supplementary Fig. S3). On
269 the other hand, as a species of commercial interest, human actions such as plants transportation,
270 as well as clonal and mass selection are over imposed to the observed patterns of genetic
271 diversity [2,22,49].

272 Historical records report that the first 'Malbec' plants were introduced from France to
273 Argentina (Mendoza province) in the 1850s [32,34]. After that, wine-producers kept introducing
274 plants into Argentina at a continuous rate, that was slightly increased during the 1990s [52]. The
275 found genetic diversity pattern could be reflecting this history. Distinguishing among genotypes
276 that have gone through alternative pathways and accumulated different somatic mutations, and
277 bringing together those with a more recent shared history. Genotypes included in Group-Fr are
278 closely related to the resequenced clones that have longer remained in Europe (C143 and C225),
279 including also all the other analyzed samples that were never grown or were recently introduced
280 into Argentina. On the other hand, genotypes from Group-Ar are closely related to the
281 resequenced clones with a longer time span of clonal propagation in Argentina (MB53 and
282 MB59), suggesting a closer link to those first plants introduced from France. Among the
283 analyzed samples, we can pinpoint those accessions never grown or more recently introduced
284 into Argentina (<30 years). However, we cannot tell with accuracy the exact time span of clonal
285 propagation for those accessions that have remained for more than 70 years in Argentina. In
286 particular, some of the latter accessions appeared included in Group-Fr (Supplementary Table
287 S6). This could be resembling intermediate times of introduction for certain accessions (as those

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.27.356790; this version posted October 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

288 from Genotype-G) or possible traceability inconsistencies. In the same direction, it is not
289 possible to track back the precise history of individual plants from the sampled mass selections;
290 the only available information relates to their vineyard of origin and productive criteria of
291 selection. In this context, it is important to highlight that we were able to retrieve the
292 phylogenetic relations among the four resequenced clones with a known history, by means of
293 the custom-designed genotyping chip.

294 The set of markers included in the chip proved useful for clonal genotypes distinction.
295 In fact, by genotyping as few as four Sh-SNVs would be enough to tell if a 'Malbec' plant is
296 closely related, either to ancestors that were early introduced in Argentina or that longer
297 remained in Europe. At the same time, CS-SNVs were essential at discovering the main clonal
298 genotypes identified here. This observation further supports the importance of combining clone
299 specific and shared variants, to enhance genotyping experiments sensitivity for clonal diversity
300 studies [[30,51]]. Moreover, custom-designed genotyping for grapevine cultivars has already been
301 proven as a valuable tool with different applications. For example, for nurseries to fill with
302 genetic evidence the historical gaps of clonal accessions [[51]], and for the wine industry for
303 traceability and authentication purposes [[53]].

304 In conclusion, we could setup an efficient workflow to identify a reliable set of clonal
305 genetic variants, that were employed to design an informative genotyping experiment. We were
306 able to distinguish several clonal genotypes within 'Malbec' and observed that clonal
307 propagation history has shaped its genetic diversity pattern. Findings add further evidence on
308 the importance of high-throughput genotyping in grapevines as baseline information, to better
309 understand cultivars' history and as a tool with industrial application.

310 **Materials & Methods**

311 **a. Biological material.**

312 To perform WGR, we obtained young leaves and shoot tips from four 'Malbec' clones. Two
313 clones were sampled at the Mercier Argentina nursery collection (Perdriel, Lujan de Cuyo,
314 Mendoza): Malbec-501 (MB53) and Cot-ENTAV-598 (C225). One clone was sampled at
315 Mercier Argentina nursery Granata vineyards (Perdriel, Lujan de Cuyo, Mendoza), Malbec-059
316 (MB59). The fourth clone: Cot-143 (C143) was sampled at the "*Finca El Encín*" ampelographic
317 collection (ESP-080, Alcala de Henares, Spain). The two accessions labeled as 'Malbec'
318 (MB53 and MB59) represent plants with long history of clonal propagation in Argentina,
319 meaning that they have been propagated in this country for more than 70 years (Mercier nursery
320 records). We also included two accessions labeled as 'Cot' (C225 and C143), with short and
321 null histories of clonal propagation in Argentina. More precisely, C225 was introduced into

