# Diploid genome assembly of the Malbec grapevine cultivar enables haplotype-aware analysis of transcriptomic differences underlying clonal phenotypic variation

**Authors:**

Calderón, Luciano[1*#]; Carbonell-Bejerano, Pablo[2,3*#]; Muñoz, Claudio[1,4]; Bree, Laura[5]; Sola, Cristobal[5]; Bergamin, Daniel[5]; Tulle, Walter [1]; Gomez-Talquenca, Sebastian[6]; Lanz, Christa[3]; Royo, Carolina[2]; Ibáñez, Javier[2]; Martinez-Zapater, José Miguel[2]; Weigel, Detlef[3] & Lijavetzky, Diego[1*]

**Affiliations:**

[1] Instituto de Biología Agrícola de Mendoza (CONICET-UNCuyo), Genetica y Genomica de Vid. (5505) Mendoza, Chacras de Coria, Argentina.

[2] Instituto de Ciencias de la Vid y del Vino, ICVV, CSIC - Universidad de La Rioja - Gobierno de La Rioja. (26007) La Rioja, Logroño, Spain

[3] Department of Molecular Biology, Max Planck Institute for Biology Tübingen. (72076) Tübingen, Germany.

[4] Facultad de Ciencias Agrarias (UNCuyo), Cátedra Fitopatología. (5505) Mendoza, Chacras de Coria, Argentina.

[5] Vivero Mercier Argentina. (5500) Mendoza, Perdriel, Argentina.

[6] Instituto Nacional de Tecnología Agropecuaria, Plant Virology Laboratory. (5534) Mendoza, Luján de Cuyo, Argentina.

**Corresponding Authors:**

* Calderón, Luciano: lcalderon@mendoza-conicet.gob.ar

* Carbonell-Bejerano, Pablo: pablo.carbonell@icvv.es

* Lijavetzky, Diego: dlijavetzky@conicet.gov.ar

**Symbols reference:** #These authors contributed equally to this work; * Corresponding authors.

## 1. Abstract

Grapevine cultivars (*Vitis vinifera* L. ssp. *vinifera*) must be clonally propagated because of their highly heterozygous genomes. Malbec, a France-originated cultivar appreciated for the production of high-quality wines, is the offspring of cultivars Prunelard and Magdeleine Noire des Charentes. Here, we have built a diploid genome assembly of Malbec, after trio binning of PacBio long reads into the two haploid complements inherited from either parent. After haplotype-aware deduplication and corrections, complete assemblies for the two haplophases were obtained with very low haplotype switch-error rate (<0.025). The haplophases alignment identified >25% of polymorphic regions. Gene annotation including RNA-seq transcriptome assembly and *ab initio* prediction evidence resulted in similar gene model numbers for both haplophases. The annotated diploid assembly was exploited in the transcriptomic comparison of four clonal accessions of Malbec that exhibited variation in berry composition traits. Analysis of the ripening pericarp transcriptome using either haplophases as reference yielded similar results, although some differences were observed. Particularly, among the differentially expressed genes in the Magdeleine-inherited haplotype, we observed an over-representation of hypothetically hemizygous genes. The higher berry anthocyanin content of clonal accession 595 was associated with increased abscisic acid responses, leading to overexpression of secondary metabolism genes and deregulation of genes associated to abiotic stress response. Overall, the results highlight the importance of producing diploid assemblies to fully represent the genomic diversity of highly heterozygous woody crop cultivars and to unveil the molecular bases of clonal phenotypic variation.

48 **2. Introduction:**

49 The cultivated grapevine (*Vitis vinifera* L. subsp. *vinifera*) was one of the first plant species
50 accounting for a reference genome [1]. To facilitate genome assembly, this reference was produced from a
51 nearly homozygous line (i.e. PN40024) obtained after several rounds of selfing of the cultivar
52 Helfensteiner, the result of Pinot noir and Schiava grossa outcross [1,2]. The PN40024 reference genome has
53 been a key tool for the study of grapevine molecular biology and genetics and both the reference assembly
54 and annotation have been continuously improved [2–4].

55 The genomes of grapevine cultivars are characterized by enormous diversity, including the many
56 differences that typically distinguish the two haplotypes of each cultivar [7]. On average one single
57 nucleotide polymorphism (SNP) every 100 bp and one Insertion-deletion (InDel) difference every 450 bp
58 [7,8]. Also, larger structural variants (SVs) ≥50 bp have been reported to affect ~17% of the genome space
59 of each haplotype within the same cultivar [9]. Hemizygosity is also rampant in grapevines, since many SVs
60 cause presence-absence variation (PAV) between haplotypes [8], which may have severe phenotypic
61 impacts [11]. In addition to the PN40024 reference genomes, several additional cultivars have seen their
62 genomes assembled from Pacific Biosciences (PacBio) long reads using the haplotype-aware assemblers
63 FALCON-unzip [6,10,12–15]. FALCON-unzip outputs pseudo-haplotypes, which usually represent a single allele
64 at each position without preserving phase information across the genome [16], although tools such as
65 HaploSync have been developed to completely phase this type of assemblies [17]. In this context, the
66 combination of multiple sequencing technologies such as Oxford Nanopore Technologies (ONT) and short
67 reads[21], the use of PacBio HiFi and newer algorithms such as HiFiasm [18], as well as information from
68 parent-offspring trios [16] have notably improved assembly phasing and quality of grapevine genomes [4,19
69 20].

70 The high and uneven heterozygosity of cultivated grapevine genomes is most likely the
71 consequence of the outcrossing nature of their dioecious wild ancestor (*V. vinifera* ssp. *sylvestris*) and
72 their vegetative way of propagation [6]. Cultivars relevant for wine making have typically resulted from the
73 outcross of preexisting cultivars; progeny with desired phenotypic traits were then vegetatively
74 propagated to retain those traits [9]. Although vegetative propagation mostly ensures cultivar uniformity,
75 it also favors the accumulation of somatic mutations in the absence of outcrossing [21]. In fact, somatic
76 variation is known to exist at the phenotypic [22], genetic [14] and transcriptomic [23] levels. Moreover, within
77 each grape cultivar growers select somatic variants to propagate them, labelling these selections as
78 clones. Therefore, somatic diversity has been relevant for viticulture as a source of trait innovation and
79 adaptation. Because the global wine market is very slow at accepting new outcross varieties, a deep

80 understanding of the variation that exists within elite varieties is crucial for intra-cultivar improvement
81 and to meet emerging challenges from climate change [24].
82 Malbec, a red cultivar appreciated for its potential to produce high-quality wines, originated in the
83 Cahors region of France (where it is known as Cot) as the result of an outcross of cultivars Magdeleine
84 Noir des Charentes (hereafter Magdeleine) and Prunelard [25]. Malbec was introduced to Argentina in the
85 1850s, becoming the most relevant cultivar for the wine industry of this country [26]. Currently, Argentina
86 harbors the largest planted surface of Malbec in the world [27] and 85% of this production is concentrated
87 in Mendoza province [28]. Since its introduction, this cultivar has shown remarkable adaptability, producing
88 wines with properties that reflect the different growing conditions [29]. The observed adaptability may be
89 linked to the phenotypic plasticity exhibited by Malbec, as clonal variability has been reported for traits
90 relevant for the industry, such as polyphenolic composition of berries [30] and phenology [31]. Also molecular
91 plasticity has been described among clones, since epigenetic [32] and genetic [33] variability has been
92 reported.
93 The main objective of this study was to produce a phased and annotated genome assembly of
94 Malbec representing both haploid complements, or haplophases, to serve as reference for intra-cultivar
95 variation studies. We used our assembly in a transcriptome analysis to investigate the molecular bases of
96 Malbec clonal variants that differ in fruit traits relevant for wine making. The transcriptomic analysis was
97 performed following a haplotype-aware scheme and the results obtained using both haplophases as
98 reference were compared. This allowed us to further evaluate potential reference biases, and to identify
99 haplotype-specific differentially expressed genes that could be correlated with clonal variation in berry
100 composition.

101 **3. Results:**

102 **The diploid *de novo* genome assembly of Malbec:** A total of 82.3 Gb of sequence coverage (168x
103 coverage of the haploid genome) was produced for Malbec in 4,117,439 PacBio CLR reads, with a mean
104 read length of 20 kb and read N50 of 33.7 kb (Table S1A). Illumina short reads were obtained for Malbec
105 (13.4 Gb) as well as the parental cultivars Prunelard (46.9 Gb) and Magdeleine (47.6 Gb) (Table S1B).
106 Malbec PacBio reads were binned into two sets of 35.9 Gb and 46.0 Gb, corresponding to the haplotypes
107 inherited from Magdeleine and Prunelard (Table S2A). The binned PacBio long reads were separately
108 assembled into two haplophases, named hereafter Malbec-Pru for the inherited from Prunelard and
109 Malbec-Mag for the inherited from Magdeleine. The draft raw assemblies of Malbec-Pru and Malbec-Mag
110 had good contiguity, with contig N50 = 6.2 Mb and 5.6 Mb, and gene completeness, with 96.2% and 95.7%
111 of BUSCO core genes represented. The draft assemblies were larger than the expected for a haploid
112 grapevine genome of ~475 Mb [2], namely 608 Mb for Malbec-Pru and 558 Mb for Malbec-Mag, and they

113 had a high percentage of duplicated BUSCO genes, 19.2% and 8.3% (Table S2B). After deduplication,
114 consensus polishing and miss-assembly corrections, the two haplophases were closer to the expected
115 size, 479.3 Mb and 479 Mb, with improved contiguity of N50 = 7.7 Mb and 6.6 Mb. Additionally, we
116 obtained a smaller percentage of duplicated BUSCO genes (i.e., 2.3% and 1.8%) without a significant
117 change in completeness: 96.3% and 95.9% (Table 1A). The polished and deduplicated haplophases were
118 reference-based scaffolded, resulting in 19 pseudomolecules, as expected for the haploid set of grapevine
119 nuclear chromosomes (Fig. S1), with a scaffold N50 = 24.4 for Malbec-Pru and 24.5 Mb for Malbec-Mag
120 (Table 1A). Some unplaced contigs remained after the scaffolding, for Malbec-Pru = 21 contigs (total
121 length = 3.3 Mb) and for Malbec-Mag = 32 contigs (total length = 5.74 Mb). Malbec-Mag corresponds to
122 the maternally inherited haplotype [25], and included one scaffold for a partial mitochondrial chromosome
123 (length = 115,997 bp).

124 Phasing accuracy and consensus quality was assessed on the final scaffolded assemblies. The
125 Malbec-Pru assembly comprised 97.3% of the Prunelard-specific *hap-mers* and only 0.13% of Magdeleine-
126 specific ones (Fig. 1A), whereas the Malbec-Mag assembly comprised 98.1% of the Magdeleine-specific
127 *hap-mers* and only 0.03% of the Prunelard-specific ones (Fig. 1B). The consensus quality score (QV) was
128 estimated for Malbec-Pru = 41.62 (error rate = 6.9 e-5) and Malbec-Mag = 42.06 (error rate = 6.2 e-5)
129 (Table 1B). Total bases placed in proper phase blocks and the phase blocks N50 for each scaffolded
130 haplophase assembly were for Malbec-Pru = 463.8 Mb (N50 = 11.96 Mb) and Malbec-Mag = 456.3 Mb
131 (N50 = 8.9 Mb) (Figs 1A and B, Fig. S2). Also, a very low haplophase switch error rate was observed, 0.025%
132 for Malbec-Pru and 0.019% for Malbec-Mag (Table 1B).

133 The estimated divergence between the two haplophases did not depend on which one was used
134 as reference, and we report the results obtained with Malbec-Pru as reference and Malbec-Mag as query.
135 Sequence differences affected 7.09% (33.55 Mb) of Malbec-Mag, this included SNPs, InDels, copy number
136 variations (CNVs), highly divergent regions (HDRs) and tandem repeats (TRs) (Fig. 1C and Table S3).
137 Considering the accumulative length, HDR was the most abundant type of sequence variation and
138 duplications (DUPS) were the most abundant type structural variation, affecting 4.9% (23.28 Mb) and
139 3.98% (18.84 Mbp) of Malbec-Mag (Fig. 2A). SNPs and InDels occurred every 147 and 1458 nt,
140 respectively; and SNPs distribution along the chromosomes was rather uneven (Fig. 1C). Several drops of
141 SNP density were detected, possibly as consequence of runs of homozygosity. The longest drop of SNPs
142 occurred on chr13 (Fig. 1C) and this overlapped with a large inversion between the haplophases (Fig. 2A),
143 which may appear as a SNP desert in the SyRI output [34]. Structural variations affected 9.97% (47.2 Mb) of
144 Malbec-Mag, more precisely inversions (INV) = 2.07%, translocations (TRANS) = 3.92% and duplications
145 (DUPS) = 3.98% (Fig. 2A and Table S3A). Finally, 76.5% (362.14 Mb) of the haplophases were identified as

146 syntenic and 14.2% (67.2 Mb) were classified as not aligned (Fig. 2A and Table S3A), altogether indicating
147 for a high heterozygosity of Malbec cultivar represented in the diploid assembly.

148 **Gene model predictions for the assembled haplophases:** Annotation and soft-masking of
149 transposable elements (TEs) was performed on the Malbec diploid assembly, after concatenating the two
150 haplophases (958 Mb). Almost half, 48.4 % (464 Mb), of the genome accounted for TEs, including
151 structurally intact and fragmented TEs (Table S4). The great majority was classified as long terminal
152 repeats (LTRs) representing 28.8% (276 Mb) of the genome, whereas Gypsy was the most abundant type
153 among the LTRs class with 13.36% (128 Mb) (Fig. 1C, Table S4). In addition, telomeric and centromeric
154 repeats were identified in the diploid assembly. Out of the 19 scaffolded pseudomolecules for each
155 haplophase, 14 and 18 featured telomeric repeats in at least one end in Malbec-Mag and Malbec-Pru (Fig.
156 1C). Centromeric repeats were also detected (between 173 and 5817 clustered copies) in most
157 pseudomolecules, with the exception of chr16 and chr18 for both haplophases, and for chr06, chr10 and
158 chr14 for Malbec-Pru. The fact that the annotated position of centromeric repeats overlapped between
159 the two haplophases, as well as with LTR-TEs, suggest that these regions might correspond to the -at least
160 partially- assembled centromeres (Fig. 1C).

161 The final weighed consensus gene models obtained for the Malbec diploid genome assembly
162 yielded similar metrics to those of the PN40024.v4 annotation [2]. The number of protein coding loci was
163 35,739 for Malbec-Pru and 35,463 for Malbec-Mag, with an average of 1.2 transcript isoforms per locus
164 (Table 2). Annotated genes were similarly distributed when comparing the pseudomolecules of both
165 haplophases. Drops of gene density overlapped with centromeric repeats, and the former were also
166 evident even in those pseudomolecules where peaks of centromeric repeats were not detected, like chr16
167 and chr18 (Fig. 1C), further supporting the putative position of the centromeres.

168 Considering all annotated isoforms, the total number of predicted proteins for Malbec
169 haplophases were higher (Malbec-Pru = 44,590 and Malbec-Mag = 44,183) than the one reported for the
170 PN40024.v4 reference genome (41,160) [2]. However, the size metrics (max., min., and average lengths, as
171 well as N50) of the predicted proteins were similar (Table 3A). BUSCO completeness for the proteins
172 annotated on each haplophase was >97%, whereas 98.5% completeness was achieved when the proteins
173 of both haplophases were jointly analyzed for BUSCO proteins (Table 3B).

