Molecular investigation of Tuscan sweet cherries sampled over three years: gene expression analysis coupled to metabolomics and proteomics

Roberto Berni[1,2], Sophie Charton[3], Sébastien Planchon[3], Sylvain Legay[4], Marco Romi[1], Claudio Cantini[5], Giampiero Cai[1], Jean-Francois Hausman[4], Jenny Renaut[3,*], Gea Guerriero[4,*]

[1] *Department of Life Sciences, University of Siena, via P.A. Mattioli 4, I-53100 Siena, Italy.*

[2] *TERRA Teaching and Research Center, Gembloux Agro-Bio Tech, University of Liège, 5030 Gembloux, Belgium.*

[3] *Environmental Research and Innovation Department, Luxembourg Institute of Science and Technology, 41, Rue du Brill, L-4422 Belvaux, Luxembourg.*

[4] *Environmental Research and Innovation Department, Luxembourg Institute of Science and Technology, 5, rue Bommel, L-4940, Hautcharage, Luxembourg.*

[5] *Istituto per la BioEconomia (IBE CNR), Dipartimento di Scienze BioAgroAlimentari, via Aurelia 49, 58022 Follonica (Italy).*

[*] E-mail: jenny.renaut@list.lu (Renaut J.), gea.guerriero@list.lu (Guerriero G.)

Abstract

Sweet cherry (*Prunus avium* L.) is a stone fruit widely consumed and appreciated for its organoleptic properties, as well as its nutraceutical potential. We here investigated the characteristics of six non-commercial Tuscan varieties of sweet cherry maintained at the Regional Germplasm Bank of the CNR-IBE in Follonica (Italy) and sampled ca. 60 days post anthesis over three consecutive years (2016-2017-2018). We adopted an approach merging genotyping and targeted gene expression profiling with metabolomics. To complement the data, a study of the soluble proteomes was also performed on two varieties showing the highest

content of flavonoids. Metabolomics identified the presence of flavanols and proanthocyanidins in highest abundance in the varieties Morellona and Crognola, while gene expression revealed that some differences were present in genes involved in the phenylpropanoid pathway during the three years and among the varieties. Finally, proteomics on Morellona and Crognola showed variations in proteins involved in stress response, primary metabolism and cell wall expansion. To the best of our knowledge, this is the first multi-pronged study focused on Tuscan sweet cherry varieties providing insights into the differential abundance of genes, proteins and metabolites.

Introduction

Prunus avium L. is a fruit-tree belonging to the genus *Prunus* within the Rosaceae family that produces stone fruits with a characteristic aroma and taste. This fruit-tree is native to many regions worldwide, with a preference for temperate climates like the Mediterranean area in Europe. It has a diploid genome of 16 chromosomes (2*n*=16) and, like other members of the Rosaceae family, sweet cherry contains toxic cyanogenic glycosides[1,2], which are present in low concentrations in the stone (0.8%)[2].

The fruits of *P. avium* are rich sources of health-promoting compounds[3,4] and have a moderate content of simple sugars (and therefore a low glycemic index), as well as organic acids. They are cholesterol-free, low in calories with a high content of water. These drupes are also rich sources of vitamins (notably vitamin C) and minerals (K, P, Ca, Mg).

Polyphenols and triterpenes are among the beneficial phytochemicals composing the rich palette of bioactives in sweet cherry fruits [3,5]. Triterpenes are present in the cuticle of the fruits and, more specifically, they are found almost exclusively associated with the intracuticular waxes[6,7].

Italy is an important producer of sweet cherries which account for an important portion of the agricultural production[8–10]; therefore, this fruit-tree plays a prominent role in the agricultural and economic landscape of Italy.

Among the different Italian regions, Tuscany is known for the high quality of its food products exported worldwide (wine, oil, cheese, meat) and for specific geographic areas within its territory that have obtained the Protected Geographical Indication (IGP) label. Such an example is Lari, where a specific variety of sweet cherry is cultivated[11], or the reference areas of Capalbio, Batignano, Campagnatico, Castiglione delle Pescaie, where the olive varieties Frantoio, Leccino, Moraiolo and Pendolino are grown[12].

Understanding more about the physiology and bioactive contents of non-commercial sweet cherry varieties of Italian collections can inspire exploitation programs valorizing these local fruits at the regional level. Such ancient local varieties have either disappeared or have been marginalized and reduced in number to a few trees, because of the introduction of new cherry varieties in crop systems[11]. Nevertheless, they constitute an important reservoir of interesting characters (e.g. morphological, organoleptic and genetic) which can contribute to the selection of new varieties through breeding programs[13].

We previously showed that six non-commercial varieties of Tuscan sweet cherries maintained at the Germplasm Bank of the CNR-IBE in Follonica (Grosseto, Italy) are high producers of pentacyclic triterpenes[5], as well as phenolics[3]. We here enrich these data by using genotyping and gene expression profiling of phenylpropanoid (hereafter abbreviated PPP) biosynthetic genes, as well as untargeted metabolomics on fruits sampled at maturity during three years (2016-2017-2018). We additionally investigate the soluble proteomes of two varieties, Crognola and Morellona, ranking as the highest producers of phenolic compounds. A commercial variety, Durone, commonly found in Italian fruit markets, was included in the study. The molecular data obtained by including this variety allow to have a comparison with

fruits found on the market. The goal of the study is to provide molecular information on the synthesis and content of phenolic compounds in the Tuscan sweet cherries and to compare the data with those obtained for a commercial counterpart.

The data pave the way to follow-up studies focused, for example, on earlier developmental stages, or on the post-harvest stability of the Tuscan fruits, which will provide an accurate evaluation of their further economic valorization.

Results and Discussion

Genotyping of six non-commercial Tuscan sweet cherries

As a first step towards the molecular characterization of the Tuscan sweet cherries, a similarity tree was generated (Figure 1). Commercial varieties originating from France, Turkey and Luxembourg were included to enrich the dataset and to better discriminate the phylogenetic relatedness of the Tuscan fruits. However, these commercial varieties were not included in the other analyses performed.

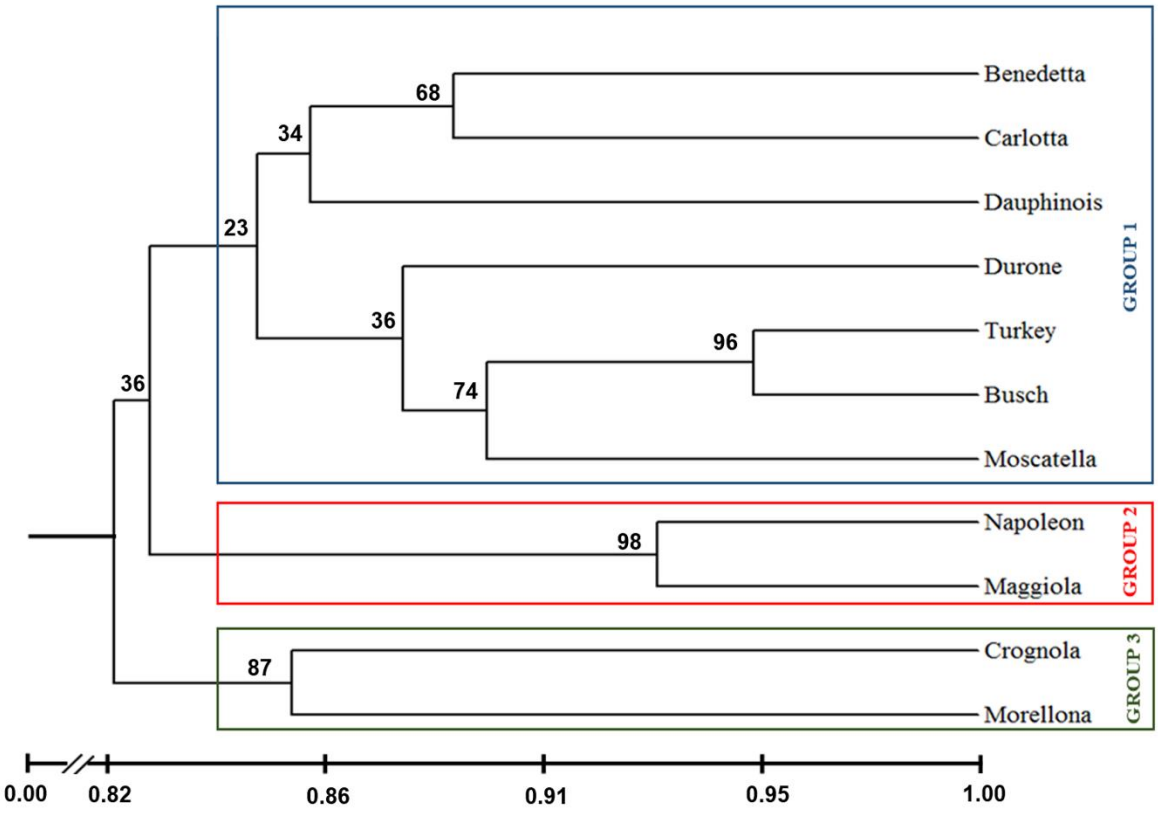


Figure 1. Dendrogram derived from the genotyping assay (UPGMA method) using SSR markers specific for genomic regions with a high coefficient of polymorphism. The relationships between the Tuscan cherries and the commercial ones from Luxembourg (Busch), Turkey, France (Napoleon and Dauphinois) and Italy (Durone) are shown. Nei & Li's similarity coefficients are displayed in the black bar below the tree. Bootstrap values are indicated above the branches (1000 replicates).

In a previous study, genotyping of Tuscan sweet cherries was used to investigate the self-incompatibility alleles (*S-alleles*) which are necessary to determine incompatibility relationships between cultivars and establish appropriate breeding programs[14]. The authors focused their attention on the tree breeding incompatibility by designing SSR markers specific for the *S* locus.

Different SSR markers were here used and the results showed 3 main genetic clusters belonging to 3 different branches of the tree. The biggest cluster included 3 ancient varieties (Benedetta, Carlotta, Moscatella) and 4 commercial ones (Durone from Italy, Dauphinois from France, commercial from Turkey and commercial from Luxembourg, referred to as Busch). The second branch comprised the ancient variety Maggiola and the French one Bigarreau Napoléon (referred to as Napoleon in Figure 1). The last cluster included the varieties Crognola and Morellona, which grouped separately from all the other Tuscan sweet cherries studied.

Despite the small number of samples here studied, genotyping was in agreement with the distribution of the varieties across the Tuscan territory: Benedetta and Carlotta share a wide distribution across the whole region, Crognola and Morellona are both from the province of Pisa, while Moscatella is the only representative of the geographical area around Siena and Maggiola of Roccalbegna (province of Grosseto).

Untargeted metabolomics

Untargeted metabolomics identified 15 differentially abundant metabolites in positive and 14 in negative mode (Table 1 and Table 2). This approach was adopted to confirm and enrich the data already present in the literature on Tuscan sweet cherries[3], by providing information on other families of molecules, namely flavanols, proanthocyanins and flavonolignans (cinchonain and deoxyhexosyl cinchonain; Table 1 and 2).

A hierarchical clustering of the heatmap was performed to identify similar patterns of abundance shared by the classes of molecules detected (Figure 2). It should be noted that the variety Benedetta only appears in 2016, as the trees did not give any fruits in the other years studied.

As previously reported in other studies focused on sweet cherries[15], the majority of the molecules detected in the Tuscan fruits belonged to the flavonoid class.

Flavonoids play different roles in plants, e.g. interaction with pollinators[16], photoprotection, reactive oxygen species (ROS) scavenging[17], response to abiotic stresses[18], as well as auxin transport[19]. Their biosynthesis is known to be affected by the genotype and the environment. A study on 27 strawberry genotypes grown in the North and South of Italy revealed a higher content of flavonols in the fruits from the northern location[20]. Additionally, autochthonous cultivars of sweet cherry from South Italy showed differences in flavonoids, thereby revealing that the genotype is responsible for statistically significant differences in the content of bioactive molecules[13].

In the Tuscan fruits examined, flavonoids varied in abundance, notably (epi)afzelechin-(epi)catechin and A/B-type flavanols. The A- and B-type flavanols observed could be fragments of bigger polymers, such as proanthocyanidins; however, the exact number of monomers is difficult to determine due to the limit of 2000 *m/z* of the MS1 scan. For these molecules, the most striking differences were observed in two varieties, i.e. Crognola and

Morellona, which ranked as the highest in abundance in all the years studied (Figure 2). This result is in agreement with the previously published data obtained using spectrophotometric assays and targeted metabolite quantification using HPLC-DAD[3,5,21]: Crognola and Morellona produced high amounts of pentacyclic triterpenes, as well as anthocyanins and flavonoids.

In contrast, the commercial cherries were among the fruits producing the lowest amounts of flavanols. Although the post-harvest storage conditions of the commercial cherries are not known and were supposedly different from those of the Tuscan fruits, they were included for comparative purposes and as representatives of the most common cherries on the Italian market.

Fewer differences among the varieties were found for the hydroxycinnamic acid coumaroyl quinic acid (that could, however, also represent a fragment of bigger molecules, such as quinic acid esters of hydroxycinnamic acids previously detected in depitted sweet cherries[22]) and for flavanones (di- and trihydroxyflavanones). These compounds were also less abundant as compared to A/B-type flavanols (Figure 2). It is worthy to note that untargeted metabolomics detected the presence of cinchonain and deoxyhexosyl cinchonain in the fruits of sweet cherries. These secondary metabolites are flavonolignans and have high antioxidant, hepatoprotective and antimicrobial activities[23,24]. Such molecules are rare in nature and found in species such as *Cinchona*, *Trichilia*, *Acer*, *Sorbus*[25]. However, a recent study detected cinchonain in Hungarian sour cherries[26], thereby confirming that this flavonolignan occurs in the fruits of members within the genus *Prunus.* From a nutraceutical point of view, Olszewska and colleagues showed an antiradical capacity of cinchonain (measured with the DPPH assay) up to four times higher than (+)-catechin[25], which is known as one of the most effective antioxidants both *in vitro* and *in vivo*[27]. Moreover, *in vitro* experiments showed an insulinotropic effect of cinchonain and it was thus proposed that the consumption of this molecule through the diet may be helpful for managing type 2 diabetes[28].

158 Crognola and Morellona produced cinchonain at higher levels (Figure 2), a finding confirming
159 the high nutraceutical potential of these two Tuscan varieties. The commercial variety Durone
160 showed among the lowest amounts of the flavonolignan.
161
162 **Table 1.** List of differentially abundant compounds in sweet cherries obtained by UHPLC-
163 DAD-HR-MS/MS in positive ESI mode. The details of the compounds are given, together with
164 the specification of the reliability class and references used for the detection; R_t, retention time.
165 MSI, Metabolomics Standards Initiative. *, only in harvests 2016 and 2017. All observed ions
166 are $[M+H]^+$.

Putative identification	$R_{t\ (min)}$	Formula	Theoretical *m/z*	Observed *m/z*	Mass error (ppm)	Main MS2 fragments	MSI reliability class with references used for the annotation
Flavanol hexoside	8.79	$C_{21}H_{24}O_{11}$	453.1391	453.1379	-2.78	139.0381	2 [29]
A-type flavanol dimer I	13.87	$C_{30}H_{24}O_{12}$	577.1341	577.1318	-3.91	245.0434	3 [29]
Coumaroyl quinic acid	16.00	$C_{16}H_{18}O_{8}$	339.1074	339.1065	-2.91	147.0433 – 119.0481 – 91.0534	3 [30–32]
(epi)afzelechin–(epi)catechin	17.65	$C_{30}H_{26}O_{11}$	563.1548	563.1534	-2.50	107.0480 – 147.0431 – 287.0544	2 [33,34]
A-type flavanol trimer I	19.02	$C_{45}H_{36}O_{18}$	865.1974	865.1968	-0.73	245.0441 – 287.0544 – 163.0375	3 [35]
B-type flavanol trimer I	19.02	$C_{45}H_{38}O_{18}$	867.2131	867.2131	0.00	245.0422 – 127.0379 – 163.0382	3 [29]
B-type flavanol tetramer	19.64	$C_{60}H_{50}O_{24}$	1155.2765	1155.2770	0.45	245.0440 - 247.0593 – 163.0375	3 [35]
A-type flavanol trimer II	19.64	$C_{45}H_{36}O_{18}$	865.1974	865.1962	-1.41	287.0539 – 247.0587 – 135.0428	3 [35]
A-type flavanol tetramer	19.95	$C_{60}H_{48}O_{24}$	1153.2608	1153.2609	0.06	135.0441 – 123.0416 – 163.0376	3 [35]
B-type flavanol trimer II *	19.98	$C_{45}H_{38}O_{18}$	867.2131	867.2117	-1.59	127.0374 – 163.0394 – 135.0409	3 [35]
Trihydroxyflavanone	20.79	$C_{15}H_{12}O_{5}$	273.0758	273.0751	-2.31	153.0167	2
B-type flavanol dimer	21.52	$C_{30}H_{26}O_{12}$	579.1497	579.1492	-0.90	123.0429 – 127.0378 – 163.0331	3 [35]
A-type flavanol dimer II	21.52	$C_{30}H_{24}O_{12}$	577.1341	577.1320	-3.56	123.0431 – 245.0442 – 135.0430	3 [31]
Cinchonain	24.52	$C_{24}H_{20}O_{9}$	453.1180	453.1168	-2.67	191.0333 – 163.0362	3 [33]
Dihydroxyflavanone hexoside	25.74	$C_{21}H_{22}O_{10}$	435.1286	435.1273	-2.85	169.0122 – 273.0757	2 [30]

167

Table 2. List of differentially abundant compounds in sweet cherries obtained by UHPLC-DAD-HR-MS/MS in negative ESI mode. The details of the compounds are given, together with the specification of the reliability class and references used for the detection; R_t, retention time. MSI, Metabolomics Standards Initiative. *, only in harvests 2016 and 2017. All observed ions are $[M-H]^-$.