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.27.356790; this version posted October 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

322 Argentina from France (ENTAV-INRA) during the 1990s (Mercier nursery records) and C143
323 was sampled in a Spanish germplasm collection, therefore it was never grown in Argentina.
324 For the genotyping analysis, shoot tips and young leaves were obtained from 219 plants.
325 We sampled 70 'Malbec' clonal accessions (including the four resequenced ones) belonging to:
326 (a) the National Institute of Agricultural Technology (INTA-Mendoza) collection (28 clones),
327 (b) Mercier Argentina Nursery collection (37 clones) (c) Mercier Granata vineyard (three
328 clones) and (d) Finca El Encin (two clones). Time span of clonal propagation in Argentina were
329 obtained from [52] and from Mercier nursery records. We also obtained 30 samples from each
330 of five different Mercier's mass selections (150 samples in total), located at Granata vineyards
331 (Perdriel, Lujan de Cuyo, Mendoza). Further details about mass selections and samples are in
332 Supplementary Tables S7 and S8 respectively.

333 **b. Whole genome resequencing, variant calling and validation.**

334 **DNA extractions and resequencing:** Whole genomic DNA from the four 'Malbec' clones:
335 MB53, MB59, C225 and C143 was isolated using the DNeasy Plant Mini Kit (Qiagen),
336 including a RNase treatment, according to manufacturer recommendations. DNA quantification
337 and quality checks were performed with NanoDrop 2000 spectrophotometer and agarose gel
338 (5%) electrophoresis. Library preparation and sequencing was performed at the Center for
339 Genomic Regulation (Barcelona, Spain) 125 bp length paired-end reads were produced using
340 the HiSeq 2000 Illumina technology with the Sequencing v4 chemistry.

341 **Reads alignment, variant calling and filtering:** Standard quality checks of the FASTQ files
342 were performed with FastQC [54]. Raw reads were pre-processed following the GATK Best
343 Practices workflow with the toolkit GenomeAnalysisTK-3.3-0 [41]. After marking Illumina
344 adapters with Picard toolkit v2.9.4 [55], sequences were aligned to *Vitis vinifera* L. reference
345 genome PN40024 [4]. We employed the Burrows-Wheeler algorithm as implemented in BWA-
346 MEM v0.7.12-r1039 [56], to align our reads to the reference genome. Mapped reads were
347 thoroughly filtered also with Picard toolkit [55] allowing only non-duplicates, unique and
348 concordant alignments with a maximum read edit distance of 1 per 25 nucleotides of query
349 sequence [57]. Filtered alignments were used as input for variant calling, comparing to
350 PN40024, using three different tools with default parameters and in the multi-allelic mode:
351 GATK UnifiedGenotyper [41], BCFTOOLS call v1.9 [42] and VARSCAN2 mpileup2cns v2.3.9
352 [43]. Produced gVCF files for each accession were intersected and only those single nucleotide
353 variants identified by all three callers were retained, while INDELs and structural variations
354 were not considered in this study. Bioinformatic procedures were adjusted using a set of SNPs
355 between 'Malbec' and PN40024 retrieved from Vitis18kSNP array results [58]. Only confident

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.27.356790; this version posted October 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