174 The orthology analysis based on predicted proteins showed, as expected, that there were more
175 orthologues between Malbec haplophases than with the PN40024 (Table S5A). When only Malbec
176 haplophases were compared, that ~96% of the proteins could be placed in orthogroups, on average 4.1%
177 of the proteins remained unassigned, and 2.45% were haplotype-specific orthogroups (Table S5B). The
178 liftoff analysis, based on annotated genes, between both haplophases showed that there were 1,915

179 genes of the Malbec-Mag that were unmapped in Malbec-Pru, and 2,296 unmapped genes for the inverse
180 comparison. Interestingly, the GO enrichment analysis of unmapped genes suggested that the
181 trihydroxystilbene synthase activity and the RNA (uridine-N3-)-methyltransferase molecular functions
182 were overrepresented in Malbec-Mag and Malbec-Pru, respectively (Fig. S3). The riparian plots for
183 synteny visualization, showed that most gene orthogroups matched in their physical location across the
184 19 pseudomolecules of Malbec haplophases and PN40024.v4 (Fig. 2B). Furthermore, many genes located
185 in the Unknown chromosome of the PN40024.v4 had orthologues in genes assigned to chrMT, chr02,
186 chr07, chr16 and chr17 in Malbec-Mag and chr02, chr03, chr07 and chr17 in Malbec-Pru (Fig. 2B and Fig.
187 S4).

188 **Malbec haplophases retrieved congruent -although not identical- transcriptomic differences,**
189 **that resembled the clonal phenotype variation:** The obtained diploid assembly was employed to further
190 understand the molecular bases of Malbec intra-cultivar phenotypic variation for berry composition traits.
191 Phenotyping of 27 clones of Malbec grown in the same plot showed significant diversity for traits
192 evaluated on mature berries. In a principal component analysis (PCA), the first component (PC1) explained
193 46.2% of the variation, mainly driven by total anthocyanins (TA), total polyphenols (TP) and must pH,
194 whereas PC2 explained 27.3% of the variation and appeared to be associated with sugar content (Bx) and
195 total acidity (Ac) (Fig. S5A). From the starting 27 phenotyped accessions, four clones (595, 596, 505 and
196 136N) were selected for further transcriptomic analysis (Fig. S5B). The greatest phenotypic differences
197 among the selected clones were associated with total polyphenols (TP) and total anthocyanin (TA)
198 concentrations, with accessions 595 and 596 having the highest and lowest values ($p<0.05$), whereas 505
199 and 136N had intermediate values (Fig. 3A). These differences among clones were consistent across
200 seasons, despite the inter-annual differences in the TP and TA absolute values (higher in 2018 than in
201 2019) (Fig. 3B). Overall, accession 595 showed the highest TP and TA concentrations in mature berries at
202 two consecutive harvest seasons.

203 The RNA-seq analysis of berry pericarp of the four selected clonal accessions produced on average
204 45.7 million raw reads (6.7 Gb) per replicate (14x raw coverage), with 93.4% of the reads having base QC
205 > 30 and an average GC content of 46% (Table S1C). After transcript assembly and read mapping, 37,807
206 and 37,556 transcripts were effectively counted to align to Malbec-Pru and Malbec-Mag haplophases.
207 Despite this slight difference, the PCA based on global transcriptomic data showed that biological
208 replicates of clone 595 were consistently differentiated from the other accessions (PC1 = 57%) (Figs 4A
209 and B). According to their smaller phenotypic variation, the remaining three accessions appeared closer
210 in the RNA-seq based PCA (PC2 = 17%) (Figs 4A and B). Differentiation of clone 595 was corroborated
211 through clone pairwise comparisons, employing heatmaps of sample-to-sample distances (Fig. 4C and Fig.

212 S6A to F). For all the described analyses the relations observed among the replicates of the four clonal
213 accession was the same, using either of the haplophases as reference (Figs 4A-C and Fig S6A-F)
214 To further investigate the differentially expressed genes (DEGs) that were driving differentiation
215 of Malbec clone 595 from the others, pairwise comparisons were performed: 595vs136N, 595vs596 and
216 595vs505. For each comparison a very similar number of DEGs was detected with either haplophase as
217 reference (Figs 5A and B; Table S6). The number of DEGs ranged between 392 and 722, with 595vs505
218 comparison producing the highest number of DEGs and 595vs596 the lowest. Hereafter, downregulated
219 and upregulated DEGs are expressed in relative terms after comparing clone 595 to the other three. In all
220 comparisons there were more downregulated than upregulated DEGs (Figs 5A and B; Table S6). In each
221 comparison, on average 82% of the DEGs were detected with both haplophases (Fig. 5C), including 236
222 DEGs that were detected by both haplophases in all three comparisons (Fig. 5D). From DEGs detected
223 uniquely with one haplophase, a higher number was detected with Malbec-Pru than with Malbec-Mag
224 (Fig. 5C), whereas there were 14 DEGs consistently detected in the three comparisons only with Malbec-
225 Pru and six DEGs detected uniquely by Malbec-Mag (Fig. 5D). On the other hand, for results obtained with
226 Malbec-Mag an average of 7.4% of all the detected DEGs had no ortholog annotated in the Malbec-Pru
227 haplophase (Table S6), this was significantly above (Fisher's exact test $p<0.05$) of the total number of
228 unmapped genes (5.4% = 1,915 genes) in the other haplophase. For Malbec-Pru the inverse scenario was
229 observed, since an average of 3% of all the detected DEGs had no orthologue in Malbec-Mag (Table S6)
230 and this was significantly below (Fisher's exact test $p<0.05$) the total percentage of unmapped genes
231 (2,296 genes = 6.4%).
232 Gene Ontology (GO) enrichment analysis of biological processes (BPs) also produced similar results
233 using either of the haplophases as reference, both for the enriched terms and the direction of change
234 relative to clone 595. GO terms associated with 'response to abiotic stress', including 'response to salt
235 stress' and 'response to oxygen-containing compound' where downregulated, whereas genes related to
236 'secondary metabolism biosynthesis activity' were upregulated in clone 595 (Figs 6A-C). Even though
237 genes in the 'response to abiotic stress' GO term were mostly downregulated in 595, a number of DEGs
238 in the same category were also upregulated, indicating for a possible deregulation of stress responses in
239 berries of this clone (Fig. 6, Tables S6 and S8). In fact, a core of eight DEGs annotated in this category were
240 consistently identified in all comparisons. Five out of the eight core DEGs in the 'response to abiotic stress'
241 were consistently downregulated in 595 (Table S8), including: i) one gene (Mb_P_chr12g17550 |
242 Mb_M_chr12g17200 | Vitvi12g02138) coding for a small heat shock protein (sHSP) assigned to the three
243 enriched 'response to abiotic stress' child categories; ii) one gene coding for an ortholog of the DELLA
244 protein GAI1 (Mb_P_chr01g05780 | Mb_M_chr01g05800 | Vitvi01g00446) that assigned to 'response to

245 osmotic stress' and 'response to oxygen-containing compound' child categories; and iii) three genes that
246 coded for a SRM1 transcription factor (Mb_M_chr10g18440 | Mb_P_chr10g18110 | Vitvi10g01533), a
247 Scarecrow-like protein 3 (Mb_P_chr06g12050 | Mb_M_chr06g12350 | Vitvi06g01133) and a C2H2-type
248 domain-containing protein (Mb_M_chr01g09990 | Mb_P_chr01g10150 | Vitvi01g00845) assigned to the
249 'response to oxygen-containing compound' child category. On the other hand, three of the 'response to
250 abiotic stress' core genes were consistently upregulated in 595 (Table S8). More precisely, two genes
251 coding for sHSPs (Mb_M_chr13g05600 | Mb_P_chr13g05620 | Vitvi13g00491 and Mb_M_chr02g00360
252 | Mb_P_chr02g00330 | Vitvi02g00025), and one for a ZAT11 zinc finger protein (Mb_M_chr06g05190 |
253 Mb_P_chr06g05230 | Vitvi06g01682). Overall, the function of the mentioned proteins is related to
254 protein-folding homeostasis in response to stress [35] or regulation of transcription in the gibberellic acid
255 signaling in response to osmotic stress [36].

256 GO terms enriched with DEGs that were upregulated in clone 595 were associated with secondary
257 metabolism, more precisely to the phenylpropanoid and beta-glucoside biosynthesis (Fig. 6A-C). A core of
258 four DEGs assigned to these categories was consistently upregulated in the three pairwise comparisons
259 and with both haplophases as reference (Tables S6 and S8), including i) two genes that were assigned to
260 the 'phenylpropanoid biosynthesis' category, one coding for an Ultraviolet-B photoreceptor (UVR8-2)
261 gene (Mb_M_chr13g04770 | Mb_P_chr13g04840 | Vitvi13g00350) and another coding an F-box/kelch
262 repeat protein (Mb_M_chr05g19550 | Mb_P_chr05g19310 | Vitvi05g01703) (Table S8) and ii) two genes
263 that were assigned to 'beta-glucoside biosynthesis' category, one annotated as putative UDP-
264 glycosyltransferase in the reference gene catalogue (https://integrape.eu/resources/),
265 (Mb_P_chr12g12840 | Mb_M_chr12g12580 | Vitvi12g01718) and a *VviUFGT* gene (Mb_P_chr16g02060
266 | Mb_M_chr16g01980 | Vitvi16g00156) encoding the UDP-glucose:anthocyanidin:flavonoid
267 glucosyltransferase (UFGT) enzyme, that limits anthocyanin accumulation in grape berries [37]. Another
268 relevant gene in the anthocyanin biosynthetic pathway, *VvAOMT* (Mb_P_chr01g17010 |
269 Mb_M_chr01g16440 | Vitvi01g04438), was upregulated in the 595vs136N and 595vs596, but not in the
270 595vs505 comparison (Table S6).

271 Among the hundreds of detected DEGs some have key regulatory functions that could be driving
272 the described gene expression pattern. Furthermore, these genes appeared in the three comparisons with
273 consistent change directions (Fig. 7). More precisely, two homeodomain-leucine zipper transcription
274 factor (TF) encoding genes, *VviHB7* (Mb_P_chr15g12080 | Mb_M_chr15g11550 | Vitvi15g00912) and
275 *VviHB12* (Mb_P_chr02g02730 | Mb_M_chr02g02810 | Vitvi02g00228), were upregulated in 595 in all
276 comparisons (Fig. 7A). Orthologs of these genes play a role in stress-responsive co-expression networks
277 mediated by abscisic acid (ABA) signaling [38]. In fact, several ABA-related DEGs were consistently identified

278 here in all comparisons, including the upregulation in Malbec 595 of the ABA biosynthesis gene *VviNCED3*
279 (Mb_P_chr19g13140 | Mb_M_chr19g12630 | Vitvi19g01356), coding for a 9-*cis*-epoxycarotenoid
280 dioxygenase that is limiting for ABA biosynthesis, and one PP2C gene (Mb_P_chr16g13850 |
281 Mb_M_chr16g14410 | Vitvi16g01985); along with the downregulation of the *VviPYL7* ABA receptor
282 (Mb_P_chr15g13250 | Mb_M_chr15g13250 | Vitvi15g00997) and an ABA-8 hydroxylase coding gene
283 involved in ABA catabolism (Mb_P_chr18g15370 | Mb_M_chr18g15400 | Vitvi18g01304) (Fig. 7A).
284 Another ABA receptor, *VviPYL1* (Mb_P_chr02g09580 | Mb_M_chr02g09810 | Vitvi02g00695), was
285 upregulated in 595vs136 and 595vs505, whereas read counts were also higher in 595 when compared to
286 596 despite not being classified as DEG in the 595vs596 comparison (Fig. S7A and B). In addition,
287 *VviMYB157* (Mb_P_chr17g10170 | Mb_M_chr17g09800 | Vitvi17g00822), a homolog of the *WEREWOLF*
288 / *AtMYB66* TF, was upregulated in 595 in all comparisons (Fig. 7A). The allelic frequency of one SNV at
289 *VviMYB157* suggested downregulation of the Magdeleine allele in the other three clones (Fig. S7C and D).
290 Overall, the detected transcriptome variation during veraison correlated with the higher TA
291 concentrations on mature berries of 595 (Figs 3, 6 and 7).

292 **4. Discussion**

293 One of the main challenges for accurate genomic and transcriptomic clonal variation analyses in
294 grapevine is to start with an appropriate reference genome for the cultivar of interest. Here we present
295 the haplotype-resolved and annotated genome for cultivar Malbec, composed of the haploid
296 complements inherited from its parental cultivars Prunelard and Magdeleine. We also employed Malbec
297 diploid assembly as reference genome to perform a haplotype-aware evaluation of the transcriptomic
298 differences underlying clonal phenotypic variation.

299 **The use of parental reads improved the phasing and evaluation of the assembly process.**

300 A haplotype-resolved genome assembly for Malbec was obtained in a trio-based approach
301 applying prior knowledge of Malbec pedigree [25]. Trio phasing and haplotype-aware deduplication yielded
302 two haplophases that were close to the expected genome size of ~475 Mbp and with low duplicated
303 BUSCO genes (Table 1A). In this regard our results were comparable to the obtained with the Merlot
304 genome assembly, also based on trio binning [20]. This approach differ from the applied for assemblies of
305 other grapevine cultivars generated as pseudo-haplotypes, which exhibit inflated primary assemblies
306 compared to the haplotigs [6,10,12–15,39]. Nonetheless, recently developed bioinformatic tools help to
307 improve the proper assignment of primary contigs to the two haplophases in those cases [17].

308 We analyzed the assembly consensus and the phasing accuracy based on *k-mers* from the
309 parentals' short-reads whole genome sequencing (Figs 1A and B). This had only been reported before in
310 grapevine for the PN40024 reference genome [2]. The estimated assembly accuracy for Malbec (QV > 40)

311 indicates that error *k*-mers in the haplophase assemblies are < 0.00001%, which is slightly above the
312 accuracy (QV=37.43) reported for the reference genome [2]. Parental-specific *k-mers* (*hap*-mers) from short
313 read sequencing of Malbec parents identified very few blocks that were apparently switched (<0.025%,
314 Fig. 1 and Table 1B), in some cases these might be true homozygous regions in Malbec that were
315 misassigned as switched. These results emphasize the importance of grapevine pedigree analysis and the
316 value of using the parental reads, considering that some commercially relevant cultivars still lack this
317 information (e.g. Nebbiolo) [40].

318 The diploid assembly also enabled a genome-wide study of the variation between the two Malbec
319 haplotypes that showed ca. 25% of polymorphic regions (Fig. 2A and Table S3). The fraction of Malbec
320 haplophases affected by SVs (12.1%) is much higher than in Zinfandel (6.5%) [14], but lower than in Nebbiolo
321 (17.8%) [10] and close to Chardonnay (15.1%) [6], with the caveat that different SV evaluation methodologies
322 were used, and that a thorough validation of reported SVs was performed only for Nebbiolo [10]. Overall,
323 the fraction of the Malbec genome affected by SVs is in agreement with the general observation that
324 grapevine genomes can be unbalanced and that hemizygosity is widespread [6].

325 Almost half of Malbec genome consists of repetitive sequences and transposable elements (TE),
326 with long terminal repeats (LTRs) representing the most abundant class of TEs (Fig. 1C and Table S4). This
327 is in agreement with other grapevine cultivars [10,19,39,41], including the T2T assembly of the PN40024
328 genotype [4]. Annotated gene number was similar in the two Malbec haplophases (Tables 2 and 3), and also
329 to the PN40024 reference [2]. Nonetheless, close to 6% of genes did not have a counterpart mapping on
330 the other haplophase. Furthermore, an enrichment of the resveratrol biosynthetic pathway in Malbec-
331 Mag specific genes was observed (Fig. S3), suggesting for a putative higher contribution of this haplotype
332 to a key secondary metabolite in grapevines, both from an industrial and biological perspective [42].

