Genetic dissection of the (poly)phenol profile of diploid strawberry (*Fragaria vesca*) fruits using a NIL collection

Maria Urrutia [a], Wilfried Schwab [b], Thomas Hoffmann [b], Amparo Monfort [a,*]

[a] IRTA, Center for Research in Agricultural Genomics (CSIC-IRTA-UAB-UB), Campus UAB, 08193 Bellaterra, Barcelona, Spain
[b] Biotechnology of Natural Products, Technische Universität München, Liesel-Beckmann-Str. 1, 85354 Freising, Germany

ABSTRACT

Over the last few years, diploid strawberry (*Fragaria vesca*) has been recognized as a model species for applied research of cultivated strawberry (*Fragaria × ananassa*) that is one of the most economically important crops. Berries, particularly strawberries, are known for their high antioxidant capacity due to a high concentration of (poly) phenolic compounds. Studies have already characterized the phenolic composition of fruits from sets of cultivated strawberries but the quantification of phenolics in a *Fragaria* mapping population has not been reported, yet. The metabolite profiling of a *F. vesca* near isogenic line (NIL) collection by LC–MS allowed the unambiguous identification of 22 (poly)-phenols, including anthocyanins, flavonols, flavan-3-ols, flavanones, hydroxycinnamic acid derivatives, and ellagic acid in the diploid strawberry fruit. The variability in the collection revealed that the genetic factor was more decisive than the environmental factor for the accumulation of 18 of the 24 compounds. Genotyping the NIL collection with the Axiom® IStraw90® SNPs array, we were able to map 76 stable QTLs controlling accumulation of the (poly)-phenolic compounds. They provide a powerful new tool to characterise candidate genes to increase the antioxidant capacity of fruits and produce healthier strawberries for consumers.

1. Introduction

Strawberries are one of the most popular fruits among consumers worldwide, breeding programs have traditionally focussed on fruit agronomic traits such as firmness or color and resistance to pests, but little attention has been paid to fruit nutritional quality. Strawberries (and other berries) have interesting nutritional properties and a very high antioxidant capacity mainly due to their high proportion of (poly)-phenolic compounds and beneficial effects on human health have been reported [1–4]. Breeding programs aiming to improve fruit nutritional quality could have a positive impact on consumer health. The (poly)-phenols that account for most of this antioxidant capacity belong to two branches of the phenylpropanoid pathway: flavonoids and hydroxycinnamic acid derivatives. Phenylpropanoids are a large family of compounds usually classified as "secondary metabolites" although they carry out important functions for plant survival and fitness such as seed dispersal and resistance to biotic and abiotic stress [5].

The phenylpropanoid biochemical pathway has been extensively studied and described in several plant species, including *Arabidopsis*, poplar, tomato, and rose [6]. Regulation of the pathway is modulated by the development stage and environmental factors, involving both structural genes and transcription factors [7] and references herein. In strawberry, many of the most relevant enzymes have been characterized and the regulation described [8,9]. Phenylpropanoids arise from the phenylalanine derived from the shikimate pathway (Fig. 1), and can lead to hydroxycinnamic acid derivatives and to flavonoids (anthocyanins, flavonols, flavanones, and flavan-3-os) whereas, chalcone synthase (CHS) represents the branching point. The (poly)-phenolic content analyses of a mapping population in diploid strawberry (*Fragaria vesca*) it has not been done before.

There are both quality and genetic reasons for using *F. vesca* as a model species for the *Fragaria* genus and as a source of variability for *Fragaria* breeding programs. The octoploid nature of commercial varieties of strawberry (mainly varieties from the *Fragaria × ananassa*) is a drawback for breeding programs. However the diploid wild relative *F. vesca* (woodland or wild strawberry) can be used as a model for breeding purposes, taking advantage of the many tools that have been developed recently for its study and which demonstrate that *F. vesca* and the commercial hybrid

* Corresponding author. Fax: +34 93563 6601.
E-mail address: amparo.monfort@irta.es (A. Monfort).
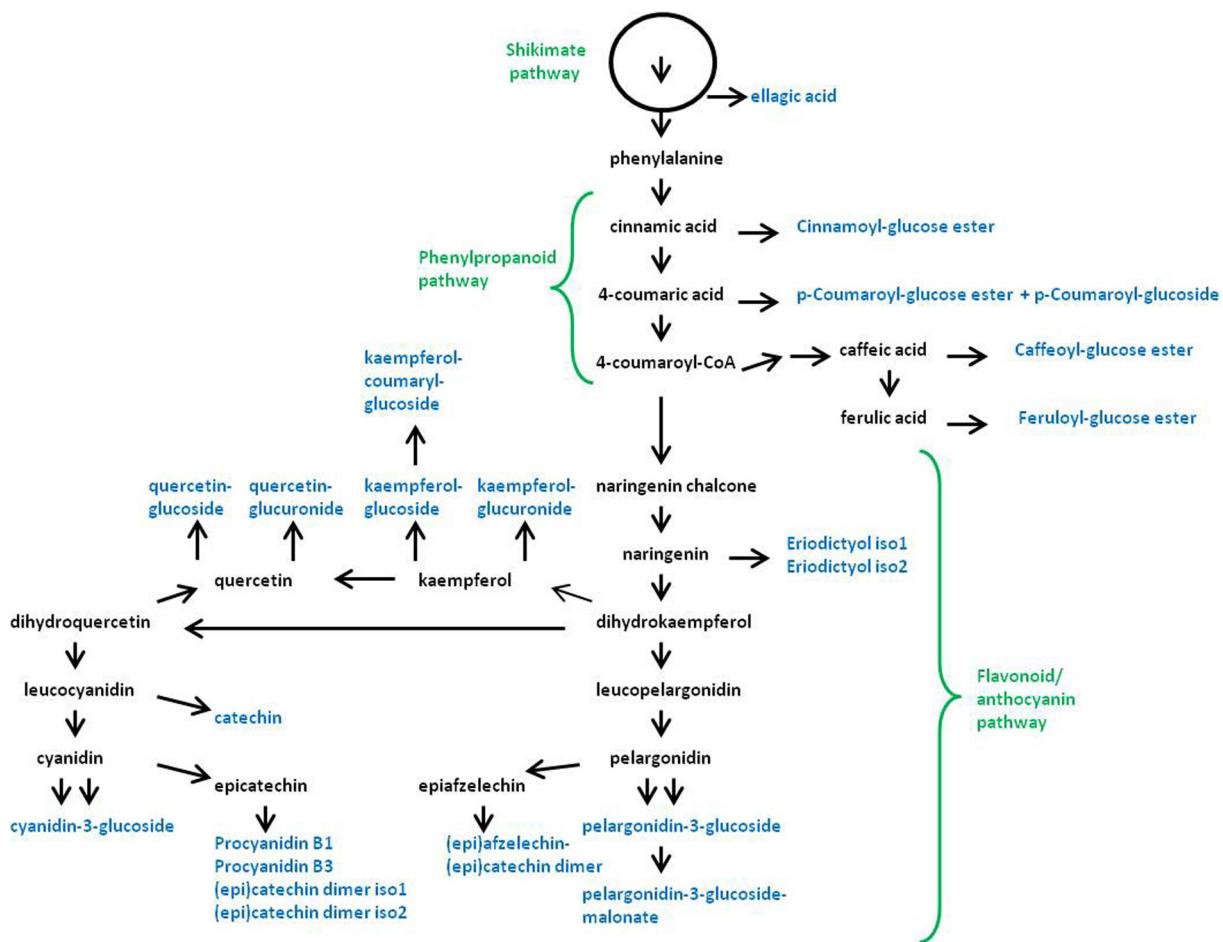
Fig. 1. General scheme of the phenylpropanoid, flavonoid and anthocyanin pathway. Metabolites that were quantified in this study are marked in blue. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

F. × ananassa share a very high degree of synteny [10]. Saturated genetic linkage maps [11] and synteny studies between *F. vesca*, *Fragaria × ananassa*, and other model species from the *Rosaceae* family (*Prunus persica* and *Malus × domestica)* available facilitate transfer of the results between species. This is especially important considering the long intergenerational period of fruit trees [12–15]. Besides, the genome and the annotation of *F. vesca* var. Hawaii 4 have been published and [16], and interesting population resources have been developed in *F. vesca*, including a bin mapping population [17] a mutant collection [18] and a near isogenic lines (NILs) collection [19] and a whole genome genotyping array IStraw90® [20]. The mapping population that we used in this study is described in detail in Urrutia et al. [19]. NIL collections are mapping populations based on introgression lines. All the lines share a common genetic background (usually from an elite variety) and have introgressions from a donor parental (usually an exotic relative). Together, all the lines must cover the background genome with overlapping exotic introgressions [21]. These introgressions can be tracked using molecular markers such as microsatellites (SSRs) or single nucleotide polymorphisms (SNPs). The main advantage of NIL collections compared to other mapping populations is that they can virtually transform the complex inheritance of quantitative traits into simple character inheritance. NIL collections have been developed and successfully used for QTL mapping in many crop species, such as tomato [22,23], melon [24–26], wheat [27], lettuce [28] and also in the model species *Arabidopsis thaliana* [29,30].

Multiple studies have characterized the phenolic composition of fruits from sets of strawberry varieties, mainly from octoploid accessions [31–33]; however, there has never been a rigorous characterization of phenolic content in a *Fragaria* mapping population. In this study we characterised in detail the phenolic composition of ripe strawberry fruit in a *F. vesca* NIL collection in order to map the metabolomic QTL (mQTL) responsible for the variability of metabolites, or groups of metabolites, and to quantify the effect of the genotype and the environment on the phenolic composition in wild strawberry as a tool for increase nutritional quality in cultivated strawberry breeding programs.

2. Materials and methods

2.1. Chemicals

All chemicals, solvents, and reference compounds were obtained from Sigma–Aldrich, Fluka, Riedel de Haën (all Taufkirchen, Germany), Merck (Darmstadt, Germany) or Roth (Karlsruhe, Germany).

2.2. Plant material and sample extraction

Six to eight plant replicates, per genotype, of the available *F. vesca* NILs collection [19], 42 lines in total (Supplemental Fig. 1), the recurrent parental *F. vesca* var. Reine des vallées (RV), the hybrid (*F. vesca × Fragaria bucharica*) F1 parental of the whole collection,

and the white fruited variety of *F. vesca* var. 'Yellow wonder' (YW), were planted in a shade greenhouse in Caldes de Montbui (latitude: 41°36′N, longitude: 2°10′E, altitude 203 m above sea level, pre-coastal Mediterranean climate) for two consecutive years (2012 and 2013). As plants were maintained in outfield conditions we can consider each harvest year as an independent experiment. Three to five biological replicates of ripe strawberry fruits were harvested. Each biological replicate, a mix of ten to 15 fruits (>5 g), was individually and immediately frozen in liquid nitrogen and ground to a fine powder. An aliquot of 500 mg of the powder from every biological replicate was used for extraction. A biochanin A solution (250 μL in methanol; 0.2 mg mL^{-1}) was added as internal standard (IS) yielding 50 μg of IS in each sample. After addition of 250 μL methanol, 1 min vortexing and sonication for 10 min, the sample was centrifuged at 16,000 × g for 10 min. The supernatant was removed and the residue re-extracted with 500 μL methanol. The supernatants were combined, concentrated to dryness in a vacuum concentrator and re-dissolved in 35 μL water. After 1 min vortexing, 10 min sonication and 10 min centrifugation at 16,000 × g, the clear supernatant was used for LC/MS-analysis. Each extract was injected twice as a technical replicate.