356 identified raw variants were retained, based on WGR recommendations of total depth (DP),
357 variant allele frequency (VAF), strand bias and distance bias (Bentley et al. 2008). Cut-off
358 values for these parameters were: DP = [15-150]; VAF(Ref) ≤ 0.025; VAF(Het) = [0.25-0.75];
359 VAF(Hom) ≥ [0.95]; P-value (strand bias) ≤ 0.0001 and P-value (distance bias) ≤ 0.0001.
360 Variant allele frequency ranges were particularly adjusted to reduce -at the minimum possible-
361 the presence of spurious variants. Chimeric mutations are frequent in grapevines, occurring
362 differentially between the L1 and L2 cell layers of the developmental tissue from the apical
363 meristem [49]. L1 layer gives rise to the epidermis and represent a smaller proportion of the total
364 tissues conforming a plant (nearly 30%) [50]. With the employed VAF filters we expected to
365 detect most of the chimeric mutations occurring in the L2 cell layer, VAF around 0.3 (half of
366 the total frequency in 60% of somatic tissues). While chimeric heterozygous mutations
367 occurring only in the L1 would be mostly excluded. We assumed that variants loss as a trade-
368 off, in the aim of reducing the false positives.
369 **Corroboration of the bioinformatic pipeline:** We employed IGV v2.3.97 [59] to manually
370 corroborate a sub-set of the identified SNVs and to isolate a ~600 bp length sequence containing
371 the target SNVs in the mid-region. These sequences were used as templates for primer design
372 to perform PCRs and Sanger sequencing of the amplicons. In order to avoid both, primer
373 annealing and later genotyping issues, we checked for the absence of variable sites on the 5'-
374 and 3'- regions of the sequence and in the proximities of the SNVs target position. Primers were
375 designed using the Primer BLAST tool [60], with an average annealing temperature (Tm) of
376 60.3ºC (range: 58.8-62.5ºC) and an average amplicon length of 447 bp (range: 300-582 bp),
377 more details in Supplementary Table S2. PCRs were conducted in a 25 µl final reaction volume
378 containing: 0,3 ul (5 U/µl) Taq Polymerase High fidelity (TransTaq); 1,25 ul (10x) Buffer GC-
379 enhancer (TransTaq); (2,5 ul) 10x PCR Buffer I (TransTaq); 1 ul (2.5 mM) dNTPs; 1 ul of each
380 (10 µM) Primer forward and reverse and 3 ul (40 ng/µl) DNA template. Cycles consisted in a
381 denaturation step of 5' at 98ºC; 35 cycles of 30" 94ºC, 30" at 60ºC and 30" at 72ºC, and final
382 extension of 7' at 72ºC. PCR products were purified using ExoSAP-IT PCR Product Cleanup
383 (Thermofisher), following the manufacturer recommendations. To validate the target SNVs,
384 electropherograms of the four resequenced clones were aligned and inspected with CODON
385 CODE ALIGNER v4.0.4 (CodonCode Corp. USA). SNVs were considered as validated if the
386 allelic state at the position of interest in the sequence, coincided with that observed in the *vcf*
387 file and in the IGV genome browser. For example, for a heterozygous CS-SNVs the clone for

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.27.356790; this version posted October 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

388 which the variant was identified must be heterozygous and for the other three clones must be
389 homozygous as the reference genotype (e.g. Supplementary Fig. S1).

390 **c. Genotyping.**

391 DNA extractions were performed employing the NucleoSpin® Plant II Plant Mini kit
392 (Macherey-Nagel). Quantification of the isolated DNAs was performed using NanoDrop 8000
393 Spectrophotometer (Thermo Fisher Scientific) and Qubit 2.0 Fluorometer (Invitrogen, Life
394 Technologies).