333 Orthologous genes between the two Malbec haplophases were mostly located in syntenic regions,
334 whereas many genes assigned to the "Unknown" chromosome in PN40024 could be placed on specific
335 chromosomes across Malbec haplophases (Fig. 2B and Fig. S4). This suggests that the Malbec diploid
336 assembly is improved in contiguity and representation of previously difficult-to-assemble regions. This is
337 further supported by the correct location of most centromeres along the scaffolded pseudomolecules(Fig.
338 1C), when compared to the centromeres position in the T2T sequenced PN40024 [4].

339 **Clone transcriptome variation in ripening berries could explain differences in berry anthocyanins**
340 **content at harvest**

341 We searched for consistent phenotypic and transcriptomic variation among Malbec clones that
342 were cultivated in the same vineyard for over 20 years. Four clones were selected for analysis based on
343 their consistent variation in berry composition during 2018 and 2019 seasons (Fig. 3 and Fig S3A).

344 According to a previous intra-cultivar genetic variation analysis, the four selected clones represent
345 different derived genotypes from the same clonal genetic lineage (Group-Fr), with clones 595, 596, 136N
346 and 505 belonging to genotypes I, B, H and E, respectively [33]. Clone 595 had reproducible higher levels of
347 total polyphenols and anthocyanins berry content at harvest (Fig. 3), whereas it also showed a clear
348 transcriptome differentiation from the other clones (Figs 4 and 5). Functional enrichment of DEGs
349 indicated for abiotic stress response deregulation in 595 compared to the other three clones (Fig. 6; Tables
350 S6 and S8). On the other hand, phenylpropanoid and anthocyanin biosynthesis genes were consistently
351 upregulated in 595 (Fig. 6; Tables S6 and S8), in agreement with the observed phenotypic differences. In
352 our analysis, statically enriched processes (Fig. 6) correlated with the phenotype of the mature berry (Fig.
353 3). Similar relations were previously observed among Malbec clones diverging in berry anthocyanin
354 content, when the transcriptional activity of post-veraison berries was studied [30].

355 Particularly, we identified ABA-related DEGs in Malbec clone 595 that might be major drivers of its
356 global transcriptional profile and phenotype (Fig. 7A). ABA treatments have been reported to induce the
357 gene *VviNCED3* and PP2C encoding genes and also to downregulate ABA receptor genes in different
358 grapevine tissues including berry skin [43]. Similarly, *VviNCED3* and an homolog of the ABA-inducible
359 repressor PP2CA phosphatase [44] were upregulated in clone 595 compared to the other three, whereas
360 the *VviPYL7* receptor gene was downregulated (Fig. 7A, Table S6). *VviHB7* and *VviHB12* class I HD-ZIP TF
361 genes were also upregulated in 595. Their homologs in *Arabidopsis thaliana* have been implicated in a
362 negative feedback loop induced by ABA and repressing ABA receptor genes [45]. These expression patterns
363 suggest that ABA over-accumulation or hypersensitivity could be a feature of clone 595. However, even
364 though ABA usually represses *PYL* receptor genes [43], *VviPYL1* was upregulated in 595 (Figs 7A, S7 and
365 Table S6). We hypothesize that somatic mutations in *VviPYL1* of clone 595 could trigger its over-expression
366 and the hyperactivated ABA gene expression responses compared to other clones. To test this hypothesis,
367 we inspected RNA-seq reads alignments for possible allele-specific over-expression of *VviPYL1* in 595, that
368 could indicate the presence of a *cis*-acting somatic mutation in a regulatory region of this gene triggering
369 its overexpression. However, both the RNA-seq reads and haplophase assembly comparison by SyRI
370 support that Malbec is homozygous for *VviPYL1*, making it difficult to identify allele-specific differences
371 among clones (Fig. S7). Therefore, further functional analyses are required to identify the underlying
372 reasons explaining clone 595 differential response to ABA signaling.

373 Both the higher accumulation of anthocyanins and the differential expression of abiotic stress
374 genes detected in clone 595 (Fig. 6, Tables S6 and S8) could be explained by an alteration in ABA
375 homeostasis (Fig. 7B) [46]. The overexpression of *VviUFGT* in 595 compared to the other three clones
376 (Tables S6 and S8) could ultimately result in high berry anthocyanin content of this clone, since UFGT plays

377 a critical role in the synthesis and accumulation of anthocyanins in grape berries [47]. ABA treatment induces
378 *VviUFGT* in grape berry skins, increasing its anthocyanin content [43]. Therefore, ABA signal hyper-activation
379 in 595 could potentially over-activate the flavonoid pathway leading to the higher phenolic character in
380 595 fruits (Fig. 7B). Transient overexpression of *VlPYL1* in grape berries induces ABA-responsive genes like
381 NCEDs, but also *VviUFGT* [48], suggesting that the overexpression of *VviPYL1* in 595 could be an upstream
382 trigger of both the ABA hypersensitivity and the high berry anthocyanin content in this clone (Fig. 7A-B).
383 In addition, the putative allele-specific overexpression of *VvMYB157/MYB66* in 595 (Fig. S7) could be
384 involved in the deregulation of the ABA signaling and the upregulation of the anthocyanin biosynthesis
385 pathways characterizing this clonal line (Fig. 7B). *VvMYB157* might be particularly relevant in Malbec
386 clonal anthocyanin content variation, since this cultivar is homozygous for *VvMAYBA1* and *VvMYBA2* loci
387 and no evidence of differential gene expression was detected here (Fig. S7). A previous study of mature
388 berries also reported the absence of variation for *VvMYBA1* expression among Malbec clones with
389 different anthocyanins content [30]. On the other hand, overexpression of a *MYB157/MYB66* homolog in
390 the fruit skin of Russet apple somatic variants was associated with activation of ABA responses and the
391 phenylpropanoid pathway [49]. Our results warrant further functional genomic analyses to confirm the role
392 of *VvMYB157* in the regulatory mechanism responsible of the high phenolic content found in Malbec clone
393 595.

394 **Phasing haplotypes helps to identify haplotype-specific gene expression.**

395 Analysis of transcriptomic data with either of the two Malbec haplophases as reference retrieved
396 similar results, however some differences were observed. In this direction, 0.58% more transcripts aligned
397 when Malbec-Pru was used as reference than with Malbec-Mag. This is in line with the fact that Malbec-
398 Pru assembly turned 0.06% larger in size (Table 1A), accounting with 0.78% more protein coding
399 annotated genes (Table 2) and 0.92% more transcripts predicted (Table 3), than Malbec-Mag. The analysis
400 of cultivar Merlot haplotype-resolved genome also reported a similar difference, regarding annotated
401 gene number between the two haplophases [20]. In our analysis, the number of detected DEGs was 2.7%
402 (±1.2% SD) higher with Malbec-Pru than with Malbec-Mag, on average for the three comparisons
403 involving Malbec clone 595 (Fig. 5A and 5B). This observation is in agreement with the expected by chance
404 based on the assembly and annotation features of both haplophases. Regarding the haplotype-specific
405 DEGs, we found that six and 11 genes were consistently detected in all pairwise comparisons only with
406 Malbec-Mag and Malbec-Pru (Fig. 5D), from which four and six had no ortholog in the other haplophase
407 (Table S7), respectively. There is indeed a significant enrichment of Malbec-Mag putative hemizygous
408 DEGs that were unmapped in the other haplophase, in contrast there was underrepresentation of Malbec-
409 Pru DEGs that were unmapped in Malbec-Mag. Furthermore, two of the haplotype-specific DEGs detected

410 only with Malbec-Pru coded for fruit development relevant functions. More precisely, an acyltransferase
411 protein involved in fruit ripening (Mb_P_chr11g08440 | Vitvi11g04184) that was upregulated in clone
412 595, and an F-box like domain protein involved in the regulation of phenylpropanoids biosynthesis
413 (Mb_P_chr08g18140 | Vitvi08g01787) that was downregulated (Table S7). Visual inspection of RNA-seq
414 alignments on Malbec-Pru reference indicate putative hemizygosity only for Mb_P_chr11g08440. On the
415 other hand, two consistent heterozygous SNPs in Mb_P_chr08g18140 indicate that Malbec might not be
416 hemizygous for this locus, even though its ortholog was not predicted in Malbec-Mag haplophase (Fig.
417 S7).

418 Overall, our findings are in agreement with the general observation that the genomes of grapevine
419 cultivars are unbalanced and that hemizygosity is widespread [6,11]. However, caution must be taken since
420 misassembly and/or misannotation might be also overestimating the reported hemizygosity. Further
421 functional experiments employing our diploid genome assembly will be required to corroborate the
422 putative allele-specific expression differences, in order to fully understand how such differences might
423 impact complex phenotypes in grapevine.

424 **5. Conclusions**

425 We have assembled the diploid genome of Malbec in two haplophases of the expected size and
426 phase. Similar results were obtained in our RNA-seq analysis with either haplophase used as reference,
427 nonetheless some haplotype-specific gene expression differences could be spotted thanks to the phased
428 genome assembly. In this direction, genes with no ortholog annotated in the other haplophase resulted
429 in the enrichment of relevant biological processes for grapevines, also a significant number of all the
430 detected DEGs with Malbec-Mag had no ortholog in Malbec-Pru. Finally, we proposed that a somatic
431 variant causing the hyperactivation of ABA signaling and the differential activity of a MYB transcription
432 factor in a Malbec clone, might be responsible for its increased berry anthocyanin content.

433 **6. Materials and Methods:**

434 **6.1 Genome *de novo* assembly and annotation:**

435 **Plant material:** For genomic sequencing and assembly, samples were collected from the Malbec
436 clone accession 136N, representing a clonal lineage with short propagation history in Argentina (i.e.,
437 Group-Fr) [33]. Accession 136N is implanted at the Vivero Mercier Argentina experimental vineyard located
438 at Lujan de Cuyo, Mendoza, Argentina (-33.09° S, -68.87° W). Samples of Malbec parental cultivars [25] were
439 requested to the *Institut Français de la Vigne et du Vin* (VInopôle Sud-Ouest, France), corresponding to
440 clone 1232 of Prunelard (reference number: 08-04-00004) and the unique available accession of
441 Magdeleine Noire des Charentes (reference number: 08-04-00075).

442 **DNA extractions, library preparation and genomic sequencing:** Single node wood samples of the
443 mentioned accessions were hydroponically grown, until the development of roots and shoots with young
444 leaves. For PCR-free Illumina sequencing, DNA was obtained from roots for all three cultivars. For the
445 parental cultivars solely roots were used, because this tissue derives from the L2 meristem cell layer (as
446 gametes do in grapevine) and its genotype should be closer to the inherited by Malbec, in case chimeric
447 somatic mutations existed. For Prunelard and Magdeleine, DNA extractions were performed with the
448 DNeasy Plant mini kit (Qiagen). For Malbec, high molecular weight (HMW) DNA was obtained with the
449 Nanobind plant nuclei kit (Circulomics-PacBio), either from roots or young leaves. Leaves were used for
450 long read sequencing because it was the only tissue yielding the amount of HMW DNA needed for library
451 preparation. In brief, nuclei were isolated from the source tissue according to Workman *et al*. [50]. Isolated
452 nuclei were the input for HMW DNA extraction with the Nanobind nuclei kit following manufacturer's
453 instructions. The HMW DNA was sheared with two strokes using a 26 G needle in a 1 mL syringe, with an
454 average fragment size of 70 kb.
455 Library preparation and genomic sequencing were performed at the Max Planck Institute for
456 Biology (Tübingen, Germany). A large insert PacBio gDNA library was prepared from the sheared HMW
457 DNA of Malbec leaves using the SMRTbell Express Template Preparation Kit 2.0. The library was size-
458 selected for >30 kb using BluePippin with a 0.75% agarose cassette (Sage Science) resulting in a 48 kb peak
459 size final library. Five PacBio Sequel I SMRT cells were loaded with the same SMRTbell library and were
460 sequenced with the PacBio Sequel I sequencer. Illumina PCR-free sequencing libraries were produced
461 from gDNA of roots, using the NxSeq® AmpFREE Low DNA Library Kit (Lucigen) and sequenced to obtain
462 paired-end short reads (150 bp), as described in Rabanal *et al.* [51].
463 **Genome de novo assembly:** The assembler Canu v1.8 [52] was employed for *de novo* genome
464 assembly using the TrioBinning module [16]. Here, parental-specific *k-mers* from the short-reads of the
465 parental cultivars (Prunelard and Magdeleine) were used as template to split long reads of the child
466 (Malbec) into the haploid complements. After adapters trimming with cutadapt v2.3 [53], Magdeleine and
467 Prunelard short-reads were loaded into Canu assembler as *-haplotype1 and -haplotype2*, respectively,
468 whereas Malbec PacBio CLR reads were loaded with option *-pacbio-raw*. The child haplotypes inherited
469 by Malbec cultivar from each parent were separately assembled from the PacBio read partition,
470 theoretically without the interference of inter-haplotype variation [16]. Canu was run in all cases with option
471 *genomeSize=490m*. No contaminant contig from non-plant genera or phyla was found from the draft
472 assemblies by running blobtools v.1.1.1 [54], comparing to the non-redundant (nr) protein database using
473 diamond v2.0.8.146 [55]. The post-processing of the raw assembled contigs representing each haplotype of
474 Malbec (hereafter: haplophases) involved a first round of polishing based on the long reads raw data using

475 the ArrowGrid wrapper [52,56] of Arrow (http://github.com/PacificBiosciences/GenomicConsensus/)
476 consensus framework, within the SMRT Analysis Software (http://github.com/PacificBiosciences/SMRT-
477 Analysis). ArrowGrid was run on the primary assembly output of Canu with the corresponding haplotype-
478 binned Malbec PacBio CLR reads for each haplophase, separately.

479 The assembly size metrics were tested along with the quality of the phasing process, to reduce
480 haplotype switch errors [16,57]. This was performed with Merqury v1.3 [57], based on the count of haplotype
481 specific *k-mers* (i.e., *hap-mers*) from the short reads of Magdeleine and Prunelard as compared to Malbec
482 short reads and the assembly *k-mers*. *Hap-mers* were used to determine phase blocks, defined as sets of
483 markers inherited from the same parental haplotype [57]. Each assembled haplophase genomic composition
484 was tested using the default values, allowing 100 marker switches in a 20 kb bounded range (0.5% switch
485 rate). Afterwards, complete contigs showing haplotype switch errors all along the contig and no-matches
486 to the own haplotype in the Merqury output, were removed using a custom script (based on Unix/Linux
487 *awk* and *grep* commands). Then, another haplotype-aware deduplication was performed based on
488 minimap2 [58] alignments of one haplophase assembly to the other using *-x asm5* alignment option. Average
489 similarity between haplotypes was computed for each contig from the *dv* value of the output PAF file.
490 Candidate haplotype-switched contigs with high similarity to the other haplotype (average dv < 0.005)
491 were extracted and used as reference to align the rest of contigs from the same haplophase with
492 minimap2 (*-ax asm20 -secondary=no*). To check for possible duplicated contigs from that alignment, the
493 coverage utility from samtools v1.10 [59] was run on the resulting *bam* file. High similarity contigs to the
494 other haplotype were removed if they showed >70% bases covered, with depth ≥1 and mean depth of
495 coverage >0.7 from the other contigs in the same haplophase. The resulting haplophases were further
496 deduplicated with purge_dups v1.2.5 [60], read depth was used to remove duplicated contigs, by mapping
497 the haplotype-binned Malbec PacBio reads to the corresponding haplophase assembly separately. The
498 purge_dups command get_seqs was run with *-s* option to break contigs instead of adding Ns at internal
499 breaks.