Putative identification	$R_{t\ (min)}$	Formula	Theoretical *m/z*	Observed *m/z*	Mass error (ppm)	Main MS2 fragments	MSI reliability class with references used for the annotation
A-type flavanol dimer I	13.85	$C_{30}H_{24}O_{12}$	575.1195	575.1176	-3.28	125.0249 – 163.0017 – 255.0300	3 [34,35]
B-type flavanol trimer I *	13.93	$C_{45}H_{38}O_{18}$	865.1985	865.1976	-1.07	125.0231 – 161.0243 – 407.0770	3 [30,35]
Coumaroyl quinic acid	15.99	$C_{16}H_{18}O_8$	337.0929	337.0928	-0.16	173.0462 – 93.0341 – 119.0494	3 [30–32]
(epi)afzelechin–(epi)catechin	17.63	$C_{30}H_{26}O_{11}$	561.1402	561.1381	-3.81	289.0733 – 245.0821 – 203.0747	2 [33,34]
A-type flavanol trimer I	19.01	$C_{45}H_{36}O_{18}$	863.1829	863.1813	-1.79	285.0370 – 125.0265 – 161.0259	3 [35]
B-type flavanol trimer II	19.01	$C_{45}H_{38}O_{18}$	865.1985	865.1988	0.35	125.0244 – 407.0797 – 161.0239	3 [35]
B-type flavanol tetramer *	19.63	$C_{60}H_{50}O_{24}$	1153.2619	1153.2570	-4.27	125.0279 – 243.0303 – 161.0254	3 [35]
A-type flavanol trimer II	19.63	$C_{45}H_{36}O_{18}$	863.1829	863.1810	-2.16	125.0230 – 161.0276 – 243.0317	3 [35]
B-type flavanol pentamer	19.93	$C_{75}H_{62}O_{30}$	1441.3253	1441.3218	-2.44	125.0219 – 243.0327 – 287.0709	3 [35]
Trihydroxyflavanone	20.78	$C_{15}H_{12}O_5$	271.0612	271.0605	-2.58	153.0167	2
B-type flavanol dimer	21.50	$C_{30}H_{26}O_{12}$	577.1352	577.1342	-1.61	125.0237 – 289.0737 – 161.0258	3 [30,31,35]
A-type flavanol dimer II	21.50	$C_{30}H_{24}O_{12}$	575.1195	575.1179	-2.79	125.0224 – 161.0292 – 177.0172	3 [31]
Deoxyhexosyl cinchonain	24.53	$C_{30}H_{30}O_{13}$	597.1614	597.1587	-4.43	341.0686 – 189.0171 – 217.0126	3 [33]
Dihydroxyflavanone hexoside	25.72	$C_{21}H_{22}O_{10}$	433.1140	433.1125	-3.40	271.0612 - 243.0652	2 [30]

173

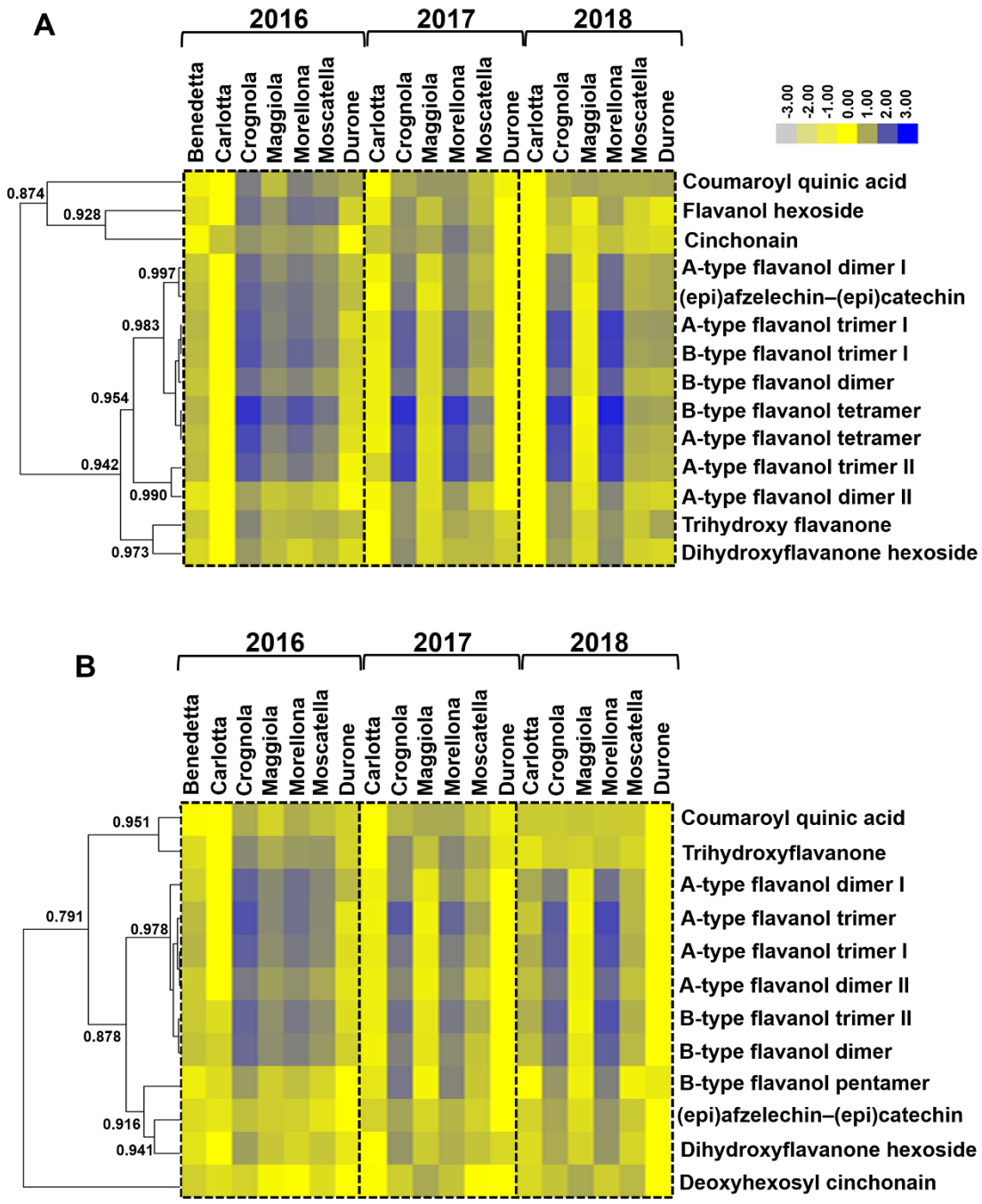


Figure 2. Heat map hierarchical clustering showing the fold change differences of the compounds identified. (A) Metabolites identified in positive and (B) in negative mode in the three years. A maximum fold change>3 in absolute value was used, together with a *p*-value<0.05 at the one-way ANOVA. Fold-changes were calculated using the means of normalized abundances. To build the heatmap, the fold change values were rescaled based on the lowest value detected per single metabolite and then log10-transformed. Numbers indicate

the Pearson correlation coefficients. The color bar indicates the log10-transformed fold change values.

Other phenolic compounds, namely neochlorogenic acid, catechin, chlorogenic acid, epicatechin and quercetin, were identified using standards (Supplementary Table 1 and 2 and Supplementary Figure 1) and hence classified as MSI reliability class 1 compounds. The data obtained for these compounds are in line with those obtained previously, especially for flavonoids and confirm Crognola and Morellona as the best producers of secondary metabolites[3].

The results obtained with metabolomics showed an impact of the genotype on the biosynthesis of flavonoids: this is well known and supported by a strong body of evidence in the literature [20,36–39]. Based on these results and those previously published[3,5,21], Crognola and Morellona appear to be genetically predisposed to produce high amounts of secondary metabolites. Interestingly, these two varieties clustered together in a separate branch of the dendrogram (Figure 1), a finding indicating differences at the genome-level with respect to all the others.

Targeted gene expression analysis

Since the biosynthesis of secondary metabolites is regulated at the gene level[40], RT-qPCR was performed to quantify the relative gene expressions of the PPP-related genes. The gene expression analysis was carried out on the three years of harvest 2016, 2017 and 2018 and on genes intervening in the PPP. The genes investigated were phenylalanine ammonia lyase-*PAL*, cinnamate-4-hydroxylase-*C4H*, 4-coumarate-coenzyme A ligase-*4CL*, chalcone synthase-*CHS*, chalcone isomerase-*CHI*, flavanone 3-hydroxylase-*F3H*, dihydroflavonol 4-reductase-*DFR*, anthocyanidin synthase-*ANS* and a UDP-glycosyltransferase-*UGT* (responsible for the glycosylation of anthocyanin aglycones).

The genes involved in the PPP are notoriously multigenic[41,42]; for *PAL*, *4CL* and *CHI*, two isoforms were analyzed because of the roles that these genes have as gatekeepers (*PAL*) and members of the general steps (*4CL*), respectively, as well as their implication in branch points (*CHI*). By studying the genes coding for isoforms, it is possible to speculate about their potential role in the provision of precursors needed for the synthesis of aromatic macromolecules. Subsequently, a hierarchical clustering of the heatmap was carried out to unveil potential correlations of expression patterns among the genes studied (Figure 3).

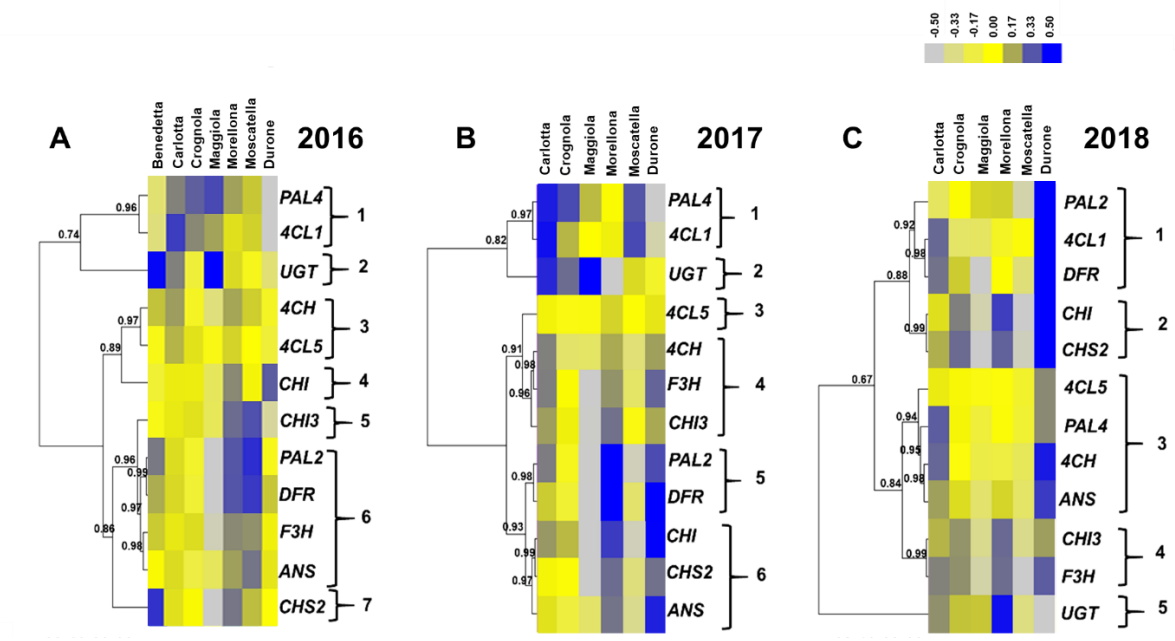


Figure 3. Heat map hierarchical clustering of the PPP-related gene expression data across the 3 years of study (A, 2016; B, 2017; C, 2018). Numbers indicate the Pearson correlation coefficients. The color bar indicates the log10-transformed normalized relative quantities. The bar graphs of the expression data are available in Supplementary Figure 2.

In 2016, seven major patterns could be distinguished by setting 0.96 as threshold value for the Pearson correlation. The first one was composed by *PAL4* and *4CL1*, the second by *UGT*, the third by *C4H* and *4CL5*, the fourth and fifth by *CHI* and *CHI3*, the sixth comprised *PAL2*, *DFR*, *F3H* and *ANS* and the last *CHS2* (Figure 3A). Besides *CHI*, the commercial fruits displayed lower expressions as compared to the ancient ones and this was particularly evident

for the genes partaking in the general phase of the PPP, i.e. the isoforms of *PAL*, *4CL*, as well as *4CH* (Figure 3A). A lower expression of these genes may indeed be directly responsible for a decreased synthesis of products shunted to the specialised branch of the PPP leading to the biosynthesis of flavonoids.

Carlotta, Crognola and Maggiola showed overall higher expression of *PAL4*, while *CHI3*, *PAL2*, *DFR*, *F3H*, *ANS* and *CHS2* were highly expressed in the varieties Morellona and Moscatella. The variety Benedetta showed high expression of *UGT* and *CHS2*.

In 2017, six major clusters could be recognized by setting a threshold value of 0.93 for the Pearson correlation (Figure 3B): the first two were the same as those of 2016, i.e. *PAL4*/*4CL1* and *UGT*, the third comprised only *4CL5*, the fourth *C4H*, *F3H* and *CHI3*, the fifth cluster included *PAL2* and *DFR*, the last *CHI*, *CHS2* and *ANS*. It was possible to observe again the clustering of *PAL4* with *4CL1* and of *PAL2* with *DFR*, as previously seen for the fruits sampled in 2016. The commercial variety showed instead differences, notably higher expression of the genes involved in the central and late stages of the PPP. Morellona confirmed the high expression of genes involved in flavonoids/anthocyanin biosynthesis. Maggiola displayed a high expression of *UGT*, as previously observed in 2016 (Figure 3A).

The year 2018 differed from the previous ones in terms of gene expression (Figure 3C). By setting a threshold value of 0.88 for the correlation coefficient, five major expression clusters were observed: the first grouped *PAL2*, *4CL1* and *DFR*, the second *CHI* and *CHS2*, the third comprised *4CL5*, *PAL4*, *C4H* and *ANS*, the fourth cluster grouped *CHI3* and *F3H* and the last one was represented by *UGT*. *PAL2* and *DFR* were in the same cluster; however, in 2018, *4CL1* was also present, differently from the previous years, where it grouped with *PAL4*.

PAL2 showed lower expression in Morellona with respect to 2016 and 2017, while the commercial fruits showed much higher expression of all the genes, with the exception of *UGT*.

It is interesting to note that the two *PAL* isoforms clustered with different genes in 2016 and 2017: *PAL4* grouped with *4CL1*, while *PAL2* with *DFR*. In thale cress, four different *PAL* isoforms were described[43] and a redundant role for *PAL1* and *PAL2* was demonstrated in flavonoid biosynthesis[44]. The sweet cherry *PAL2* may be involved in the PPP branch shunting precursors towards the synthesis of flavonoids. This however awaits experimental confirmation. RT-qPCR also revealed a higher expression of the genes involved in the central and late stages of the PPP in Morellona (Figure 3).

The expression of PPP-related genes is subjected to transcriptional regulation. V-MYB myeloblastosis viral oncogene homolog (MYB) transcription factors (TFs) are master regulators of the PPP and activate the branches leading to the biosynthesis of monolignols, flavonoids and anthocyanins[45]. Two genes encoding MYB TFs were here investigated using RT-qPCR, i.e. *MYB10.1* and *MYB11*. The choice of these genes is motivated by their role in the biosynthesis of anthocyanins[46,47] and flavonols[48], respectively.

In 2016, *MYB10.1* showed the highest expression in the varieties Moscatella and the lowest in Crognola, while in 2017 and 2018 the commercial fruits showed the highest expression of the gene (Figure 4). Despite the variations in expression across the three years, Crognola was always among the varieties expressing low levels of *MYB10.1*, while Morellona showed higher expression of the transcript. Both varieties are characterized by a red color of the skin[49]; however, Morellona is the only variety with a red pulp. Therefore, the high content of anthocyanins reported previously[3,21] can be explained by a higher gene expression. Durone showed also high expression of *MYB10.1*, especially in 2017 and 2018: this can be explained by the intense red color of both the skin and the pulp, features that make these commercial fruits particularly appealing to consumers. Although Moscatella does not display intense red pigmentation, our results show that this variety ranks among the highest producers of

proanthocyanidins (Figure 2): this result can be explained by a high expression of *MYB10.1* in this variety.

The *MYB10.1* gene was sequenced in the varieties Morellona (deposited in GenBank with the accession number MH545964) and Crognola (partial sequence in Supplementary Figure 3) to check the occurrence of the previously reported alleles *MYB10.1a* and *b* responsible for the red and blush color of the skin[46]. The *a* allele was cloned from Crognola and the *b* allele from Morellona; it remains to be verified whether the *MYB10.1a* allele occurs together with *MYB10.1b* in Morellona, as is expected on the basis of the strong red color observed in the fruits of this variety.

The gene expression patterns of *MYB11* showed the largest variations among the Tuscan cherries in 2017, where the varieties Crognola and Maggiola were the highest and Durone the lowest. Interestingly, Durone always showed the lowest expression of the TF in all the years studied.

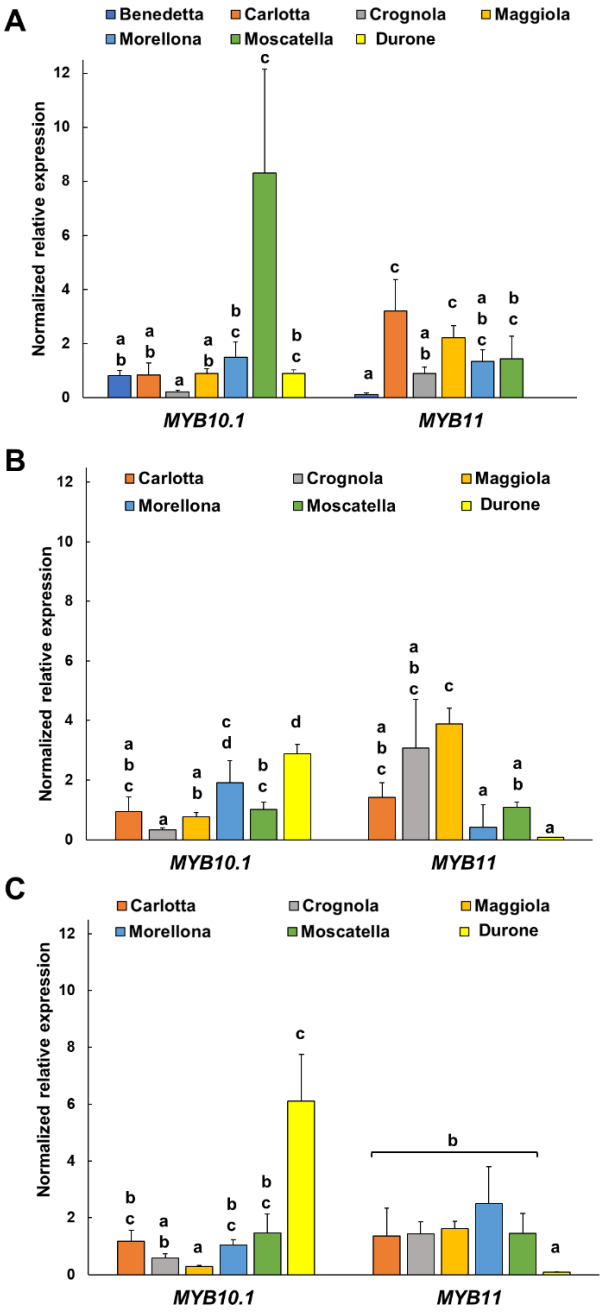


Figure 4. Gene expression analysis (indicated as Normalized relative expression) of the MYB TF-encoding genes. (A) results relative to 2016, (B) results obtained on the samples harvested in 2017, (C) results relative to the year 2018. Error bars correspond to the standard deviation (n=4). Different letters represent statistical significance (p-value<0.05) among the groups of data. If a letter is shared, the difference is not significant. A one-way ANOVA followed by Tukey's post-hoc test was performed on genes showing homogeneity and normal distribution;

for the others, a Kruskal-Wallis test followed by Dunn's post-hoc test was used. The statistical parameters in A are *MYB10.1* F(6,20)=34.37, *p*-value=0.000; *MYB11* F(5,18)=28.26, *p*-value=0.000, B: *MYB10.1* F(5,18)=35.69, *p*-value=0.000; *MYB11* X^2(5)=17.46, *p*-value=0.004 and in C: *MYB10.1* X^2(5)=18.44, *p*-value=0.002; *MYB11* F(5,17)=17.84, *p*-value=0.000.

Overall, the RT-qPCR data showed that the PPP-related genes were differentially expressed in the Tuscan varieties. It was not possible to link the abundance of the phenolic compounds (Figure 2, Table 1 and 2) with gene expression profiles (Figure 3, Supplementary Figure 2), since the highest producing varieties Crognola and Morellona did not always show high expression of PPP-biosynthetic genes. The reason for a lack of correlation may be linked to post-transcriptional and post-translational events: for example, it was reported that a Kelch repeat F-box protein, SAGL1, regulates the PPP in thale cress by interacting with PAL1 and mediating its proteasome-dependent degradation[50]. The field conditions are also known to affect the gene expression pattern and this explains the variations observed in the Tuscan fruits. To understand the environmental causes that could be (at least in part) responsible for the differential gene expression in the Tuscan varieties, the daily temperatures from March (blooming period) to May (fruit sampling), as well as the precipitation and humidity maximum/minimum averages were retrieved from the LaMMA meteorological station in Grosseto (http://lamma.eu/en). As can be seen in Supplementary Figure 4, the year 2018 had, in average, warmer minimum temperatures. Variations in the humidity averages were also recorded across the years, with 2016 recording lower average maximum humidity and 2018 higher average minimum humidity (Supplementary Figure 4).