395 SNVs chosen as genetic markers to build the genotyping chip accomplished the following
396 criteria. We included 42 CS-SNVs and 6 Sh-SNVs, only heterozygous alternative variants were
397 selected from the deep-filtered list, based on their ability to discriminate among the four
398 resequenced clones. For the CS-SNVs, equivalent number of variants for each clone were
399 chosen. Sh-SNVs were picked for their ability to differentiate between the resequenced clones
400 with a long history of clonal propagation in Argentina, from those with a short or null history
401 in this country. Since we were particularly interested in identifying genetic markers that could
402 consistently resemble that historical aspect across our samples. We also intended the chosen
403 SNVs to be distributed across different chromosomes to better represent the genome-wide
404 diversity. In total, 48 sequences containing one SNVs of interest (Supplementary Table S3)
405 were provided to the Genomics Service Sequencing and Genotyping Unit (UPV/EHU)
406 (Bizkaia, Spain) to design probes for a Fluidigm chip (https://www.fluidigm.com/) and perform
407 the genotyping. Each experiment allowed to simultaneously genotype 48 samples using 48
408 SNVs, in a two steps reaction. In the first step, the target region containing the position to be
409 genotyped is amplified using two pre-amplification primers (locus-specific primer and specific
410 target amplification). In the second step, an additional PCR amplifies a portion of that target
411 SNVs region, using the locus-specific primer and two fluorescently labeled allele-specific
412 primers, which are internal primers containing either the first or the second allele respectively.
413 Finally, the genotype is determined by measuring the fluorescence intensity of both alleles using
414 the Fluidigm genotyping analysis software.

415 **d. Inter-clonal genetic diversity analyses**

416 To assess the degree of genetic variation among 'Malbec' clones, biallelic genotypes were
417 coded to sequences in the *fasta* format. In first place, we performed a phylogenetic analysis
418 with MEGA v7.0.26 [61] including only the four resequenced clones and using PN40024
419 genotype as outgroup. A Neighbor-Joining tree was estimated based on the deep-filtered list of
420 SNVs, using uncorrected *p*-distances and nodes' support were obtained after 200 bootstrap
421 iterations.

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.27.356790; this version posted October 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

422 In second place, we performed a Median-Joining Network [62] analysis with POPART
423 software [63]; to screen the diversity across all the genotyped samples, identify the number of
424 different genotypes, their frequencies and phylogenetic relations. Afterwards, we obtained a
425 single representative sequence for each of the identified genotypes to reconstruct a Neighbor-
426 Joining tree with MEGA v7.0.26 [61], using the same parameters described above. Genetic
427 diversity was also analyzed considering each SNVs position as an independent marker, by
428 estimating codominant-genotypic distances among the identified genotypes with GenAlEx v6.5
429 [64]. Genetic distances among genotypes were analyzed with a model-free approach of Principal
430 Coordinate Analysis (PCoA), to detect potential groups of genotypes closer related among each
431 other. We estimated the proportion of the total molecular variance that is explained by the
432 variance between groups through an AMOVA, we obtained the Phi_{PT} parameter recommended
433 for distances obtained from codominant genotypic data, *p*-value was obtained after 900
434 bootstrap iterations. Both PCoA and AMOVA were also performed with GenAlEx v6.5.