500 To perform miss-assembly corrections, RaGOO v1.1 [61] was used to split contigs that may generate
501 spurious structural variations using the PN40024 12X.v2 [3] as reference, with options *-b -r 300000 -c*
502 *15000000 -d 30000000 -g 100 -s -C*. Afterwards, another round of assembly consensus correction based
503 on long reads data was performed with ArrowGrid, in this case using all the Malbec PacBio CLR subreads
504 to correct deduplicated contigs of the two haplophase assemblies concatenated. To overcome the high
505 error rate of long reads, PCR-free Illumina short-reads of Malbec were used to polish the concatenated
506 haplophases contigs, using Pilon v1.23 [62] with options *--fix snps,indels --minmq 3*. Then, haplophases were

507 split according to their original contig IDs; and another round of haplotype-switch error detection was
508 performed based on Merqury *hap-mer* analysis.

509 A final manual deduplication was conducted based on the gene content of the contigs spotted as
510 haplotype switches by Merqury, and on BUSCO duplicated genes identified using the orthologue plant
511 core genes database (eudicotyledons_odb10) and BUSCO v5.4.4 [63]. When ≥2 BUSCO genes with different
512 IDs were duplicated in the same order in another assembly region that involved a haplotype-switch error
513 spotted by Merqury, this situation was considered as an artifactual haplotype duplication. Therefore, the
514 switched region of the contig was removed, by running the *complement* and the *getfasta* applications of
515 bedtools v2.27.1, using the contigs assembly (*fasta* file) and the duplicated coordinates (*bed* file). The
516 boundaries of the coordinates of the duplicated regions to be removed were refined according to
517 minimap2 alignments, with option *--secondary=no -x asm20 --MD* of each pair of duplicated contigs to the
518 other haplophase contigs assembly. When no duplicated BUSCO genes were observed, then the contig
519 spotted as switched by Merqury was consider as a likely homozygous region and was kept in the final
520 assembly. After running the deduplication pipeline, another round of RaGOO miss-assembly correction
521 was run as described before leading to only one break between chr2 and chr7 in Prunelard haplophase
522 and none in Magdeleine haplophase.

523 RagTag v1.0.1 [64], using the PN40024 12X.v2 assembly as reference (including the 19 scaffolded
524 chromosomes of grapevine, as well as the plastid and mitochondrial genomes), was employed for the final
525 scaffolding of the deduplicated haplophases into the pseudo-molecules that represent the 19
526 chromosomes of grapevine haploid complement. The ragtag.py script was run with options *--remove-*
527 *small -f 20000 -q 10 -d 500000 -i 0.5 -a 0.05 -s 0.3 -u --mm2-params "-x asm20*. Seqkit [65] was used to obtain
528 assembly summary statistics of size and contiguity. SyRI v1.3 [34] was used to quantify the synteny and
529 diversity between Malbec assembly haplophases. To run SyRI, the two haplophase assemblies were
530 aligned with nucmer using options *-maxmatch -l 100 -c 500 -b 500* and delta-filter with options *-m -l 88 -l*
531 *100*. Nucmer and delta-filter were used from mummer v4.0.4 [66]. Plots were produced with plotsr from
532 SyRI.

533 **Genome assembly annotation:** Prior to gene annotation, EDTA v1.9.6 [67] was employed for the
534 annotation of repeats and transposable elements (TEs), and to soft-mask the repeated regions. A repeat
535 library was generated by running the EDTA.pl script with the options *--sensitive 1 --anno 1* on the
536 concatenated assembly (Malbec scaffolded haplophases), while non-TE plant proteins were excluded
537 from the repeat library. First, blastx from blast v2.2.29 was run with options *-evalue 1e-10 -*
538 *num_descriptions 10* to identify overlaps between the repeat library and a curated plant protein database
539 obtained from the MAKER-P manual [68]. Then, the overlaps were excluded from the EDTA repeat library

540 by running ProtExcluder.pl [68]. Finally, the curated repeat library was used to annotate TEs in the diploid
541 assembly by running again EDTA.pl with options *--step anno --sensitive 1 --anno 1*.
542 For gene annotation a pipeline combining gene evidence obtained from different sources based
543 on EVidenceModeler v1.1.1 [69] was implemented, using as input genome the soft-masked Malbec diploid
544 assembly (concatenated and scaffolded haplophases). Here, *de novo* assembled transcripts, *ab initio*
545 predictions and gene lift-over data were combined, and obtained as follows.
546 Transcriptome assembly was produced from an RNA-seq dataset obtained from berry pericarp of
547 eight different Malbec clones and their biological replicates. Clones 225, 228, 59 and 53, were sampled
548 during 2016 and employed solely for the annotation pipeline (Table S1C). For these samples total RNA
549 extractions were performed from mature berries pericarp using a CTAB based protocol [70]. Three to four
550 berries were grounded with liquid nitrogen discarding the seeds, to obtain the desired amount of RNA
551 (>3,5 ug). Afterwards, a column-based purification step was conducted with the kit *SV total RNA isolation*
552 *System* (Promega). Extracted RNA integrity, concentration and purity were tested in 2% agarose gel and
553 by spectrophotometry (AmpliQuant AQ-07). Samples were shipped to CRG facilities (Barcelona, Spain) for
554 library preparation and sequencing. Library preparation was performed with a TruSeq Stranded mRNA LT
555 Sample Prep Kit (Illumina) and included rRNA depletion using the Ribo-Zero kit (Illumina). Sequencing was
556 performed with Illumina HiSeq 2500 and paired-end fragments of 125 bp were obtained. The four clones
557 employed for transcripts *de novo* assembly, were also used for the clonal transcriptomic variability
558 analysis (see that section for further details on total RNA obtention). Overall, a batch of 22 RNA-seq
559 samples was employed for transcripts *de novo* assembly, adding ~150 Gb of raw sequence (Table S1C).
560 The Illumina sequencing adaptors and low quality reads were trimmed with TrimGalore v0.6.4 (Babraham
561 Bioinformatics) that includes cutadapt v2.3 for read trimming and FastQC [71] for sequence quality checks.
562 Two rounds of TrimGalore were run, the first to remove adaptors and low-quality reads with options *-q*
563 *20 --paired --length 30 --max_n 1 --trim-n* and the second to remove polyA tails and run the final QC with
564 options *--paired --length 30 --polyA --fastqc*. For transcriptome assembly, the trimmed RNA-seq reads
565 were aligned to the diploid Malbec genome assembly using HiSat2 v2.2.1 [72] with options *--rna-strandness*
566 *RF --dta-cufflinks --max-intronlen 20000*. The obtained alignments (*bam* files) were merged into a single
567 *bam* using the merge application of samtools v1.10 [59] and assembled into potential transcripts with
568 StringTie v2.2.1 [73], using the options *-t -c 2 -f 0.1 -s 4.75 -M 0.95 -m 80 -j 2 -A --rf*. Starting from a genome-
569 based transcript structure obtained with StringTie, the software TransDecoder v5.7.0 (Haas,
570 BJ. https://github.com/TransDecoder/TransDecoder) was used to identify the likely coding sequences
571 (CDs), adding them to the annotation files.

572 On the other hand, *ab initio* gene predictions were performed with Augustus v3.5.0 [74]. Before
573 running Augustus, Braker2 v2.1.6 [75] was run for the Malbec diploid assembly using as strand-aware hints
574 the *bam* files produced by HiSat2 from the malbec RNA-seq data. This was performed to generate the
575 retraining configuration file to be used as input in the *--species* option of Augustus. The EDTA soft-masked
576 diploid assembly was used as input genome to run both Augustus and Braker2. To run Augustus, *extrinsic*
577 *M.RM.E.T* information configuration was used as described in the *extrinsic.cfg* file (File S1) and the UTR
578 prediction was inactivated *--UTR=off*. The hints input for Augustus was a *gff* file with exons annotated by
579 TransDecoder from the StringTie transcriptome assembly of the merged Malbec RNA-seq (*gtf* file). Gene
580 models resulting from Augustus were kept as the final *ab initio* gene predictions, but gene models were
581 only kept for loci that showed some overlap with gene models that were also predicted in both, the final
582 output of Braker2 and the output of Genemark that Braker2 uses for its retraining. The intersection
583 between the three gene prediction version *gff* files was obtained by running a custom script based on
584 *comm*, *grep* and the *intersect* -c command of bedtools v2.27.1 [76].
585 Also, a lift-over of the PN40024.v4 gene annotation to the Malbec haplophases assemblies was
586 conducted with Liftoff v1.6.1 [77], with options*: -a 0.5 -s 0.5 -copies -sc 0.99 -exclude_partial -polish*. This
587 included a final polishing step and transferred genes were only kept if similarity and coverage were >50%
588 for the primary gene copy, whereas for secondary copies the similarity threshold was increased to 99%.
589 The PN40024.v4.1 gene annotations of the improved grapevine reference genome assembly [2] were
590 retrieved on 15.11.2021, from the Integrape Cost Action CA17111 portal (https://integrape.eu).
591 The weighed consensus gene models for Malbec diploid and unmasked genome assembly were
592 obtained with EVidenceModeler, using the evidence obtained from the described sources. The final
593 combination of evidence inputs and weights was selected after testing more than 50 EVidenceModeler
594 runs and comparing the obtained gene models to the deeply curated PN40024.v4 [2]. For EVidenceModeler,
595 the RNA-seq transcriptome assembly obtained with StringTie was used as TRANSCRIPT input with the
596 maximum weight = 4, the lifted-over of PN40024.v4.1 reference annotations obtained with Liftoff was
597 loaded as OTHER PREDICTION (weight = 3), and the gene model produced by Augustus was used as AB
598 INITIO PREDICTION with the minimum weight = 1.
599 Finally, the *annotation update* function of PASA v2.4.1 [69,78] was used to polish exon boundaries
600 and to update the 5' and 3' UTRs to the genes structures and alternative isoform annotations, previously
601 predicted by EVidenceModeler. Here, the input transcripts were the ones assembled with StringTie from
602 the Malbec RNA-seq samples and the transcripts annotated in the primary (REF) and alternative (ALT)
603 haplotypes of the PN40024.v4.1. All these transcripts were cleaned from polyA, low quality or vector
604 sequences using the *seqclean* script of PASA. Cleaned transcripts were aligned to the Malbec diploid

605 assembly with Blat [79] and Gmap [80] aligners using the *Launch_PASA_pipeline.pl* script with –
606 *transcribed_is_aligned_orient* option and the following options in the alignAssembly.config file:
607 validate_alignments_in_db.dbi:--MIN_PERCENT_ALIGNED=90; validate_alignments_in_db.dbi:--
608 MIN_AVG_PER_ID=95; subcluster_builder.dbi:-m=50; run_spliced_aligners.pl:-N=5;
609 validate_alignments_in_db.dbi:--MAX_INTRON_LENGTH=25000. Two consecutive rounds of PASA
610 updates from the aligned clean transcripts were run using again the *Launch_PASA_pipeline.pl* script, by
611 loading firstly the EVidenceModeler annotations and secondly the updated annotations resulting from
612 the first round. Genes were renamed according to assembly chromosome coordinates using the script
613 retrieved from https://github.com/andreaminio/AnnotationPipeline-EVM_based-DClab [81].
614 To perform QC at each annotation step, the general metrics of the predicted gene models were
615 computed using the *agat_sp_statistics.pl* script from AGAT v0.8.0 [82]. Transcripts and translated proteins
616 were extracted from *gff* annotation files with Gffread v0.11.7 [83] and the general stats were computed
617 with Seqkit v0.12.0 [65]. Completeness and duplication levels for the annotated protein sets were obtained
618 running BUSCO v3.0.2 (on protein mode), using the eudicotyledons_odb10 dataset lineage [75].
619 Functional annotation to predict gene ontology (GO) classes and gene functional description (DE)
620 was performed with Panzzer2 [86], using as input the *fasta* files containing all the protein sequences of each
621 haplophase. The correspondence between the obtained gene model predictions for each haplophase,
622 PN40024.v4 and VCostv3 annotations was performed using the proteins amino acid sequences, with
623 OrthoFinder [87]. Afterwards, GENESPACE [88] was used to evaluate the order and location of the predicted
624 gene models along the scaffolded pseudomolecules of the assembled haplophases, compared to the
625 PN40024.v4.
626 To annotate centromeres and telomeres, a search for the 107 nt monomer (CEN107=
627 AGTACCGAAAAAGGGTCGAATCAGTGTGAGTACCGAAAAATGGTAGAATCCGGGCGAGTACCGGGAAAAGGTA
628 GAATCCGTGCGAGTATCGAAAAACTGTCCGGGCG) of centromeric repeats and the 7 nt monomer
629 (TTTAGGG) of telomeric repeats was performed [89]. EDTA v1.9.6 [67] was used to annotate centromeric
630 repeats on the Malbec diploid assembly by running the EDTA.pl command with *--sensitive* 1 *--anno* 1
631 parameters from a repeat library containing only the CEN107 monomer. Telomeric repeat positions were
632 identified on the diploid assembly using BLAST v2.2.29+ [90]. Blast alignment was run with oligonucleotide
633 query parameters (*blastn* -task *blastn-short*) using as query two consecutive copies of the 7-mer
634 monomer. Finally, the density of all the annotated elements was plotted along the assembled
635 pseudomolecules using ggplot2, by taking the coordinates of the start position of each annotated element
636 as well as of variants detected between Malbec haplophases.
637 **6.2. Analysis of the phenotypic and transcriptomic clonal variation:**

638 **Clonal phenotypic variation analysis:** A survey focused on berry composition variation was
639 performed on 27 Malbec clones, with three biological replicates, resulting in 81 plants. All analyzed plants
640 were located in the same plot at Vivero Mercier Argentina experimental vineyard (Luján de Cuyo,
641 Mendoza, Argentina), implanted since 2002. Therefore, the analyzed plants have been exposed to the
642 same cultural treatments and environmental conditions ever since. Phenotypic measurements were
643 conducted during 2018 on mature berries. For each biological replicate a random sample of 50 berries
644 was obtained and the pulp was separated from the skin. The pulp was used for analytical measures (pH,
645 total acidity and total soluble solids) and the skin was used for spectrometry analyses (total anthocyanins
646 and total polyphenols). The pH was directly measured with a Denver Instrument (UB-10, pH/mV Meter).
647 Total acidity (expressed in g/L of tartaric acid) was measured through acid-base titration with 0.1 N NaOH
648 (pH = 8.2), using 5 mL of juice diluted with 10 mL of distilled $H_2O$ and containing bromothymol blue
649 (0.04%). Total soluble solids were measured in wt/wt of sucrose, expressed in Brix degrees (Bx), using an
650 ATAGO refractometer (PAL-1). The skins were dried and weighed to perform total polyphenols extraction
651 using a ethanol and water solution (12:88, v/v), containing 5 g/L of tartaric acid, according to Muñoz *et al.*
652 [30]. Total polyphenols (TP) and total anthocyanins (TA) were measured with a UV-VIS spectrophotometer
653 (Spectrum SP-2000UV) at wavelengths of 280 and 520 nm respectively, and expressed in AU/g of skin. The
654 same procedures were repeated for the phenotypic analyses in 2019, only for the clonal accessions
655 selected for transcriptomic analyses.
656 **Plant material selection for clonal diversity transcriptomic experiments:** Four Malbec clones
657 (595, 596, 136N and 505) were selected after a ranking-based procedure, based on phenotypic data
658 obtained for the 27 clones in 2018. The selected samples combined high, mid and low values for the
659 surveyed traits, except for sugar content trait where the selected clones had either high or low values,
660 aiming to represent as much as possible of the observed variation (Fig. S4B). ANOVAs and Tukey HSD tests
661 were performed for pairwise comparisons between the selected clones, to test for significant differences
662 among the analyzed traits ($p$<0.05). Berries used for total RNA extraction were sampled during *veraison*
663 stage (switch point from berry development to ripening, when 75% of berries were colored); a
664 transcriptionally active stage for genes involved in secondary metabolism pathways [91]. Sampling was
665 performed previous to 2019 harvest, within the same day (01.15.2019) and during morning time (between
666 8-10 a.m.); the collected berries were kept in liquid nitrogen until final storage in ultra-freezer (-80°C).
667 **Total RNA extraction, library preparation and sequencing:** Total RNA extractions and quality
668 checks were performed as previously described (annotation section). RNA samples were shipped to
669 Novogene (Beijing, China) for sequencing. At arrival, samples were tested before library preparation and
670 all passed the QC performed for quantitation (NanoDrop, ThermoFisher), degradation (agarose gel) and

671 integrity (Agilent 2100, Bioanalyzer systems). Library preparation was performed with a TruSeq Stranded
672 mRNA LT Sample Prep Kit (Illumina) and included rRNA depletion using the Ribo-Zero kit (Illumina). All
673 samples passed the library QC performed for concentration (Qubit 2.0, ThermoFisher), insert size (Agilent
674 2100) and library effective concentration (qPCR). An Illumina HiSeq 2500 instrument was used for
675 sequencing, to obtain 150 bp paired-end reads.