As discussed in the previous section on metabolomics, an influence of the environment on the expression of PPP-related genes is known[51]. The functional quality of strawberries was enhanced under mild drought salinity stress, since the content of phenolics, anthocyanins and

316 ascorbic acid increased[52]. Likewise, in grapevine, seasonal water deficit was shown to affect
317 anthocyanin biosynthesis during ripening by upregulating both genes and metabolites[53].
318 The Tuscan varieties here investigated thrive in wild conditions and in soils with minimal
319 human intervention; given the non-controlled conditions of growth, it is not surprising that
320 gene expression showed such a high variability across the years of study.
321
322 **Analysis of the soluble proteomes of two ancient varieties**
323 An analysis of the proteome was carried out on the varieties Morellona and Crognola. The goal
324 was to highlight differences in the abundance of soluble proteins that could explain the
325 metabolite and gene expression variations observed in the three years.
326 The two-dimensional difference gel electrophoresis (2D-DIGE) experiments showed 166
327 differentially abundant protein spots. The spots were selected according to the following
328 parameters: max fold change>2 and *p*-value <0.01.
329 The spot identification was done through peptide sequence searches in the MASCOT engine.
330 The search was carried out against the NCBI non-redundant protein database restricted to the
331 *P. avium* entries. From an initial number of 166 proteins, after the identification of the same
332 protein isoforms, 51 proteins were retrieved with a good e-value and similarity sequence score,
333 according to MASCOT (Table 3). The proteins were classified by function and category, on
334 the basis of their known involvement in specific pathways: “Stress”, “Cell wall”, “Proteasome-
335 related”, “Primary metabolism” and “Other” (Table 3). Considering all the proteins detected,
336 the three categories with the highest number of differentially abundant proteins are “Stress”,
337 “Cell wall” and “Primary metabolism” (Table 3).
338
339 **Table 3.** Details of the spot numbers, accession numbers, annotations and *p*-values of the
340 identified proteins. The *p*-values<0.01 are highlighted in light green.

Spot N°	Accession N°	Protein Name	Function	Category	*p*-value (year)	*p*-value (variety)
753	XP_021804554.1	Thioredoxin reductase NTRB	Oxidoreductase activity	STRESS	0.12515	0.00297
996	XP_021801592.1	Superoxide dismutase	Oxidoreductase activity	STRESS	0.00965	0.14682
1004	XP_021816866.1	2-Cys peroxiredoxin BAS1	Cell redox homeostasis/Stress response	STRESS	0.7117	0.00026
988	XP_021823852.1	Glutathione *S*-transferase F11	Response to oxidative stress	STRESS	0.66672	0.00001
976	XP_021813812.1	Glutathione *S*-transferase DHAR2	Oxidoreductase activity	STRESS	0.00007	0.32618
1006	XP_021832122.1	Glutathione *S*-transferase	Oxidoreductase activity	STRESS	0.24001	0.00043
1023	XP_021808495.1	NAD(P)H dehydrogenase (quinone) FQR1	Response to auxin/Oxidoreductase activity/Stress response	STRESS	0.00036	0
378	XP_021806563.1	Protein disulfide-isomerase	Cell redox homeostasis	STRESS	0.84562	0
299	XP_021815855.1	Stromal 70 kDa heat shock-related	Stress response	STRESS	0.17303	0.00017
948	XP_021820221.1	Low-temperature-induced cysteine proteinase	Stress response	STRESS	0.81184	0.00003
632	XP_021832997.1	Protein SRC2	Stress response	STRESS	0.27232	0.00004
845	XP_021820409.1	Glycine-rich protein 2	Response to water deficit. ABA/Cell wall	STRESS	0.0056	0
1163	XP_021824432.1	17.1 kDa class II heat shock protein	Stress response	STRESS	0.00695	0.00163
985	XP_021805019.1	20 kDa chaperonin	Stress response	STRESS	0.0128	0.00019
674	XP_021804100.1	Plastoglobulin-1	Lipid metabolism of chloroplasts/Stress response	STRESS	0.05678	0.00082
1238	XP_021831271.1	Phosphatidylglycerol/phosphatidyl inositol transfer protein	Lipid binding (recognition of pathogen related products)	STRESS	0.15061	0.00045
792	XP_021812280.1	Xyloglucan endotransglucosylase 31	Cell wall metabolism	CELL WALL	0.4535	0.00001
477	XP_021823827.1	UTP-glucose-1-phosphate uridylyltransferase	Cell wall metabolism	CELL WALL	0.40482	0.00437
822	XP_021802094.1	Xyloglucan endotransglucosylase 6	Cell wall metabolism	CELL WALL	0.01025	0.00207
347	AAA91166.1	Beta-glucosidase	Carbohydrate metabolism/Cell wall metabolism	CELL WALL	0.00037	0.56739

221	XP_021814417.1	Alpha-xylosidase 1	Cell wall metabolism	CELL WALL	0.08851	0.00005
201	XP_021823469.1	Acid beta-fructofuranosidase 1	Carbohydrate metabolism/Cell wall metabolism	CELL WALL	0.08196	0.00561
511	XP_021814422.1	Ubiquitin receptor RAD23c	Protein catabolic process	PROTEASOME-RELATED PATHWAY	0.00018	0.91013
1012	XP_021825182.1	Proteasome subunit beta type-6	Protein catabolic process	PROTEASOME-RELATED PATHWAY	0.0371	0.00475
936	XP_021828441.1	Proteasome subunit alpha type-6	Protein catabolic process	PROTEASOME-RELATED PATHWAY	0.00762	0.03074
258	XP_021832122.1	Phosphoenolpyruvate carboxykinase	Decarboxylase activity	PRIMARY METAB. GLUCONEOG.	0.00066	0.00002
161	XP_021822858.1	Aconitate hydratase	Lyase activity	PRIMARY METAB, TCA CYCLE	0.00664	0.41726
1453	XP_021823850.1	Isocitrate dehydrogenase	Oxidoreductase activity	PRIMARY METAB, TCA CYCLE	0.06206	0.00213
1293	XP_021813779.1	2-oxoglutarate dehydrogenase	Oxidoreductase activity	PRIMARY METAB, TCA CYCLE	0.00439	0.00327
709	XP_021804616.1	Malate dehydrogenase	Oxidoreductase activity	PRIMARY METAB, TCA CYCLE	0.0004	0.00121
411	XP_021825040.1	Glucose-6-phosphate 1-dehydrogenase	Oxidoreductase activity	PRIMARY METAB, PENTOSE PHOSPHATE PATHWAY	0.15166	0.00721
505	XP_021817392.1	6-phosphogluconate dehydrogenase. decarboxylating 3	Oxidoreductase activity	PRIMARY METAB, PENTOSE PHOSPHATE PATHWAY	0.16838	0.00277
883	XP_021831593.1	Gamma carbonic anhydrase 1	Carbonate dehydratase activity	PRIMARY METAB, PHOTORESP	0.39206	0
624	XP_021810138.1	Fructose-bisphosphate aldolase 1	Lyase activity	PRIMARY METAB, GLYCOLYSIS	0.01983	0.00042
251	XP_021834633.1	5-methyltetrahydropteroyltriglutamate homocysteine methyltransferase	Methyltransferase	PRIMARY METAB, AMINO ACID BIOSYNTH	0.02626	0.000000
209	XP_021818418.1	Eukaryotic translation initiation factor 3 subunit B	Translation regulation	PRIMARY METAB, PROTEIN BIOSYNTH	0.01917	0.00332
188	XP_021828563.1	Elongation factor 2	Polypeptide chain elongation	PRIMARY METAB, PROTEIN BIOSYNTH	0.06334	0.00455
761	XP_021821654.1	Elongation factor 1	Polypeptide chain elongation	PRIMARY METAB, PROTEIN BIOSYNTH.	0.45155	0.00002
910	XP_021819098.1	Soluble inorganic pyrophosphatase 4	Hydrolase activity	PRIMARY METAB, PO_4^{3-}-CONTAINING	0.04654	0.00001

				COMPOUND METABOLIC PROCESS		
66	XP_021824244.1	ATP synthase subunit beta	ATP synthesis	PRIMARY METAB, ENERGY PRODUCTION	0.14432	0.00849
439	XP_021830782.1	Polyphenol oxidase	Oxidoreductase activity (pigment biosynthesis)	OTHER	0.00005	0.80611
957	XP_021826507.1	Ferritin-4	Iron binding	OTHER	0.03802	0.00003
944	XP_021820122.1	Ferritin-3	Iron binding	OTHER	0.00021	0.00003
772	XP_021825752.1	Annexin-like protein RJ4	Ca^{2+}-dependent phosphlipid binding	OTHER	0.00198	0
1024	XP_021804938.1	Auxin-binding protein ABP19a	Auxin receptor	OTHER	0.00007	0.00033
887	XP_021828712.1	GTP-binding nuclear protein Ran-3	GTPase activity	OTHER	0.58551	0.00403
46	XP_021826137.1	Patellin-3-like	Cell cycle/cell division	OTHER	0.26719	0.00001
1009	XP_021804963.1	Uncharacterized protein LOC110749212	Nutrient reservoir activity /Response to ABA	OTHER	0.43022	0.00502
814	XP_021820843.1	Short-chain dehydrogenase TIC 32	Oxidoreductase activity	OTHER	0.14667	0.00616
725	XP_021806982.1	Voltage-gated potassium channel subunit beta	Ion transport	OTHER	0.00007	0
802	XP_021807453.1	Mitochondrial outer membrane protein porin of 36 kDa	Ion transport	OTHER	0.43668	0.00712

341

The pattern of protein abundances between the two varieties and the three years is represented as a heatmap hierarchical clustering in Figure 5. By choosing a Pearson correlation coefficient >0.94, two major clusters could be distinguished which correspond to proteins that were more abundant in Morellona or Crognola. Nevertheless, variability in the abundance of some proteins was observed in Crognola in 2016.

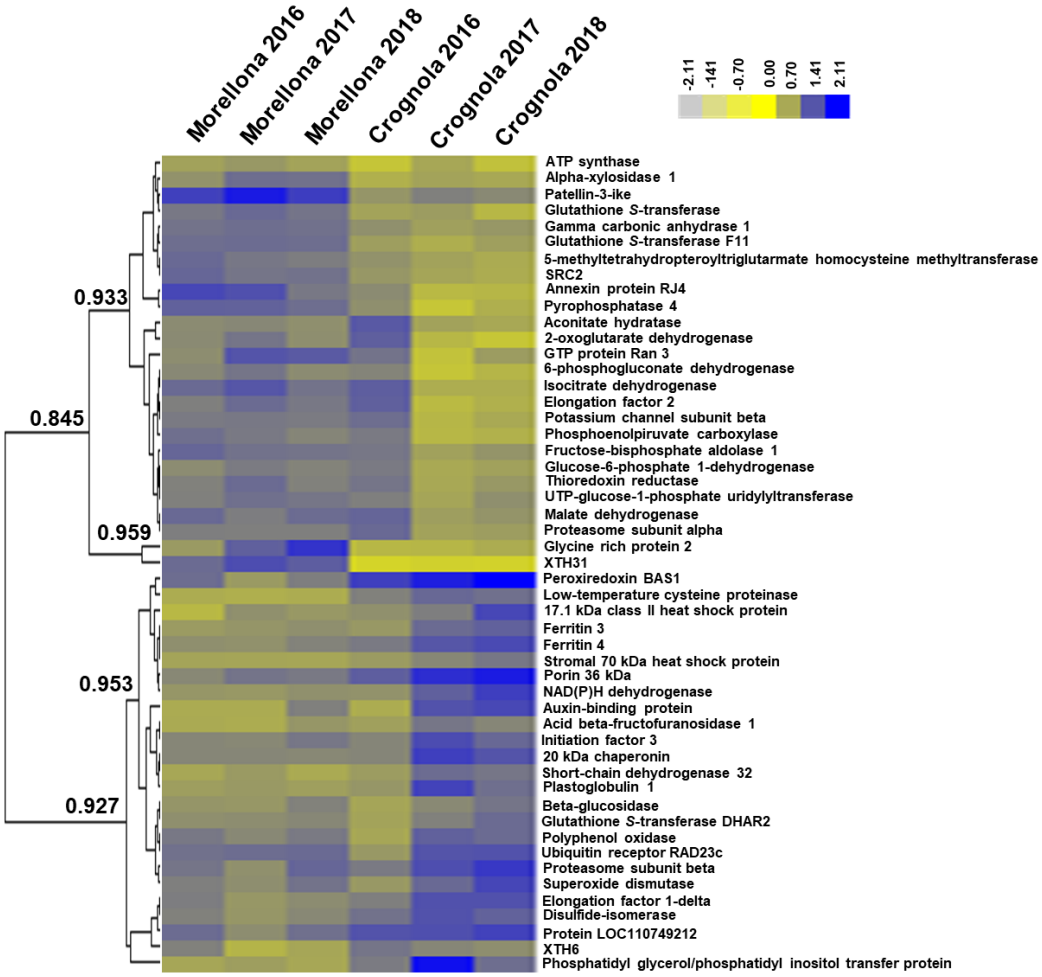


Figure 5. Heat map hierarchical clustering of the 51 proteins changing significantly between Crognola and Morellona across the three years of study. The color bar indicates the log10-transformed relative protein abundances. The numbers indicate the Pearson's correlation coefficients.

Hereafter, each protein category is discussed separately.

353 ***Proteins related to stress response***

354 In the category "Stress", seven proteins related to the maintenance of the redox status were
355 found, namely a thioredoxin reductase (TrxR), a superoxide dismutase (SOD), a peroxiredoxin
356 (Prx), three glutathione-*S*-transferases (GSTs) and the quinone reductase FQR1 (Table 3).
357 Additionally, a protein disulfide isomerase (PDI) was detected, which in plants is involved in
358 the redox control of proteins' disulfide bonds, thus likely acting as a chaperone in response to
359 stress[54]. The differences were mainly related to the variety (Table 3), a finding suggesting that
360 the two varieties respond differently to exogenous cues.

361 Within plant cells, during normal growth and development, there exists a balance between
362 oxidants and antioxidants[55]. ROS are needed for normal growth but, at high levels, they cause
363 premature senescence by oxidative stress[56] that in fruits triggers a loss of texture, flavour and
364 a decrease in health beneficial molecules.

365 Among the proteins devoted to the plant's defense against (a)biotic stresses, Glutathione S-
366 transferase DHAR2 (DHAR2) is present in a similar concentration between the varieties
367 Morellona and Crognola. On the contrary, the amount of this protein appeared to increase in
368 relation with the different years of harvest. Indeed, in 2018 an increase of DHAR2 can be
369 noticed in both varieties, maybe due to the variable environmental conditions. DHAR2 belongs
370 to a subclass of enzymes in the wide group of GSTs and is able to catalyze the reduction of
371 dehydroascorbate to ascorbate with the concomitant oxidation of reduced GSH to glutathione
372 disulfide[57,58].

373 GST F11 belongs to another GST group carrying out different reactions compared to the main
374 GST class, due to the lack of a serine in the active site. Instead, GST F11 seems to have a role
375 in glucosinolate metabolism[57,59]. GST F11 was highly abundant in the variety Morellona,
376 thereby showing a dependency on the genotype. The higher abundance of GST F11, involved
377 in the glucosinolate pathway[60], is interesting in light of the health-related effects of

isothiocyanates. These are indeed molecules obtained through the action of myrosinase on glucosinolates and displaying anticancerogenic activity[61]. It remains to be verified whether glucosinolates are present in sweet cherry (the protocol here used is indeed not optimized to extract this class of compounds) and whether Morellona produces more glucosinolates than Crognola. The differences observed may also be due to biotic stress events, since glucosinolates are typically synthesized in response to herbivores' attack[62].

Prx is a protein that is commonly responsible for the signalling related to ROS[63] and acts together with TrxR and by using NADPH as a source of reducing power[64,65]. The Prx BAS1 here identified as more abundant in Crognola was reported to be involved in the protection against oxidative stress by participating, together with the Trx CDSP32, in the reduction of alkyl hydroperoxides[66].

Statistically significant changes between the varieties were obtained also for the quinone reductase FQR1 which showed higher levels in Crognola: the corresponding gene was shown to be induced by auxin despite the absence of auxin-responsive elements in its promoter and to be involved in stress response by regulating oxidative stress together with GST[67].

In the category "Stress", proteins related to the response to external cues (temperature stress) were identified. More specifically, SRC2 (the product of *soybean gene regulated by cold-2*), a low-temperature-induced cysteine proteinase, a stromal 70 kDa heat shock-related protein (HSP), a 17.1 kDa class II and a 70 kDa HSPs, together with a 20 kDa chaperonin showed differences between the varieties (Table 3).

SRC2 is considered a cold stress marker: an increase of this protein was indeed reported in plant tissues exposed to low temperatures[68].

Glycine-rich protein 2 (GRP2) was identified in 10 different spots. Two of them were more abundant in Morellona and varied across the years (Table 3 and Figure 5). Domain analysis of sweet cherry GRP2 with Motif Scan (https://myhits.isb-sib.ch/cgi-bin/motif_scan) revealed the

presence of a glycine-rich, as well as, an ABA/WDS domain (Abscisic Acid/Water Deficit Stress) domain (e-value=2.9e-45). The ABA/WDS domain was found in the dual transcription factor/chaperone protein ASR1 (ABA, Stress and Ripening) which was induced in tomato upon drought and was expressed during ripening[69]. GRP2 is also linked to fruit maturation. For example, in pear, this protein increased in abundance after gibberellin application[70]. This plant growth regulator induces fruit expansion and GRPs are known to act at the cell wall level by providing a scaffold for the deposition of cell wall constituents[71].

Crognola showed higher abundance of HSP70, 20 and 17.1 (members of small HSPs; Table 3) which are involved in the tolerance to abiotic stresses[72,73], as well as of a low-temperature-induced cysteine proteinase showing the presence of granulin domains (e-values=1.2e-10, 2.1e-06).

Proteins related to the cell wall

The proteomic analysis revealed six proteins related to the cell wall: a xyloglucan endotransglucosylase/hydrolase 31 (XTH31), a UTP-glucose-1-phosphate uridylyltransferase. a xyloglucan endotransglucosylase/hydrolase 6 (XTH6), an alpha-xylosidase 1, a β-glucosidase and an acid beta-fructofuranosidase 1 (Table 3).

Xyloglucans bridge cellulose microfibril, thereby contributing to the mechanical properties of the cell walls and to morphogenesis[74]. XTHs display both xyloglucan *endo*-transglucosylase (XET, cutting and rejoining xyloglucan chains) and xyloglucan *endo*-hydrolase (XEH, hydrolysis of xyloglucan) activities[75–78]. The majority of XTHs enzyme kinetics data showed the predominant presence of XET activity; a bioinformatic analysis coupled to structural data and enzymology predicted AtXTH31 and 32 from thale cress as potential hydrolases belonging to clade III-A[74]. Subsequent studies showed that AtXTH31 accounts for the majority of XET activity in *Arabidopsis thaliana* roots and has a pivotal role under Al stress[79].

The phylogenetic analysis of thale cress, poplar, tomato and nasturtium XTHs showed that the cherry XTH31 clustered in group III-A, together with AtXTH31, the paralog AtXTH32 and nasturtium TmNXG1 (a predominant xyloglucan hydrolase[74]) (Supplementary Figure 5).

It was demonstrated that thale cress XTHs from group III-A are endohydrolases involved in tissue expansion and are dispensable for normal growth[80].

XTH31 was identified in 6 different spots; one of them showed a significant decrease in abundance in Morellona. The statistically significant difference detected in this spot is interesting if one considers the sizes of the fruits produced by the 2 varieties: Morellona is significantly bigger than Crognola ($p<0.05$, n=10, diameter Morellona=1.77±0.10 cm, diameter Crognola=1.64±0.09 cm; height Morellona=1.95±0.05 cm, height Crognola=1.76±0.10 cm).

The other XTH detected in the soluble proteomes of the two Tuscan varieties is XTH6, which clusters together with AtXTH6 (Supplementary Figure 5). The abundance of XTH6 increased in thale cress shoots and roots under heat stress in response to cytokinin[81], while the transcript was downregulated upon drought stress in 6 different accessions of *A. thaliana*[82]. It is therefore reasonable to assume that, in Crognola, the higher abundance of XTH6 is linked to environmental cues to which the variety reacted

Alpha-xylosidases are involved in xyloglucan remodelling[83]; the sweet cherry XYL1 identified via proteomics is orthologous to thale cress XYL1. The higher abundance in Morellona is indicative of a higher xyloglucan remodelling at maturity. A previous study showed that *xyl1*/*axy3* mutants displayed reduced silique length and altered xyloglucan structure, where the hemicellulose was less tightly bound to other cell wall components[84].