435 **References**

436 1. Gaut, B. S., Seymour, D. K., Liu, Q. & Zhou, Y. Demography and its effects on
437 genomic variation in crop domestication. *Nat. Plants* **4**, 512–520 (2018).
438 2. This, P., Lacombe, T. & Thomas, M. Historical origins and genetic diversity of wine
439 grapes. *Trends Genet.* **22**, 511–519 (2006).
440 3. OIV, F. FAO-OIV Focus Table and Dried Grapes. *Vol. I7042 FAO* 64 (2016).
441 4. Jaillon, O. *et al.* The grapevine genome sequence suggests ancestral hexaploidization
442 in major angiosperm phyla. *nature* **449**, 463 (2007).
443 5. Gambino, G. *et al.* Whole-genome sequencing and SNV genotyping of 'Nebbiolo'
444 (Vitis vinifera L.) clones. *Sci. Rep.* **7**, 17294 (2017).
445 6. Vondras, A. M. *et al.* The genomic diversification of grapevine clones. *BMC*
446 *Genomics* **20**, 972 (2019).
447 7. Zhou, Y. *et al.* The population genetics of structural variants in grapevine
448 domestication. *Nat. Plants* **5**, 965–979 (2019).
449 8. Kobayashi, S. Retrotransposon-Induced Mutations in Grape Skin Color. *Science* **304**,
450 982–982 (2004).
451 9. Lijavetzky, D. *et al.* Molecular genetics of berry colour variation in table grape. *Mol.*
452 *Genet. Genomics* **276**, 427–435 (2006).
453 10. Carbonell-Bejerano, P. *et al.* Catastrophic Unbalanced Genome Rearrangements Cause
454 Somatic Loss of Berry Color in Grapevine. *Plant Physiol.* **175**, 786–801 (2017).
455 11. Emanuelli, F. *et al.* A candidate gene association study on muscat flavor in grapevine
456 (Vitis viniferaL.). *BMC Plant Biol.* **10**, 241 (2010).
457 12. Fernandez, L., Torregrosa, L., Segura, V., Bouquet, A. & Martinez-Zapater, J. M.
458 Transposon-induced gene activation as a mechanism generating cluster shape somatic
459 variation in grapevine. *Plant J.* **61**, 545–557 (2010).
460 13. Fernandez, L. *et al.* Haplotype diversity of VvTFL1A gene and association with
461 cluster traits ingrapevine (V. vinifera). *BMC Plant Biol.* **14**, 209 (2014).
462 14. Boss, P. K. & Thomas, M. R. Association of dwarfism and floral induction with a
463 grape 'green revolution' mutation. *Nature* **416**, 847–850 (2002).
464 15. Royo, C. *et al.* The major origin of seedless grapes is associated with a missense
465 mutation in the MADS-box gene VviAGL11. *Plant Physiol.* **177**, 1234–1253 (2018).

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.27.356790; this version posted October 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