676 **Differentially expressed genes and functional enrichment analyses:** The base quality of the raw
677 reads was checked with FastQC [71]. Trimming of sequencing adaptors and low quality sequences was
678 performed with Trimmomatic [92]. rRNA and tRNA sequences were removed from the trimmed Illumina
679 reads using Bowtie2 [93] aligner, by mapping the reads to a specific grapevine database [94]. All forthcoming
680 bioinformatic procedures were performed following a haplotype-aware scheme, using in parallel both
681 Malbec haplophases as reference genomes. Alignment of the trimmed reads was conducted with STAR [95],
682 the assembly of the aligned reads with StringTie [73], and GffCompare [83] was used for extraction and
683 annotation of transcripts. The read counting of the aligned *bam* files was performed with FeatureCount
684 [96], according to Chialva *et al*. [97].

685 Heatmaps and PCAs displaying the global gene expression were built in R, based on the regularized
686 log transformation of the normalized counts (rld), obtained with DESeq2 [98]. To detect differentially
687 expressed genes (DEG) with DESeq2 [98], pairwise comparisons among the four clones were performed. The
688 threshold to consider a gene to be differentially expressed (DEG) was set to a *p*-value<0.01 (FDR-adjusted)
689 and an absolute value of $\log_2$ fold-change $|LFC|\geq 1$. Functional enrichment analyses were conducted using
690 the list of DEGs with g:Profiler2 [99], to find statistically significant gene ontology (GO) terms related to
691 Biological Processes (BP) and Molecular Functions (MF). For all GO enrichment analysis the g:GOSt
692 function of g:Profiler was implemented using the PN40024.v4 functional annotation as reference. Aiming
693 to compare DEGs and BPs identified with each of Malbec haplophases, the orthologue genes obtained
694 with Orthofinder were employed. For all performed GO terms enrichment analyses, only manually curated
695 annotations of experimental and computational studies were considered, and electronically inferred
696 annotations were dismissed to avoid sporous results. This was performed based on a *p*-value <0.05 (g:SCS-
697 corrected for multiple comparisons [99]). The aggregate scores and z-scores where obtained and graphically
698 represented with the R package GeneTonic [100]. Z-scores summarize the direction of change based on the
699 LFC values of the gene set involved in the enriched process. Positive z-scores (color red) indicate that a BP
700 GO term is upregulated, whereas negative z-scores (color blue) indicate downregulation of a BP [100].
701 GeneTonic function *gs_overlap* = 0.9 was employed to summarize redundant GO terms and
702 *gs_summarize_overview_pair* was employed to make dumbbell plots, to visualize simultaneously the
703 enriched GO terms detected with both haplophases. The web tools UpSetR

704 (https://gehlenborglab.shinyapps.io/upsetr/) and jVenn [101] were employed to perform upset plots and
705 Venn diagrams for DEGs pairwise comparisons.

706 **7. Acknowledgements**

707 This work was supported by Agencia Nacional de Promoción Científica y Tecnológica (ANPCyT):
708 PICT2018-02381, MINCyT/ANPCyT (FONTAR)-CDTI IBEROGEN; CONICET (Bilateral PCB-II, CONICET-CSIC;
709 MINECO BIO2017-86375-R), and the Max Planck Society. L.C. was supported by a CONICET travel grant for
710 a short-term staying at the Max Planck (Tübingen). We are grateful to Ilja Bezrukov for answering
711 bioinformatics question, and to Silvina van Houten for collaboration with field work. We thank Olivier
712 Yobregat (Institut Français de la Vigne et du Vin, IFV) for providing plant material from Malbec parental
713 cultivars. This project received funding from the European Union's Horizon 2020 research and innovation
714 programme under the Marie Sklodowska-Curie grant agreement No 797460

715 **8. Author contributions**

716 L.C. coordinated the project and wrote the article. Produced and analyzed phenotypic,
717 transcriptomic and genomic data. P.C.B. laboratory work for genomic sequencing and designed the
718 bioinformatic pipelines for genome assembly and annotation, global data interpretation. C.M. phenotypic
719 and transcriptomic diversity analysis. L.B. field and laboratory work for phenotypic diversity analysis. D.B.
720 and C.S. project design, provided insight and access to the analyzed clonal accessions. W.T. bioinformatic
721 analysis, set-up and maintenance of genome browse at IBAM-CONICET web-page. C.L. laboratory work
722 for genomic data obtention. S.G.T., J.I. and J.M.M.Z. project design for genomic and transcriptomic
723 experiments. C.R., J.I. and J.M.M.Z. obtained parental cultivars plant material and provided resources for
724 whole DNA extractions. D.W. provided resources, facilities and advice for genomic laboratory work and
725 bioinformatic analyses. D.L. designed and coordinated the entire project, performed transcriptomic
726 analysis. All authors read and improved this manuscript

727 **9. Data availability statement**

728 Malbec-Mag and Malbec-Pru assemblies are available at the NCBI public repository under
729 BioProjects: PRJNA1036636 and PRJNA1036637, respectively. All raw data is available at NCBI under
730 BioProject: PRJNA1037531, including genomic Illumina short-reads obtained for cultivars Malbec,
731 Prunelard and Magdeleine, PacBio long reads for Malbec and a batch of 22 RNA-Seq from eight Malbec
732 clones and replicates. Malbec annotated assemblies are also available for visualization and gene browse
733 at two public repositories: IBAM-CONICET (http://ibam.mendoza-conicet.gob.ar/resources/genome-
734 visualizer/) and Gramene (https://www.gramene.org/). A gene correspondence between Malbec-Pru and
735 Malbec-Mag annotated genes, alongside with PN40024.v4 and VCost.v3 annotations is provided,
736 therefore gene queries can be performed using any of the mentioned formats.

## 10. Conflicts of interests

The authors declare no conflicts of interest.

## 11. Supplementary data

Supplementary data is available at Horticulture Research online.

## 12. References


1. Jaillon, O. *et al.* The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *nature* **449**, 463 (2007).
2. Velt, A. *et al.* An improved reference of the grapevine genome reasserts the origin of the PN40024 highly homozygous genotype. *G3 Genes|Genomes|Genetics* jkad067 (2023) doi:10.1093/g3journal/jkad067.
3. Canaguier, A. *et al.* A new version of the grapevine reference genome assembly (12X.v2) and of its annotation (VCost.v3). *Genomics Data* **14**, 56–62 (2017).
4. Shi, X. *et al.* The complete reference genome for grapevine (Vitis vinifera L.) genetics and breeding. *Horticulture Research* uhad061 (2023) doi:10.1093/hr/uhad061.
5. Da Silva, C. *et al.* 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 (2013).
6. Zhou, Y. *et al.* The population genetics of structural variants in grapevine domestication. *Nat. Plants* **5**, 965–979 (2019).
7. Velasco, R. *et al.* A High Quality Draft Consensus Sequence of the Genome of a Heterozygous Grapevine Variety. *PLoS ONE* **2**, e1326 (2007).
8. Lijavetzky, D., Cabezas, J., Ibáñez, A., Rodríguez, V. & Martínez-Zapater, J. M. High throughput SNP discovery and genotyping in grapevine (Vitis vinifera L.) by combining a re-sequencing approach and SNPlex technology. *BMC Genomics* **8**, 424 (2007).
9. This, P., Lacombe, T. & Thomas, M. Historical origins and genetic diversity of wine grapes. *Trends in Genetics* **22**, 511–519 (2006).
10. Maestri, S. *et al.* 'Nebbiolo' genome assembly allows surveying the occurrence and functional implications of genomic structural variations in grapevines (Vitis vinifera L.). *BMC Genomics* **23**, 159 (2022).
11. Carbonell-Bejerano, P. *et al.* Catastrophic Unbalanced Genome Rearrangements Cause Somatic Loss of Berry Color in Grapevine. *Plant Physiol.* **175**, 786–801 (2017).
12. Chin, C.-S. *et al.* Phased diploid genome assembly with single-molecule real-time sequencing. *Nature Methods* **13**, 1050–1054 (2016).
13. Roach, M. J. *et al.* Population sequencing reveals clonal diversity and ancestral inbreeding in the grapevine cultivar Chardonnay. *PLoS Genet* **14**, e1007807 (2018).
14. Vondras, A. M. *et al.* The genomic diversification of grapevine clones. *BMC Genomics* **20**, 972 (2019).
15. Massonnet, M. *et al.* The genetic basis of sex determination in grapes. *Nat Commun* **11**, 2902 (2020).
16. Koren, S. *et al.* De novo assembly of haplotype-resolved genomes with trio binning. *Nat Biotechnol* **36**, 1174–1182 (2018).
17. Minio, A., Cochetel, N., Vondras, A. M., Massonnet, M. & Cantu, D. Assembly of complete diploid-phased chromosomes from draft genome sequences. *G3 (Bethesda)* **12**, jkac143 (2022).
18. Cheng, H., Concepcion, G. T., Feng, X., Zhang, H. & Li, H. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nat Methods* **18**, 170–175 (2021).
19. Onetto, C. A., Ward, C. M. & Borneman, A. R. The phased diploid genome assembly of Vitis vinifera cv. Shiraz. 2022.12.12.520163 Preprint at https://doi.org/10.1101/2022.12.12.520163 (2022).
20. Sichel, V. *et al.* Chimeras in Merlot grapevine revealed by phased assembly. (2022).

21. Carbonell-Bejerano, P., Royo, C., Mauri, N., Ibáñez, J. & Miguel Martínez Zapater, J. Somatic Variation and Cultivar Innovation in Grapevine. in *Advances in Grape and Wine Biotechnology* (eds. Morata, A. & Loira, I.) (IntechOpen, 2019). doi:10.5772/intechopen.86443.
22. Arrizabalaga, M. *et al.* Tempranillo clones differ in the response of berry sugar and anthocyanin accumulation to elevated temperature. *Plant Science* **267**, 74–83 (2018).
23. Grimplet, J., Ibáñez, S., Baroja, E., Tello, J. & Ibáñez, J. Phenotypic, Hormonal, and Genomic Variation Among Vitis vinifera Clones With Different Cluster Compactness and Reproductive Performance. *Front. Plant Sci.* **9**, 1917 (2019).
24. Wolkovich, E. M., García de Cortázar-Atauri, I., Morales-Castilla, I., Nicholas, K. A. & Lacombe, T. From Pinot to Xinomavro in the world's future wine-growing regions. *Nature Clim Change* **8**, 29–37 (2018).
25. Boursiquot, J.-M. *et al.* Parentage of Merlot and related winegrape cultivars of southwestern France: discovery of the missing link. *Australian Journal of Grape and Wine Research* **15**, 144–155 (2009).
26. Lacoste, P. El Malbec de Francia: la Denominación de Origen Controlada 'Cahors'. Historia y perspectivas. *Idesia* **33**, 113–124 (2015).
27. OIV. Distribution of the world's grapevine varieties. (2017).
28. Instituto Nacional del Vino (Argentina). Informe variedad Malbec. (2019).
29. Urvieta, R., Jones, G., Buscema, F., Bottini, R. & Fontana, A. Terroir and vintage discrimination of Malbec wines based on phenolic composition across multiple sites in Mendoza, Argentina. *Scientific Reports* **11**, 2863 (2021).
30. Muñoz, C. *et al.* Relationships among Gene Expression and Anthocyanin Composition of Malbec Grapevine Clones. *J. Agric. Food Chem.* **62**, 6716–6725 (2014).
31. van Houten, S. *et al.* Natural Genetic Variation for Grapevine Phenology as a Tool for Climate Change Adaptation. *Applied Sciences* **10**, 5573 (2020).
32. Varela, A. *et al.* Vineyard environments influence Malbec grapevine phenotypic traits and DNA methylation patterns in a clone-dependent way. *Plant Cell Rep* **40**, 111–125 (2021).
33. Calderón, L. *et al.* Whole genome resequencing and custom genotyping unveil clonal lineages in 'Malbec' grapevines ( Vitis vinifera L.). *Scientific Reports* **11**, 7775 (2021).
34. Goel, M., Sun, H., Jiao, W.-B. & Schneeberger, K. SyRI: finding genomic rearrangements and local sequence differences from whole-genome assemblies. *Genome Biol* **20**, 277 (2019).
35. Waters, E. R. & Vierling, E. Plant small heat shock proteins – evolutionary and functional diversity. *New Phytologist* **227**, 24–37 (2020).
36. Shu, K., Zhou, W., Chen, F., Luo, X. & Yang, W. Abscisic Acid and Gibberellins Antagonistically Mediate Plant Development and Abiotic Stress Responses. *Frontiers in Plant Science* **9**, (2018).
37. Kobayashi, S. Retrotransposon-Induced Mutations in Grape Skin Color. *Science* **304**, 982–982 (2004).
38. Wong, D. C. J., Lopez Gutierrez, R., Gambetta, G. A. & Castellarin, S. D. Genome-wide analysis of cis-regulatory element structure and discovery of motif-driven gene co-expression networks in grapevine. *DNA Res* **24**, 311–326 (2017).
39. Minio, A., Massonnet, M., Figueroa-Balderas, R., Castro, A. & Cantu, D. Diploid Genome Assembly of the Wine Grape Carménère. *G3* **9**, 1331–1337 (2019).
40. Raimondi, S. *et al.* DNA-based genealogy reconstruction of Nebbiolo, Barbera and other ancient grapevine cultivars from northwestern Italy. *Sci Rep* **10**, 15782 (2020).
41. Minio, A. *et al.* Iso-Seq Allows Genome-Independent Transcriptome Profiling of Grape Berry Development. *G3 Genes|Genomes|Genetics* **9**, 755–767 (2019).
42. Chialva, C. *et al.* Differential Expression Patterns Within the Grapevine Stilbene Synthase Gene Family Revealed Through Their Regulatory Regions. *Plant Mol Biol Rep* **36**, 225–238 (2018).
43. Pilati, S. *et al.* Abscisic Acid Is a Major Regulator of Grape Berry Ripening Onset: New Insights into ABA Signaling Network. *Front Plant Sci* **8**, 1093 (2017).