2.3. LC–ESI–MS^n analysis

Samples were analyzed on an Agilent 1100HPLC/UV-system (Agilent Technologies, Waldbronn, Germany) equipped with a reversed phase column (Luna 3 μ C18(2) 100A 150 × 2 mm, Phenomenex, Aschaffenburg, Germany) and connected to a Bruker esquire3000$_{plus}$ ion trap mass spectrometer (Bruker Daltonics, Bremen, Germany). Solutions of 0.1% formic acid in water (A) and 0.1% formic acid in methanol (B) were used as mobile phases. The injection volume was 5 μL and the flow rate 0.2 mL min^{-1}. The gradient was from 0% B to 50% B in 30 min. Then to 100% B in the next 5 min, maintained at these conditions for 15 min and decreased to 0% B over 5 min. The initial conditions were held for 10 min for system equilibration. The electrospray ionization voltage of the capillary was set to −4000 V and the end plate to −500 V. Nitrogen was used as nebulizer gas at 30 psi and as dry gas at 330 °C and 9 L min^{-1}. MS spectra were recorded in alternating polarity mode. The full scan ranged from 100 to 800 m/z^{-1}. Ions were accumulated until an ion charge control (ICC) target, 20,000 (positive mode) and 10,000 (negative mode), was achieved or the maximum time of 200 ms was reached. Tandem mass spectrometry was carried out using helium as the collision gas at 4×10^{-6} mbar and a collision voltage of 1 V. Data were analyzed with Data analysis 5.1 software (Bruker Daltonics, Bremen, Germany). Metabolites were identified by comparing their retention times and mass spectra (MS and MS2) with that of measured authentic reference compounds. The major known phenolic metabolites were quantified in the positive and negative MS mode by the internal standard method, using QuantAnalysis2.0 (Bruker Daltonics, Bremen, Germany), and expressed as part per ten thousand equivalents of fresh weight (fw) assuming a response factor of 1.

2.4. Genotyping analysis

DNA from the NIL collection and the parental lines (*F. vesca* var. Reine des Vallées, *F. bucharica* FDP60147 and F_1), was extracted using DNeasy Plant Mini kit (Qiagen, Hilden, Germany). A total of 60 diploid samples (47 NILs, parentals and the hybrid per duplicate and the 7 individuals of BIN set from the F_2 mapping population *F. vesca* var. 815 × *F. bucharica* FDP601 [17] that were added to increase the allelic variability) were hybridized with the IStraw90® SNP array (Affymetrix, CA, USA) [20] at CEGEN (Santiago de Compostela, Spain).

Genotyping analysis was performed using Genotyping console^{TM} and SNPpolisher© (both from Affymetrix, CA, USA) following manufacturers recommendations and an inbred penalty of 3 for the NIL collection. Double recombinants for one single SNP were considered as genotyping errors and the flanking allele was assumed.

2.5. Data analysis

All lines that set fruit each year were processed and analyzed by LC–MS and taken into account for QTL mapping, however, for the exploratory analysis only those genotypes that gave more than 5 g fruit production both years were considered. All statistical analyses and graphical representations were done using the free source software R v.2.15.1 [34] with the Rstudio v.0.92.501 interface [35], unless otherwise specified. Pearson's correlation between the metabolites and the genotypes was calculated using the *rcorr* function from the Hmisc package [36]. The analysis of variance (ANOVA), fitting the model G + E + G × E, was calculated using the *Anova* function from the car package for R [37], and the omega squared values (ω^2) were obtained from ANOVA parameters according to the following formula: $\frac{SS_i - df_i \times MS_{Error}}{SS_t + MS_{Serror}}$

Principal components analysis (PCA) was calculated with the *prcomp* function and scaled values. Hierarchical clustering analysis (HCA) was calculated considering Euclidean distance and complete linkage clustering method. Cluster network analysis (CNA) was using the correlation values with the *qgraph* function from the qgraph R package [38]. The significance tests were recursively calculated using the *t.test* function, comparing each genotype mean value with the RV mean value in the same harvest. The *p*-values obtained were corrected by false discovery rate (FDR) with the *p.adjust* function. Significant threshold was set at *p*-value < 0.05 after correction. QTLs were mapped to a specific bin only when all the NILs carrying an exotic introgression in this bin had a significant effect, and in the same direction, for the specific metabolite of study. SSR genotypes have been used as segregation data [19]. To confirm these QTLs and estimate their effect, an interval mapping analysis was performed using MapQTL v.6 [39] for advanced backcross inbred lines. Stable regions that explained around 20% or more of the variability and had LOD scores >1.8 were considered strong QTLs. The graphical representation of the mQTLs was done using MapChart 2.2 [40]. Candidate genes were based in the *F. vesca* reference genome v1.1 (genome and annotations were downloaded from the *Fragaria* repository in the Genomic Rosaceae Database, GDR, https://www.rosaceae.org/species/fragaria/fragaria_vesca).

3. Results

3.1. Global metabolite profiling analysis

To evaluate the different levels of phenolic compounds identified by LC–MS and detect genomic regions responsible for their control the metabolite profiles of fully ripe strawberry fruits from 42 interspecific introgression lines of a *F. vesca* NIL collection [19] have been analized. The recurrent parental, *F. vesca* var. Reine des vallées (RV) was used as population control. Additionally the white-fruited *F. vesca* var. Yellow wonder (YW) served as an external control or out-group. Over a period of two consecutive years (2012 and 2013), the metabolite profiling analyses and the QTL mapping was carried out with all the genotypes that set enough fruit each year. However, only those 25 genotypes that could be analyzed with at least three biological replicates in both years were taken into account for the exploratory statistical analysis (Supplemental Fig. 1). These lines cover 86.4% of genome [19].
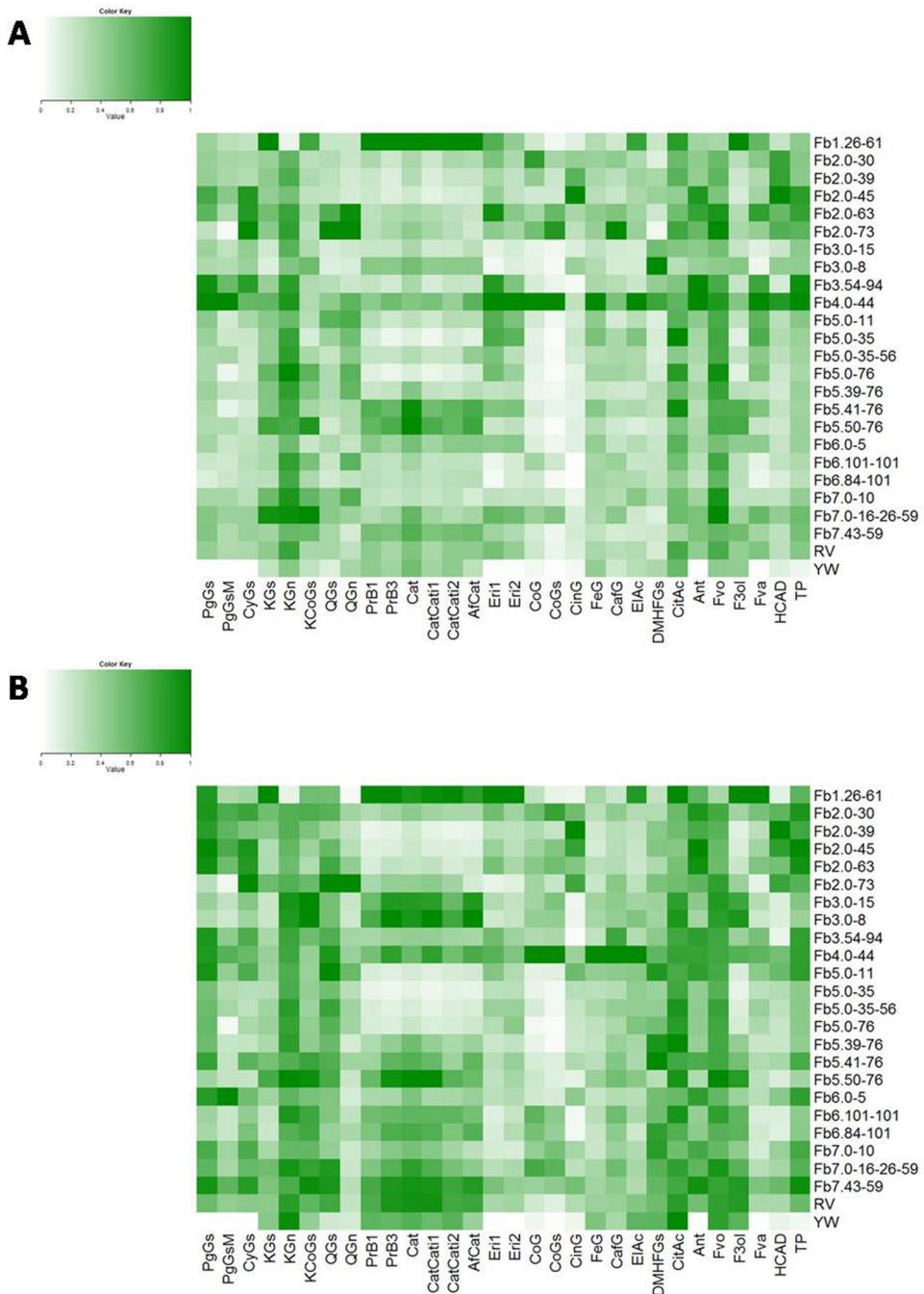
Fig. 2. Relative metabolite profiling in 2012 (A) and 2013 (B). Detected compounds are at the bottom and the NIL analyzed in both harvests on the right. The white (0%) to green (100%) color scale represents the relative accumulation of the compound in each NIL, RV, and YW line. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Among the 936 putative metabolites observed (data not shown), 24 were unambiguously identified by comparison with commercial standards run under the same conditions. The known compounds were quantified in positive and negative MS mode with an internal standard, assuming a response factor of 1. The results are expressed as part per ten thousand (mg/10 g) of fresh weight (fw) (Supplemental Table 1) and graphically represented in Fig. 2. Various compounds were identified. This included three anthocyanins, pelargonidin-3-glucoside, pelargonidin-3-glucoside-malonate, and cyanidin-3-glucoside; five flavonols, kaempferol- and quercetin- glucosides and glucuronides, and kaempferol-coumaryl-glucoside; six flavan-3-ols, procyanidin -B1

and -B3, catechin, two isomers of (epi) catechin dimers and a (epi) afzelechin – (epi) catechin dimer; two flavanones, isomers of eriodictyol; five hydroxycinnamic acid derivatives, cinnamoyl-glucose, p-coumaroyl-glucose and – glucoside, feruloyl-glucose and caffeoyl-glucose. In addition, there were three other compounds previously described in strawberry that not belonging to any of the previous families: ellagic acid, 4-hydroxy-2,5-dimethyl-3(2H)-furanyl glucoside and citric acid [31–33,41–44]. See Table 1 for metabolites abbreviation. The content of chemical families by summing up the metabolites identified: anthocyanins, flavonols, flavanones, flavan-3-ols, and hydroxycinnamic acid derivatives has been quantified. Additionally the total phenolics accumulation was

Table 1
Relative phenolic accumulation: Percentage contribution of chemical families and ellagic acid (first column) to total phenolics content and percentage contribution of specific metabolites to their chemical families (second column) in 2012 and 2013 harvests for RV, average NIL collection and YW.