466 16. Muñoz, C. *et al.* Relationships among Gene Expression and Anthocyanin Composition
467 of Malbec Grapevine Clones. *J. Agric. Food Chem.* **62**, 6716–6725 (2014).
468 17. Grimplet, J., Ibáñez, S., Baroja, E., Tello, J. & Ibáñez, J. Phenotypic, Hormonal, and
469 Genomic Variation Among Vitis vinifera Clones With Different Cluster Compactness and
470 Reproductive Performance. *Front. Plant Sci.* **9**, 1917 (2019).
471 18. Rossmann, S. *et al.* Mutations in the miR396 binding site of the growth-regulating
472 factor gene *VvGRF4* modulate inflorescence architecture in grapevine. *Plant J.* **101**, 1234–
473 1248 (2020).
474 19. Myles, S. Improving fruit and wine: what does genomics have to offer? *Trends Genet.*
475 **29**, 190–196 (2013).
476 20. Yang, S. *et al.* A next-generation marker genotyping platform (AmpSeq) in
477 heterozygous crops: a case study for marker-assisted selection in grapevine. *Hortic. Res.* **3**,
478 16002 (2016).
479 21. Myles, S. *et al.* Genetic structure and domestication history of the grape. *Proc. Natl.*
480 *Acad. Sci.* **108**, 3530–3535 (2011).
481 22. Bacilieri, R. *et al.* Genetic structure in cultivated grapevines is linked to geography
482 and human selection. *BMC Plant Biol.* **13**, 25 (2013).
483 23. Marrano, A. *et al.* SNP-Discovery by RAD-Sequencing in a Germplasm Collection of
484 Wild and Cultivated Grapevines (V. vinifera L.). *PLOS ONE* **12**, e0170655 (2017).
485 24. Liang, Z. *et al.* Whole-genome resequencing of 472 Vitis accessions for grapevine
486 diversity and demographic history analyses. *Nat. Commun.* **10**, 1190 (2019).
487 25. Imazio, S. *et al.* Molecular tools for clone identification: the case of the grapevine
488 cultivar 'Traminer'. *Plant Breed.* **121**, 531–535 (2002).
489 26. Ibáñez, J., de Andrés, M. T., Molino, A. & Borrego, J. Genetic study of key Spanish
490 grapevine varieties using microsatellite analysis. *Am. J. Enol. Vitic.* **54**, 22–30 (2003).
491 27. de Lorenzis, G. *et al.* Study of intra-varietal diversity in biotypes of Aglianico and
492 Muscat of Alexandria (*Vitis vinifera* L.) cultivars: Intra-varietal diversity assessment in
493 grapevine. *Aust. J. Grape Wine Res.* **23**, 132–142 (2017).
494 28. Mercati, F. *et al.* High-throughput 18K SNP array to assess genetic variability of the
495 main grapevine cultivars from Sicily. *Tree Genet. Genomes* **12**, 59 (2016).
496 29. Carrier, G. *et al.* Transposable Elements Are a Major Cause of Somatic Polymorphism
497 in Vitis vinifera L. *PLoS ONE* **7**, e32973 (2012).
498 30. Roach, M. J. *et al.* Population sequencing reveals clonal diversity and ancestral
499 inbreeding in the grapevine cultivar Chardonnay. *PLOS Genet.* **14**, e1007807 (2018).
500 31. Maul et al. VIVC. http://www.vivc.de/index.php?r=site%2Findex.
501 32. Lacoste, P. El Malbec de Francia: la Denominación de Origen Controlada 'Cahors'.
502 Historia y perspectivas. *Idesia Arica* **33**, 113–124 (2015).
503 33. Boursiquot, J.-M. *et al.* Parentage of Merlot and related winegrape cultivars of
504 southwestern France: discovery of the missing link. *Aust. J. Grape Wine Res.* **15**, 144–155
505 (2009).
506 34. Beezley, W. H. La senda del Malbec: la cepa emblemática de Argentina. *Univers.*
507 *Talca* **20**, (2005).
508 35. International Organisation of Vine and Wine (OIV). Distribution of the world's
509 grapevine varieties. (2017).
510 36. Fanzone, M., Peña-Neira, Á., Jofré, V., Assof, M. & Zamora, F. Phenolic
511 Characterization of Malbec Wines from Mendoza Province (Argentina). *J. Agric. Food Chem.*
512 **58**, 2388–2397 (2010).
513 37. Instituto Nacional del Vino (Argentina). Informe variedad Malbec. (2019).
514 38. Zhou, Y., Massonnet, M., Sanjak, J. S., Cantu, D. & Gaut, B. S. Evolutionary
515 genomics of grape (*Vitis vinifera* ssp. *vinifera*) domestication. *Proc. Natl. Acad. Sci.* **114**,

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.27.356790; this version posted October 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