Differences in the abundance of a UTP-glucose-1-phosphate uridylyltransferase were also detected between the two Tuscan varieties. A BLASTp analysis against thale cress revealed sequence similarity with UGP2, one of the two genes contributing to sucrose and cell wall

biosynthesis[85]. The higher expression in Morellona may indicate an involvement in cell wall-related processes and in accommodating the request of nucleotide sugars during fruit maturation. The bigger size of Morellona cherries as compared to Crognola may indeed require a higher provision of precursors for cell wall biosynthesis.

A β-glucosidase (BGLU) with sequence similarity to the apoplast-localized *A. thaliana* BGLU15 was also identified as more abundant in Crognola, with respect to Morellona. Interestingly, despite the apoplastic localization, this protein is not related to cell wall processes, but to the hydrolysis of flavonol 3-*O*-β-glucoside-7-*O*-α-rhamnoside (a flavonol bisglycoside acting as antioxidant and reducing ROS damage), which occurs in thale cress during recovery from synergistic abiotic stresses (i.e. N deficiency, low temperature, high light intensity, UV light)[86,87]. Therefore, the BGLU protein seems to be linked to stress-related pathways.

Proteins related to primary metabolism

In the category "Primary metabolism", 15 differentially abundant proteins were identified (Table 3), the majority of which was more abundant in Morellona in 2017 and 2018 (Figure 5). The differences detected may be due to the different genotypes; however, plant primary metabolism is also significantly influenced by environmental conditions, like biotic[88] and abiotic constraints[89] encountered in the field during the different years studied.

Three different proteins related to carbohydrate metabolism were identified (Table 3), i.e. glucose-6-phosphate dehydrogenase (G6PDH), 6-phosphogluconate dehydrogenase (6PGDH) and fructose-bisphosphate aldolase (FBA), showed significant differences between the two cherry varieties. The abundance of these 3 proteins was generally higher and steady in the variety Morellona during the 3 years. On the contrary, Crognola showed a variable abundance: indeed, as shown in Figure 5, the levels of G6PDH, 6PGDH and FBA were higher in 2016 than

2017 and 2018. G6PDH and 6PGDH are involved in the oxidative pentose phosphate pathway and their role, despite linked to glucose oxidation, is anabolic, rather than catabolic. Additionally, G6PDH sustains nitrogen assimilation[90] and counteracts stress conditions[91].

Four proteins related to the tricarboxylic acid cycle (TCA) were also detected: aconitate hydratase (ACO), isocitrate dehydrogenase (IDH), 2-oxoglutarate dehydrogenase (OGDH) and malate dehydrogenase (MDH). In 2017 and 2018, the abundance of these proteins, as well as phosphoenolpyruvate carboxykinase (PEPCK), was higher in Morellona (Figure 5). Fruit maturity is linked with primary metabolism and the production of organic acids. In sweet cherry, malate accumulates at the highest levels during stage III (coinciding with expansion and ripening) and is used for gluconeogenesis by the action of PEPCK[92]. Therefore, from the results obtained, it appears that the fruits of Morellona put in place biochemical processes linked with fruit maturation earlier than Crognola. This finding is supported by the previously described higher abundance of GRP2 which is involved in cell wall-related processes accompanying ripening, as well as of annexin RJ4 (in the category "Other". see below), typically expressed during fruit ripening in strawberry[93].

Proteins related to the proteasome and other functions

In order to cope with the organism's demands and to maintain the normal functions, cells require a continuous turnover of proteins, operated, among other actors, by the proteasome system[94].

The ubiquitin receptor RAD23c, the proteasome subunits alpha type-6 and beta type-6 were found to be differentially abundant in dependence of the years and the varieties (Table 3).

Interestingly, for the variety Crognola, the alpha type-6 subunit was more abundant in 2016, while in 2017 and 2018, the beta type-6 subunit was higher in abundance (probably coinciding with different moments of the protein complex turn-over).

503 It should be noted that polyphenol oxidase (PPO) was the only detected protein in the category
504 "Other" related to the metabolism of phenolic compounds: this enzyme catalyzes the
505 polymerization of quinones formed through the oxidation of phenols[95] to produce brown
506 pigments. PPO plays also a role against biotic stresses, such as insect attacks and in defense
507 mechanisms related to altered environmental conditions[95,96]. Although a dependence on the
508 year of harvest was detected (Figure 5), Crognola showed the highest amount of PPO. Future
509 studies will confirm if a higher amount of PPO in fruits may confer additional defense
510 properties in relation to stress conditions in this variety.
511 Two ferritins were also more abundant in Crognola in 2017 and 2018 (Figure 5). A ferritin was
512 previously also identified via proteomics in peach during its development and its abundance
513 was higher in the mesocarp[97]. Besides playing a role in iron storage, ferritins are also involved
514 in ROS metabolism and the maintenance of the redox balance within plant tissues[98].
515 Considering their secondary role in ROS detoxification, the results obtained for ferritins can
516 be compared with those previously discussed for proteins related to the stress response. Ferritin
517 3 and 4 showed a higher abundance in Crognola, especially in 2017 and 2018, similarly to what
518 described for BAS1, FQR1 and HSPs. Therefore, Crognola and Morellona showed different
519 levels of proteins related to stress response. More specifically, Crognola had an overall higher
520 abundance of these proteins.
521 ***Gene expression analysis on some targets identified with proteomics***
522 The expression of genes belonging to the categories "Stress" and "Cell wall" was measured to
523 find a correlation with the abundances highlighted by proteomics. The expression of *PPO*
524 within the category "Other" was also studied. Nineteen primers were designed on the targets
525 reported in Table 4.
526

527 **Table 4**. Target proteins on whose corresponding genes primers for RT-qPCR were designed.
528 Their categories are also indicated.

Targets	Abbreviation	Category
Polyphenol oxidase	*PPO*	OTHER
SRC2	*SRC2*	STRESS
Glutathione *S*-transferase F11	*GST*	STRESS
Superoxide dismutase	*SOD*	STRESS
Thioredoxin reductase NTRB	*NTR*	STRESS
Glutathione *S*-transferase DHAR2	*DHAR2*	STRESS
2-Cys peroxiredoxin BAS1	*BAS1*	STRESS
Stromal 70 kDa heat shock-related protein	*70HS*	STRESS
Heat shock cognate 70 kDa protein 2	*70HS2*	STRESS
Low-temperature-induced cysteine proteinase	*LTP*	STRESS
17.1 kDa class II heat shock protein	*HSP17*	STRESS
Glycine-rich protein 2	*RBG2*	STRESS
20 kDa chaperonin	*CPN20*	STRESS
Beta-glucosidase	*BGL*	CELL WALL
Xyloglucan endotransglucosylase 31	*XTH31*	CELL WALL
Xyloglucan endotransglucosylase 6	*XTH6*	CELL WALL
Alpha-xylosidase 1	*XYL1*	CELL WALL
Acid beta-fructofuranosidase 1	*VI1*	CELL WALL
UTP-glucose-1-phosphate uridylyltransferase	*UGP2*	CELL WALL

529 The gene expression graph is given in Figure 6. Hereafter, the genes confirming the trend
530 observed in proteomics are described with more emphasis.
531 The gene encoding the 2-Cys peroxiredoxin BAS1 showed no difference among the three years
532 in Crognola, while in Morellona sampled in 2017 it displayed the lowest expression. Generally,
533 *BAS1* showed lower expression in Morellona, thereby confirming the results obtained with
534 proteomics (Figure 5 and Table 3).
535 *GST F11* showed higher expression in Morellona: despite the low expression in 2018, it was
536 highly expressed in this variety in the years 2016 and 2017, thus following the trend of the
537 protein.

538 The gene *HSP17.1* varied in expression among the years of harvest, mostly in the variety
539 Crognola. According to Table 3, the protein HSP17.1 showed statistically significant variations
540 across the years of study, confirming the gene expression results. Moreover, the high protein
541 abundance in Crognola in 2018 was in agreement with the high expression in Figure 5.
542 The gene *PPO* showed variations in the 3 years in Morellona, especially in 2016 and 2018. In
543 accordance with these data, the relative protein showed statistically significant variations
544 across the years (Figure 5).
545 A clear trend could be observed for *SRC2*, expressed at higher levels in the variety Morellona.
546 Notably, the relative protein was also more abundant in Morellona (Figure 5).
547 In the category "Cell wall", the gene coding for XTH31 was expressed at higher levels in
548 Morellona (Figure 6), as previously seen also for the protein abundances (Figure 5). Differently
549 from what observed with proteomics, *XTH6* did not show a statistically significant difference
550 between the two varieties. In 2018, *XTH6* was induced in Morellona (Figure 6).
551 The partial agreement of the RT-qPCR data with proteomics is not surprising: it is known that
552 gene expression changes are not always accompanied by a similar trend in the corresponding
553 proteins[99]. This can be due to post-transcriptional modifications or to other protein processing
554 events. However, it was possible to confirm, at the gene level, that *XTH31* was upregulated in
555 Morellona at the sampling time-point.

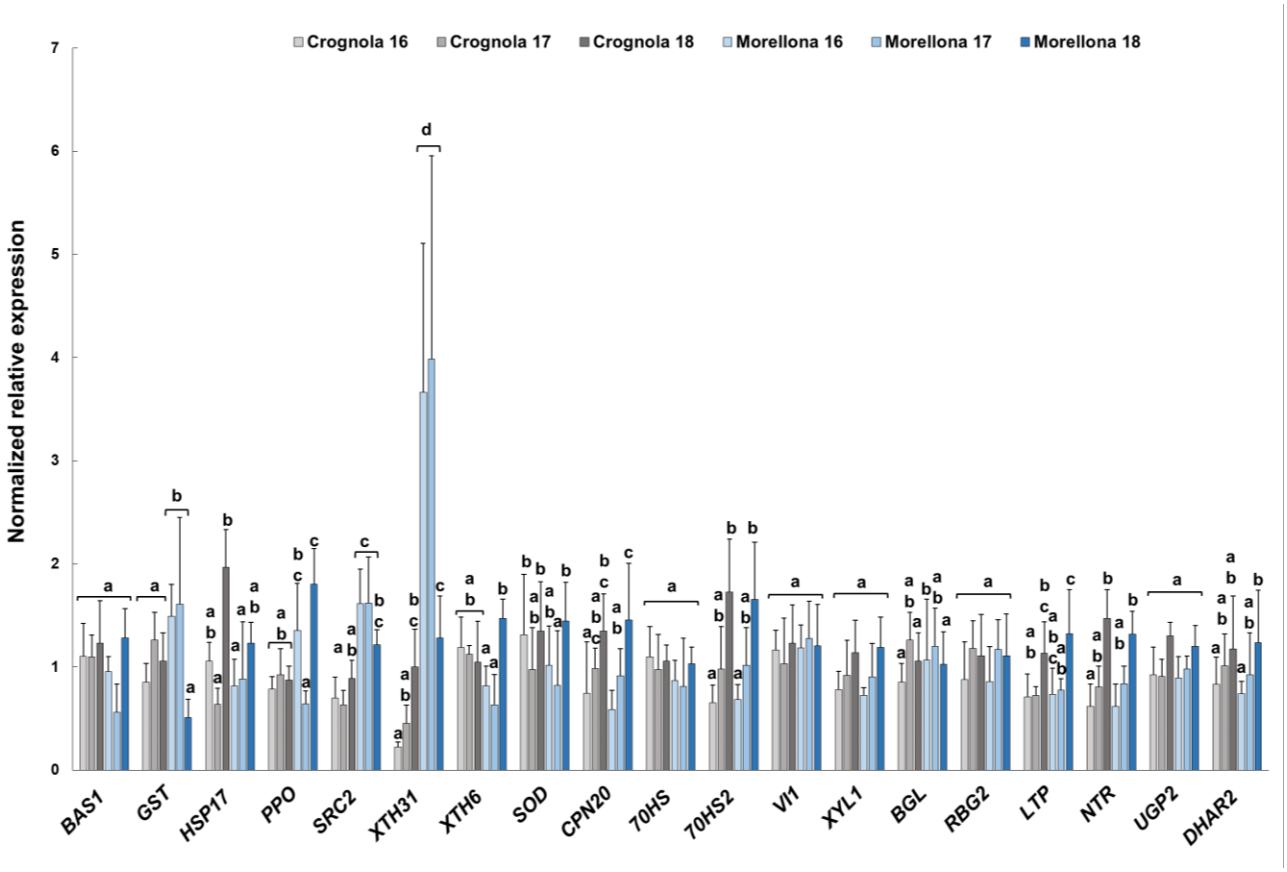


Figure 6. Relative expression (indicated as Normalized relative expression) of some genes coding for differentially abundant proteins in the two Tuscan varieties. Error bars refer to the standard deviation (n=4). Different letters represent the statistical significance (p<0.05) present among the groups of data obtained. If a letter is shared, the difference is not significant. A one-way ANOVA followed by Tukey's post-hoc test was performed on genes showing homogeneity and normal distribution; for the others, a Kruskal-Wallis test followed by Dunn's post-hoc test was used. The statistical parameters are: *BAS1* $X^2(5)=2.98$, p-value=0.702*; GST* $F(5,17)=4.16$, p-value=0.012; *HSP17* $F(5,17)=7.28$, p-value=0.001; *PPO* $F(5,17)=9.96$, p-value=0.000; *SRC2* $F(5,17)=12.88$, p-value=0.000; *XTH31* $F(5,17)=34.95$, p-value=0.000; *XTH6* $F(5,17)=0.3449$, p-value=0.846*; SOD* $F(5,17)=6.33$, p-value=0.002; *CPN20* $F(5,17)=7.07$, p-value=0.001; *70HS* $X^2(5)=1.067$, p-value=0.957; *70HS2* $F(5,17)=9.14$, p-value=0.000; *VI1* $F(5,17)=1.55$, p-value=0.225; *XYL1* $F(5,17)=1.300$, p-value=0.310*; BGL* $X^2(5)=14.27$, p-value=0.014; *RBG2* $X^2(5)=4.083$, p-value=0.537; *LTP* $F(5,17)=8.73$, p-

570 value=0.000; *NTR* F(5,17)=4.51, *p*-value=0.008; *UGP2* F(5,17)=2.49, *p*-value=0.072; *DHAR2*
571 F(5,17)=1.539, *p*-value=0.230.

572 **Conclusions**

573 This is the first study providing a multi-angle molecular analysis of non-commercial sweet
574 cherry fruits from Tuscany. From the results obtained with metabolomics, it emerges that the
575 Tuscan sweet cherries are interesting from a nutraceutical point of view. In particular, the
576 varieties Crognola and Morellona are the most valuable in terms of bioactive content and show
577 genetic features that distinguish them from all the others. Besides being rich sources of
578 flavonoids, the two varieties were found to produce higher amounts of the rare flavonolignan
579 cinchonain which has interesting health properties. The RT-qPCR analysis revealed what was
580 already shown for other fruits, i.e. that the expression of the genes differed among the varieties
581 and across the years and could not always explain the differences observed in the content of
582 secondary metabolites. Proteomics revealed differences between the two most interesting
583 varieties in terms of flavonoids' abundance, Morellona and Crognola.
584 Despite the differences detected across the years for gene expression, it is possible to resume
585 the key finding as follows: a) Morellona and Crognola showed the highest contents of phenolic
586 compounds, b) the higher production of metabolites in Morellona was accompanied by a high
587 expression of genes involved in the late phases of the PPP in 2016 and 2017 and c) the 2 top
588 producers of phenolics show different proteome signatures, i.e. stress-related proteins were
589 more abundant in Crognola, while Morellona showed higher abundance of proteins related with
590 primary metabolic functions, fruit maturation and cell wall remodelling.
591 These data open the way to future investigations aimed at studying the Tuscan fruits at different
592 developmental stages and/or at assessing the post-harvest stability of the cherries produced by
593 Morellona and Crognola. It may be that the two varieties show a distinct post-harvest behaviour
594 which makes one of them more suitable for a commercial valorization.

Materials and methods

Sample collection

The sampling of the 6 Tuscan sweet cherry varieties Morellona, Moscatella, Maggiola, Crognola, Carlotta and Benedetta was carried out in the morning (between 9:00-10:00 am) on May the 16th 2016, May 19th 2017 and May 18th 2018 (temperatures: min 12 °C-max 22 °C in 2016, min 13 °C-max 26 °C in 2017, min 8 °C-max 24 °C in 2018). Samples were analyzed in four biological replicates, each consisting of a pool of 4-6 fruits for a total of 20-24 fruits coming from at least 3 different trees per variety. Fruits were sampled ca. 60 dpa (days post anthesis) from the same trees for each of the years investigated. The experimental field coordinates, growth conditions and total number of trees for each variety have been previously reported[49]. All the phenotypical aspects of the genotypes are reported in the Tuscan germplasm website (http://germoplasma.regione.toscana.it/index.php?option=com_content&view=article&id=5&Itemid=110).

One variety, Benedetta, gave fruits only in 2016. The commercial variety Durone, here included for comparative purposes, was purchased at a local grocery shop in Siena. After removing the stem, the fruits were immersed in liquid nitrogen and stored at −80 °C.

Genotyping

DNA was extracted from fruits including the exocarp and the mesocarp. The samples were reduced to a fine powder with liquid nitrogen and DNA was extracted with the QIAGEN DNeasy Plant Mini Kit, following the manufacturer's instructions. Each sample extraction was repeated three times. After extraction, the concentration values were measured at the Nanodrop. Fourteen primer pairs were used (Table 5) and were taken from the literature[100–105]. The primers were previously tested on species belonging to the same family, i.e. *P. cerasus* (sour cherry),

620 *P. armeniaca* (apricot) and *P. persica* (peach). SSR markers were selected for their high
621 polymorphism (Table 5). PCR reactions were prepared in a final volume of 20 µL using
622 genomic DNA at 2ng/µL, Q5® Hot Start High-Fidelity 2X Master Mix and 5 µM of primers
623 forward and reverse.
624 The PCR parameters were as follows: initial cycle at 98 ºC for 30 sec, 35 cycles at 98 ºC for
625 10 sec, 1 min at 60 ºC and 30 sec at 72º C, final extension at 72 ºC for 2 min.
626 PCR products were first visualized on agarose gels. Then, they were multiplexed, diluted in
627 double distilled water (1:50 v/v) and futher analyzed on an ABI3500 Genetic Analyzer (Life
628 Technologies, Waltham, MA, United States). Subsequently, fragment analysis was carried out
629 using the Genemapper 5.0 software. The sweet cherry allelic profiles obtained by genotyping
630 were used to generate a phylogenetic tree using an Unweighted Pair-Group Method with
631 Arithmetic mean (UPGMA) and Sequential Agglomerative Hierarchal Nested (SAHN) cluster
632 analysis with the NTSYSpc software version 2.2 (Exeter Software, USA).
633
634 **Table 5. Sequences of the 14 primer pairs used for genotyping and relative details.**
635 Fluorophore, size range, melting temperature, number of alleles and references are detailed.