Chemical families	Abv	(poly)-phenolic compounds	RV		NIL collection		YW	
			2012	2013	2012	2013	2012	2013
Anthocyanins	Ant		0.78	0.88	0.80	0.91	0.00	0.00
	PgGs	Pelargonidin-3-glucoside	0.54	0.46	0.51	0.47	–	–
	PgGsM	Pelargonidin-3-glucoside-malonate	0.11	0.12	0.10	0.10	–	–
	CyGs	Cyanidin-3-glucoside	0.34	0.41	0.39	0.43	–	–
Flavonols	Fvo		0.11	0.05	0.10	0.04	0.44	0.56
	KGs	Kaempferol-glucoside	0.07	0.08	0.08	0.08	0.06	0.08
	KGn	Kaempferol-glucuronide	0.70	0.70	0.656	0.68	0.70	0.80
	KCoGs	Kaempferol-coumaryl-glucoside	0.05	0.11	0.06	0.12	0.06	0.08
	QGs	Quercetin-glucoside	0.14	0.08	0.14	0.08	0.09	0.02
	QGn	Quercetin-glucuronide	0.04	0.03	0.06	0.04	0.09	0.02
Flavan-3-ols	F3ol		0.07	0.04	0.05	0.02	0.47	0.33
	PrB1	Procyanidin B1	0.22	0.18	0.21	0.18	0.22	0.22
	PrB3	Procyanidin B3	0.15	0.18	0.16	0.18	0.15	0.16
	Cat	Catechin	0.21	0.23	0.22	0.24	0.23	0.22
	CatCati1	(epi)Catechin dimers iso1	0.15	0.19	0.16	0.19	0.16	0.16
	CatCati2	(epi)Catechin dimers iso2	0.24	0.19	0.22	0.19	0.23	0.22
	AfCat	(epi)Afzelechin-(epi)catechin dimers	0.03	0.03	0.03	0.03	0.02	0.02
Flavanones	Fva		0.02	0.01	0.02	0.01	0.00	0.00
	Eri1	Eriodictyol iso1	0.78	0.79	0.78	0.82	0.62	1.00
	Eri2	Eriodictyol iso2	0.22	0.21	0.22	0.18	0.38	0.00
Hydroxycinnamic.ac.deriv	HCAD		0.01	0.03	0.02	0.03	0.05	0.09
	CoG	*p*-Coumaroyl-glucose ester	0.26	0.05	0.24	0.07	0.15	0.05
	CoGs	*p*-Coumaroyl-glucoside	0.04	0.01	0.07	0.03	0.01	0.01
	CinG	Cinnamoyl-glucose ester	0.46	0.87	0.53	0.85	0.34	0.68
	FeG	Feruloyl-glucose ester	0.13	0.04	0.09	0.03	0.35	0.19
	CafG	Caffeoyl-glucose ester	0.11	0.03	0.08	0.02	0.16	0.08
Ellagic acid	ElAc		0.01	0.00	0.01	0.00	0.03	0.03

evaluated by summing all identified compounds, except for citric acid and HDMF-glucoside because they are not phenolic compounds.

3.2. General distribution of metabolites in F.vesca RV

The general distribution parameters (mean, standard deviation, and coefficient of variation) of the metabolites in independent harvests (2012 and 2013) were evaluated for RV and the NIL collection separately (Supplementary Table 2). The range calculated for the NIL collection was from undetectable concentrations to a 9-fold change in the mean of the collection. Correlation coefficients between the compounds in different harvests were very high (Table 2). Moreover 21 out of 24 (87%) of them had a significant positive correlation between the two years. Significant correlation values ranged between 0.42 and 0.77, with 17 out of the 21 (80.1%) above 0.6 and the median greater than 0.70. Considering the chemical families, there was no significant correlation only with the flavonols. These results indicate a genotypic effect in the pattern of inheritance of phenolic expression.

The relative contribution of the individual compounds to their chemical family, for both harvests independently, was calculated (Table 1) and the relative contribution of the chemical families and ellagic acid to total phenolics content was also evaluated. Both absolute and relative values for phenolics concentration were very similar between RV and the average *F. vesca* NIL collection within each harvest; as expected, considering that all lines from the NIL collection share their genetic background with the recurrent parental RV (Supplemental Fig. 1) [19]. The global RV (poly)-phenol compounds distribution was the reference to evaluate the significance of the differences observed in the NILs. Anthocyanins were the major contributors to total phenolic content, representing 78% in 2012 and 88% in 2013 (Table 1). There was a difference in the anthocyanin

accumulation in absolute terms (16 mg/10 g and 28 mg/10 g of fw, respectively), (Supplemental Table 2), emphasizing a possible environmental effect. The concentration of individual anthocyanins (pelargonidin- and cyanidin-3-glucoside, and pelargonidin-3-glucoside-malonate) was also higher in the 2013 harvest, although their proportional contribution to total anthocyanin accumulation was the same in both years (pelargonidin-3-glucoside, 52%, cyanidin-3-glucoside, 37% and pelargonidin-glucoside-malonate, 11% of total anthocyanins on average), indicating a genotype control for a specific metabolite balance (Table 1).

Flavonols were the second major contributors to total phenolic content (11% in 2012 and 5% in 2013) (Table 1). The accumulation was significantly different in the two years (Supplemental Table 2) but, as in the case of anthocyanins, the individual flavonol ratios remained constant. Kaempferol-glucuronide is the major flavonol, accounting for an average 70% of total flavonols while the kaempferol-, kaempferol-coumaryl- and quercetin-glucoside each accounted for around 10% of total flavonols. The least abundant flavonol was quercetin-glucuronide, representing about 3% of total flavonols (Table 1). This finding is in agreement with results in *F. vesca* published by Muñoz et al. [32] but contrasts with other publications where quercetin-glucuronide has been described as the most abundant flavonol in octoploid strawberry varieties [31,33].

The contribution of other phenolics (flavan-3-ols, flavanones, hydroxycinnamic acid derivatives, and ellagic acid) to total phenolics concentration is minor (usually <5%, Table 1). The average concentration of flavan-3-ols was slightly higher in 2012 (1.38 mg/10 g of fw) than in 2013 (1.24 mg/10 g of fw). The individual flavan-3-ols contributed equally to total flavan-3-ols accumulation (around 20% each) except the (epi) afzelechin-(epi) catechin dimers, which accounted for 3% (Table 1). There were higher concentrations of flavanones eriodictyol isomers in 2012 than in 2013 (Supplemental Table 2), but their relative abundance was stable 80:20 and favorable to isomer 1.

Table 2
Compound correlation between harvests and G + E + G × E Omega squared values (ω^2). Correlation values were calculated using average genotype values (Supplemental Table1) for both harvests. Percentage of the variance accounted for by each of the studied factors (G, E, and G × E), calculated using the omega squared values from the ANOVA results.

compound	corr	p-Value	ω^2G (%)	ω^2E (%)	ω^2G x E (%)
Pelargonidin-3-glucoside	0.71	***	32.38	15.24	3.61
Pelargonidin-3-glucoside-malonate	0.54	**	35.80	12.37	9.53
Cyanidin-3-glucoside	0.76	***	34.60	19.21	3.32
Kaempferol-glucoside	0.74	***	35.23	7.76	5.88
Kaempferol-glucuronide	0.50	*	17.68	19.87	7.78
Kaempferol-coumaryl-glucoside	0.42	*	28.96	6.40	8.58
Quercetin-glucoside	0.64	***	25.12	31.38	18.66
Quercetin-glucuronide	0.61	**	28.39	14.23	10.59
Procyanidin B1	0.74	***	38.61	16.39	6.86
Procyanidin B3	0.73	***	41.76	4.80	6.32
Catechin	0.71	***	41.04	7.80	7.55
(epi)Catechin dimers iso1	0.71	***	41.58	4.19	6.55
(epi)Catechin dimers iso2	0.73	***	37.08	16.21	6.91
(epi)Afzelechin-(epi)catechin dimers	0.63	***	38.54	8.71	8.22
Eriodictyol iso1	0.70	***	37.63	11.71	9.50
Eriodictyol iso2	0.43	*	19.45	23.89	13.03
p-Coumaroyl-glucose ester	0.77	***	49.42	4.69	11.31
p-Coumaroyl-glucoside	0.61	**	43.92	0.00	11.94
Cinnamoyl-glucose ester	0.71	***	30.69	17.60	9.04
Feruloyl-glucose ester	0.75	***	26.11	11.21	2.20
Caffeoyl-glucose ester	0.25	-	19.32	10.16	16.62
Ellagic acid	0.67	***	19.93	25.48	10.59
DMHF-glucoside	0.00	-	4.53	15.03	4.41
Citric acid	0.03	-	7.37	35.36	9.76
Anthocyanins	0.68	***	29.88	19.69	3.90
Flavonols	0.21	-	10.76	23.84	10.55
Flavan-3-ols	0.73	***	33.93	15.00	10.00
Flavanones	0.67	***	41.00	9.90	6.90
Hydroxycinnamic.ac.deriv	0.80	***	37.80	14.09	6.73
Total phenolics	0.69	***	32.49	9.99	4.14

Asterisk indicates significance thresholds for correlation coefficients p-value.
* p-Value<0.05.

The average concentration of the hydroxycinnamic acid derivatives was higher in 2013 (0.87 mg/10 g of fw) than in 2012 (0.31 mg/10 g of fw). This was mainly due to cinnamoyl-glucose ester that increased its concentration from 0.14 mg/10 g of fw in 2012 to 0.76 mg/10 g of fw in 2013. In contrast, the content and relative accumulation of the other hydroxycinnamic acid derivatives (p-coumaroyl-glucose ester and -glucoside, feruloyl- and caffeoyl-glucose ester) was reduced or maintained (Supplemental Table 2) in 2013 compared with 2012 (Table 1). As minor components ellagic acid was more abundant in the 2012, HDMF-glucoside was more abundant in the 2013 harvest than in 2012 and finally, although citric acid is not a phenolic compound, it was among the most abundant metabolites. Its absolute average concentration significantly varied between years (10.09 mg/10 g of fw in 2012 and 5.97 mg/10 g of fw in 2013).

The phenolic compounds accumulation profile in the yellow wild strawberry, out-group *F. vesca* YW, differed compared with RV or the NIL collection (Table 1), characterized by the absence of anthocyanins, flavanones, and a low content in p-coumaroyls and cinnamoyl-glucose (Fig. 2).

3.3. Genotypic and environmental effect over metabolite accumulation

Concentrations of all detected metabolites (except p-coumaroyl-glucoside) varied between years in RV and in the overall NIL collection, highlighting an environmental factor affecting accumulation. However, the relative contribution to chemical families remained constant for anthocyanins, flavonols, flavan-3-ols, and flavanones. To examine if the accumulation of metabolites was affected by the genotype (G) and the environment (E), we further evaluated their significance and quantified their effect

and interaction on the accumulation of phenolics. The analysis of variance (ANOVA), following the model G + E + G × E (harvests in different years are considered different environments), revealed that the metabolites were significantly affected by both factors and their interaction (p-value < 0.05), except for p-coumaroyl-glucoside that did not have significant variation between years (Supplemental Table 3). Although both factors contributed significantly to variance, their influence differed between metabolites (Table 2). Those most influenced by environment were citric acid and quercetin-glucoside while the genotype had most effect on p-coumaroyl-glucose ester and—glucoside (Fig. 3).