516 11715–11720 (2017).
517 39. Poplin, R. *et al.* A universal SNP and small-indel variant caller using deep neural
518 networks. *Nat. Biotechnol.* **36**, 983–987 (2018).
519 40. Xu, C. A review of somatic single nucleotide variant calling algorithms for next-
520 generation sequencing data. *Comput. Struct. Biotechnol. J.* **16**, 15–24 (2018).
521 41. McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for
522 analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).
523 42. Danecek, P., Schiffels, S. & Durbin, R. *Multiallelic calling model in bcftools (-m).*
524 (June, 2014).
525 43. Koboldt, D. C. *et al.* VarScan 2: somatic mutation and copy number alteration
526 discovery in cancer by exome sequencing. *Genome Res.* **22**, 568–576 (2012).
527 44. Da Silva, C. *et al.* The High Polyphenol Content of Grapevine Cultivar Tannat Berries
528 Is Conferred Primarily by Genes That Are Not Shared with the Reference Genome. *Plant Cell*
529 **25**, 4777–4788 (2013).
530 45. Di Genova, A. *et al.* Whole genome comparison between table and wine grapes
531 reveals a comprehensive catalog of structural variants. *BMC Plant Biol.* **14**, 7 (2014).
532 46. Minio, A., Massonnet, M., Figueroa-Balderas, R., Castro, A. & Cantu, D. Diploid
533 Genome Assembly of the Wine Grape Carménère. *G3amp58 GenesGenomesGenetics* **9**,
534 1331–1337 (2019).
535 47. Li, H. Toward better understanding of artifacts in variant calling from high-coverage
536 samples. *Bioinformatics* **30**, 2843–2851 (2014).
537 48. Goodwin, S., McPherson, J. D. & McCombie, W. R. Coming of age: ten years of next-
538 generation sequencing technologies. *Nat. Rev. Genet.* **17**, 333–351 (2016).
539 49. Pelsy, F. Molecular and cellular mechanisms of diversity within grapevine varieties.
540 *Heredity* **104**, 331–340 (2010).
541 50. Torregrosa, L. *et al.* Origins and consequences of somatic variation in grapevine.
542 *Genet. Genomics Breed. Grapes* 68–92 (2011).
543 51. Roach, M. J., Borneman, A. R. & Schmidt, S. A. Origin of Chardonnay clones with
544 historical significance in Australia and California. *Aust. J. Grape Wine Res.* **n/a**,.
545 52. Scarsi, H. A. Caracterización vitícola y enológica de 19 clones de Malbec argentinos
546 obtenidos por selección clonal y sanitaria. *Master Sci. Thesis Mendoza Argent.* (2002).
547 53. Boccacci, P., Chitarra, W., Schneider, A., Rolle, L. & Gambino, G. Single-nucleotide
548 polymorphism (SNP) genotyping assays for the varietal authentication of 'Nebbiolo' musts
549 and wines. *Food Chem.* **312**, 126100 (2020).
550 54. Andrews, S. FastQC a quality-control tool for high-throughput sequence data
551 http://www. *Bioinformaticsbabraham Ac Ukprojectsfastqc* (2014).
552 55. Picard Tools - By Broad Institute. http://broadinstitute.github.io/picard/.
553 56. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler
554 transform. *bioinformatics* **25**, 1754–1760 (2009).
555 57. Royo, C. *et al.* Developmental, transcriptome, and genetic alterations associated with
556 parthenocarpy in the grapevine seedless somatic variant Corinto bianco. *J. Exp. Bot.* **67**, 259–
557 273 (2016).
558 58. Laucou, V. *et al.* Extended diversity analysis of cultivated grapevine Vitis vinifera
559 with 10K genome-wide SNPs. *PLOS ONE* **13**, e0192540 (2018).
560 59. Robinson, J. T., Thorvaldsdóttir, H., Wenger, A. M., Zehir, A. & Mesirov, J. P.
561 Variant review with the integrative genomics viewer. *Cancer Res.* **77**, e31–e34 (2017).
562 60. Ye, J. *et al.* Primer-BLAST: A tool to design target-specific primers for polymerase
563 chain reaction. *BMC Bioinformatics* **13**, 134 (2012).
564 61. Kumar, S., Stecher, G. & Tamura, K. MEGA7: molecular evolutionary genetics
565 analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* **33**, 1870–1874 (2016).

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.27.356790; this version posted October 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

566 62. Bandelt, H.-J., Forster, P. & Röhl, A. Median-joining networks for inferring
567 intraspecific phylogenies. *Mol. Biol. Evol.* **16**, 37–48 (1999).
568 63. Leigh, J. W. & Bryant, D. POPART: full-feature software for haplotype network
569 construction. *Methods Ecol. Evol.* **6**, 1110–1116 (2015).
570 64. Peakall, R. O. D. & Smouse, P. E. GENALEX 6: genetic analysis in Excel. Population
571 genetic software for teaching and research. *Mol. Ecol. Notes* **6**, 288–295 (2006).
572

573 **Acknowledgments**

574 This work was supported by Agencia Nacional de Promoción Científica y Tecnológica
575 (ANPCyT): PICT2015-0822, MINCyT/ANPCyT (FONTAR)-CDTI 'IBEROGEN';
576 CONICET (Bilateral PCB-II, CONICET–CSIC; MINECO BIO2017-86375-R). COST Action
577 CA17111. We thank to M. Victoria Bertoldi for the assistance with laboratory work at IBAM-
578 CONICET.