Primer name	5' labelling	Sequence (5'→3')	Size range (bp)	T_m (°C)	Allele N°	Reference
UDP98411 Fwd	Dragonfly Orange	AAGCCATCCACTCAGCACTC	155–179	57	4	[100,101]
UDP98411 Rev		CCAAAAACCAAAACCAAAGG				
UDP98412 Fwd	FAM	AGGGAAAGTTTCTGCTGCAC	110-124	57	4	[100]
UDP98412 Rev		GCTGAAGACGACGATGATGA				
UDAp420 Fwd	FAM	TTCCTTGCTTCCCTTCATTG	166-172	56	3	[102]
UDAp420 Rev		CCCAGAACTTGATTCTGACC				
BPPCT039 Fwd	Dragonfly Orange	ATTACGTACCCTAAAGCTTCTGC	135-150	55	4	[103]
BPPCT039 Rev		GATGTCATGAAGATTGGAGAGG				
AMPA101 Fwd	FAM	CAGTTTGATTTGTGTGCCTCTC	184-192	56	4	[104]
AMPA101 Rev		GATCCACCCTTTGCATAAAATC				
UDAp-414 Fwd	HEX	CAAGCACAAGCGAACAAAAT	142-146	56	3	[102]
UDAp-414 Rev		GGTGGTTTCTTATCCGATG				
UDAp-415 Fwd	Dragonfly Orange	AACTGATGAGAAGGGGCTTG	157-161	56	2	[102]
UDAp-415 Rev		ACTCCCGACATTTGTGCTTC				
UDP96008 Fwd	Dragonfly Orange	TTGTACACACCCTCAGCCTG	148-152	60	3	[100,101]
UDP96008 Rev		TGCTGAGGTTCAGGTGAGTG				
BPPCT034 Fwd	Dragonfly Orange	CTACCTGAAATAAGCAGAGCCAT	221-255	57	7	[103]
BPPCT034 Rev		CAATGGAGAATGGGGTGC				
BPPCT040 Fwd	Dragonfly Orange	ATGAGGACGTGTCTGAATGG	133-148	57	5	[103]
BPPCT040 Rev		AGCCAAACCCCTCTTATACG				
EMPaJ15 Fwd	FAM	TTTTGGTCAATCTGCTGCTG	216-253	60	6	[105]
EMPaJ15 Rev		CTCTCATCTTCCCCCTCCTC				
EMPaS11 Fwd	HEX	ACCACMGAGGAACTTGGG	59-103	60	5	[105]
EMPaS11 Rev		CTGCCTGGAAGAGCAATAAC				
EMPaS12 Fwd	FAM	TGTGCTAATGCCAAAAATACC	135-145	60	4	[105]
EMPaS12 Rev		ACATGCATTTCAACCCACTC				
UCD-CH17 Fwd	HEX	TGGACTTCACTCATTTCAGAGA	188-212	60	6	[105]
UCD-CH17 Rev		ACTGCAGAGAATTTCCACAACCA				

636

Sample preparation for untargeted metabolomics

Whole fruits without the stones were ground using a mortar and a pestle in liquid nitrogen and approximately 500 mg of frozen powders were aliquoted and lyophilized in a freeze-dryer (Christ, Osterode, Germany). Lyophilized samples were then accurately weighed (approximately 15 mg) and stored at -80 °C until extraction. Before adding the extraction solvent, 2 µL of chloramphenicol (5mg/mL, Sigma-Aldrich), used as internal standard, were put directly on the powders and then 998 µL of the extraction solvent (MeOH:H_2O 80%; v/v) were added to the samples. The samples were vortexed thoroughly and then shaken in a Thermomixer (Eppendorf, Hamburg, Germany) at 1400 rpm for 4 hours at 21 °C. The samples were vortexed again and centrifuged for 30 min at 20000 g at 4 °C. The supernatants (750 µL) were collected and completely evaporated using a CentriVap Vacuum Concentrator (LABCONCO, Kansas City, MO, US). Finally, the samples were resuspended in 188 µL of MeOH:H_2O 5% (v/v) with 0.1% of formic acid (FA) and filtered through 0.22 µm polytetrafluoroethylene (PTFE) filters (Merck Millipore. Darmstadt. Germany). The harvest years 2016 and 2017 were analyzed in 2018 and the 2018 harvest year was analyzed in 2019, with the same protocol and analytical conditions.

Untargeted metabolomics analysis with UHPLC-DAD-HR-MS/MS

The separation of molecules was achieved using an Acquity UPLC I-class UHPLC (Waters, Milford, MA, US) with a PDA detector, coupled to a hybrid quadrupole-time of flight (Q-TOF) mass spectrometer TripleTOF 6600 (SCIEX Instruments, Concord, ON, Canada) with a DuoSpray Ion Source operating in negative and positive ion mode. Five µL of the samples in random order were injected in the UHPLC system and analyzed in a run time of 60 min. The separation was performed on a reverse-phase Acquity UPLC BEH C18 column (2.1 × 100 mm, 1.7-µm particle size) (Waters, Milford, MA, US). The solvents used were A) water + 0.1% FA and B) acetonitrile (ACN) + 0.1% FA, all LC-MS grade and the column was maintained at 50 °C during all the run time. The gradient

was as follows: 0 min, 1% B; 4 min, 1% B; 16 min, 5% B; 35 min, 40% B; 45 min, 100% B; 50 min, 100% B; 54 min, 1% B; 60 min, 1% B, at 0.5 mL/min flow rate. UV-visible spectra were also acquired between 190 and 800 nm at a rate of 10 points/sec.

The ESI parameters were set as follows: source temperature of 650 °C, ion spray voltage of −4500 V and 4500 V, for the negative and positive mode respectively, curtain gas (nitrogen) of 30 psi, nebulizer gas (air) of 55 psi and turbine gas (air) of 50 psi. The declustering potential was set at −60 eV in negative mode and 60 eV in positive mode. The precursor charge state selection was set at 1. For information-dependent acquisition in high sensitivity mode, survey scans were acquired in 175 msec and the 10 most abundant product ion scans were collected during 200 msec if exceeding a threshold of 100 counts/sec, the total cycle time being 2.225 sec. A sweeping collision energy setting of 15 eV below and above 15 eV was applied to all precursor ions. The dynamic exclusion was set for two sec after three occurrences before the precursor could be fragmented again. For MS1, full HR-MS spectra between 100 and 2000 mass-to-charge ratio (*m/z*) were recorded. MS2 scans were recorded between 50 and 2000 *m/z*, in profile mode.

Data analysis

The software Progenesis QI (v2.3.6275.47962, Nonlinear Dynamics, Waters, Newcastle, UK) was used to generate a list of potentially differentially abundant compounds for each ionization mode (a compound in Progenesis QI is a combination of retention time and *m/z* ratio deduced from isotopes and adducts ions) for the data corresponding to 2016 and 2017. All possible adducts and automatic processing were selected, with default automatic sensitivity and no chromatographic peak width indicated.

For the positive ionization mode, the number of compounds for the 3 years of harvest was 14776, 2583 of which had experimental MS2 data. Out of these 2583 compounds, 206 respected the statistical criteria fixed. Progenesis QI gave an identification proposal for 197 compounds, but 15 metabolites were in the end identified.

For the negative ionization mode, 19295 compounds were obtained with Progenesis QI, 4350 of which had experimental MS2 data. Out of these 4350 compounds, 446 respected the statistical criteria fixed. Progenesis gave an identification proposal for 396 compounds, but 14 metabolites were in the end identified. There was no signal-to-noise selection, except for the parameter "default" for the sensitivity of peak picking in Progenesis QI.

Alignment and peak picking were done with default parameters, then all adducts between 0 and 50 min and only compounds with available MS2 data were kept for further statistical analysis. Once the statistics had been performed with R, we also took advantage of Progenesis QI plugins to tentatively identify compounds in the databases Pubchem, MassBank, NIST, ChemSpider and ChEBI, as well as in an in-house database.

R[106] (v3.6.0 64-bit) was used to normalize the abundances using internal standard and dry weight and to perform a one-way analysis of variance (ANOVA) with genotype as factor on the abundances of the compounds in the first two harvests data to establish a list of compounds to search in databases. The criteria of compounds' choice were p-value ANOVA <0.05 and maximum fold-change >3.

PeakView (v 1.2.0.3, SCIEX, Concord. ON, Canada) and database Metlin in addition to the other databases available in Progenesis QI were used to perform manual identification checking.

Annotations and identifications were classified in accordance with the levels of Metabolomics Standards Initiative (MSI)[107]. Compounds in class 1 were identified by comparison with standards analyzed in the same analytical conditions, based on exact mass, retention time, MS2 fragmentation pattern and UV-visible spectrum, compounds in class 2 were identified based on the same criteria by comparison with data in databases and/or literature. Class 3 was assigned to compounds with the same information as class 2 when they allowed only chemical class determination, typically when the molecule identified is a fragment of a bigger not fully determined molecule.

Calculation of fold-changes and average abundance per genotype were performed again on the 3 harvests for the already putatively identified molecules when harvest 2018 data were available.

The hierarchical clustering of the heatmaps relative to metabolomics, as well as gene expression and proteomics were obtained with Cluster 3.0 (available at: http://bonsai.hgc.jp/~mdehoon/software/cluster/software.htm) and Java TreeView (available at: http://jtreeview.sourceforge.net/).

Raw data were deposited in the repository MetaboLights, under the study number MTBLS1803 (http://www.ebi.ac.uk/metabolights/).

Bioinformatics and primer design

The genes of interest were obtained by blasting the thale cress protein sequences in NCBI, as well as by querying the Genome Database for Rosaceae (GDR; available at https://www.rosaceae.org/) and the cherry database (available at http://cherry.kazusa.or.jp). Multiple alignments were performed in CLUSTAL-Ω (http://www.ebi.ac.uk/Tools/msa/clustalo/). The primers were designed with Primer3Plus (http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi) and checked with OligoAnalyzer 3.1 (http://eu.idtdna.com/calc/analyzer). All the primers and their relative features were previously described[5], or are reported in Table 6. The maximum likelihood phylogenetic tree of XTHs was constructed from full-length protein sequences from sweet cherry and other species, namely poplar, tomato, thale cress, nasturtium[74]. The sequences were aligned with CLUSTAL-Ω and the tree obtained by using IQ-TREE web server (http://iqtree.cibiv.univie.ac.at/) with the "Auto" parameter to identify the best-fit substitution model[108]. The tree was rooted with *Bacillus licheniformis* lichenase (accession n° CAA40547).

Table 6. List of primers used for gene expression analysis. Details relative to the sequences of the target genes, together with primer efficiency %, melting temperature, amplicon sizes and accession numbers are provided.

Name	Primer sequence (5' →3')	Tm (° C)	Efficiency (%)	Amplicon size (bp)	Accession number
Phenylpropanoid pathway					
PavPAL2 Fwd	CTGCGAGGGAAAGATTATCG	83.9	92.04	114	XM_021948624
PavPAL2 Rev	AGTGGAATGGAATGCAGCAC				
PvPAL4 Fwd	AGCCTCTTCCTTTCCCATTC	79.1	95.88	155	XM_021971014
PvPAL4 Rev	AATGCCAAACTTGACGAACC				
Pav4CH Fwd	TCCGCATTTTTCCTCTGC	84.8	88.32	111	GU990522.1
Pav4CH Rev	ATGATGGCGATGAAGAGACC				
Pav4CL2 Fwd	GTTGCGATGCCGTATTCTTC	85.7	100.45	105	XM_021954366.1
Pav4CL2 Red	TTCTCCCCATCCACTTGTTG				
Pav4CL5 Fwd	TGATGGTGAGGAAGGAAAGG	84.2	109.59	147	XM_021954558.1
Pav4CL5 Rev	TCAGATTCTTGTGCGACGAC				
PavCHS2 Fwd	GTACCAACAAGGCTGTTTTGC	88.1	93.28	146	KP347499.1
PavCHS2 Rev	TGTCAAGGTGGGTATCACTGG				
PavCHI3 Fwd	ATAGATTGGCAGCCGATGAC	80.9	91.46	143	XM_021945901.1
PavCHI3 Rev	AATCTCAGCAGTGGCAGAAG				
PavCHI Fwd	TTTCCACCGTCAGTCAAACC	85.7	96.42	102	KP347511.1
PavCHI Rev	TCACGAAGTTCCCCTGAATC				
PavF3H Fwd	GAAGATTGTGGAGGCTTGTG	78.8	87.1	72	XM_021960102.1
PavF3H Rev	ATGAGCTTGGCATCAACTCC				
PavDFR Fwd	GCCCATTTCTCATGTCATCC	81.9	86.6	116	XM_021975874.1
PavDFR Rev	TCGTCCAAGTGAACGAACTG				
PavANS Fwd	ATGGGCAGTTTTCTGTGAGC	83.27	88.78	100	XM_021947877.1
PavANS Rev	GTTCTTGGTGGGAAGATTGG				
PavUGT Fwd	ACAACTTGGGCACCTCAAAC	81.26	93.55	80	XM_021947368.1
PavUGT Rev	AGTGAACTCCAACCGCAATG				
PavPPO Fwd	ATGCGAGCCTTACCAGATG	86.07	110.51	104	XM_021975090.1
PavPPO Rev	AGATCCGAATACCCGACTTG				
Cell wall					
PavBGL Fwd	AGAATGGCATGGACGAGTTC	79.64	92.17	97	XM_021950416.1
PavBGL Rev	TAACAGAGGTGGCGATAGCAG				
PavXTH31 Fwd	TCTCTGGTTTGACCCAACAC	78.84	98.12	95	XM_021956588.1
PavXTH31 Rev	TTCTTACGGGCACATCATCC				
PavXTH6 Fwd	GTGCGTGATGAGCTAGACTTTG	81.92	96.09	103	XM_021946402.1
PavXTH6 Rev	TTTGCTCCCTGTTACCCTTC				
PavXYL1 Fwd	CTTCCTCAACCCGAAAACTG	82.32	91.02	100	XM_021958725.1
PavXYL1 Rev	AGCCTCGTTCATGTCAATCC				
Stress response					
PavVI1 Fwd	CCTGCTTGGTTGGATCAATG	80.71	91.66	83	XM_021967777.1
PavVI1 Rev	TCCTTGGAATGGTCTGAAGG				
Pav70HS Fwd	TGATGAGGTGGAAAGGATGG	82.05	92.87	93	XM_021960163.1

Pav70HS Rev	TCAGCCTGGTTCTTTGTGTC				
Pav70HS2 Fwd	TGGGAGGAGAGGATTTTGAC	78.97	95.27	91	XM_021958066.1
Pav70HS2 Rev	CTTGGGTTTCCGATGATGTC				
PavUGP2 Fwd	TCGTCTCTCGTTATGTCAGTGG	81.38	96.85	104	XM_021968135.1
PavUGP2 Rev	AGGTGCCAAGCCATCATAAG				
PavSRC2 Fwd	TGACGTTAAGGTTACGATTAGGG	85.67	98.85	108	XM_021977305.1
PavSRC2 Rev	GGCGGAGCAGGATAATCTC				
PavLTP Fwd	TGTATTTACGGGACGGTGTG	80.45	94.21	108	XM_021964529.1
PavLTP Rev	CCACCCCATGAATTTCTCAC				
PavGST Fwd	TATTGGGAACAACCCTGGAG	81.65	96.09	108	XM_021968160.1
PavGST Rev	GCACCAGAAGTTGAAGTACCAG				
PavHSP17 Fwd	AGGACGACAATGTGCTTCTG	83.66	96.02	107	XM_021968740.1
PavHSP17 Rev	TAAACTTGCCGACTCTCCTCTC				
PavRBG2 Fwd	CGAGTCAAAGAAAGACCCAGAG	84.2	92.14	104	XM_021964717.1
PavRBG2 Rev	GATGCTCATGGAAGGCAAAC				
PavSOD Fwd	AGAAGCACCAGACTTACG	84.47	82.12	92	XM_021945900.1
PavSOD Rev	AACAACAGCAGCAGCATCAC				
PavCPN20 Fwd	TCAAGGTTGCTGAGGTTGAG	82.32	90.81	82	XM_021949327.1
PavCPN20 Rev	GTGCCAATCGAAGGTTTCTC				
PavNTR Fwd	GAGCAACCCGAAAATCAGAG	84.2	85.51	107	XM_021948862.1
PavNTR Rev	CCCCAGTCACCAAATTCTTC				
PavDHAR2 Fwd	CTCAGCGACAAACCCAAATG	82.59	104.4	97	XM_021958120.1
PavDHAR2 Rev	TCACGTCAGAATCAGCCAAC				
PavBAS1 Fwd	GTTTGCCCCACAGAAATCAC	82.32	90.81	106	XM_021961174.1
PavBAS1 Rev	CAAGGTGCGAAAACACACTG				
MYB transcription factors					
PavMYB10.1 Fwd	TTAGGTGACGAGGATGCTTT	81.12	107	132	KP455680.1
PavMYB10.1 Rev	TTAGTCCTTCTGAACATTGG				
PavMYB11 Fwd	TTGTCGAAGCAGGACATGAG	83.93	108	73	XM_021949436.1
PavMYB11 Rev	TCCTCCCACAACCAAGAAAG				

RNA extraction, reverse transcription and RT-qPCR

RNA extraction from whole fruits comprising the exocarp and the mesocarp (excluding the stones), purity/integrity measurement, cDNA synthesis and RT-qPCR were performed as previously described [49]. A melt curve analysis was performed at the end of the PCR cycles to check the specificity of the primers. The primer efficiencies were determined using a calibration curve consisting of a serial dilution of 6 points (10–2–0.4–0.08–0.016–0.0032 ng/µL).

744 The expression values were calculated with qBase^{PLUS} (version 3.2, Biogazelle, Ghent, Belgium)[109]
745 by using *PavACT7* and *PaveTIF4E* as reference genes. which were sufficient for normalization
746 according to geNORM[110].
747 The log10-transformed NRQ (Normalized Relative Quantities) results were analyzed with IBM SPSS
748 Statistics v20 (IBM SPSS, Chicago, IL, USA). Normal distribution of the data was checked with a
749 Shapiro-Wilk test and graphically with a Q-Q plot. Homogeneity was checked with the Homogeneity
750 of Variance Test. For data following normal distribution and homogeneous, a one-way ANOVA with
751 Tukey's post-hoc text was performed. For data not following normal distribution and/or not
752 homogeneous, a Kruskal-Wallis test was performed with Dunn's post-hoc test.
753
754 **Cloning and sequencing of *MYB10.1***
755 Genomic DNA was extracted from the fruits (devoid of the stones) of Crognola and Morellona using
756 the QIAGEN DNeasy Plant Mini Kit, as previously described.
757 PCRs were performed using 50 ng DNA and the Q5 Hot Start High-Fidelity 2X Master Mix,
758 following the manufacturer's instructions. The final volume of the reactions was 50 µL and the
759 primers were used at the final concentration of 0.5 µM. The PCR program consisted of an initial
760 denaturation at 98 °C for 1 minute, followed by 35 cycles at 98 °C for 10 sec, 62 °C for 30 sec and
761 72 °C for 1 min, then a final extension at 72 °C for 2 min was performed and the reaction kept at
762 4 °C. The following primers were used: MYB10.1 Fwd ATGGAGGGCTATAACTTGGGTG
763 MYB10.1 Rev TTAGTCCTTCTGAACATTGGTACA. PCR products were run on a 2% agarose gel,
764 then they were purified using the PCR purification kit from QIAGEN, following the manufacturer's
765 instructions. The eluted products were ligated into the pGEM-T Easy vector, according to the
766 manufacturer's recommendations and cloned into JM109 chemically competent cells. Three positive
767 clones for each variety were grown o/n at 37 °C in LB medium supplemented with ampicillin 100
768 µg/ml. The following day, plasmids were extracted with the QIAGEN plasmid miniprep kit and
769 sequenced on an Applied Biosystems 3500 Genetic Analyzer using the BigDye Terminator v3.1

Cycle Sequencing and the BigDye XTerminator Purification kits, according to the manufacturer's instructions.

Protein extraction

Whole cherry fruits comprising the exocarp and the mesocarp (excluding the stones) were ground in liquid nitrogen with a mortar and a pestle. The fine powder (1 g) was resuspended in 1 mL of cold acetone with 10% of trichloroacetic acid (TCA) and 0.07% of dithiothreitol (DTT)[111]. The samples were vortexed and left at -20 °C for 60 min, then centrifuged 5 min at 10000 g. The pellets thus obtained were washed twice with cold acetone and, then, dried at room temperature overnight. The dried pellets were resuspended in 0.8 mL of phenol (Tris-buffer, pH 8.0) and 0.8 mL of SDS buffer [30% (w/v) sucrose, 2% (v/v) SDS, 0.1 M Tris-HCl pH 8.0, 5% (v/v) 2-mercaptoethanol]. The mixtures were thoroughly vortexed and centrifuged for 3 min at 10000 g. The upper phase (300 µL) was transferred in a new 2 mL-tube, diluted in 5 volumes of cold ammonium acetate ($NH_4CH_3CO_2$) in MeOH and left at -20 °C for 30 min. The precipitated samples were washed twice with the same solution, removing the supernatants each time. Finally, the samples were washed with 80% (v/v) acetone 2 times and the pellets were dried. The dried pellets were dissolved in a buffer of urea 7M, thiourea 2M, Tris 30 mM and CHAPS 4% (w/v). The extracted proteins were quantified using the Bradford method[112], using BSA for the standard curve.