The influence of genotype on total phenolics accumulation was high, and in general environmental effect was low (Table 2). Between family compounds, only flavonols were highly influenced by environment, confirming the non-significant correlation data of this family. The other chemical families were more highly influenced by the genotype. Anthocyanins and Flavan-3-ols variation was highly influenced by the genotype while the effect of the environment was mild for all compounds and the GxE effect was low (<10%) (Supplemental Fig. 2). The effect of the three factors was very different between the flavonols, with kaempferol-glucoside and kaempferol-coumaroyl-glucoside mainly influenced by the genotype (35 and 29%, respectively) with little effect of the environment, while kaempferol-glucuronide was equally affected by the genotype (18%) and the environment (20%). In contrast, quercetin-glucoside and quercetin-glucuronide were both mildly affected by the GxE factor (19 and 11%, respectively), and more by the genotype (25 and 28%, respectively), but quercetin glucoside was more influenced by the environment (32%) than quercetin glucuronide (14%). There was also a great difference in how the factors affected the accumulation of flavanones. Eriodictyol isomer 1 was more influenced by the genotype and isomer 2 more by the environ-
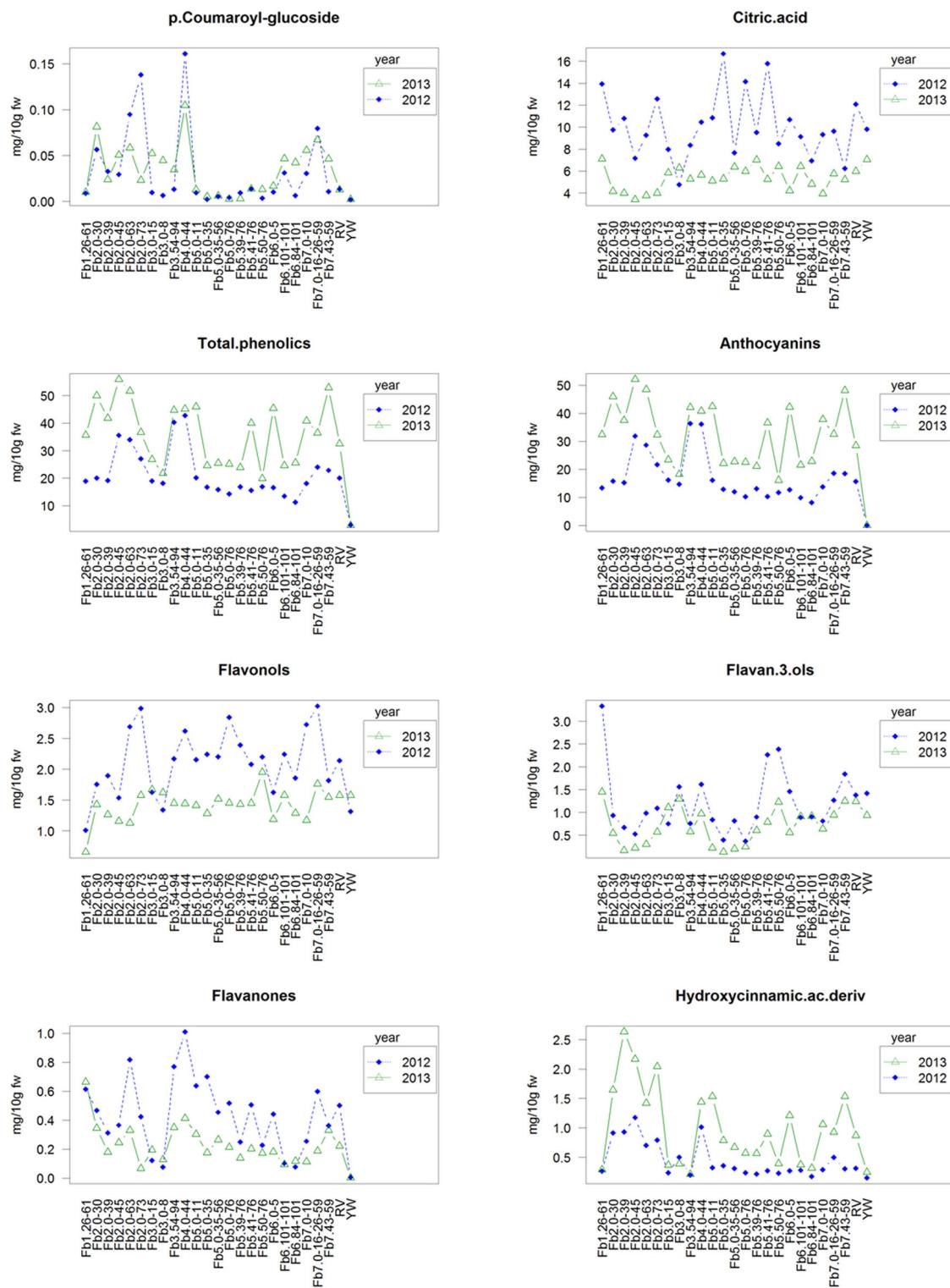
Fig. 3. Interaction plots: Individual plots for p-coumaroyl-glucoside, citric acid and chemical-families representing average values (mg/10 g of fw) for each line in the NIL collection for both harvests (blue, 2012; green, 2013). Parallel lines indicate a major effect of the genotype. Non parallel lines reflect the effect of the environment, and crosses of the lines for specific genotypes reflect the effect of GxE interaction. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

ment. The effect of GxE was similar in both cases. Various behaviors were also observed among hydroxycinnamic acid derivatives. Both *p*-coumaroyl-glucose ester and glucoside were highly affected by the genotype and very little affected by environment. The influence of all three factors on the level of caffeoyl-glucose ester was mild, with the genotype and the $G \times E$ accounting for most of the variability. Citric acid and DMHF-glucoside accumulation were more affected by the environment than by the genotype, although DMHF-glucoside variation was barely affected by either of these factors.
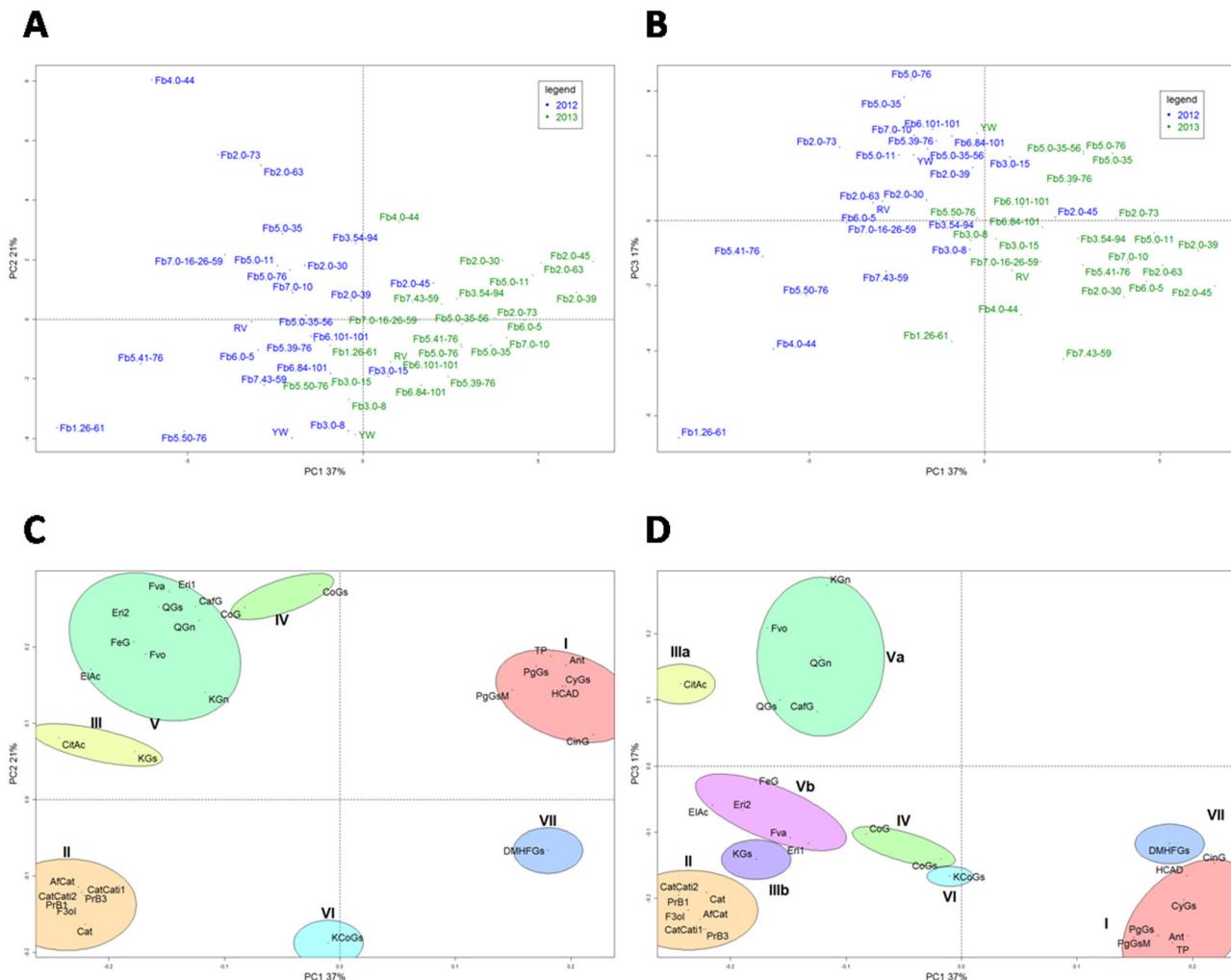
Fig. 4. PCA scores and loadings plot: PCA scores plot for PC1 vs. PC2 (A) and PC1 vs. PC3 (B). Genotypes represented in blue correspond to the 2012 harvest and in green to the 2013 harvest. PCA loadings plot for PC1 vs. PC2 (C) and PC1 vs. PC3 (D). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

3.4. Variance and relation of metabolites among the F. vesca NIL collection

To further understand the structure of the phenolic data set, a principal component analysis (PCA) was carried out. Genotypes were distributed in the first three components (PC1, PC2, and PC3 explaining 37%, 21%, and 17% of the variance, respectively), without forming defined groups, gathering around the axis and the recurrent parental (RV). This is in accordance with the isogenicity of the NIL collection, as they share most of their genetic background [19]. Samples tended to separate according to the different harvest years, which points to an environmental effect (Fig. 4A and B). However, genotypes had a very high correlation between harvests (0.95 on average), demonstrating that the year to year accumulation pattern was similar within each NIL. Only three lines had a correlation data below 0.9, NILs Fb**5**:0−35 and Fb**5**:0−76 with 0.85 and the NIL Fb**5**:41−76 with a minor value of 0.78.

Compounds contributing to samples differentiation formed 7 clear groups according to principal components 1 and 2 (Fig. 4C). Group I included anthocyanins and cinnamoyl-glucose; group II contained flavan-3-ols; group III was formed by citric acid and kaempferol glucoside; groups IV, V, and VI gathered hydroxy-cinnamic acid derivatives, flavanones most flavonols and ellagic and citric acids. There were also two additional single-compound

groups, the kaempferol-coumaryl-glucoside (group VI) and DMHF glucoside (group VII). Both groups III and V were split into two sub-groups attending to PC3 (Fig. 4D). Citric acid and kaempferol glucoside formed subgroups IIIa and IIIb, respectively. Quercetin derivatives, kaempferol glucuronide, and caffeoyl glucose formed subgroup Va, while eriodictyol isomers, ellagic acid, and feruloyl glucose ester formed subgroup Vb.

In general, the 2013 harvest was globally characterized by the accumulation of compounds from group I and VII, and the 2012 harvest by compounds from group IIIa and Va (Fig. 4). Compounds from groups I, IIIa, Va, and VII were the most conditioned by the environment.

As there is an important genetic factor in phenolics accumulation in strawberry fruits, and the detailed regulation mechanism of plant secondary metabolism has not yet been fully understood, we wanted to study the relationships between these metabolites in order to detect possible co-regulation. The correlation between metabolites of the independent harvests was studied (Supplemental Table 4) and represented as a correlation network analysis (CNA), (Fig. 5). Correlation between metabolites was revealed to be higher and more stable between compounds belonging to the same chemical family or sharing known biological pathways (Fig. 1), such as correlations within anthocyanins, flavan-3-ols or flavanones that were positive and strong between all the compounds in the chemi-
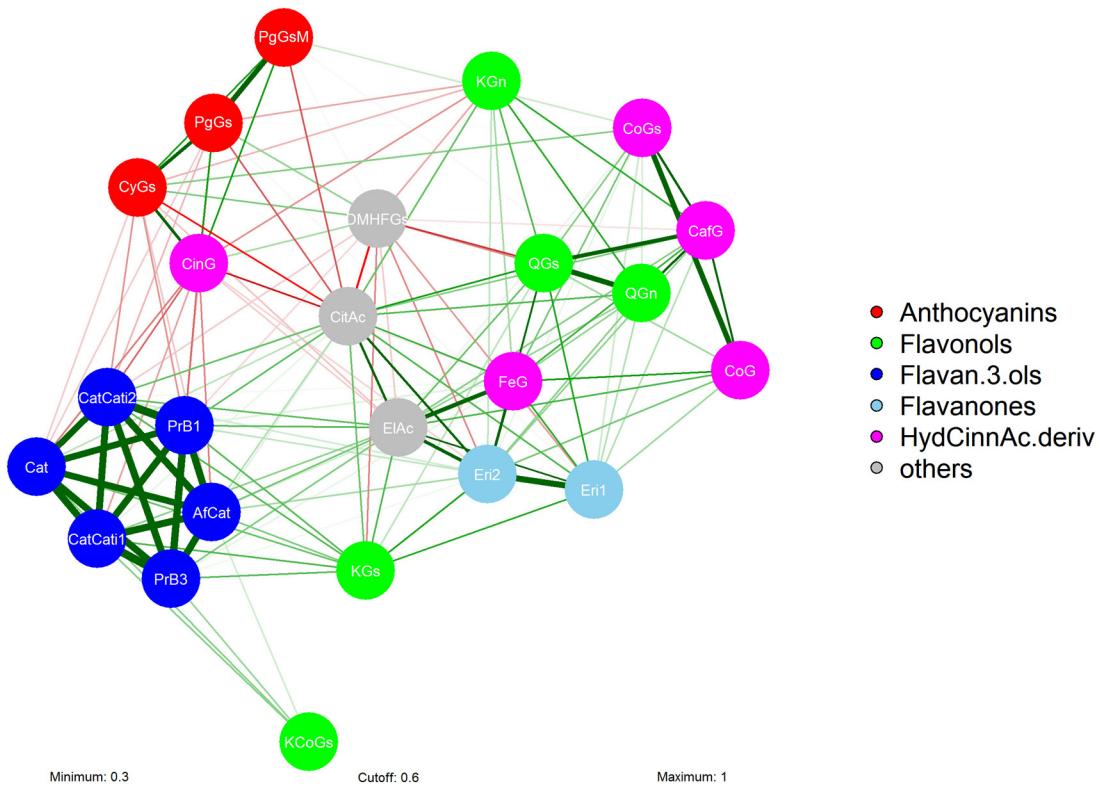
Fig. 5. Cluster network analysis (CNA): Graphical representation of the correlation between phenolic compounds. Metabolites are represented as nodes, colored according to chemical families as specified by the legend. Positive (green) and negative (red) correlations with absolute values >|0.3| are represented as links between the nodes. Links representing absolute correlations >|0.6| are wider the stronger they are and have the maximum color saturation. Absolute correlation <|0.6| are lighter the weaker they are and have the narrowest width. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