838 44. Tischer, S. V. *et al.* Combinatorial interaction network of abscisic acid receptors and
839 coreceptors from Arabidopsis thaliana. *Proceedings of the National Academy of Sciences* **114**,
840 10280–10285 (2017).
841 45. Valdés, A. E., Övernäs, E., Johansson, H., Rada-Iglesias, A. & Engström, P. The
842 homeodomain-leucine zipper (HD-Zip) class I transcription factors ATHB7 and ATHB12 modulate
843 abscisic acid signalling by regulating protein phosphatase 2C and abscisic acid receptor gene
844 activities. *Plant Mol Biol* **80**, 405–418 (2012).
845 46. Koyama, K., Sadamatsu, K. & Goto-Yamamoto, N. Abscisic acid stimulated ripening and
846 gene expression in berry skins of the Cabernet Sauvignon grape. *Funct Integr Genomics* **10**, 367–
847 381 (2010).
848 47. Walker, A. R. *et al.* White grapes arose through the mutation of two similar and adjacent
849 regulatory genes: White grape genes. *The Plant Journal* **49**, 772–785 (2007).
850 48. Gao, Z. *et al.* Characterization of the ABA Receptor VlPYL1 That Regulates Anthocyanin
851 Accumulation in Grape Berry Skin. *Front Plant Sci* **9**, 592 (2018).
852 49. Falginella, L. *et al.* Differential regulation of triterpene biosynthesis induced by an early
853 failure in cuticle formation in apple. *Hortic Res* **8**, 1–15 (2021).
854 50. Workman, R. *et al.* High molecular weight DNA extraction from recalcitrant plant species
855 for third generation sequencing. (2018).
856 51. Rabanal, F. A. *et al.* Pushing the limits of HiFi assemblies reveals centromere diversity
857 between two Arabidopsis thaliana genomes. *Nucleic Acids Research* **50**, 12309–12327 (2022).
858 52. Koren, S. *et al.* Canu: scalable and accurate long-read assembly via adaptive k-mer
859 weighting and repeat separation. *Genome Res.* **27**, 722–736 (2017).
860 53. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads.
861 *EMBnet.journal* **17**, 10–12 (2011).
862 54. Laetsch, D. R. & Blaxter, M. L. BlobTools: Interrogation of genome assemblies. Preprint at
863 https://doi.org/10.12688/f1000research.12232.1 (2017).
864 55. Buchfink, B., Reuter, K. & Drost, H.-G. Sensitive protein alignments at tree-of-life scale
865 using DIAMOND. *Nat Methods* **18**, 366–368 (2021).
866 56. Chin, C.-S. *et al.* Nonhybrid, finished microbial genome assemblies from long-read SMRT
867 sequencing data. *Nat Methods* **10**, 563–569 (2013).
868 57. Rhie, A., Walenz, B. P., Koren, S. & Phillippy, A. M. Merqury: reference-free quality,
869 completeness, and phasing assessment for genome assemblies. *Genome Biol* **21**, 245 (2020).
870 58. Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–
871 3100 (2018).
872 59. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–
873 2079 (2009).
874 60. Guan, D. *et al.* Identifying and removing haplotypic duplication in primary genome
875 assemblies. *Bioinformatics* **36**, 2896–2898 (2020).
876 61. Alonge, M. *et al.* RaGOO: fast and accurate reference-guided scaffolding of draft genomes.
877 *Genome Biol* **20**, 224 (2019).
878 62. Walker, B. J. *et al.* Pilon: An Integrated Tool for Comprehensive Microbial Variant
879 Detection and Genome Assembly Improvement. *PLOS ONE* **9**, e112963 (2014).
880 63. Seppey, M., Manni, M. & Zdobnov, E. M. BUSCO: Assessing Genome Assembly and
881 Annotation Completeness. in *Gene Prediction: Methods and Protocols* (ed. Kollmar, M.) 227–245
882 (Springer, 2019). doi:10.1007/978-1-4939-9173-0_14.
883 64. Alonge, M. *et al.* Automated assembly scaffolding using RagTag elevates a new tomato
884 system for high-throughput genome editing. *Genome Biology* **23**, 258 (2022).
885 65. Shen, W., Le, S., Li, Y. & Hu, F. SeqKit: A Cross-Platform and Ultrafast Toolkit for
886 FASTA/Q File Manipulation. *PLoS ONE* **11**, e0163962 (2016).
887 66. Marçais, G. *et al.* MUMmer4: A fast and versatile genome alignment system. *PLOS*
888 *Computational Biology* **14**, e1005944 (2018).
889 67. Ou, S. *et al.* Benchmarking transposable element annotation methods for creation of a

890 streamlined, comprehensive pipeline. *Genome Biology* **20**, 275 (2019).
891 68. Campbell, M. S. *et al.* MAKER-P: A Tool Kit for the Rapid Creation, Management, and
892 Quality Control of Plant Genome Annotations. *Plant Physiology* **164**, 513–524 (2014).
893 69. Haas, B. J. *et al.* Automated eukaryotic gene structure annotation using EVidenceModeler
894 and the Program to Assemble Spliced Alignments. *Genome Biol* **9**, R7 (2008).
895 70. Reid, K. E., Olsson, N., Schlosser, J., Peng, F. & Lund, S. T. An optimized grapevine RNA
896 isolation procedure and statistical determination of reference genes for real-time RT-PCR during
897 berry development. *BMC Plant Biol* **6**, 27 (2006).
898 71. Andrews, S. FastQC a quality-control tool for high-throughput sequence data http://www.
899 *Bioinformaticsbabraham. ac. uk/projects/fastqc* (2014).
900 72. Kim, D., Paggi, J. M., Park, C., Bennett, C. & Salzberg, S. L. Graph-based genome
901 alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol* **37**, 907–915 (2019).
902 73. Pertea, M. *et al.* StringTie enables improved reconstruction of a transcriptome from RNA-
903 seq reads. *Nat Biotechnol* **33**, 290–295 (2015).
904 74. Stanke, M., Tzvetkova, A. & Morgenstern, B. AUGUSTUS at EGASP: using EST, protein
905 and genomic alignments for improved gene prediction in the human genome. *Genome Biology* 8
906 (2006).
907 75. Brůna, T., Hoff, K. J., Lomsadze, A., Stanke, M. & Borodovsky, M. BRAKER2: automatic
908 eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein
909 database. *NAR Genomics and Bioinformatics* **3**, lqaa108 (2021).
910 76. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic
911 features. *Bioinformatics* **26**, 841–842 (2010).
912 77. Shumate, A. & Salzberg, S. L. Liftoff: accurate mapping of gene annotations.
913 *Bioinformatics* **37**, 1639–1643 (2021).
914 78. Haas, B. J., Zeng, Q., Pearson, M. D., Cuomo, C. A. & Wortman, J. R. Approaches to
915 Fungal Genome Annotation. *Mycology* **2**, 118–141 (2011).
916 79. Kent, W. J. BLAT—The BLAST-Like Alignment Tool. *Genome Res* **12**, 656–664 (2002).
917 80. Wu, T. D. & Watanabe, C. K. GMAP: a genomic mapping and alignment program for
918 mRNA and EST sequences. *Bioinformatics* **21**, 1859–1875 (2005).
919 81. Cochetel, N. *et al.* Diploid chromosome-scale assembly of the Muscadinia rotundifolia
920 genome supports chromosome fusion and disease resistance gene expansion during Vitis and
921 Muscadinia divergence. *G3 Genes|Genomes|Genetics* **11**, jkab033 (2021).
922 82. Dainat, J., Hereñú, D., LucileSol & pascal-git. NBISweden/AGAT: AGAT-v0.8.1. (2022)
923 doi:10.5281/zenodo.5834795.
924 83. Pertea, G. & Pertea, M. GFF Utilities: GffRead and GffCompare. *F1000Res* **9**, ISCB Comm
925 J-304 (2020).
926 84. Waterhouse, R. M. *et al.* BUSCO Applications from Quality Assessments to Gene
927 Prediction and Phylogenomics. *Mol Biol Evol* **35**, 543–548 (2018).
928 85. Manni, M., Berkeley, M. R., Seppey, M., Simão, F. A. & Zdobnov, E. M. BUSCO Update:
929 Novel and Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for
930 Scoring of Eukaryotic, Prokaryotic, and Viral Genomes. *Molecular Biology and Evolution* **38**,
931 4647–4654 (2021).
932 86. Törönen, P. & Holm, L. PANNZER—A practical tool for protein function prediction.
933 *Protein Science* **31**, 118–128 (2022).
934 87. Emms, D. M. & Kelly, S. OrthoFinder: phylogenetic orthology inference for comparative
935 genomics. *Genome Biol* **20**, 238 (2019).
936 88. Lovell, J. T. *et al.* GENESPACE: syntenic pan-genome annotations for eukaryotes.
937 2022.03.09.483468 Preprint at https://doi.org/10.1101/2022.03.09.483468 (2022).
938 89. Di Gaspero, G. & Foria, S. 2 - Molecular grapevine breeding techniques. in *Grapevine*
939 *Breeding Programs for the Wine Industry* (ed. Reynolds, A.) 23–37 (Woodhead Publishing, 2015).
940 doi:10.1016/B978-1-78242-075-0.00002-8.
941 90. Camacho, C. *et al.* BLAST+: architecture and applications. *BMC Bioinformatics* **10**, 421

942 (2009).
943 91. Massonnet, M. *et al.* Ripening Transcriptomic Program in Red and White Grapevine
944 Varieties Correlates with Berry Skin Anthocyanin Accumulation. *Plant Physiol.* **174**, 2376–2396
945 (2017).
946 92. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina
947 sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
948 93. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**,
949 357–359 (2012).
950 94. Chan, P. P. & Lowe, T. M. GtRNAdb: a database of transfer RNA genes detected in
951 genomic sequence. *Nucleic Acids Res* **37**, D93–D97 (2009).
952 95. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21
953 (2013).
954 96. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for
955 assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).
956 97. Chialva, C., Blein, T., Crespi, M. & Lijavetzky, D. Insights into long non-coding RNA
957 regulation of anthocyanin carrot root pigmentation. *Scientific Reports* **11**, 4093 (2021).
958 98. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for
959 RNA-seq data with DESeq2. *Genome Biology* **15**, 550 (2014).
960 99. Raudvere, U. *et al.* g:Profiler: a web server for functional enrichment analysis and
961 conversions of gene lists (2019 update). *Nucleic Acids Research* **47**, W191–W198 (2019).
962 100. Marini, F., Ludt, A., Linke, J. & Strauch, K. GeneTonic: an R/Bioconductor package for
963 streamlining the interpretation of RNA-seq data. *BMC Bioinformatics* **22**, 610 (2021).
964 101. Bardou, P., Mariette, J., Escudié, F., Djemiel, C. & Klopp, C. jvenn: an interactive Venn
965 diagram viewer. *BMC Bioinformatics* **15**, 293 (2014).
966

967

968 **Table 1**. A) Summary statistics obtained for each haplophase after deduplication pipeline for
969 contigs and scaffolds. BUSCO percentages are expressed over a total of 2121 genes. (B) Evaluation
970 of the phasing process, performed by comparing *k-mers* between the Malbec haplophase
971 assemblies and raw short-reads from the parental cultivars.
972

973 **A**

| | Summary stats | Malbec-Pru | Malbec-Mag |
|---|---|---|---|
| Deduplicated contigs | Total length (Mb) | 479.36 | 479.06 |
| | Number | 163 | 179 |
| | Longest (Mb) | 18.3 | 17.6 |
| | Mean length (Mb) | 2.94 | 2.67 |
| | N50 (Mb) | 7.69 | 6.64 |
| BUSCO genes (%) | Completeness | 96.5 | 95.9 |
| | Duplicated | 2.3 | 1.8 |
| | Fragmented | 1.3 | 1.7 |
| | Missing | 2.2 | 2.4 |
| Scaffolds | Total length (Mb) | 479.35 | 479.05 |
| | Number | 47 | 52 |
| | Longest (Mb) | 38.8 | 38.8 |
| | Mean length (Mb) | 10.19 | 9.21 |
| | N50 (Mb) | 24.42 | 24.48 |
| | N counts | 115 | 127 |
| | Gaps | 11,500 | 12,700 |
| BUSCO genes (%) | Completeness | 96.1 | 95.8 |
| | Duplicated | 1.9 | 1.8 |
| | Fragmented | 1.2 | 1.6 |
| | Missing | 2.7 | 2.6 |

974 **B**

| Haplophase | Blocks Nº | Total bases In blocks | Block N50 size (bp) | Nº of switched makers | Total nº of markers | Switch error rate |
|---|---|---|---|---|---|---|
| Malbec-Pru | 198 | 463,853,342 | 11,965,117 | 5,589 | 22,596,691 | 0.025% |
| Malbec-Mag | 198 | 456,349,354 | 8,939,504 | 3,842 | 20,345,345 | 0.019% |

975
976

977 **Table 2.** Coding protein sequences predicted for Malbec haplophases (considering isoforms),
978 compared to the PN40024.v4.2 (REF) annotation. For the three assemblies the average number of
979 mRNAs per gene was 1.2.
980

| | Malbec-Pru | | Malbec-Mag | | PN40024.v4.2 (REF) | |
|---|---|---|---|---|---|---|
| | Total number | Mean length (bp) | Total number | Mean length (bp) | Total number | Mean length (bp) |
| Protein coding genes | 35,739 | 5,086 | 35,463 | 5,071 | 35,197 | 4,742 |
| Transcripts | 44,590 | 5,025 | 44,183 | 5,017 | 41,160 | 4,689 |
| Exons | 236,022 | 319 | 233,585 | 320 | 208,719 | 283 |
| CDS | 224,357 | 1,182 | 221,965 | 1,180 | 200,059 | 1,126 |
| 5' UTRs | 21,386 | 371 | 21,221 | 372 | 17,478 | 280 |
| 3' UTRs | 22,138 | 670 | 21,818 | 676 | 18,344 | 440 |

981
982
983 **Table 3.** (A) Summary statistics of the total predicted proteins. (B) BUSCO proteins retrieved for
984 each Malbec haplophase and for the diploid assembly. In both tables results are compared to the
985 PN40024.v4.2 (REF) annotation.
986
987 **A**

| Predicted proteins | Malbec-Pru | Malbec-Mag | PN40024.v4.2 (REF) |
|---|---|---|---|
| Total proteins | 44,590 | 44,183 | 41,160 |
| Min. length (aa) | 37 | 41 | 50 |
| Average length (aa) | 393.2 | 392.5 | 375.3 |
| Max. length (aa) | 5,480 | 5,488 | 5,481 |
| N50 (aa) | 532 | 534 | 519 |