2D-DIGE

A volume of sample equivalent to 50 µg of proteins was labelled for DIGE analysis. The biological replicates of each sample were split and marked half with CyDye 3 fluorochrome and half with CyDye 5. The CyDye 2 fluorochrome was added to the internal standard, which is a mixture of all samples in equal amount. The labelling was done by addition of 400 pmol of dye, followed by a 30 min incubation on ice in the dark. Then, 1 µL of lysine 10 mM was added to stop the reaction and the samples were incubated 10 more min is the same conditions. The samples were combined as follows:

1 Cy3-labelled, 1 Cy5 labelled and 1 internal standard. They were then loaded on strips (pH 3-10 non-linear, 24 cm) for the first dimension, using the passive rehydration method. Nine µL of ampholytes and 2.7 µL of destreak reagent were added to 450 µL of sample in buffer solution [urea 7M, thiourea 2M, Tris 30 mM and CHAPS 0.5% (w/v)]. The strips have been rehydrated with the samples overnight. The isoelectric focusing (IEF) was performed with an Ettan IPGphor 3 system (GE Healthcare). A gradual increase of the voltage was used to reach a total of ca. 90000 V h within 25 h, through 5 steps planned as follows: 0-3 h 100 V, 3-7 h ramping to 1000 V, 7-14 h 1000 V, 14-20 h ramping to 10000 V and 20-25 h 10000 V. The second dimension was run on precast 12% flatbed gels (25x20 cm) in a horizontal electrophoresis tower (HPE FlatTOP Tower, Serva) following the manufacturer's instructions. The 2D gels were scanned using a laser scanner (Typhoon FLA 9500, GE Healthcare) and consequently analyzed with the software SameSpots (http://totallab.com/home/samespots/).

Spot picking and mass spectrometry

Spots of interest were selected using the SameSpots program using two filters, *p*-value <0.01 and max fold change >2 (a total of 166 proteins was obtained). The gel spots were picked with an Ettan Spot Picker (GE Healthcare) and trypsinized using an EVO2 workstation (Tecan). The dried samples were solubilized in 0.7 µL of an α-cyano-4-hydroxycinnamate solution (7 mg/mL in 50% acetonitrile and 0.1% trifluoroacetic acid) and spotted onto a MALDI plate. A MALDI mass spectrum was acquired using the SCIEX 5800 TOF/TOF (Sciex). The 10 most intense peaks, excluding known contaminants, were automatically selected and fragmented. MS and MS2 were submitted to an in-house MASCOT server (version 2.6.1; Matrix Science, www.matrixscience.com) for database-dependent identifications against the NCBI non-redundant protein sequence database (NCBInr) limited to the taxonomy *P. avium* (taxID4229; 10 July 2019; 35758 sequences). The parameters were as follows: peptide mass tolerance 100 ppm, fragment mass tolerance 0.5 Da, cysteine carbamidomethylation as fixed modification (alkylation was performed during the equilibration step

between IEF and second dimension) and methionine or tryptophan oxidation, double oxidation of tryptophan and tryptophan to kynurenine as variable modifications. Kynurenine, resulting from tryptophan oxidation, is an artefact often observed during automatic digestion in the laboratory where the analysis was performed (Luxembourg Institute of Science and Technology-LIST). Up to two miscleavages were allowed. All identifications were manually validated.

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD019468 (that can be accessed with the following details: Username: reviewer47554@ebi.ac.uk and Password: XxL99W9a).

Acknowledgments

During the acquisition of the data presented in this study, R.B. was in receipt of the PhD fellowship "Pegaso" financed by the Region Tuscany. Aude Corvisy and Laurent Solinhac are gratefully acknowledged for technical help in sequencing and genotyping.

Authors' contributions

R.B.: investigation, formal analysis, writing-original draft. S.C.: methodology, formal analysis, data curation, writing-review & editing. S.P.: methodology, formal analysis, data curation, writing-review & editing, S.L.: methodology, data curation, writing-review & editing. M.R.: project administration, validation, supervision, writing-review & editing. C.C.: resources, project administration, validation, supervision, writing-review & editing. G.C.: resources, project administration, validation, supervision, writing-review & editing. J-F.H.: resources, writing-review & editing. J.R.: resources, writing-review & editing. G.G.: conceptualization, investigation, formal analysis, writing-original draft.

Competing financial interests

The authors declare no competing financial interests.

References

1 Del Cueto J, Ionescu IA, Pičmanová M *et al.* Cyanogenic Glucosides and Derivatives in Almond and Sweet Cherry Flower Buds from Dormancy to Flowering. *Front Plant Sci* 2017; **8**: 800.

2 Püssa T. *Principles of Food Toxicology*. CRC Press, 2007.

3 Berni R, Romi M, Cantini C, Hausman J-F, Guerriero G, Cai G. Functional molecules in locally-adapted crops: the case study of tomatoes, onions and sweet cherry fruits from Tuscany in Italy. *Frontiers in Plant Science* 2018; **9**: 1983.

4 Ballistreri G, Continella A, Gentile A, Amenta M, Fabroni S, Rapisarda P. Fruit quality and bioactive compounds relevant to human health of sweet cherry (*Prunus avium* L.) cultivars grown in Italy. *Food Chemistry* 2013; **140**: 630–638.

5 Berni R, Hoque MZ, Legay S *et al.* Tuscan Varieties of Sweet Cherry Are Rich Sources of Ursolic and Oleanolic Acid: Protein Modeling Coupled to Targeted Gene Expression and Metabolite Analyses. *Molecules* 2019; **24**: 1590.

6 Szakiel A, Pączkowski C, Pensec F, Bertsch C. Fruit cuticular waxes as a source of biologically active triterpenoids. *Phytochem Rev* 2012; **11**: 263–284.

7 Buschhaus C, Jetter R. Composition differences between epicuticular and intracuticular wax substructures: how do plants seal their epidermal surfaces? *J Exp Bot* 2011; **62**: 841–853.

8 Quero-García J, Iezzoni A, Pulawska J, Lang GA. *Cherries: Botany, Production and Uses*. CABI, 2017.

9 Lugli S, Musacchi S, Grandi M, Bassi G, Franchini S, Zago M. The sweet cherry production in northern Italy: innovative rootstocks and emerging high-density plantings. *Inovacije u voćarstvu*

III savetovanje, Tema Savetovanja: Unapređenje proizvodnje trešnje i višnje, Beograd, Srbija, 10 februar 2011 godine Zbornik radova 2011; 75–92.

10 Tricase C, Rana R, Andriano AM, Ingrao C. An input flow analysis for improved environmental sustainability and management of cherry orchards: A case study in the Apulia region. *Journal of Cleaner Production* 2017; **156**: 766–774.

11 Taiti C, Caparrotta S, Mancuso S, Masi E. Morpho-chemical and aroma investigations on autochthonous and highly-prized sweet cherry varieties grown in Tuscany. Advances in Horticultural Science. 2017; **31**: 121–129.

12 Girelli CR, Del Coco L, Zelasco S *et al.* Traceability of "Tuscan PGI" Extra Virgin Olive Oils by 1H NMR Metabolic Profiles Collection and Analysis. *Metabolites* 2018; **8**: 60.

13 Di Matteo A, Russo R, Graziani G, Ritieni A, Di Vaio C. Characterization of autochthonous sweet cherry cultivars (*Prunus avium* L.) of southern Italy for fruit quality, bioactive compounds and antioxidant activity. *J Sci Food Agric* 2017; **97**: 2782–2794.

14 Marchese A, Giovannini D, Leone A *et al.* S-genotype identification, genetic diversity and structure analysis of Italian sweet cherry germplasm. *Tree Genetics & Genomes* 2017; **13**: 93.

15 Martini S, Conte A, Tagliazucchi D. Phenolic compounds profile and antioxidant properties of six sweet cherry (*Prunus avium*) cultivars. *Food Research International* 2017; **97**: 15–26.

16 Mathesius U. Flavonoid Functions in Plants and Their Interactions with Other Organisms. *Plants (Basel)* 2018; **7**: 30.

17 Falcone Ferreyra ML, Rius S, Casati P. Flavonoids: biosynthesis, biological functions, and biotechnological applications. *Front Plant Sci* 2012; **3**: 222.

18 Ramakrishna A, Ravishankar GA. Influence of abiotic stress signals on secondary metabolites in plants. *Plant Signal Behav* 2011; **6**: 1720–1731.

19 Kuhn BM, Geisler M, Bigler L, Ringli C. Flavonols Accumulate Asymmetrically and Affect Auxin Transport in *Arabidopsis*. *Plant Physiology* 2011; **156**: 585–595.

20 Carbone F, Preuss A, De Vos RCH *et al.* Developmental, genetic and environmental factors affect the expression of flavonoid genes, enzymes and metabolites in strawberry fruits. *Plant Cell Environ* 2009; **32**: 1117–1131.

21 Berni R, Cai G, Xu X, Hausman J-F, Guerriero G. Identification of Jasmonic Acid Biosynthetic Genes in Sweet Cherry and Expression Analysis in Four Ancient Varieties from Tuscany. *Int J Mol Sci* 2019; **20**: 3569.

22 Möller B, Herrmann K. Quinic acid esters of hydroxycinnamic acids in stone and pome fruit. *Phytochemistry* 1983; **22**: 477–481.

23 Sobeh M, Mahmoud MF, Sabry OM *et al.* HPLC-PDA-MS/MS Characterization of Bioactive Secondary Metabolites from *Turraea fischeri* Bark Extract and Its Antioxidant and Hepatoprotective Activities In Vivo. *Molecules* 2017; **22**: 2089.

24 Pizzolatti MG, Venson AF, Smânia A, Smânia E de FA, Braz-Filho R. Two epimeric flavalignans from *Trichilia catigua* (Meliaceae) with antimicrobial activity. *Z Naturforsch, C, J Biosci* 2002; **57**: 483–488.

25 Olszewska MA, Kolodziejczyk-Czepas J, Rutkowska M *et al.* The Effect of Standardised Flower Extracts of *Sorbus aucuparia* L. on Proinflammatory Enzymes, Multiple Oxidants, and Oxidative/Nitrative Damage of Human Plasma Components In Vitro. *Oxid Med Cell Longev* 2019; **2019**: 9746358.

26 Nemes A, Szőllősi E, Stündl L *et al.* Determination of Flavonoid and Proanthocyanidin Profile of Hungarian Sour Cherry. *Molecules* 2018; **23**: 3278.

27 Grzesik M, Naparło K, Bartosz G, Sadowska-Bartosz I. Antioxidant properties of catechins: Comparison with other antioxidants. *Food Chem* 2018; **241**: 480–492.

28 Qa'dan F, Verspohl EJ, Nahrstedt A, Petereit F, Matalka KZ. Cinchonain Ib isolated from *Eriobotrya japonica* induces insulin secretion in vitro and in vivo. *J Ethnopharmacol* 2009; **124**: 224–227.

29 Li H-J, Deinzer ML. Tandem Mass Spectrometry for Sequencing Proanthocyanidins. *Anal Chem* 2007; **79**: 1739–1748.

30 Senica M, Stampar F, Veberic R, Mikulic-Petkovsek M. Transition of phenolics and cyanogenic glycosides from apricot and cherry fruit kernels into liqueur. *Food Chem* 2016; **203**: 483–490.

31 Jaiswal R, Jayasinghe L, Kuhnert N. Identification and characterization of proanthocyanidins of 16 members of the *Rhododendron* genus (Ericaceae) by tandem LC-MS. *J Mass Spectrom* 2012; **47**: 502–515.

32 Sokół-Łętowska A, Kucharska AZ, Szumny A, Wińska K, Nawirska-Olszańska A. Phenolic Composition Stability and Antioxidant Activity of Sour Cherry Liqueurs. *Molecules* 2018; **23**: 2156.

33 Gu W-Y, Li N, Leung EL-H *et al.* Metabolites software-assisted flavonoid hunting in plants using ultra-high performance liquid chromatography-quadrupole-time of flight mass spectrometry. *Molecules* 2015; **20**: 3955–3971.

34 de Souza LM, Cipriani TR, Iacomini M, Gorin PAJ, Sassaki GL. HPLC/ESI-MS and NMR analysis of flavonoids and tannins in bioactive extract from leaves of *Maytenus ilicifolia*. *J Pharm Biomed Anal* 2008; **47**: 59–67.

35 Lv Q, Luo F, Zhao X *et al.* Identification of Proanthocyanidins from Litchi (*Litchi chinensis* Sonn.) Pulp by LC-ESI-Q-TOF-MS and Their Antioxidant Activity. *PLOS ONE* 2015; **10**: e0120480.

36 Kavitha P, Shivashankara KS, Rao VK, Sadashiva AT, Ravishankar KV, Sathish GJ. Genotypic variability for antioxidant and quality parameters among tomato cultivars, hybrids, cherry tomatoes and wild species. *J Sci Food Agric* 2014; **94**: 993–999.

37 Minoggio M, Bramati L, Simonetti P *et al.* Polyphenol pattern and antioxidant activity of different tomato lines and cultivars. *Ann Nutr Metab* 2003; **47**: 64–69.

38 Gómez JD, Vital CE, Oliveira MGA, Ramos HJO. Broad range flavonoid profiling by LC/MS of soybean genotypes contrasting for resistance to *Anticarsia gemmatalis* (Lepidoptera: Noctuidae). *PLOS ONE* 2018; **13**: e0205010.

39 Shamloo M, Babawale EA, Furtado A, Henry RJ, Eck PK, Jones PJH. Effects of genotype and temperature on accumulation of plant secondary metabolites in Canadian and Australian wheat grown under controlled environments. *Sci Rep* 2017; **7**: 9133.

40 Yang C-Q, Fang X, Wu X-M, Mao Y-B, Wang L-J, Chen X-Y. Transcriptional Regulation of Plant Secondary MetabolismF. *Journal of Integrative Plant Biology* 2012; **54**: 703–712.

41 Vogt T. Phenylpropanoid Biosynthesis. *Molecular Plant* 2010; **3**: 2–20.

42 Tohge T, Watanabe M, Hoefgen R, Fernie AR. The evolution of phenylpropanoid metabolism in the green lineage. *Critical Reviews in Biochemistry and Molecular Biology* 2013; **48**: 123–152.

43 Cochrane FC, Davin LB, Lewis NG. The Arabidopsis phenylalanine ammonia lyase gene family: kinetic characterization of the four PAL isoforms. *Phytochemistry* 2004; **65**: 1557–1564.

44 Huang J, Gu M, Lai Z *et al.* Functional analysis of the Arabidopsis PAL gene family in plant growth, development, and response to environmental stress. *Plant Physiology* 2010; **153**: 1526–1538.

45 Liu J, Osbourn A, Ma P. MYB Transcription Factors as Regulators of Phenylpropanoid Metabolism in Plants. *Mol Plant* 2015; **8**: 689–708.

46 Jin W, Wang H, Li M *et al.* The R2R3 MYB transcription factor PavMYB10.1 involves in anthocyanin biosynthesis and determines fruit skin colour in sweet cherry (*Prunus avium* L.). *Plant Biotechnol J* 2016; **14**: 2120–2133.

47 Starkevič P, Paukštytė J, Kazanavičiūtė V *et al.* Expression and Anthocyanin Biosynthesis-Modulating Potential of Sweet Cherry (*Prunus avium* L.) MYB10 and bHLH Genes. *PLOS ONE* 2015; **10**: e0126991.

48 Stracke R, Ishihara H, Huep G *et al.* Differential regulation of closely related R2R3-MYB transcription factors controls flavonol accumulation in different parts of the *Arabidopsis thaliana* seedling. *The Plant Journal* 2007; **50**: 660–677.

49 Berni R, Piasecki E, Legay S *et al.* Identification of the laccase-like multicopper oxidase gene family of sweet cherry (*Prunus avium* L.) and expression analysis in six ancient Tuscan varieties. *Sci Rep* 2019; **9**: 1–14.

50 Yu S, Kim H, Yun D-J, Suh MC, Lee B. Post-translational and transcriptional regulation of phenylpropanoid biosynthesis pathway by Kelch repeat F-box protein SAGL1. *Plant Mol Biol* 2019; **99**: 135–148.

51 Ncube B, Finnie JF, Van Staden J. Quality from the field: The impact of environmental factors as quality determinants in medicinal plants. *South African Journal of Botany* 2012; **82**: 11–20.

52 Perin EC, da Silva Messias R, Borowski JM *et al.* ABA-dependent salt and drought stress improve strawberry fruit quality. *Food Chem* 2019; **271**: 516–526.

53 Castellarin SD, Pfeiffer A, Sivilotti P, Degan M, Peterlunger E, DI Gaspero G. Transcriptional regulation of anthocyanin biosynthesis in ripening fruits of grapevine under seasonal water deficit. *Plant Cell Environ* 2007; **30**: 1381–1399.

54 Selles B, Jacquot J-P, Rouhier N. Comparative genomic study of protein disulfide isomerases from photosynthetic organisms. *Genomics* 2011; **97**: 37–50.

55 Berni R, Luyckx M, Xu X *et al.* Reactive Oxygen Species and heavy metal stress in plants: impact on the cell wall and secondary metabolism. *Environ Exp Bot* 2018; **161**: 98–106.

56 Tian S, Qin G, Li B. Reactive oxygen species involved in regulating fruit senescence and fungal pathogenicity. *Plant Mol Biol* 2013; **82**: 593–602.

57 Dixon DP, Edwards R. Glutathione Transferases. *Arabidopsis Book* 2010; **8**.

58 Gallie DR. The role of L-ascorbic acid recycling in responding to environmental stress and in promoting plant growth. *J Exp Bot* 2013; **64**: 433–443.

59 Hirai MY. A robust omics-based approach for the identification of glucosinolate biosynthetic genes. *Phytochem Rev* 2009; **8**: 15–23.

60 Bell L. The Biosynthesis of Glucosinolates: Insights, Inconsistencies, and Unknowns. In: *Annual Plant Reviews online*. American Cancer Society, 2019, pp 1–31.

61 Soundararajan P, Kim JS. Anti-Carcinogenic Glucosinolates in Cruciferous Vegetables and Their Antagonistic Effects on Prevention of Cancers. *Molecules* 2018; **23**: 2983.

62 Textor S, Gershenzon J. Herbivore induction of the glucosinolate–myrosinase defense system: major trends, biochemical bases and ecological significance. *Phytochem Rev* 2009; **8**: 149–170.

63 Dietz K-J, Jacob S, Oelze M-L *et al.* The function of peroxiredoxins in plant organelle redox metabolism. *J Exp Bot* 2006; **57**: 1697–1709.

64 Asada K. Production and Scavenging of Reactive Oxygen Species in Chloroplasts and Their Functions. *Plant Physiology* 2006; **141**: 391–396.

65 Nikkanen L, Rintamäki E. Thioredoxin-dependent regulatory networks in chloroplasts under fluctuating light conditions. *Philos Trans R Soc Lond, B, Biol Sci* 2014; **369**: 20130224.

66 Broin M, Cuiné S, Eymery F, Rey P. The Plastidic 2-Cysteine Peroxiredoxin Is a Target for a Thioredoxin Involved in the Protection of the Photosynthetic Apparatus against Oxidative Damage. *Plant Cell* 2002; **14**: 1417–1432.

67 Laskowski MJ, Dreher KA, Gehring MA, Abel S, Gensler AL, Sussex IM. FQR1, a Novel Primary Auxin-Response Gene, Encodes a Flavin Mononucleotide-Binding Quinone Reductase. *Plant Physiol* 2002; **128**: 578–590.

68 Kawarazaki T, Kimura S, Iizuka A *et al.* A low temperature-inducible protein AtSRC2 enhances the ROS-producing activity of NADPH oxidase AtRbohF. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* 2013; **1833**: 2775–2780.