cal families. Flavan-3-ols shared correlations ranging from 0.88 to 1, and, in the case of anthocyanins, correlation was stronger between the pelargonidins (avg. 0.81) and weaker between PgGsM and CyGs. However, HCAD and flavonols did not show clear a positive relation between all their metabolites. Among HCAD, cinnamoyl-glucose ester did not correlate with the rest of the compounds in the chemical family. Especially strong correlations were detected between CoG and CoGs (avg. 0.88). The flavonols family established the weakest relations between its compounds, as there was only significant correlation in both harvests between quercetins (avg. 0.72). Stable relations between compounds belonging to different chemical families (Fig. 5) were also found. All flavan-3-ols were positively correlated with kaempferol-glucoside and kaempferol-coumaryl-glucoside, and kaempferol-glucoside was also positively linked to flavanones, and ellagic and citric acids. Anthocyanins correlated positively with cinnamoyl-glucose ester and negatively with citric acid. There was a positive correlation between HCAD caffeoyl-glucose ester and the quercetin derivatives, while feruloyl-glucose ester was linked to eriodictyol isomer 2 and ellagic acid (Supplemental Table 4). According to hierarchical clustering analysis (HCA), metabolites grouped in pairs in four main clusters. The closest compounds forming couples were those that shared high positive correlations, clustering similarly to principal component analysis.

3.5. Fine genotyping of NILs by IStraw90® Axiom Array

Genotyping results of NILs collection hybridized with the IStraw90® [20] revealed that the vast majority of SNPs (79,922 from the initial 95,063 SNPs, 84.07%) were homozygous in all analyzed lines and therefore not suitable for genotyping the NIL collection. From the remaining 15,141 SNPs, 3327 (3.50%) did not surpass qual-

ity criteria (<95% call rate) and were discarded. Additionally, 5747 (6.04%) were classified as off target variants and probably represent allelic variants not considered by the SNP probes and 3079 (3.24%) did not present both homozygous alleles and were not considered for the analysis. Finally, the remaining 2988 (3.14%) were polymorphic SNPs that segregate among the set of analyzed samples and were used for genotyping. After genotyping, the BIN set was not further considered and the analysis continued only with the NIL collection, the parental lines and the hybrid (RV, FB, and F_1). Additional filtering steps were performed including removal of non consistent duplicates and non-segregating among the NIL collection SNPs this resulted in 1510 SNPs.

Genotyping results allowed a more precise drawing of the introgression's boundaries (Supplementary Table 5) and revealed that most NIL were "clean" and did not harbor (or harbor only short) undesired introgressions from *F. bucharica* in the genetic background of *F. vesca*. Several discrepancies between the expected introgressions according to the genetic map [19] and the obtained genotypes (Supplemental Table 5) suggested that there could be some miss-assemblies in the *F. vesca* reference genome v1.1 to which IStraw90® array is anchored, but comparing with the new v.2.0 assembly [10] discrepancies disappear. In term of Mb, the NILs collection covers 192.7 Mb that represents 92% over genome.

3.6. Metabolic QTL analysis

Phenolic metabolites accumulation segregates among the NIL collection. The profile of every NIL had a different accumulation pattern compared with the RV profile (Fig. 2). Differences in the introgressed regions are presumably caused by the *F. bucharica* alleles, indicating that phenolic accumulation is to some extent

genetically controlled. This variability is represented in relative terms (Fig. 2) and in absolute terms (Supplemental Table 1). Significant differences in phenolic compounds accumulation between any of the NILs and the control RV were evaluated by independent *t*-test corrected for false discovery rate (FDR). These significant differences have been accepted as mQTL for a specific metabolite, when all the studied lines harboring the same introgression had a significant effect. Additionally, to calculate the percentage variability explained by it and to confirm the effect of the QTL, a QTL interval mapping analysis has been performed for both harvests independently (Table 3). In total, 100 mQTL for the 24 different compounds have been mapped, the five chemical families and total polyphenols distributed among the seven linkage groups. In 39 of the mQTL the alleles of the *F. bucharica* introgressions caused over-accumulation of the compounds compared with RV, and in the other 61 mQTL the exotic introgressed regions led to a reduction of the level of the metabolites. Among all the mQTL mapped, 76 were stable in two harvests and 24 only appear in one harvest. There were 49 strong mQTLs (LOD score >1.8, observed variability >20%) that have been graphically represented in Fig. 6, 25 of which were stable in both harvests and cited as major QTL.

Anthocyanins were the most abundant phenolics in absolute values, and pelargonidin-3-glucoside the most abundant anthocyanin in almost all the genotypes of both years, followed by cyanidin-3-glucoside. Lines with introgressions in LG2 (Fb**2**:0–45, Fb**2**:0–63), LG3 (Fb**3**:0–8, Fb**3**:0–15) and LG6 (Fb**6**:84–101, Fb**6**:101–101) had a different balance between anthocyanins, accumulating similar levels of both pelargonidin-3-glucoside and cyanidin-3-glucoside, while the line Fb**2**:0–73 preferentially accumulated cyanidin-3-glucoside (Supplemental Table 1). QTL mapping revealed a single, major mQTL at region LG2:45–63 cM, over-accumulating cyanidin-3-glucoside that explained 25–69% of the variability (Table 3), and another major mQTL for the under-accumulation of pelargonidin-3-glucoside-malonate at LG5:41–50, that explained 9–27% of the variability. There were also two strong positive mQTLs explaining 28 and 52% of the variability for pelargonidin-3-glucoside and pelargonidin-3-glucoside-malonate, respectively, at LG4:9–20 but this appeared only in the 2012 harvest. Other three minor mQTL were mapped (Table 3). Total anthocyanins were mapped to three different one-year mQTL, one overlapping with the mQTLs over-accumulating pelargonidins in 2012 at region LG4:9–20, one at the region LG5:50–76 next to the mQTL for under-accumulation of pelargonidin-3-glucoside–malonate, and one that did not co-localize with mQTLs for individual anthocyanins at LG3:54–94. This is probably due to the fact that there was a slight increase in all individual anthocyanins in LG3:54–94 (Fig. 2), not sufficient to map a significant mQTL but the addition of these small increments led to a mQTL for total anthocyanins.

Distribution of total flavonols was fairly homogeneous among the NIL collection, except for a significant decline in NIL Fb**1**:26–61. Regarding specific flavonols, all genotypes had high relative values of kaempferol-glucuronide (the most abundant flavonol in absolute terms) and low relative values of kaempferol-glucoside, in contrast to the behavior of NIL Fb**1**:26–61 (Fig. 2). Absolute flavonol values confirmed this trend. This is mainly due to a significant decrease in kaempferol-glucuronide concentration that was not compensated by the increased concentration of kaempferol-glucoside (Supplemental Table 1). This resulted in a different ratio of metabolites for this specific genotype. According to significance tests, region LG1:26–61 cM gathered three major mQTL for the over-accumulation of kaempferol-glucoside (23–41% of explained variability), the under-accumulation of kaempferol-glucuronide (41–50% of explained variability) and total flavonols (16–34% of explained variabil-

ity) (Fig. 6). Positive, major mQTL were also mapped for quercetins at LG2:63–73. These mQTLs explained between 5 and 50% of the observed variability for quercetin–glucoside and 23–35% for quercetin-glucuronide. In addition, another major mQTL (18–38% of the explained variability) was mapped for the over-accumulation of kaempferol-coumaryl-glucoside at region LG7:43–59. Other minor mQTLs were mapped for individual flavonols (Table 3).

All identified flavan-3-ols were significantly under-accumulated in NILs with introgressions at the beginning of LG5 (Fb**5**:0–11, Fb**5**:0–35, Fb**5**:0–35–56, Fb**5**:0–76) compared with RV in both years. In addition, NIL Fb**1**:26–61 and lines harboring introgressions in the central region of LG7 (LG7:10–26 cM) accumulated higher concentrations of most of the flavan-3-ols (Supplemental Table 1), but this increase was significant with respect to RV only in 2012 (Table 3). Therefore, stable mQTL for total and specific flavan-3-ols were mapped to region LG5:0–11 cM. This mQTL explained between 12 and 32% of the variability of total flavan-3-ols. One-year mQTL were mapped at regions LG1:26–61 and LG7:10–26. Additionally, 8–22% of the observed variability of afzelechin-catechin dimers was explained by a mQTL located at LG2:0–30 cM.

Flavanones (eriodictyol isomers) where under-accumulated in NILs Fb**3**:0–8, Fb**3**:0–15, Fb**6**:84–101 and Fb**6**:101–101. These last two lines, with introgressions in LG6, accumulated significantly lower levels of eriodictyol isomer 1 resulting in a different ratio of isomers (Supplemental Table 1). Considering the significance tests, minor mQTLs have been mapped for under-accumulation of total flavanones and eriodictyol isomer 1 and 2. Strong mQTLs were stable in only one harvest for eriodictyol isomer 2 at region LG1:26–61 and LG4:9–20 (Fig. 6).

Relative accumulation of hydroxycinnamic acid derivatives was higher in lines with introgressions in LG2 and LG4 (Fig. 2). Lines with introgressions in LG2 presented significantly higher relative and absolute concentrations of cinnamoyl-glucose ester (Supplemental Table 1), the most important contributor to total hydroxycinnamic acid derivatives (Table 1). Consequently, positive major stable QTLs for cinnamoyl-glucose ester (37–52% of variability explained) and total HCAD (40–58% of variability explained) to region LG2:0–30 cM (Fig. 6) have been mapped. The same regions also harbor positive minor mQTLs for *p*-coumaroyl-glucose ester and — glucoside (Table 3). Lines harboring introgressions at the beginning of LG4 accumulated higher levels of *p*-coumaroyl-glucose ester, *p*-coumaroyl-glucoside and feruloyl-glucoside, and strong and stable QTLs were mapped for these compounds at region LG4:9-20 cM (Fig. 6). In addition, a strong one-year QTL was mapped for caffeoyl-glucoside in the same region. Despite these compounds being minor contributors to the total hydroxycinnamic acid-derivatives accumulation, the sum of the small increments was a significantly higher overall concentration of HCAD mapped as a minor mQTL to region LG4:9–20 cM (Table 3). The opposite behavior was shown by lines with introgressions in LG6 (Fb**6**:84–101 and Fb**6**:101–101) where the significant decrease in cinnamoyl-glucose ester accumulation led to a significant under-accumulation of the total HCAD (Table 3) although the rest of the HCAD retained average values.