579 **Authors contributions**

580 L.C. wrote the manuscript and performed: samples collection, laboratory work for library
581 preparation and SNVs validation; bioinformatic and statistical analyses on the final set of
582 genomic data. N.M. conducted bioinformatic analyses of raw data processing, variant calling,
583 filtering and SNVs classification. C.M. collaborated with samples collection and laboratory
584 work. P.C.B. provided insight on the SNPs and SNVs variant calling analysis and filtering. L.B.
585 and C.S. collaborated with samples collection at Mercier nursery and provided information
586 about historical background of the analyzed accessions. S.G.T. collaborated with samples
587 collection at INTA. C.R. contributed to obtain and genotype samples from Europe. J.I. and
588 J.M.M.Z provided insight on the design of the genotyping experiments. D.L. designed and
589 coordinated the entire project. All authors carefully read and helped to improve the final content
590 of the manuscript.

591 **Additional information**

592 **Data access:** SRA files containing raw genomic data for the four resequenced 'Malbec'
593 clones is available at NCBI, BioProject: PRJNAXXXXXX.

594 **Supplementary information**: Figures and tables complementing this paper are available in
595 separate files

596 **Competing Interests:** The authors declare no competing interests.

597 **Figures legends**

598 **Figure 1.** Total single nucleotide polymorphisms (SNPs) and variants (SNVs) identified in
599 'Malbec'. SNPs distinguish 'Malbec' from PN40024 and SNVs occurred differentially among
600 the four resequenced clones. SNVs are classified based on the clone for which they were

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.27.356790; this version posted October 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

601 identified and also according to their genotype, relative to PN40024. CS-SNVs are diagnostic
602 for variation in a single clone and Sh-SNVs are shared between two clones in different
603 combinations.

604 **Figure 2.** Phylogenetic relations among four resequenced 'Malbec' clones. Neighbor-Joining
605 tree based on p-distances and employing 941 SNVs. Percentages on the nodes represent
606 bootstrap supports after 200 iterations, only values >50% are shown. PN40024 genotype was
607 used as outgroup.

608 **Figure 3.** Intra-cultivar clonal genotypic diversity estimated with a custom designed
609 genotyping chip. Median-Joining network was built with the genotypes obtained for 214 clones
610 at 41 SNVs loci. Each circle represents a genotype and its size is proportional to the genotype
611 frequency. In total 14 clonal genotypes were found, nine were represented by multiple samples
612 (named from A to I) and five were singletons (C143, C225, MB53, MB59 and Perd_121). The
613 hashmarks crossing the connecting lines indicate the number of point mutational steps
614 differentiating genotypes. Color code represents Groups Fr (orange) and Ar (purple).

615 **Figure 4.** Phylogenetic relationships among the identified clonal genotypes in 'Malbec'. **(a)**
616 Neighbor-Joining tree based on *p*-distances among the identified genotypes (based on 41
617 SNVs), nodes bootstrap supports values >50% are shown. The orange clade (Group-Fr)
618 included genotypes closer related to the resequenced clones that have longer remained in
619 Europe (C143 and C225). Purple clade (Group-Ar) included genotypes closer related to the
620 resequenced clones with more than 70 years of clonal propagation in Argentina (MB53 and
621 MB59). **(b)** Principal coordinates analyses (PCoA) based on genetic distances among
622 genotypes. PCoA recovered the same relations than the phylogeny among the identified
623 genotypes, differentiating between Group-Ar (purple dots) and Group-Fr (orange dots).