69 Maskin L, Gudesblat GE, Moreno JE *et al.* Differential expression of the members of the Asr gene family in tomato (*Lycopersicon esculentum*). *Plant Science* 2001; **161**: 739–746.

70 Li J, Yu X, Lou Y *et al.* Proteomic analysis of the effects of gibberellin on increased fruit sink strength in Asian pear (*Pyrus pyrifolia*). *Scientia Horticulturae* 2015; **195**: 25–36.

71 Mangeon A, Junqueira RM, Sachetto-Martins G. Functional diversity of the plant glycine-rich proteins superfamily. *Plant Signaling & Behavior* 2010; **5**: 99–104.

72 Cho SM, Lee H, Jo H *et al.* Comparative transcriptome analysis of field- and chamber-grown samples of *Colobanthus quitensis* (Kunth) Bartl, an Antarctic flowering plant. *Sci Rep* 2018; **8**: 1–14.

73 Sun W, Van Montagu M, Verbruggen N. Small heat shock proteins and stress tolerance in plants. *Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression* 2002; **1577**: 1–9.

74 Baumann MJ, Eklöf JM, Michel G *et al.* Structural Evidence for the Evolution of Xyloglucanase Activity from Xyloglucan Endo-Transglycosylases: Biological Implications for Cell Wall Metabolism. *The Plant Cell* 2007; **19**: 1947–1963.

75 Nishitani K, Tominaga R. Endo-xyloglucan transferase, a novel class of glycosyltransferase that catalyzes transfer of a segment of xyloglucan molecule to another xyloglucan molecule. *J Biol Chem* 1992; **267**: 21058–21064.

76 Fry SC, Smith RC, Renwick KF, Martin DJ, Hodge SK, Matthews KJ. Xyloglucan endotransglycosylase, a new wall-loosening enzyme activity from plants. *Biochem J* 1992; **282**: 821–828.

77 Thompson JE, Fry SC. Restructuring of wall-bound xyloglucan by transglycosylation in living plant cells. *Plant J* 2001; **26**: 23–34.

78 Zhu XF, Shi YZ, Lei GJ *et al.* XTH31, encoding an in vitro XEH/XET-active enzyme, regulates aluminum sensitivity by modulating in vivo XET action, cell wall xyloglucan content, and aluminum binding capacity in *Arabidopsis*. *Plant Cell* 2012; **24**: 4731–4747.

79 Yang JL, Zhu XF, Peng YX *et al.* Cell Wall Hemicellulose Contributes Significantly to Aluminum Adsorption and Root Growth in *Arabidopsis*. *Plant Physiology* 2011; **155**: 1885–1892.

80 Kaewthai N, Gendre D, Eklöf JM *et al.* Group III-A XTH Genes of *Arabidopsis* Encode Predominant Xyloglucan Endohydrolases That Are Dispensable for Normal Growth. *Plant Physiol* 2013; **161**: 440–454.

81 Skalák J, Černý M, Jedelský P, et al. Stimulation of ipt overexpression as a tool to elucidate the role of cytokinins in high temperature responses of *Arabidopsis thaliana*. *J Exp Bot* 2016; **67**: 2861–73.

82 Clauw P, Coppens F, De Beuf K *et al.* Leaf Responses to Mild Drought Stress in Natural Variants of *Arabidopsis*. *Plant Physiol* 2015; **167**: 800–816.

83 Sampedro J, Sieiro C, Revilla G, González-Villa T, Zarra I. Cloning and Expression Pattern of a Gene Encoding an α-Xylosidase Active against Xyloglucan Oligosaccharides from *Arabidopsis*. *Plant Physiol* 2001; **126**: 910–920.

84 Günl M, Pauly M. AXY3 encodes a α-xylosidase that impacts the structure and accessibility of the hemicellulose xyloglucan in *Arabidopsis* plant cell walls. *Planta* 2011; **233**: 707–719.

85 Meng M, Geisler M, Johansson H *et al.* UDP-glucose pyrophosphorylase is not rate limiting, but is essential in *Arabidopsis*. *Plant Cell Physiol* 2009; **50**: 998–1011.

86 Roepke J, Gordon HOW, Neil KJA *et al.* An Apoplastic β-Glucosidase is Essential for the Degradation of Flavonol 3-O-β-Glucoside-7-O-α-Rhamnosides in *Arabidopsis*. *Plant Cell Physiol* 2017; **58**: 1030–1047.

87 Roepke J, Bozzo GG. *Arabidopsis thaliana* β-glucosidase BGLU15 attacks flavonol 3-*O*-β-glucoside-7-*O*-α-rhamnosides. *Phytochemistry* 2015;**109**:14–24.

88 Zhou S, Lou Y-R, Tzin V, Jander G. Alteration of Plant Primary Metabolism in Response to Insect Herbivory. *Plant Physiol* 2015; **169**: 1488–1498.

89 Mundim FM, Pringle EG. Whole-Plant Metabolic Allocation Under Water Stress. *Front Plant Sci* 2018; **9**: 852.

90 Esposito S, Guerriero G, Vona V, Di Martino Rigano V, Carfagna S, Rigano C. Glutamate synthase activities and protein changes in relation to nitrogen nutrition in barley: the dependence on different plastidic glucose-6P dehydrogenase isoforms. *J Exp Bot* 2005; **56**: 55–64.

91 Cardi M, Castiglia D, Ferrara M, Guerriero G, Chiurazzi M, Esposito S. The effects of salt stress cause a diversion of basal metabolism in barley roots: possible different roles for glucose-6-phosphate dehydrogenase isoforms. *Plant Physiol Biochem* 2015; **86**: 44–54.

92 Walker RP, Battistelli A, Moscatello S, Chen Z-H, Leegood RC, Famiani F. Phosphoenolpyruvate carboxykinase in cherry (*Prunus avium* L.) fruit during development. *J Exp Bot* 2011; **62**: 5357–5365.

93 Wilkinson JQ, Lanahan MB, Conner TW, Klee HJ. Identification of mRNAs with enhanced expression in ripening strawberry fruit using polymerase chain reaction differential display. *Plant Mol Biol* 1995; **27**: 1097–1108.

94 Araújo WL, Tohge T, Ishizaki K, Leaver CJ, Fernie AR. Protein degradation – an alternative respiratory substrate for stressed plants. *Trends in Plant Science* 2011; **16**: 489–498.

95 War AR, Paulraj MG, Ahmad T *et al.* Mechanisms of plant defense against insect herbivores. *Plant Signal Behav* 2012; **7**: 1306–1320.

96 Boeckx T, Winters AL, Webb KJ, Kingston-Smith AH. Polyphenol oxidase in leaves: is there any significance to the chloroplastic localization? *J Exp Bot* 2015; **66**: 3571–3579.

97 Hu H, Liu Y, Shi G-L *et al.* Proteomic analysis of peach endocarp and mesocarp during early fruit development. *Physiologia Plantarum* 2011; **142**: 390–406.

98 Briat J-F, Ravet K, Arnaud N *et al.* New insights into ferritin synthesis and function highlight a link between iron homeostasis and oxidative stress in plants. *Ann Bot* 2010; **105**: 811–822.

99 Stare T, Stare K, Weckwerth W, Wienkoop S, Gruden K. Comparison between Proteome and Transcriptome Response in Potato (*Solanum tuberosum* L.) Leaves Following Potato Virus Y (PVY) Infection. *Proteomes* 2017; **5**: 14.

100 Testolin R, Marrazzo T, Cipriani G *et al.* Microsatellite DNA in peach (*Prunus persica* L. Batsch) and its use in fingerprinting and testing the genetic origin of cultivars. *Genome* 2000; **43**: 512–520.

101 Cipriani G, Lot G, Huang W-G, Marrazzo MT, Peterlunger E, Testolin R. AC/GT and AG/CT microsatellite repeats in peach [*Prunus persica* (L) Batsch]: isolation, characterisation and cross-species amplification in Prunus. *Theor Appl Genet* 1999; **99**: 65–72.

102 Messina R, Lain O, Marrazzo MT, Cipriani G, Testolin R. New set of microsatellite loci isolated in apricot. *Molecular Ecology Notes* 2004; **4**: 432–434.

103 Dirlewanger E, Cosson P, Tavaud M *et al.* Development of microsatellite markers in peach [*Prunus persica* (L.) Batsch] and their use in genetic diversity analysis in peach and sweet cherry (*Prunus avium* L.). *Theor Appl Genet* 2002; **105**: 127–138.

104 Hagen LS, Chaib J, Fady B *et al.* Genomic and cDNA microsatellites from apricot (*Prunus armeniaca* L.). *Molecular Ecology Notes* 2004; **4**: 742–745.

105 Xuan H, Wang R, Büchele M, Möller O, Hartmann W. Microsatellite markers (SSR) as a tool to assist in identification of sweet (*Prunus avium*) and sour cherry (*Prunus cerasus*). *Acta Hortic* 2009; **839**: 507–514.

106 R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. .

107 Sumner LW, Amberg A, Barrett D *et al.* Proposed minimum reporting standards for chemical analysis Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI). *Metabolomics* 2007; **3**: 211–221.

108 Trifinopoulos J, Nguyen L-T, von Haeseler A, Minh BQ. W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. *Nucleic Acids Res* 2016; **44**: W232–W235.

109 Hellemans J, Mortier G, De Paepe A, Speleman F, Vandesompele J. qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data. *Genome Biology* 2007; **8**: R19.

110 Vandesompele J, De Preter K, Pattyn F *et al.* Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biology* 2002; **3**: research0034.1.

111 Wu X, Xiong E, Wang W, Scali M, Cresti M. Universal sample preparation method integrating trichloroacetic acid/acetone precipitation with phenol extraction for crop proteomic analysis. *Nat Protoc* 2014; **9**: 362–374.

112 Bradford MM. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Analytical Biochemistry* 1976; **72**: 248–254.

Figure legends

Figure 1. Dendrogram derived from the genotyping assay (UPGMA method) using SSR markers specific for genomic regions with a high coefficient of polymorphism. The relationships between the Tuscan cherries and the commercial ones from Luxembourg (Busch), Turkey, France (Napoleon and Dauphinois) and Italy (Durone) are shown. Nei & Li's similarity coefficients are displayed in the black bar below the tree. Bootstrap values are indicated above the branches (1000 replicates).

Figure 2. Heat map hierarchical clustering showing the fold change differences of the compounds identified. (A) Metabolites identified in positive and (B) in negative mode in the three years. A maximum fold change>3 in absolute value was used, together with a *p*-value<0.05 at the one-way ANOVA. Fold-changes were calculated using the means of normalized abundances. To build the heatmap, the fold change values were rescaled based on the lowest value detected per single metabolite and then log10-transformed. Numbers indicate the Pearson correlation coefficients. The color bar indicates the log10-transformed fold change values.

Figure 3. Heat map hierarchical clustering of the PPP-related gene expression data across the 3 years of study (A, 2016; B, 2017; C, 2018). Numbers indicate the Pearson correlation coefficients. The color bar indicates the log10-transformed normalized relative quantities. The bar graphs of the expression data are available in Supplementary Figure 2.

Figure 4. Gene expression analysis (indicated as Normalized relative expression) of the MYB TF-encoding genes. (A) results relative to 2016, (B) results obtained on the samples harvested in 2017, (C) results relative to the year 2018. Error bars correspond to the standard deviation (n=4). Different letters represent statistical significance (p-value<0.05) among the groups of data. If a letter is shared, the difference is not significant. A one-way ANOVA followed by Tukey's post-hoc test was performed on genes showing homogeneity and normal distribution; for the others, a Kruskal-Wallis test followed by Dunn's post-hoc test was used. The statistical parameters in A are *MYB10.1* F(6,20)=34.37, p-value=0.000; *MYB11* F(5,18)=28.26, p-value=0.000, B: *MYB10.1* F(5,18)=35.69, p-value=0.000; *MYB11* X^2(5)=17.46, p-value=0.004 and in C: *MYB10.1* X^2(5)=18.44, p-value=0.002; *MYB11* F(5,17)=17.84, p-value=0.000.

Figure 5. Heat map hierarchical clustering of the 51 proteins changing significantly between Crognola and Morellona across the three years of study. The color bar indicates the log10-transformed relative protein abundances. The numbers indicate the Pearson's correlation coefficients.

Figure 6. Relative expression (indicated as Normalized relative expression) of some genes coding for differentially abundant proteins in the two Tuscan varieties. Error bars refer to the standard deviation (n=4). Different letters represent the statistical significance (p<0.05) present among the groups of data obtained. If a letter is shared, the difference is not significant. A one-way ANOVA followed by Tukey's post-hoc test was performed on genes showing homogeneity and normal distribution; for the others, a Kruskal-Wallis test followed by Dunn's post-hoc test was used. The statistical parameters are: *BAS1* X^2(5)=2.98, p-value=0.702*; GST* F(5,17)=4.16, p-value=0.012; *HSP17* F(5,17)=7.28, p-value=0.001; *PPO* F(5,17)=9.96, p-value=0.000; *SRC2* F(5,17)=12.88, p-value=0.000; *XTH31* F(5,17)=34.95, p-value=0.000; *XTH6* F(5,17)=0.3449, p-value=0.846*; SOD* F(5,17)=6.33, p-value=0.002; *CPN20* F(5,17)=7.07, p-value=0.001; *70HS* X^2(5)=1.067, p-value=0.957; *70HS2* F(5,17)=9.14, p-value=0.000; *VI1* F(5,17)=1.55, p-value=0.225; *XYL1* F(5,17)=1.300, p-value=0.310*; BGL* X^2(5)=14.27, p-value=0.014; *RBG2* X^2(5)=4.083, p-

value=0.537; *LTP* F(5,17)=8.73, *p*-value=0.000; *NTR* F(5,17)=4.51, *p*-value=0.008; *UGP2* F(5,17)=2.49, *p*-value=0.072; *DHAR2* F(5,17)=1.539, *p*-value=0.230.

Legends to Tables

Table 1. List of differentially abundant compounds in sweet cherries obtained by UHPLC-DAD-HR-MS/MS in positive ESI mode. The details of the compounds are given, together with the specification of the reliability class and references used for the detection; R_t, retention time. MSI, Metabolomics Standards Initiative. *, only in harvests 2016 and 2017. All observed ions are $[M+H]^+$.

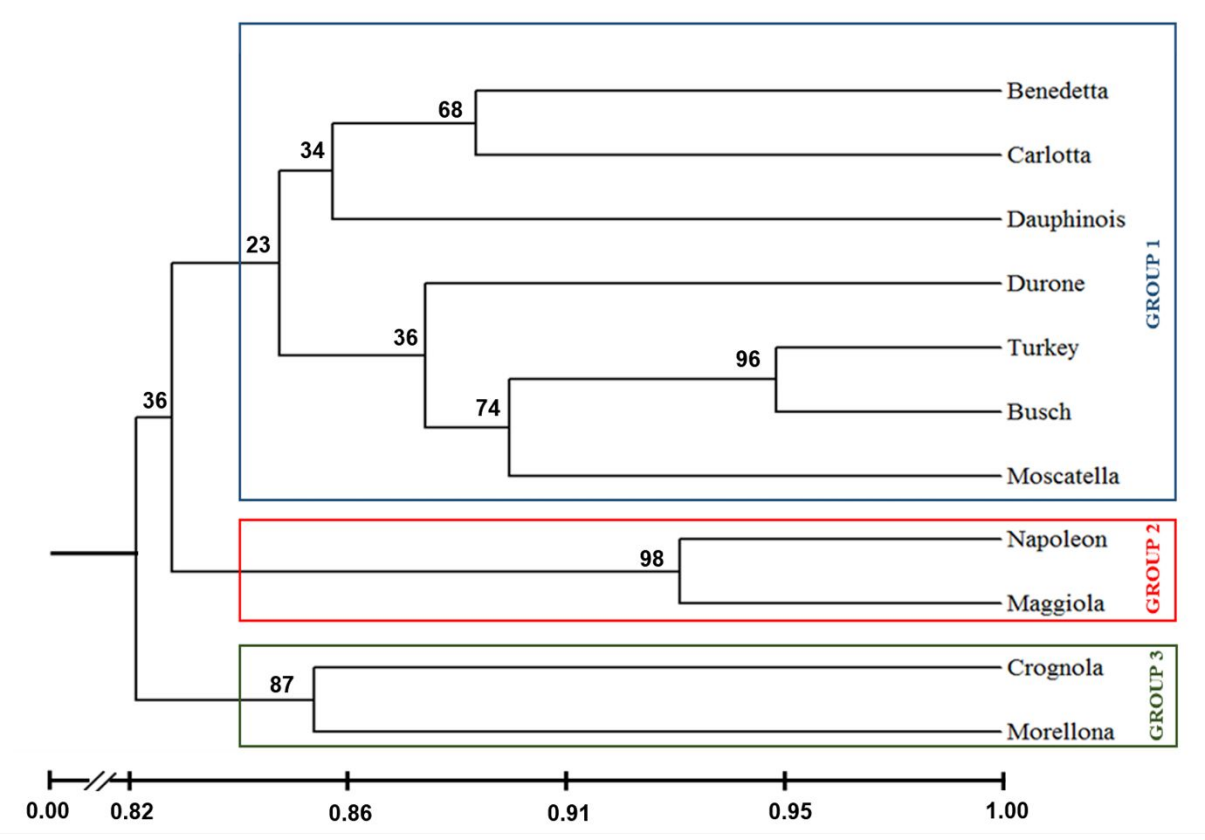
Table 2. List of differentially abundant compounds in sweet cherries obtained by UHPLC-DAD-HR-MS/MS in negative ESI mode. The details of the compounds are given, together with the specification of the reliability class and references used for the detection; R_t, retention time. MSI, Metabolomics Standards Initiative. *, only in harvests 2016 and 2017. All observed ions are $[M-H]^-$.

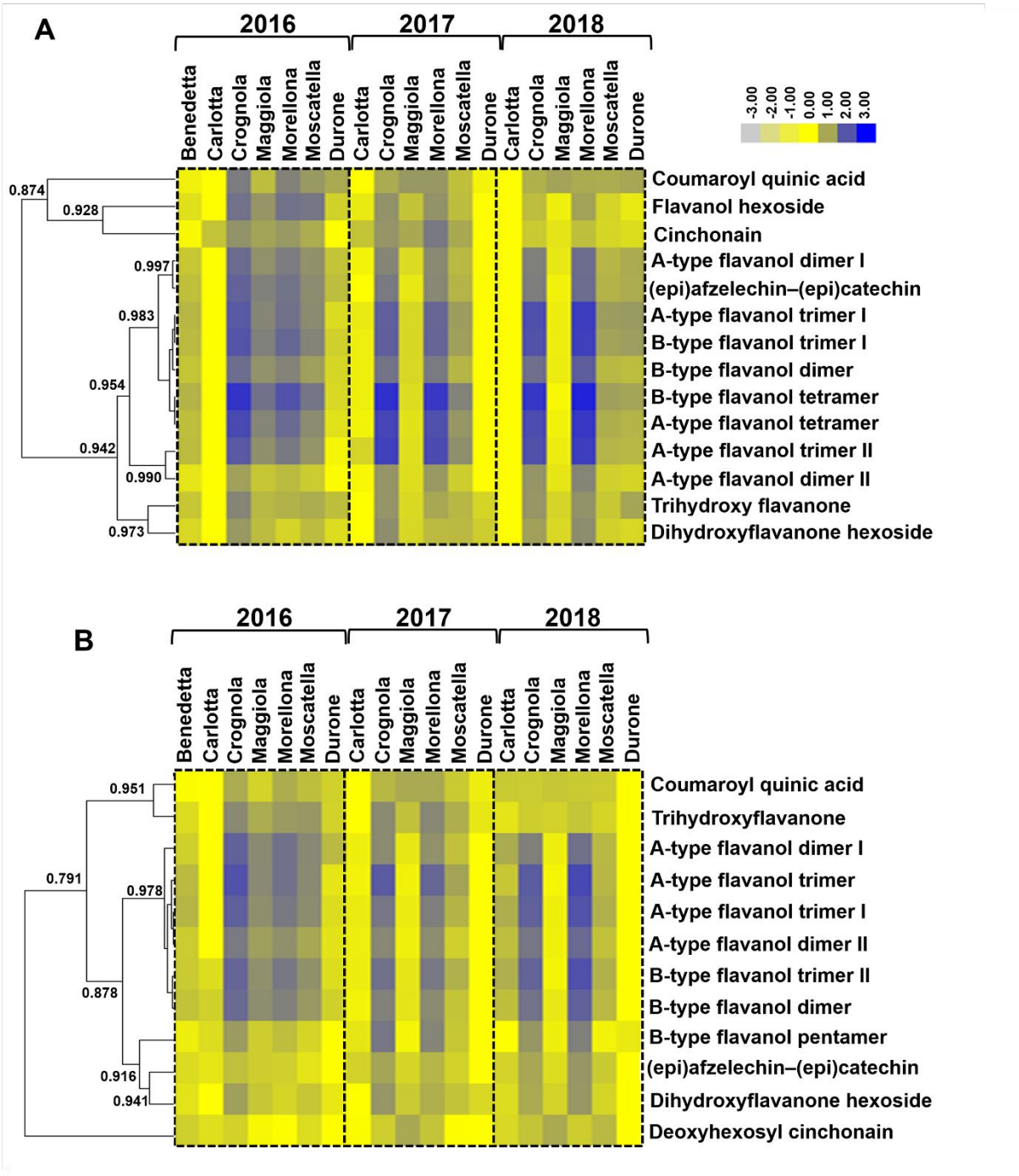
Table 3. Details of the spot numbers, accession numbers, annotations and *p*-values of the identified proteins. The *p*-values<0.01 are highlighted in light green.