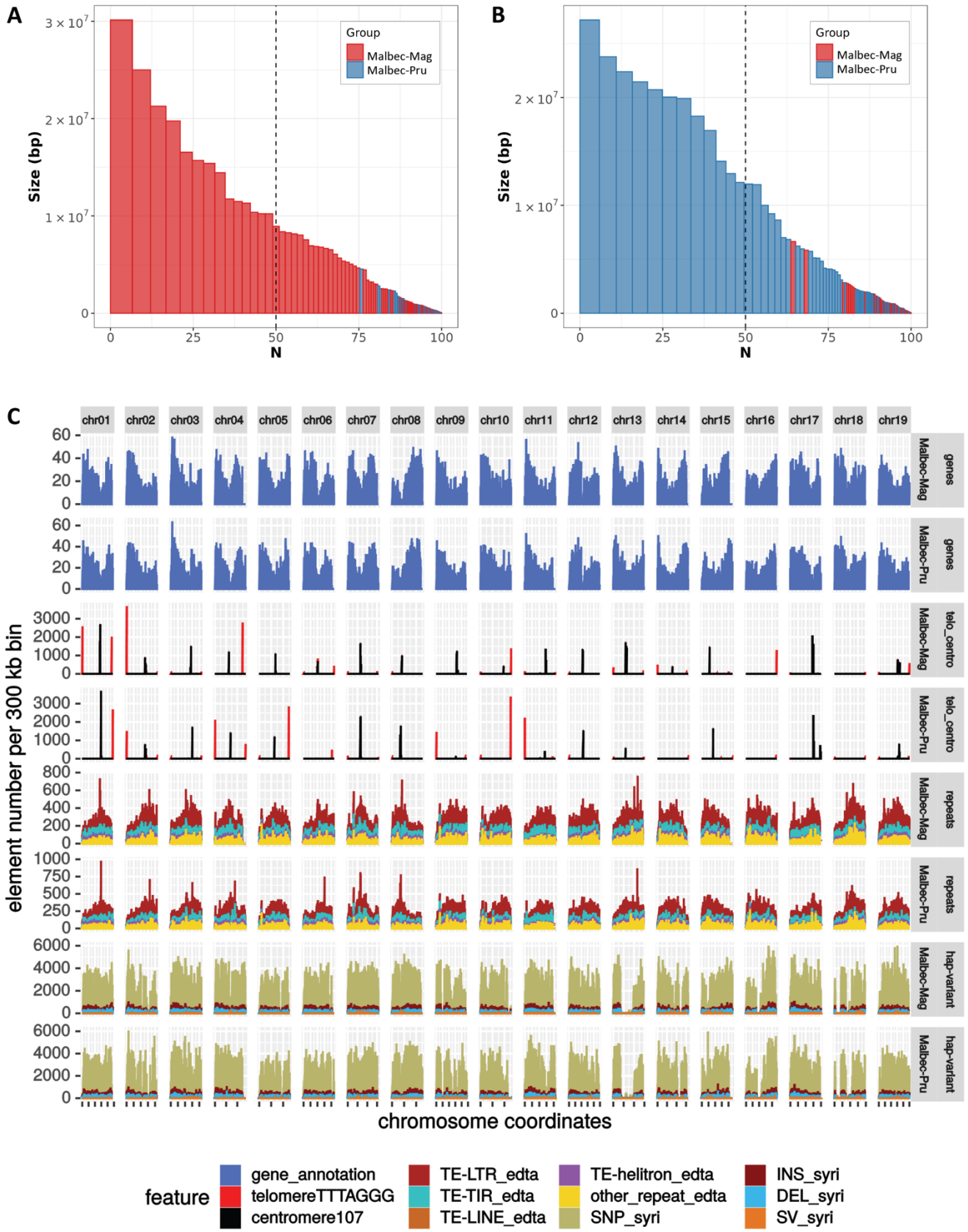
988
989 **B**

| BUSCO proteins (%) | Malbec-Pru | Malbec-Mag | Malbec diploid | PN40024.v4.2 (REF) |
|---|---|---|---|---|
| Completeness | 97.3 | 97.2 | 98.5 | 99.2 |
| Fragmented | 1.3 | 1.5 | 0.7 | 0.25 |
| Missing | 1.4 | 1.4 | 0.8 | 0.34 |