The distribution of ellagic acid was fairly homogeneous along the NIL collection, except for significant increases in lines harboring introgressions at LG1 and LG4 (Fig. 1, Supplemental Table 1). These led to the mapping of two positive stable mQTLs at regions LG1:26–61 cM and LG4:9–20 cM that contributed to 29–50% and 15–22% of the observed variability, respectively.

Finally, stable QTLs were not mapped for total (poly)-phenols accumulation but one-year mQTLs co-located with mQTLs detected for anthocyanins accumulation. This was expected considering

Table 3

mQTL for phenolic compounds in *F. vesca*. All detected mQTL are shown. The under or over accumulation effect of the mQTL is indicated by (−) or (+) symbol, respectively. The genetic position of the mQTL expressed in cM and the NIL with the shorter introgression harboring the mQTL is given. Finally, the significance test results (*t*-test *p*-value after correction and LOD score), the % of the variability explained by the QTL and the stability (1 or 2 harvests) are indicated. High variability and stability are signaled as major QTL (*).

	Compound	Direction	QTL (cM)	Smallest NIL	*t*-Test (corrected *p*-value)	LOD	Expl. var. (%)	Stable QTL	major QTL
Anthocyanins	Pelargonidin-3-glucoside	−	2:63–73	Fb2.0–73	<0.05	<1.8	3–25%	2	
	Pelargonidin-3-glucoside	+	4:9–20	Fb4.0–20	<0.05	2.47	28%	1	
	Pelargonidin-3-glucoside-malonate	−	2:63–73	Fb2.0–73	<0.05	<1.8	8–10%	2	
	Pelargonidin-3-glucoside-malonate	−	5:41–50	Fb5.41–76	<0.05	2.38	9–27%	2	*
	Pelargonidin-3-glucoside-malonate	−	6:101–101	Fb6.101–101	<0.05	<1.8	1–10%	2	
	Pelargonidin-3-glucoside-malonate	+	4:9–20	Fb4.0–20	<0.05	5.59	52%	1	
	Cyanidin-3-glucoside	+	2:45–63	Fb2.39–63	<0.05	7.2	25–69%	2	*
Flavonols	Kaempferol-glucoside	+	1:26–61	Fb1.26–61	<0.05	3.91	23–41%	2	*
	Kaempferol-glucoside	+	5:50–76	Fb5.50–76	<0.05	<1.8	3–4%	2	
	Kaempferol-glucoside	−	3:0–8	Fb3.0–8	<0.05	<1.8	10–20%	2	
	Kaempferol-glucoside	−	6:84–101	Fb6.84–101	<0.05	<1.8	6–8%	2	
	Kaempferol-glucuronide	−	1:26–61	Fb1.26–61	<0.05	5.12	41–50%	2	*
	Kaempferol-glucuronide	−	2:39–45	Fb2.0–45	<0.05	<1.8	3%	2	
	Kaempferol-coumaryl-glucoside	+	7:43–59	Fb7:43–59	<0.05	3.57	18–38%	2	*
	Kaempferol-coumaryl-glucoside	+	5:50–76	Fb5.50–76	<0.05	<1.8	4%	2	
	Quercetin-glucoside	+	2:63–73	Fb2.0–73	<0.05	5.45	4–52%	2	*
	Quercetin-glucoside	−	1:26–61	Fb1.26–61	<0.05	<1.8	1–5%	2	
	Quercetin-glucoside	−	6:101–101	Fb6.101–101	<0.05	<1.8	5–14%	2	
	Quercetin-glucuronide	+	2:63–73	Fb2.0–73	<0.05	3.22	23–35%	2	*
Flavan-3-ols	Procyanidin B1	−	5:0–11	Fb5.0–11	<0.05	2.48	9–28%	2	*
	Procyanidin B1	−	2:0–30	Fb2.0–30	<0.05	<1.8	7–16%	2	
	Procyanidin B1	−	5:39–41	Fb5.39–76	<0.05	<1.8	4–6%	2	
	Procyanidin B1	−	3:54–94	Fb3.54–94	<0.05	<1.8	1–2%	2	
	Procyanidin B1	+	1:26–61	Fb1.26–61	<0.05	2.46	30%	1	
	Procyanidin B1	+	7:10–26	Fb7:0–27	<0.05	2.61	30%	1	
	Procyanidin B3	−	5:0–11	Fb5.0–11	<0.05	2.82	12–31%	2	*
	Procyanidin B3	−	2:0–30	Fb2.0–30	<0.05	<1.8	8–14%	2	
	Procyanidin B3	−	3:54–94	Fb3.54–94	<0.05	<1.8	2%	2	
	Procyanidin B3	+	1:26–61	Fb1.26–61	<0.05	2.37	27%	1	
	Procyanidin B3	+	7:10–26	Fb7:0–27	<0.05	1.81	22%	1	
	Catechin	−	5:0–11	Fb5.0–11	<0.05	3.05	16–33%	2	*
	Catechin	−	2:0–30	Fb2.0–30	<0.05	<1.8	9–15%	2	
	Catechin	−	3:54–94	Fb3.54–94	<0.05	<1.8	1–3%	2	
	Catechin	+	7:10–26	Fb7:0–27	<0.05	2.39	28%	1	
	(epi)catechin dimers iso1	−	5:0–11	Fb5.0–11	<0.05	2.83	12–31%	2	*
	(epi)catechin dimers iso1	−	2:0–30	Fb2.0–30	<0.05	<1.8	7–13%	2	
	(epi)catechin dimers iso1	−	3:54–94	Fb3.54–94	<0.05	<1.8	2%	2	
	(epi)catechin dimers iso1	−	7:0–10	Fb7.0–10	<0.05	<1.8	1%	2	
	(epi)catechin dimers iso1	+	1:26–61	Fb1.26–61	<0.05	2.59	27%	1	
	(epi)catechin dimers iso1	+	7:10–26	Fb7:0–27	<0.05	1.8	21%	1	
	(epi)catechin dimers iso2	−	5:0–11	Fb5.0–11	<0.05	2.59	9–29%	2	*
	(epi)catechin dimers iso2	−	2:30–39	Fb2.0–39	<0.05	<1.8	7–15%	2	
	(epi)catechin dimers iso2	−	5:39–41	Fb5.39–76	<0.05	<1.8	4–6%	2	
	(epi)catechin dimers iso2	−	3:54–94	Fb3.54–94	<0.05	<1.8	1–2%	2	
	(epi)catechin dimers iso2	−	7:0–10	Fb7.0–10	<0.05	<1.8	1%	2	
	(epi)catechin dimers iso2	+	1:26–61	Fb1.26–61	<0.05	2.75	31%	1	
	(epi)catechin dimers iso2	+	7:10–26	Fb7:0–27	<0.05	2.6	30%	1	
	(epi)afzelechin-(epi)catechin dimers	−	2:0–30	Fb2.0–30	<0.05	1.86	8–22%	2	*
	(epi)afzelechin-(epi)catechin dimers	−	5:0–11	Fb5.0–11	<0.05	1.86	3–22%	2	*
	(epi)afzelechin-(epi)catechin dimers	−	3:54–94	Fb3.54–94	<0.05	<1.8	1–2%	2	
	(epi)afzelechin-(epi)catechin dimers	−	5:39–41	Fb5.39–76	<0.05	<1.8	1–2%	2	
	(epi)afzelechin-(epi)catechin dimers	−	7:0–10	Fb7.0–10	<0.05	<1.8	3%	2	
	(epi)afzelechin-(epi)catechin dimers	+	1:26–61	Fb1.26–61	<0.05	1.98	22%	1	
	(epi)afzelechin-(epi)catechin dimers	+	7:10–26	Fb7:0–27	<0.05	2.01	24%	1	
Flavanones	Eriodictyol iso1	−	6:101–101	Fb6.101–101	<0.05	<1.8	11–18%	2	
	Eriodictyol iso1	−	5:39–41	Fb5.39–76	<0.05	<1.8	1–10%	2	
	Eriodictyol iso1	−	7:0–10	Fb7.0–10	<0.05	<1.8	1–5%	2	
	Eriodictyol iso2	−	3:0–8	Fb3.0–8	<0.05	<1.8	5–11%	2	
	Eriodictyol iso2	−	7:0–10	Fb7.0–10	<0.05	<1.8	1–5%	2	
	Eriodictyol iso2	+	4:9–20	Fb4.0–20	<0.05	3.87	41%	1	
	Eriodictyol iso2	+	1:26–61	Fb1.26–61	<0.05	2.44	28%	1	
HCAD	*p*-Coumaroyl-glucose ester	+	4:9–20	Fb4.0–20	<0.05	3.77	28–40%	2	*
	p-Coumaroyl-glucose ester	−	5:39–41	Fb5.39–76	<0.05	1.86	13–22%	2	*
	p-Coumaroyl-glucose ester	−	5:11–35	Fb5.0–35	<0.05	<1.8	6–19%	2	
	p-Coumaroyl-glucose ester	+	2:0–30	Fb2.0–30	<0.05	<1.8	4–13%	2	
	p-Coumaroyl-glucoside ester	+	4:9–20	Fb4.0–20	<0.05	3.31	29–36%	2	*
	p-Coumaroyl-glucoside ester	+	2:0–30	Fb2.0–30	<0.05	<1.8	6–15%	2	
	Cinnamoyl-glucose ester	+	2:0–30	Fb2.0–30	<0.05	5.5	37–52%	2	*

Table 3 (*Continued*)

	Compound	Direction	QTL (cM)	Smallest NIL	*t*-Test (corrected *p*-value)	LOD	Expl. var. (%)	Stable QTL	major QTL
	Cinnamoyl-glucose ester	−	3:54–94	Fb3.54–94	<0.05	<1.8	2–3%	2	
	Cinnamoyl-glucose ester	−	6:101–101	Fb6.101–101	<0.05	<1.8	6–8%	2	
	Feruloyl-glucose ester	+	4:9–20	Fb4.0–20	<0.05	9.26	60–71%	2	*
	Caffeoyl-glucose ester	+	4:9–20	Fb4.0–20	<0.05	9.96	74%	1	
	Caffeoyl-glucose ester	+	2:63–73	Fb2.0–73	<0.05	5.5	52%	1	
	Ellagic acid	+	4:9–20	Fb4.0–20	<0.05	5.11	29–50%	2	*
Others	Ellagic acid	+	1:26–61	Fb1.26–61	<0.05	1.87	15–22%	2	*
	Ellagic acid	−	7:43–59	Fb7.43–59	<0.05	<1.8	1–5%	2	
	DMHF-glucoside	+	5:39–50	Fb5.39–76	<0.05	2.67	30%	1	
	Citric acid	−	2:39–45	Fb2.0–45	<0.05	3.09	34%	2	*
Chemical families	Anthocyanins	+	3:54–94	Fb3.54–94	<0.05	1.99	24%	1	
	Anthocyanins	+	4:9–20	Fb4.0–20	<0.05	1.94	23%	1	
	Anthocyanins	−	5:50–76	Fb5.50–76	<0.05	1.88	22%	1	
	Flavonols	−	1:26–61	Fb1.26–61	<0.05	3.11	16–34%	2	*
	Flavonols	−	2:39–45	Fb2.0–45	<0.05	<1.8	2–3%	2	
	Flavan-3-ols	−	5:0–11	Fb5.0–11	<0.05	2.92	12–32%	2	*
	Flavan-3-ols	−	2:0–30	Fb2.0–30	<0.05	<1.8	8–16%	2	
	Flavan-3-ols	−	3:54–94	Fb3.54–94	<0.05	<1.8	1–2%	2	
	Flavan-3-ols	−	7:0–10	Fb7.0–10	<0.05	<1.8	1%	2	
	Flavan-3-ols	−	5:39–41	Fb5.39–76	<0.05	<1.8	3–5%	2	
	Flavan-3-ols	+	7:10–26	Fb7.0–27	nd	2.32	27%	1	
	Flavanones	−	6:101–101	Fb6.101–101	<0.05	<1.8	10–17%	2	
	Flavanones	−	3:0–8	Fb3.0–8	<0.05	<1.8	4–16%	2	
	Flavanones	−	5:39–41	Fb5.39–76	<0.05	<1.8	1–10%	2	
	Flavanones	−	7:0–10	Fb7.0–10	<0.05	<1.8	1–5%	2	
	Hydroxycinnamic.ac.deriv	+	2:0–30	Fb2.0–30	<0.05	6.45	40–58%	2	*
	Hydroxycinnamic.ac.deriv	+	4:0–44	Fb4.0–44	<0.05	<1.8	1–13%	2	
	Hydroxycinnamic.ac.deriv	−	5:50–76	Fb5.50–76	<0.05	<1.8	6–8%	2	
	Hydroxycinnamic.ac.deriv	−	3:54–94	Fb3.54–94	<0.05	<1.8	2–3%	2	
	Hydroxycinnamic.ac.deriv	−	6:84–101	Fb6.84–101	<0.05	<1.8	4–6%	2	
	Total polyphenols	+	4:9–20	Fb4.0–20	<0.05	2.27	26%	1	
	Total polyphenols	−	5:50–76	Fb5.50–76	<0.05	1.84	22%	1	

that anthocyanins account for around 80% of the total (poly)-phenols.