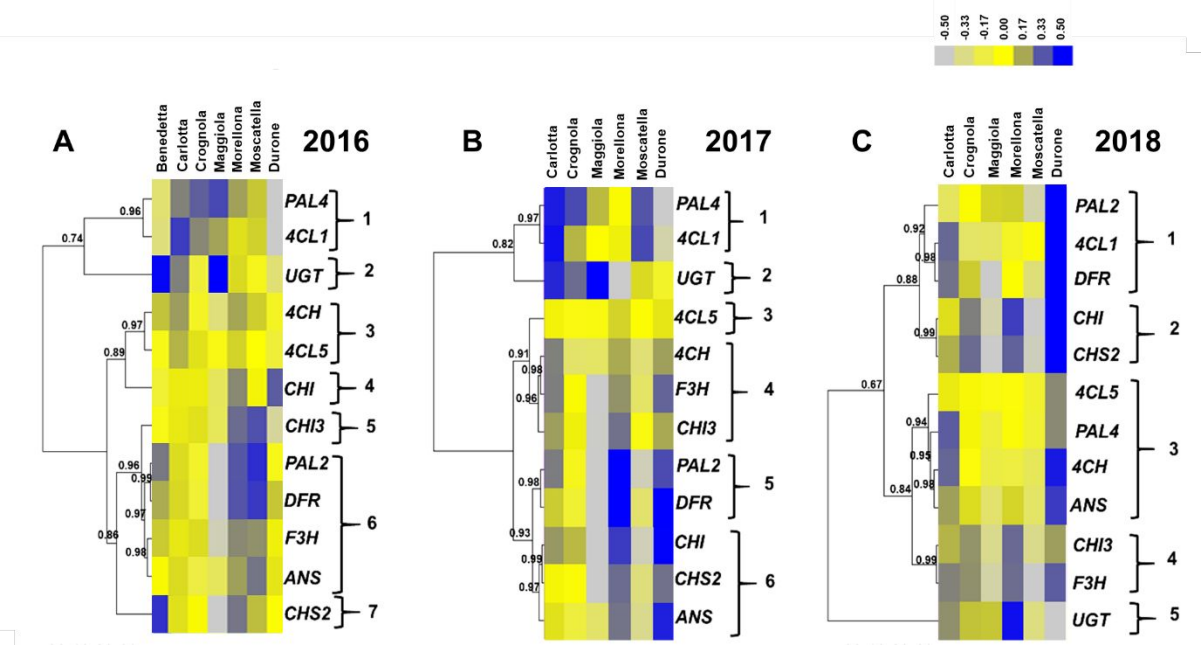
Table 4. Target proteins on whose corresponding genes primers for RT-qPCR were designed. Their categories are also indicated.

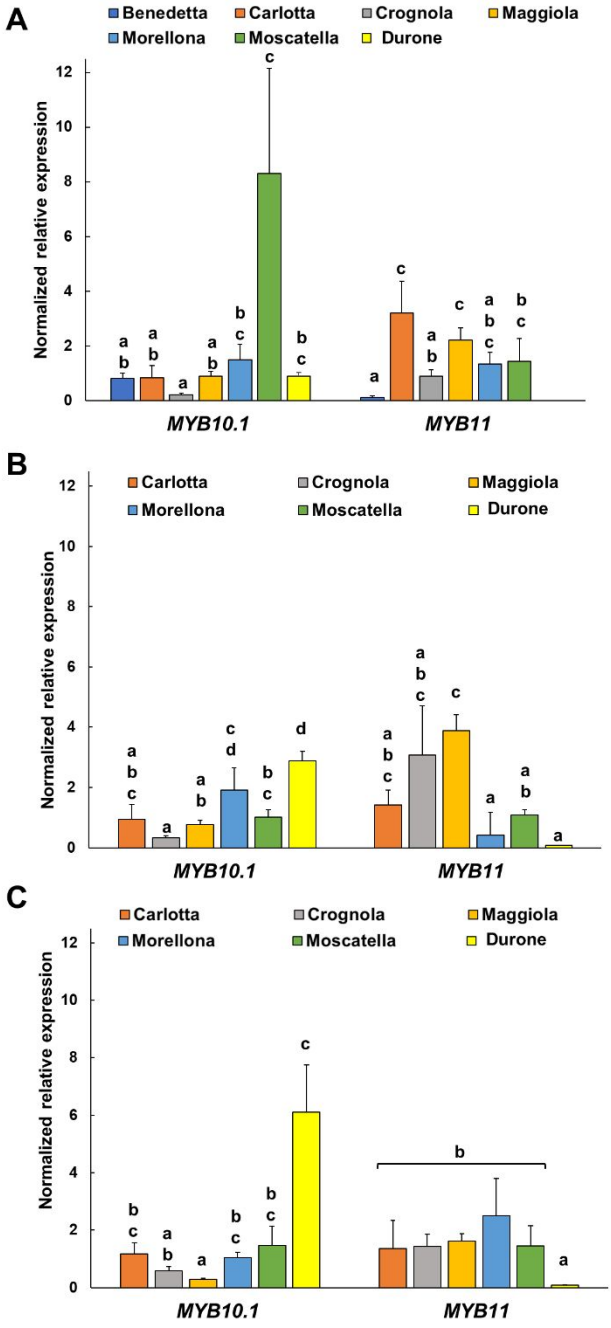
Table 5. Sequences of the 14 primer pairs used for genotyping and relative details. Fluorophore, size range, melting temperature, number of alleles and references are detailed.

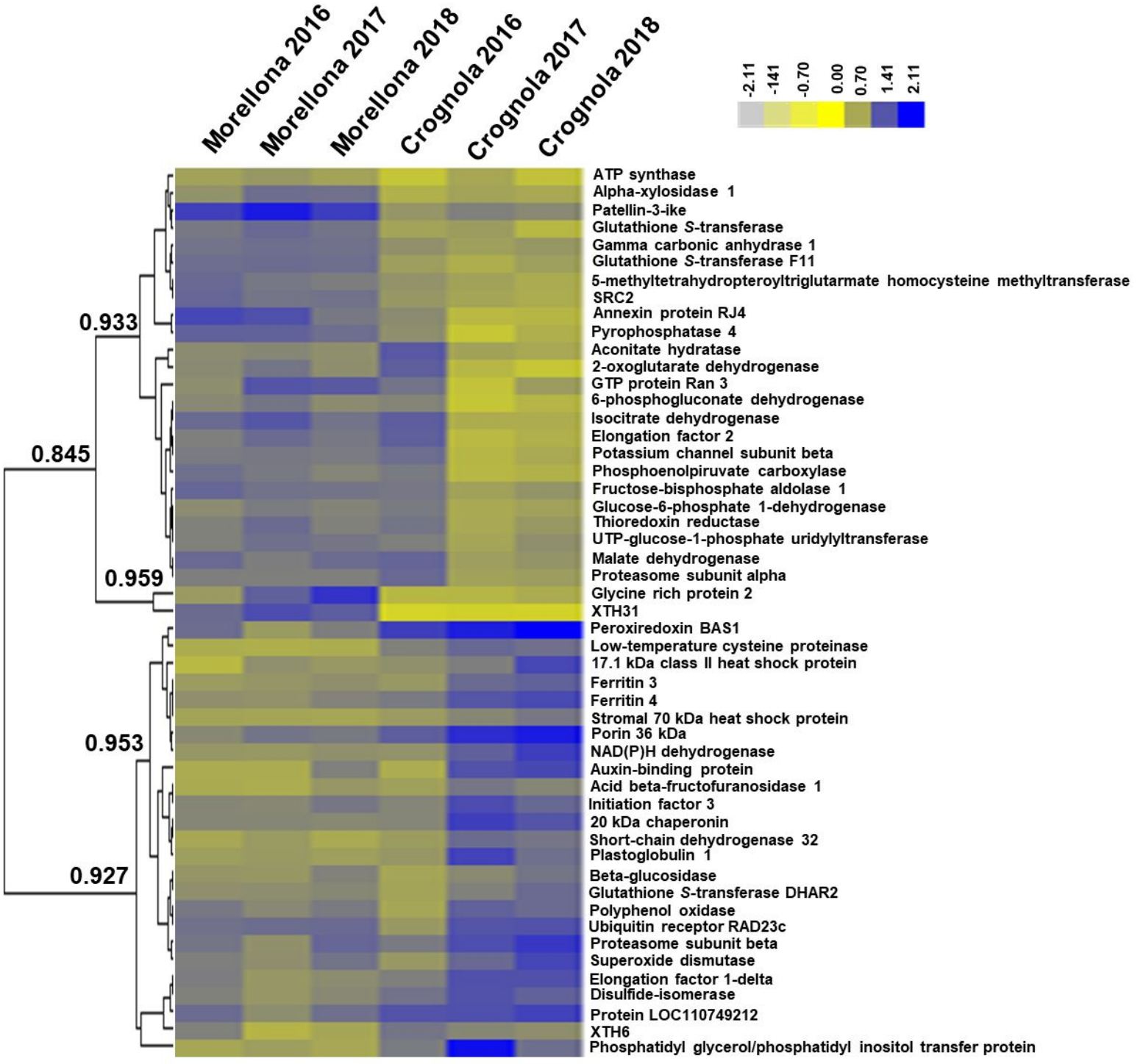
Table 6. List of primers used for gene expression analysis. Details relative to the sequences of the target genes, together with primer efficiency %, melting temperature, amplicon sizes and accession numbers are provided.

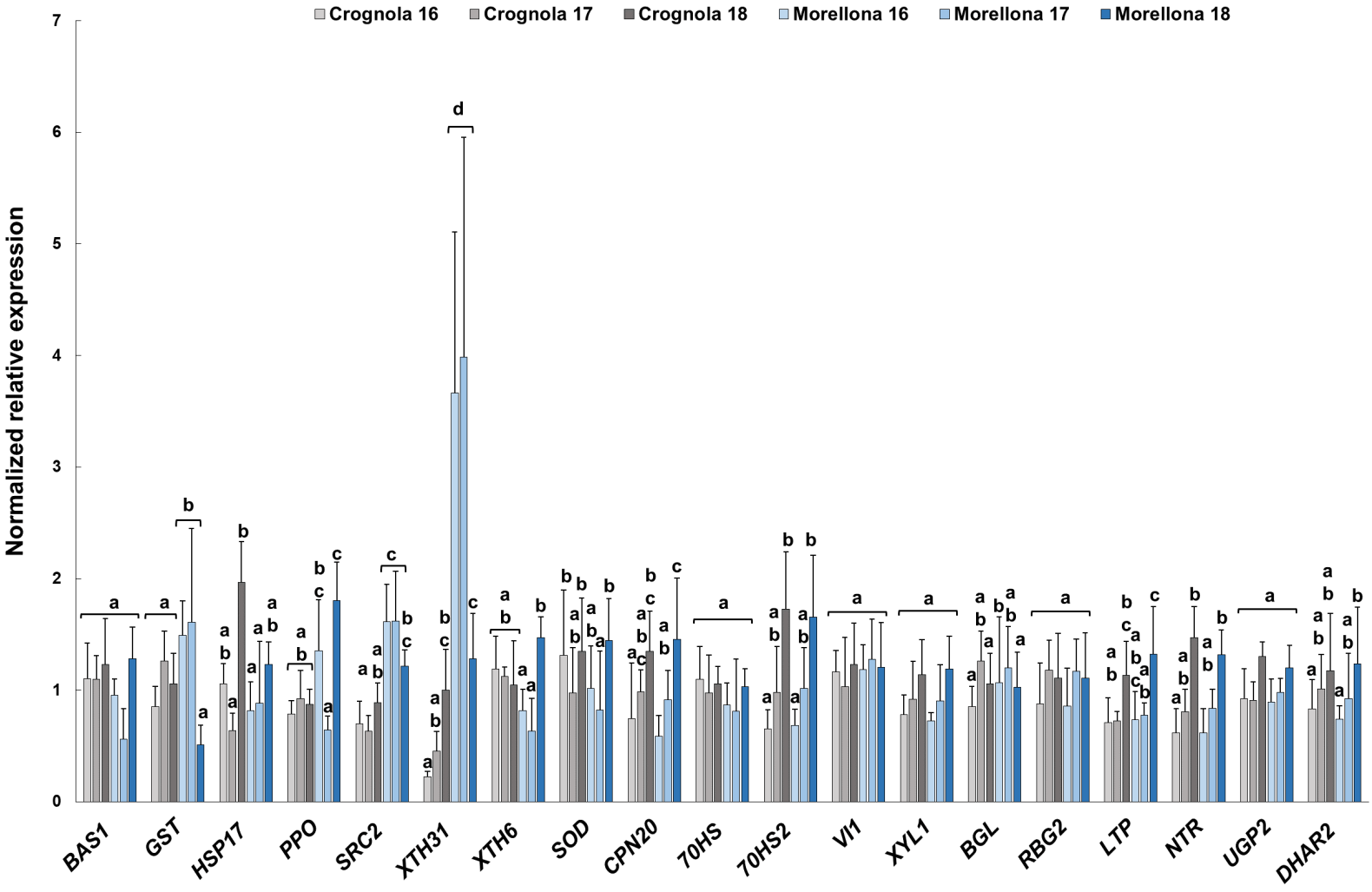

Benedetta
Carlotta
Dauphinois
Durone
Turkey
Busch
Moscatella
Napoleon
Maggiola
Crognola
Morellona
GROUP 1
GROUP 2
GROUP 3
68
34
23
36
96
74
36
98
87
0.00
0.82
0.86
0.91
0.95
1.00


A
2016
2017
2018
Benedetta
Carlotta
Crognola
Maggiola
Morellona
Moscatella
Durone
Carlotta
Crognola
Maggiola
Morellona
Moscatella
Durone
Carlotta
Crognola
Maggiola
Morellona
Moscatella
Durone
-3.00
-2.00
-1.00
0.00
1.00
2.00
3.00
0.874
0.928
0.997
0.983
0.954
0.942
0.990
0.973
Coumaroyl quinic acid
Flavanol hexoside
Cinchonain
A-type flavanol dimer I
(epi)afzelechin–(epi)catechin
A-type flavanol trimer I
B-type flavanol trimer I
B-type flavanol dimer
B-type flavanol tetramer
A-type flavanol tetramer
A-type flavanol trimer II
A-type flavanol dimer II
Trihydroxy flavanone
Dihydroxyflavanone hexoside
B
2016
2017
2018
Benedetta
Carlotta
Crognola
Maggiola
Morellona
Moscatella
Durone
Carlotta
Crognola
Maggiola
Morellona
Moscatella
Durone
Carlotta
Crognola
Maggiola
Morellona
Moscatella
Durone
0.951
0.791
0.978
0.878
0.916
0.941
Coumaroyl quinic acid
Trihydroxyflavanone
A-type flavanol dimer I
A-type flavanol trimer
A-type flavanol trimer I
A-type flavanol dimer II
B-type flavanol trimer II
B-type flavanol dimer
B-type flavanol pentamer
(epi)afzelechin–(epi)catechin
Dihydroxyflavanone hexoside
Deoxyhexosyl cinchonain


-0.50
-0.33
-0.17
0.00
0.17
0.33
0.50
A
2016
Benedetta
Carlotta
Crognola
Maggiola
Morellona
Moscatella
Durone
PAL4
4CL1
UGT
4CH
4CL5
CHI
CHI3
PAL2
DFR
F3H
ANS
CHS2
1
2
3
4
5
6
7
B
2017
Carlotta
Crognola
Maggiola
Morellona
Moscatella
Durone
PAL4
4CL1
UGT
4CL5
4CH
F3H
CHI3
PAL2
DFR
CHI
CHS2
ANS
C
2018
Carlotta
Crognola
Maggiola
Morellona
Moscatella
Durone
PAL2
4CL1
DFR
CHI
CHS2
4CL5
PAL4
4CH
ANS
CHI3
F3H
UGT


A
Benedetta
Carlotta
Crognola
Maggiola
Morellona
Moscatella
Durone
Normalized relative expression
MYB10.1
MYB11
B
Carlotta
Crognola
Maggiola
Morellona
Moscatella
Durone
Normalized relative expression
MYB10.1
MYB11
C
Carlotta
Crognola
Maggiola
Morellona
Moscatella
Durone
Normalized relative expression
MYB10.1
MYB11


Morellona 2016
Morellona 2017
Morellona 2018
Crognola 2016
Crognola 2017
Crognola 2018
-2.11
-141
-0.70
0.00
0.70
1.41
2.11
0.933
0.845
0.959
0.953
0.927
ATP synthase
Alpha-xylosidase 1
Patellin-3-ike
Glutathione S-transferase
Gamma carbonic anhydrase 1
Glutathione S-transferase F11
5-methyltetrahydropteroyltriglutarmate homocysteine methyltransferase
SRC2
Annexin protein RJ4
Pyrophosphatase 4
Aconitate hydratase
2-oxoglutarate dehydrogenase
GTP protein Ran 3
6-phosphogluconate dehydrogenase
Isocitrate dehydrogenase
Elongation factor 2
Potassium channel subunit beta
Phosphoenolpiruvate carboxylase
Fructose-bisphosphate aldolase 1
Glucose-6-phosphate 1-dehydrogenase
Thioredoxin reductase
UTP-glucose-1-phosphate uridylyltransferase
Malate dehydrogenase
Proteasome subunit alpha
Glycine rich protein 2
XTH31
Peroxiredoxin BAS1
Low-temperature cysteine proteinase
17.1 kDa class II heat shock protein
Ferritin 3
Ferritin 4
Stromal 70 kDa heat shock protein
Porin 36 kDa
NAD(P)H dehydrogenase
Auxin-binding protein
Acid beta-fructofuranosidase 1
Initiation factor 3
20 kDa chaperonin
Short-chain dehydrogenase 32
Plastoglobulin 1
Beta-glucosidase
Glutathione S-transferase DHAR2
Polyphenol oxidase
Ubiquitin receptor RAD23c
Proteasome subunit beta
Superoxide dismutase
Elongation factor 1-delta
Disulfide-isomerase
Protein LOC110749212
XTH6
Phosphatidyl glycerol/phosphatidyl inositol transfer protein


Crognola 16
Crognola 17
Crognola 18
Morellona 16
Morellona 17
Morellona 18
Normalized relative expression
0
1
2
3
4
5
6
7
BAS1
GST
HSP17
PPO
SRC2
XTH31
XTH6
SOD
CPN20
70HS
70HS2
VI1
XYL1
BGL
RBG2
LTP
NTR
UGP2
DHAR2