990
991
992
993

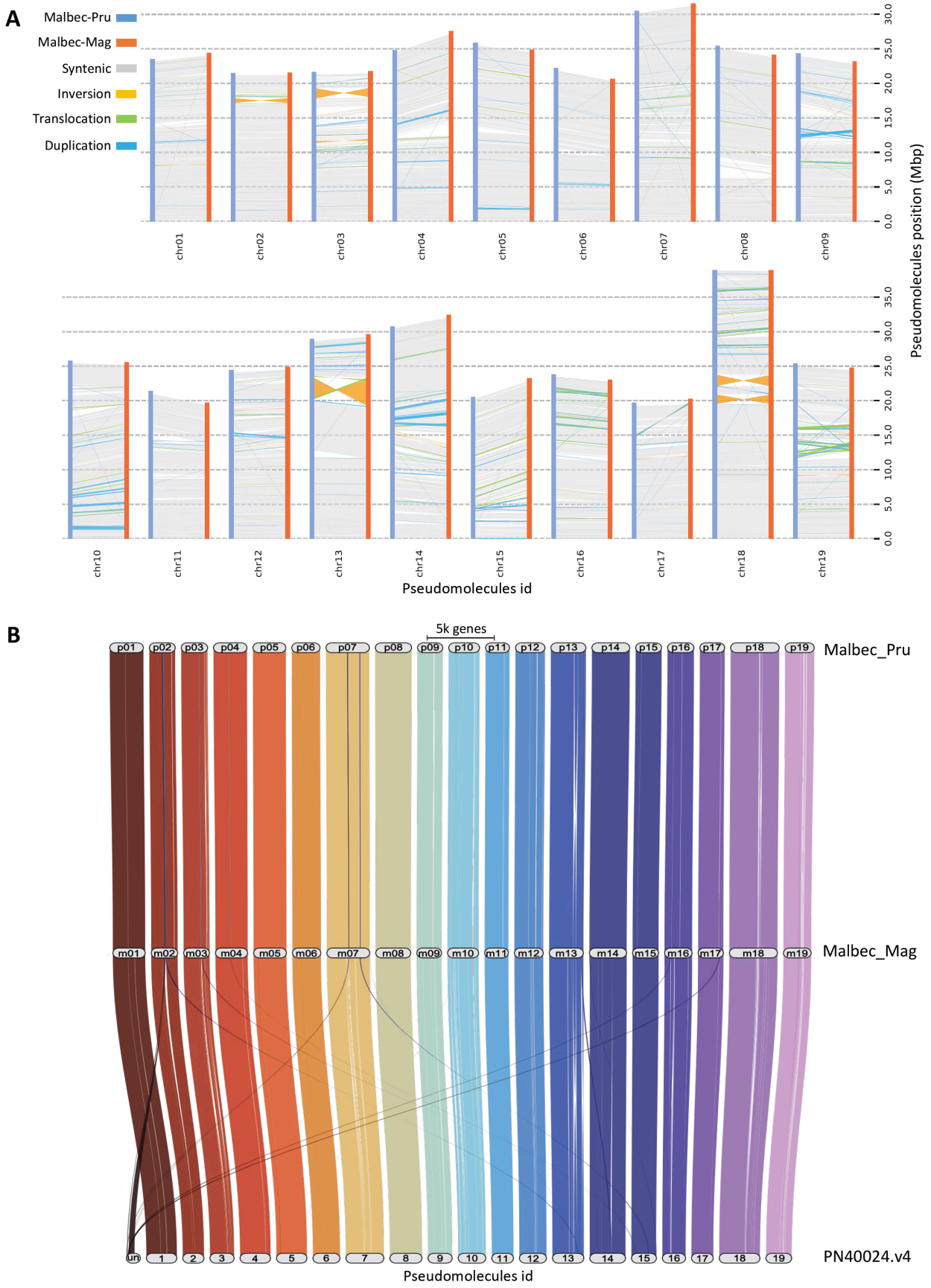
994 **Figure 1**

995




996

997 **Figure 2**

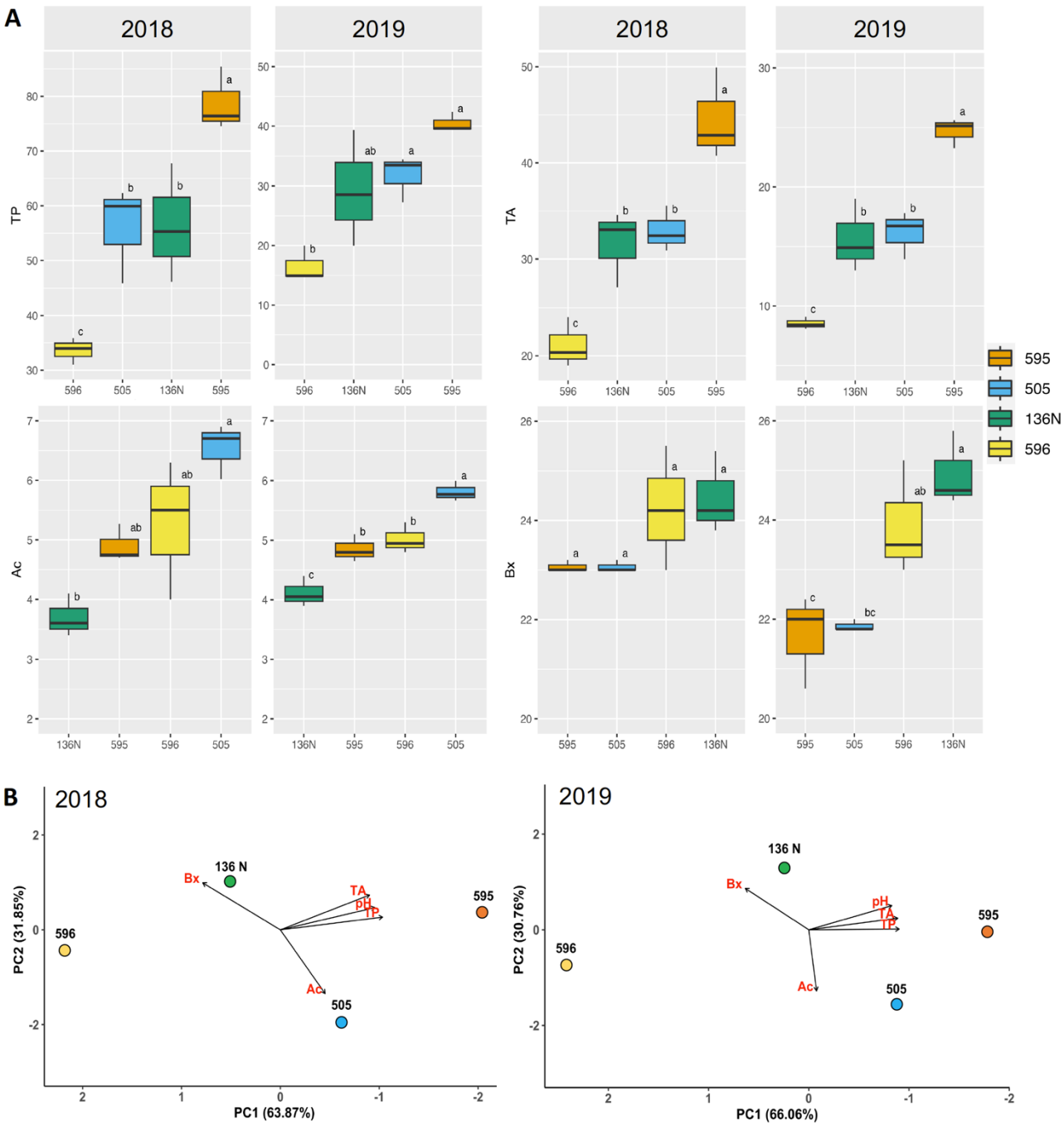



998

999

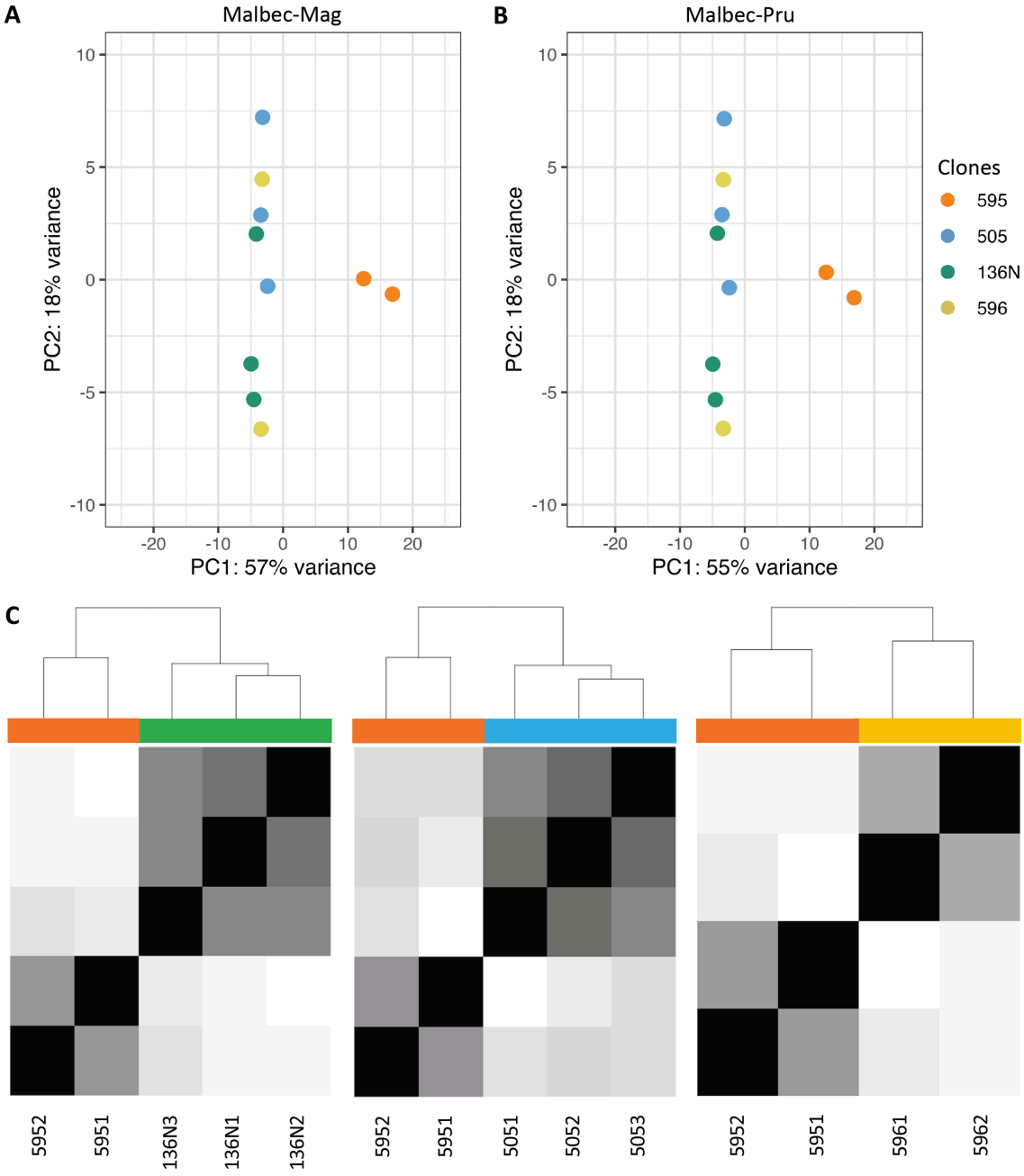
1000

1001 **Figure 3**



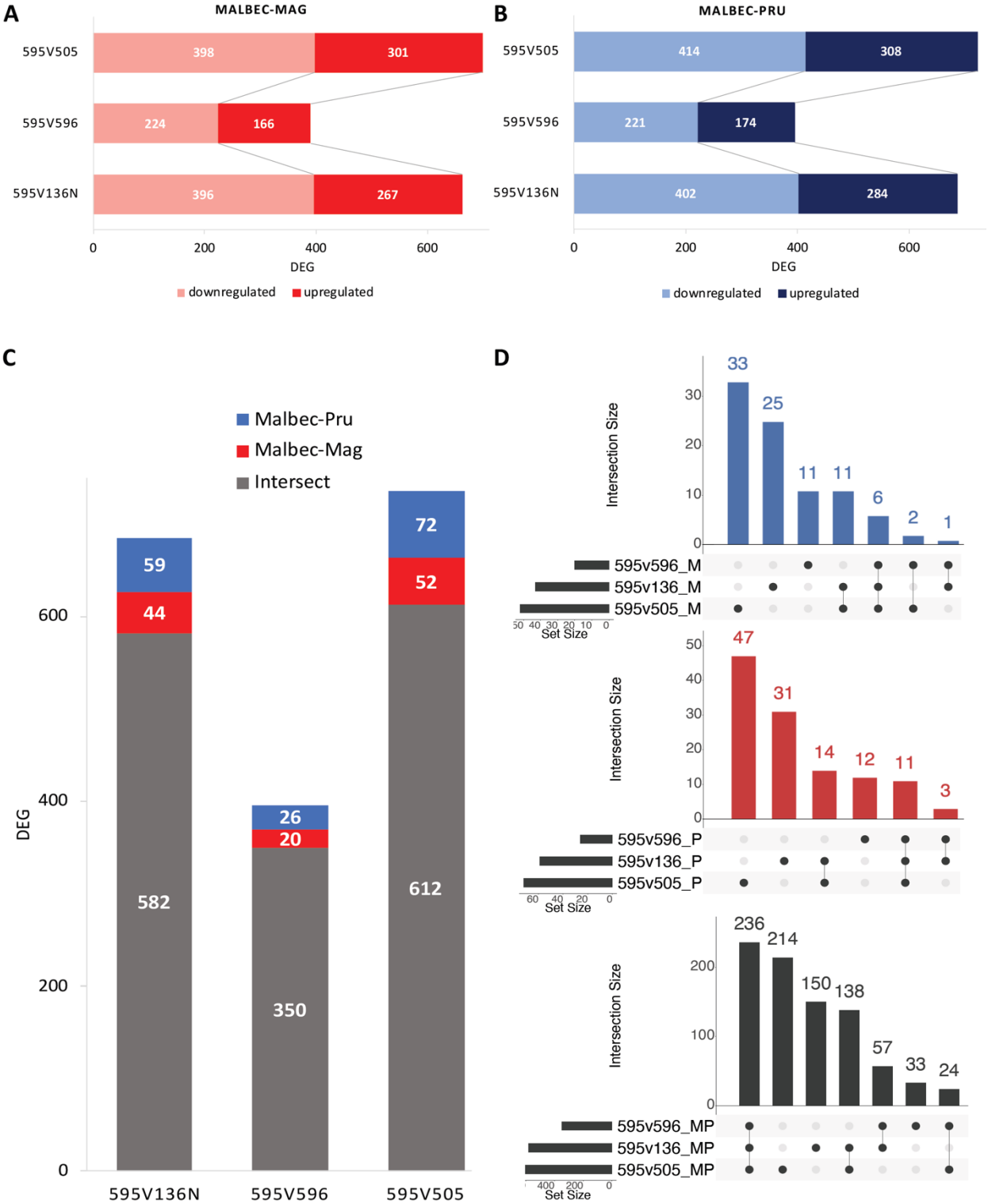

1002

1003 **Figure 4**




1004
1005

1006 **Figure 5**

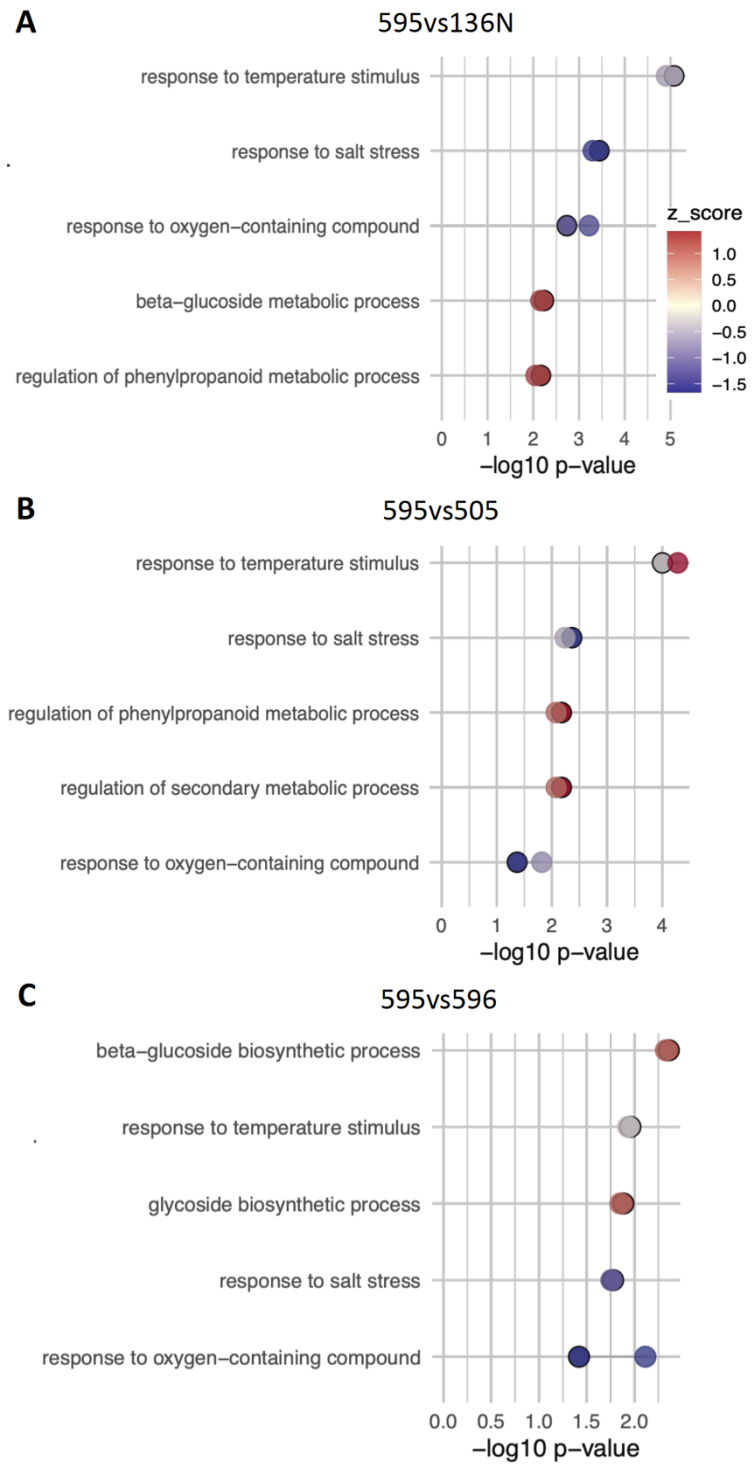



1007

1008 **Figure 6**

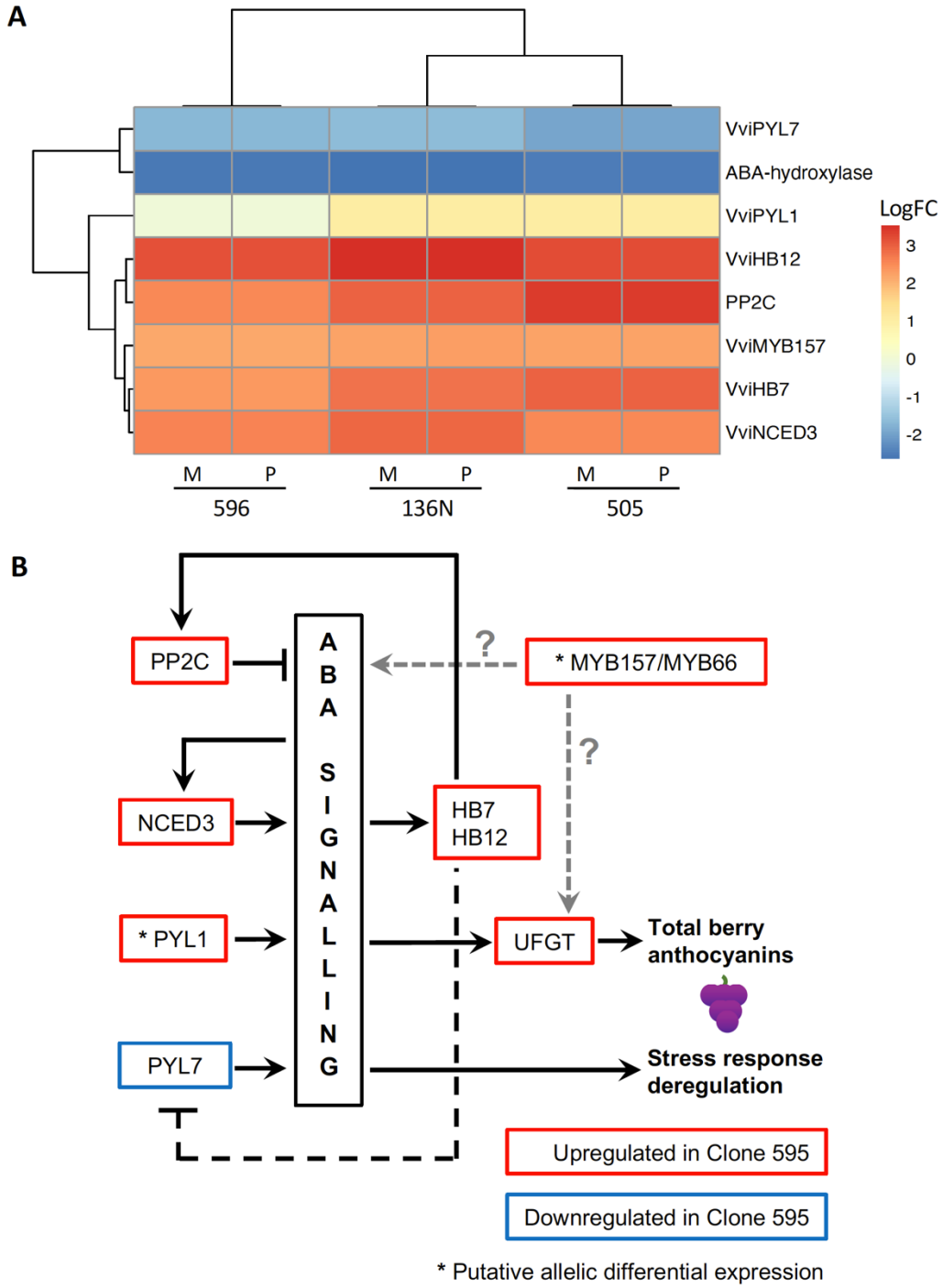
1009

1010



1011
1012

**Figure 7**



1013 **13. Figures legends**

1014 **Figure 1.** (A) Barplots of the phased blocks sorted by size to visualize the efficiency of the phasing process
1015 for Malbec-Pru and Malbec-Mag haplophases. For the former the blue bars are the own blocks and the
1016 red bars the switched blocks, for the latter the color code is inverted. Vertical bars represent phase blocks,
1017 defined as a continuous set of *k-mers* inherited from the same haplotype (i.e., *hap-mers*). The *X*-axis is the
1018 percentage of the genome size covered by blocks of this size of larger, expressed in base pairs (*Y*-axis). (B)
1019 Distribution of the annotated features along the 19 scaffolded pseudomolecules for Malbec haplophases,
1020 including: gene density, centromeres, telomeres, transposable elements (LTRs, TIR, LINE and Helitron) and
1021 haplotype variants (SNPs, INDELs, SVs).

1022

1023 **Figure 2.** Synteny analysis based on the comparison of the 19 pseudomolecules assembled for Malbec
1024 haplophases. (A) Structural variations (SVs) distribution along pseudomolecules, with Malbec-Pru (blue)
1025 considered as the reference and Malbec-Mag (orange) as the query. Results are based on the percentages
1026 estimated for the query, 76.5% (362.14 Mb) of the assembly resulted syntenic and 9.97% (47.2 Mb) was
1027 affected by different types of SVs (inversions, translocations and duplications). (B) Riparian plots
1028 representing the synteny of orthologous genes along the 19 scaffolded pseudomolecules for each haploid
1029 complement of Malbec. A strong synteny was observed among the assembled pseudomolecules of
1030 Malbec haplophases and the grapevine reference genome (PN40024.v4). Also, genes contained in the
1031 "Unknown" (Un) chromosome of the PN40024.v4 exhibited orthologous genes assigned to different
1032 pseudomolecules in Malbec haplophases.

1033

1034 **Figure 3.** Phenotypic analyses conducted on the four clones selected to perform RNAseq experiments. All
1035 phenotypic data was obtained from mature berries during two consecutive years, 2018 and 2019.
1036 Variables analyzed were total acidity (Ac), Brix (Bx), total anthocyanins (TA), total polyphenols (TP). (A)
1037 Box-plots with median values and standard error bars. ANOVA results based on three biological replicates,
1038 with the different letters representing significant differences, Tukey HSD ($p$<0.05). (B) PCAs summarizing
1039 the phenotypic differences of the four selected clones for each analyzed year the greatest differences are
1040 driven by pH, TP and TA represented by PC1.

1041

1042 **Figure 4.** Principal components analysis based on global transcriptomic data based on total aligned
1043 transcripts to (A) Malbec-Mag and (B) Malbec-Pru. Both analyses showed consistent results, with the
1044 biological replicates of clone 595 differentiating from the rest based on PC1. The biological replicates of
1045 the other clones (136N, 505 and 506) differentiated less among each other based on PC2. (C) Pairwise

1046 comparison of sample-to-sample distance between 595 and the other three clones also showed clear
1047 differentiation. This analysis was performed with both haplophases and retrieved identical results;
1048 therefore, only Malbec-Mag results are shown.
1049
1050 **Figure 5.** Bar charts showing the total numbers of the statistically significant ($p$<0.01) differentially
1051 expressed genes (DEG). (A) and (B) show the total number of DEGs detected in each clonal pairwise
1052 comparisons with Malbec-Mag and Malbec-Pru, respectively. Inside each bar downregulated (LFC $\leq$ 1)
1053 and upregulated (LFC $\geq$ 1) genes are distinguished based on the light and dark color code. These are the
1054 genes that were employed downstream for GO terms enrichment analyses. (C) Intersection of the
1055 orthologous genes detected with both haplophases, for each clonal comparison. The grey portion of the
1056 bar represents the DEGs detected with both haplophases, the blue and red portions of the bar represent
1057 haplotype-specific DEGs detected with Malbec-Pru and Malbec-Mag, respectively. **d)** UpSet plots for the
1058 intersection among the three pairwise comparisons of the DEGs consistently detected with each
1059 haplophase, red bar-chart for Malbec-Mag, blue bar-chart for Malbec-Pru and black bar-chart for DEGs
1060 detected with both haplophases.
1061
1062 **Figure 6.** Dumbbell plots showing the three pairwise comparisons of the biological processes enrichment
1063 analysis: (A) 595 vs 136N, (B) 595 vs 505 and (C) 595vs 596. For each process two dots are shown,
1064 indicating that the same process has been detected with both haplophases. Lighter dots represent the
1065 process detected with Malbec-Pru and darker dots with hard outline Malbec-Mag, in some cases dots are
1066 overlapped (and seen as a single dot) because the process has been detected with very similar *p-value*.
1067 Dots are colored based on its z-score value, blue indicates downregulation of the process, red is for
1068 upregulation and grey indicate values close to zero. The significance of the detected processes is
1069 expressed as the -$Log_{10}$ p-value.
1070
1071 **Figure 7**. (A) Heatmap showing the change direction of DEGs with regulatory functions associated to the
1072 ABA signaling pathways. Log fold change of each gene is expressed in relative terms to 595. Results
1073 obtained with both haplophases (M = Malbec-Mag, P = Malbec-Pru) for the three pairwise comparisons
1074 are shown. (B) Proposed model based on the identified DEGs for the origin of the higher berry anthocyanin
1075 content in Malbec clone 595**.** Clonal variant phenotypes observed in 595 berries are indicated in bold.
1076 Genes upregulated in clone 595 compared to the other three are represented in red boxes, whereas the
1077 downregulated gene is in a blue box. Arrow and flat end lines indicate for positive and negative regulation.
1078 Relations that are suggested but not proven are indicated with dashed lines. Genes with possible allele-

1079 specific overexpression in Malbec 595 suggesting for *cis*-acting regulatory somatic mutations are labelled
1080 with asterisks as candidate genes triggering all other downstream responses.
1081