3.7. Observed candidate genes in mQTL regions

Considering that we were able to map a number of mQTL and that the sequence and annotation of the strawberry genome are publicly available [16], we aimed to identify interesting candidate genes in the mQTL regions described. mQTL genetic intervals have been more precisely defined considering the physical position of introgresions boundaries defined by SNPs after Axiom® IStraw90® SNPs Array hybridization (Table 4). After a complete genome search, 313 genes were detected on the genetic intervals defined for the mQTLs. The list of 99 biochemically characterized *Fragaria* genes has been checked (summarized in supplemental Table 10, [16]) and found 69 of them located in genetic regions covered by our mQTLs. These included 14 genes that have been described as structural genes directly involved in pigment formation and phenylpropanoid and flavonoid metabolism (*FaCHS, FaCHS2, FaCHS3, FaCHS4 FaDFR, FaDFR1, FaF3H, FaFLS, FaLAR, FaANS, FaANR, FaGT1, FaGT2,* and *FaGT7*), one transcription factor (*FaMYB1*) and three other non-structural protein-coding genes (*Fraa1A, Fraa2, Fraa3*) related to flavonoid biosynthesis (Supplemental Table 6). Then the list of 32,831 annotated predicted genes from the diploid strawberry reference genome has been checked for genes annotated *in silico* as enzymes in the phenolic synthesis pathway (*CHS, CHI, FHT, DFR, ANS, LAR, ANR, FLS,* and *GT*), and for others that could be involved with substrates and products of our interest (anthocyanidin, flavonoid, flavonol, flavanone, chalcone, cinnamate, coumarate). As a result, a list of 340 candidate genes (not biochemically characterized) spread over all chromo-

somes, and some not anchored to the genome was drew up. The 255 located in regions covered by our mQTLs were selected (Supplemental Table 6). Finally, the phenotypic changes and the list of candidate genes associated with each QTL interval, have been studied case by case, to find any possible link between them. Located in the LG1:26-61 cM region was the biochemically characterized gene 14,611, known as *FaF3H* or *FaFHT* [8], characterized as a flavanone-3-hydroxylase and that could be related to the special balance of flavonols in NILs carrying *F. bucharica* introgression in this region. The *FaCAD1* and *FaCAD2* genes, characterized as cinnamyl alcohol dehydrogenases, were also located in this region. In addition there was one predicted glucosyltransferase (gene14947) whose expression has been inversely related to kaempferol glucuronide accumulation in previous studies with octoploid varies of *Fragaria* (data not shown). There were a three further predicted monooxygenases (gene12513, gene12514, and gene12515) that could be involved with the observed mQTLs.

In section LG2:0–30 cM the gene 11,126, biochemically characterized as flavonol synthase (*FaFLS*) [8], has been found, and four other genes annotated as FLS but without empirical evidence (gene01063, gene01034, gene23698, and gene11130). These could cause the decrease in flavan-3-ols as they compete for the same substrates [8]. The functions of other genes, annotated as coumarate — CoA ligase (gene24683), CHS (gene10965, gene10966), and CHI (gene27804), are implicated in the early steps of the phenylpropanoid pathway and so could be related to the increased accumulation of *p*-coumaroyls and cinnamoyl-glucose.

QTLs at region LG2:45–63 cM co-localize with the dihydroflavonol 4-reductases *FaDFR* and *FaDFR1* (gene15174) [8], and another gene annotated as DFR (gene15176) could be related to the effect of the mQTL that over-accumulates cyanidin-
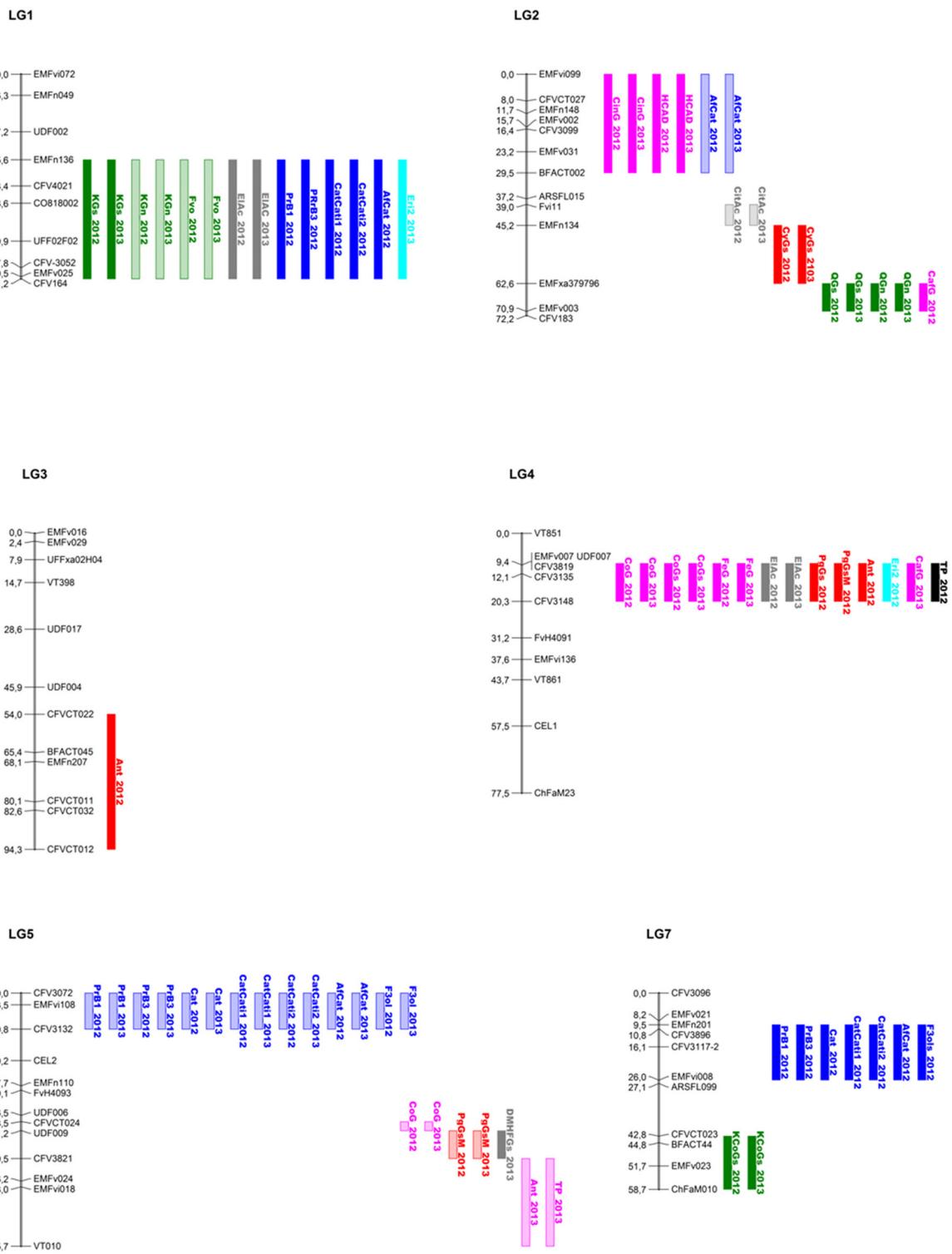
Fig. 6. mQTL: Graphical representation of strong mapped mQTL (in the same direction, significantly different from RV by *t*-test with a *p*-value after correction <0.05 in at least one harvest and all lines harboring the same introgressed region, LOD > 1.8, >20% of variability explained). Linkage groups (LG) are shown by a black line specifying name and position (in cM) of the SSRs used for genotyping. mQTLs are represented by colored bars according to chemical families as in Fig. 4 (red, anthocyanins; green, flavonols; dark blue, flavan-3-ols; pink, hydroxycinnamic acid derivatives; sky blue, flavanones; black, total (poly) phenols; grey, others). The effect of the introgression on the accumulation of the compound is shown by solid (increased accumulation) or striped (decreased accumulation) bars. The name of the specific metabolite/chemical family is on the right of the bar followed by the harvest year in which it was detected. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

glucoside. In a neighboring region (LG2:63–73 cM) there was another mQTL over-accumulating quercetin glucoside (derived from dihydroquercetin as the cyanidin-glucoside) and two mQTL under-accumulating pelargonidin-based anthocyanins (derived

from dihydrokaempferol). In this region (0.79 Mb) one gene predicted as a leucoanthocyanidin dioxygenaseanthocyanin (LDOX) (gene15078) that may be related with the observed mQTLs has been

Table 4

mQTL location and predicted genes. Genetic (cM) and physical (bp) mQTL position. Predicted genes (total number) and genes putatively involved in phenolic synthesis enclosed in the mQTL regions and a selection of candidate genes.

mQTL	LG	Possition (cM)		Possition (bp)		Size	Predicted genes		Candidate genes
		Start	End	Start	End	(Mb)	Total	Phenolic-pathway	
1:26–61	1	26	61	3,315,998	20,747,404	17.43	2629	29	gene14611 (FaF3H), gene20700 (FaCAD1, FaCAD2), gene14947, gene12513, gene12514, gene12515
2:0–30	2	0	30	0	16,919,611	16.92	2619	32	gene11126 (FaFLS), gene01063, gene01034, gene23698, gene11130, gene24683,gene10965, gene10966, gene27804
2:30–39	2	30	39	16,919,611	18,950,307	2.03	398	4	
2:39–45	2	39	45	18,950,307	21,299,464	2.35	488	7	
2:45–63	2	45	63	22,329,904	22,663,853	0.33	81	3	gene15174 (FaDFR, FaDFR1), gene15176
2:63–73	2	63	73	21,489,169	22,279,702	0.79	163	3	gene15078
3:0–8	3	0	8	315,550	1,632,631	1.32	358	8	
3:54–94	3	54	94	5,668,456	31,977,233	26.31	3812	33	gene27138, gene28898, gene28093, gene00126
4:0–9	4	0	9	0	1,708,726	1.91	235	3	
4:9–20	4	9	20	1,708,726	22,629,816	20.92	2855	25	gene07080 (Fraa1A), gene07065 (Fraa2), gene07082 (Fraa3)
4:20–44	4	20	44	22,629,816	26,554,022	3.92	607	9	
5:0–11	5	0	11	0	2,346,373	2.35	471	12	gene32347 (FaANS), gene24665 (FaANR)
5:11–35	5	11	35	2,346,373	6,412,800	4.07	741	15	
5:39–41	5	39	41	6,928,955	7,246,484	0.32	75	1	
5:41–50	5	41	50	7,246,484	11,167,415	3.92	715	15	gene25801, gene09407 (FaMYB1)
5:50–76	5	50	76	11,167,415	28,401,116	17.23	2528	31	gene29344, gene31464, gene31465, gene02203, gene13530, gene22073
6:84–101	6	84	101	31,889,065	32,907,471	1.02	189	1	
6:101–101	6	101	101	32,907,471	39,317,498	6.41	1234	36	gene26265 (FaGT2), gene26344 (FaGT7), gene 24019, gene 28428, gene 26403, gene24025, gene26301, gene26302
7:0–10	7	0	10	0	14,053,681	14.05	2169	38	gene26825 (FaCHS, FaCHS2, FaCHS3), gene26826 (FaCHS4)
7:10–26	7	10	26	14,053,681	17,678,937	3.63	652	16	gene20980, gene23397
7:43–59	7	43	59	19,453,937	21,046,215	1.59	325	4	gene12591 (FaFGT)

detected. The closeness of these two regions leads us to think that the end of LG2 is important for anthocyanin regulation.

The LG3:54–94 cM genetic interval is very long (26.31 Mb) and harbors 32 candidate genes (none of them biochemically characterized). According to the annotated functions, three genes might be involved in early steps of the pathway (gene27138, gene28898, and gene28093 putatively annotated as chalcone isomerase, coumarate-CoA ligase and trans-cinnamate 4-monooxygenase, respectively), and the gene00126, putatively annotated as anthocyanin glucosyltransferase that could be involved in the increased accumulation of anthocyanins.

Region LG4:9–20 cM covers a central region of LG4 and harbors genes such as *Fraa1A*, *Fraa2,* and *Fraa3* that have been reported to be important in the control of the flavonoid pathway and pigment accumulation during fruit ripening. It has been seen that expression

affects flavonoids and phenylpropanoids accumulation [45]. This means that alleles introduced by *F. bucharica* introgression may affect the expression of these genes and favor the accumulation of hydroxycinnamic acid derivatives and anthocyanins.

At the LG5:0–11 cM genetic section, there are nine candidate genes. Among them, *FaANS* (gene32347) and *FaANR* (gene24665) have been functionally characterized as anthocyanin synthase and reductase, respectively [8]. *FaANR* catalyses the transformation of anthocyanidins (pelargonidin and cyanidin) into flavan-3-ols (epi-catechin and epiafzelechin), and could therefore be responsible for the under-accumulation of flavan-3-ols in lines carrying a *F. bucharica* introgression in this region.

Gene25801 has been characterized as a flavonoid 3′hydroxylase, adding hydroxyl groups in position 3′ to dihydrokaempferol, kaempferol, apigenin, and naringenin, and has been shown to be

related to anthocyanin compounds balance [46]. It co-locates with the QTL described in LG5:41–50 cM, so could be involved in the decreased accumulation of pelargonidin-glucoside-malonate. In addition, the transcription factor *FaMYB1*, related to the production of flavonols and anthocyanins during fruit ripening, has been localised in this same region [8].

The genetic region in LG5:50–76 cM gathers 29 genes annotated as being related to the phenylpropanoid and flavonoid pathway, but none have been biochemically characterized. Among them, there are four genes predicted as DFR (gene29344, gene31464, gene31465, gene02203), and one as flavonol glucosyltransferase (gene13530) that could be involved in the increased accumulation of kaempferol-glucosides. Gene22073, annotated as a putative leucoanthocyanidin dioxygenase that could be related to the observed QTLs involving total anthocyanin and total (poly)-phenols accumulation, has also been located in this region.

In the LG6:101–101 cM interval, there were 33 candidate genes. Among them, 25 are annotated as probable glucosyltransferases and two have been functionally characterized: *FaGT2* (gene26265) [47] and *FaGT7* [48]. They could be related with the general decrease in glucosides. Other genes in this region, gene 24,019 (GT), gene 28,428 (DFR) and gene 26,403 (GT), are presumably related to pelargonidin-3-glucoside-malonate and kaempferol-coumaryl-glucoside accumulation in octoploid varieties (data not shown) and could here be related with the mQTL for PgGsM and KGs mapped in this region. There were also three predicted cinnamyl alcohol dehydrogenase (gene24025, gene26301 and gene26302), that could be related to the decrease in accumulation of cinnamoyl-glucose.

Localised at region LG7:0–10 cM are four biochemically characterized chalcone synthases: *FaCHS*, *FaCHS2*, *FaCHS3,* and *FaCHS4,* the first three corresponding to gene26825 and the last to gene26826 [8]. These enzymes have been shown to play a central role in the flavonoid metabolic pathway, so could be involved in the observed under-accumulation of flavanones and flavan-3-ols. Other interesting annotated genes (without empirical evidence) in this region include two probable LAR and four DFR that could be related to the lower accumulation of flavan-3-ols, and one flavonoid monooxygenase and two flavanone glucosyltransferases that could be involved in the decrease of eriodictyol isomers.

Within the boundaries of the region LG7:10–26 cM there are 16 candidate genes but none have been biochemically characterized. Functions predicted for these genes are as leucoanthocyanidin dioxigenase (gene 20,980) and chalcone isomerase (gene 23,397) that could be related to the general increased accumulation of flavan-3-ols.

Region LG7:43–59 cM includes one functionally characterized anthocyanidin glucosyltransferase (gene12,591 named *FaFGT* after Almeida et al. [8] or *FaGT1* after Griesser et al. [48]) and three other genes annotated as flavonoid glucosyltransferases.

4. Discussion

F. vesca, a source of variability for poly-phenol content in strawberry breeding programs

There are many factors contributing to (poly)-phenol compounds concentration in strawberry fruits, such as genotype, environmental conditions and growing practices, the degree of maturity at harvest, the procedures and solvents used for extraction, the use of fresh or dry material and the quantification method. So many variable parameters make comparisons between studies difficult. To our knowledge, no previous publication analyses the (poly)-phenolic content of a mapping population in diploid strawberry (*F. vesca*). One previous study reported on the phenolic content of a single variety of *F. vesca* [32] but quantification

was using dry material (DW) and a different internal standard, so absolute values are not comparable with our data, however the proportions between metabolites that we report in this work are similar to theirs. For instance, they reported similar average concentrations of pelargonidin-3-glucoside and cyanidin-3-glucoside in contrast to studies on octoploid varieties that found much higher average concentrations of pelargonidin-3-glucoside. They also report lower values of quercetin-3-glucuronide than the other flavonols detected, contrasting with studies on octoploid material where quercetin-3-glucuronide is the major flavonol [31,33,41,43]. This suggests that the relative accumulation of (poly)-phenolic compounds is species-specific. Analyses of correlation, clustering and principal components between metabolites in *F. vesca* NIL collection demonstrate that co-regulation of metabolites is controlled by proximal genetic regions.

Comparing our results with studies using fresh material from octoploid varieties, we found that the average concentration for flavonoids was similar to (ellagic acid, flavonols — except for kaempferol-glucuronide) or higher than (anthocyanins, kaempferol-glucuronide) that reported for octoploid varieties. But even more interesting is that the range of concentrations was much broader within our NIL collection than the variability found between octoploid varieties [31,33]. Fewer differences were found between the average concentrations of hydroxycinnamic acid derivatives. The concentration data summarized here are similar to those reported by other studies with octoploid varieties [31,33,41,49].

As a result, we can consider that *F. vesca* is a good source of variability in terms of quantity and balance, and could be used to improve the (poly)-phenolic content and composition of *Fragaria × ananassa* in breeding programs. It has been shown that *F. vesca* accumulates higher average concentrations of several (poly)-phenols, mainly anthocyanins but also ellagic acid, and still has great potential for improvement as the range of variation is wide within our population.

The genotype and environment effects on (poly)-phenols accumulation

The accumulation pattern of (poly)-phenolic compounds in strawberry fruits in a NIL collection revealed the important effect of the genotype and the environmental conditions on their accumulation. A previous study by Carbone et al. [9], with a *Fragaria × ananassa* breeding population segregating for flavonoid production, found significant genotype-dependent differences at each developmental stage for most analyzed flavonoids. The accumulation of flavonols was also most affected by the environment. In our NIL collection a major effect of genotype on flavonoid and phenylpropanoid accumulation was observed, and the chemical family most affected by the environment was total flavonols, although other chemical families were also affected. Analysis of data revealed that the balance between the metabolites of each class was more stable than the absolute quantification for anthocyanins, flavonols, flavan-3-ols and flavanones, suggesting that the relative abundance of compounds is more tightly regulated than their total accumulation. This was not the case for hydroxycinnamic acid derivatives, where an increase in cinnamoyl-glucose ester not coupled with any variation in the accumulation of the other hydroxycinnamic acid derivatives was observed, suggesting these compounds are not co-regulated (Table 2).

Besides the general genotype effect on the (poly)-phenolic accumulation, not all were equally affected by G, E or G × E. These parameters should be taken into account for selective (poly)-phenolic accumulation breeding as they allow us to discard those compounds with more unpredictable behavior and concentrate efforts on those highly influenced by the genotype and little

affected by the environment. For instance, accumulation of *p*-coumaroyl-glucoside was independent of the environment and had a single positive mQTL at LG2:0–30 cM, so it could be a good candidate compound to improve HCAD content. Another good target might be the accumulation of the anthocyanins PgGs and CyGs, that represent a high percentage of total (poly)-phenols, are highly affected by the genotype, and have a single mQTL each.

A. NIL collection as a tool for detecting variability in poly-phenol compounds

Introgression line collections have previously been used to successfully map metabolic QTL in tomato [50–52], melon [53] and *Arabidopsis* [54]. The great degree of variability observed in the (poly)-phenolic profiling among the *F. vesca* NIL collection and the 72 mQTLs mapped reveal that introgressions from the exotic species *F. bucharica* have a significant effect in the accumulation of (poly)-phenols in the *F. vesca* genetic background. This, together with a previous study with introgression lines in tomato (introgression from *S. pennellii* in the genetic background of *S. lycopersicum*) that allowed identification of candidate genes for phenolics accumulation [55], leads us to consider the *F. vesca* NIL collection an appropriate tool for genetic dissection and candidate gene identification in biochemical pathways controlling (poly)-phenols synthesis. The global view of the mapped mQTLs suggests that some introgressed regions could be related to a specific biological function. Anthocyanin accumulation is crucial for strawberry color and total polyphenol content. Pelargonidin-3-glucoside is known to be the majoritary Anthocyanin in *Fragaria × ananassa*, however *F. vesca* accumulates also significative levels of cyanidin-3-glucoside. A previous work by Thill et al. [46] has pointed out to *F3'H* gene as causal for the lower perlargonidin derivatives accumulation in *F. vesca*. This gene co-locates with a major QTL under-accumulating pelargonidin-3-glucosside-malonate in the NIL collection (LG5:41–50). Additionally, a major QTL for the over-accumulation of cyanidin-3-glucoside was detected in LG2:45–63 and NILs harboring a *F. bucharica* introgression on this bin accumulate similar levels of cyanidin-3-glucoside and pelargonidin-3-glucoside. This may indicate that there are unexplored sources of variability for anthocyanin composition revealed by the *F. bucharica* alleles. *FaDFR1* and other putative DFR genes annotated in this region are proposed as starting point for functional analysis. In addition, as cyanidin-3-glucoside represents around the 40% of anthocyanins, by far the most abundant family of (poly)-phenols, line Fb**2**:39–63, harboring the LG2:45–63 cM mQTL, may represent a direct improvement in nutritional quality over *F. vesca* RV.

Differences between *F. vesca* and *Fragaria × ananassa* in flavonols accumulations are clear, being kaempferol- and quercetin-glucuronide the most abundant flavonols respectively [31,33,41,43]. The non previously described major QTL that explain kaempferol-glucuronide and kaempferol-glucoside accumulation in LG1:26–76 and quercetin-glucuronide and quercetin-glucoside in LG2:63–73 might help to elucidate kaempferol and quercetin derivatives regulation in *F. vesca*.

Other genetic regions also harbor mQTL of interest. NIL with exotic introgressions in the central region of LG4 (LG4:9–20) have an interesting (poly)-phenolic phenotype, rich in hydroxycinnamic acid derivatives and ellagic acid, and, under specific environmental conditions, can over-accumulate total (poly)-phenols. Lines harboring introgressions at region LG6:101–101 cM had an under-accumulation of several glucosides, with 25 of the candidate genes in this region being glucosyltransferases (including two biochemically characterized by Lunkenbein et al. [47] and Griesser et al. [48]). This finding points to an effect of the alleles from *F.*

bucharica on glucosyltransferase activity. Another interesting set of mQTLs mapped in LG7:0–10 cM, where the under-accumulation of both flavanones and flavan-3-ols suggests that the exotic introgressed alleles could hinder the transformation from naringenin to eriodictyol and block the transformation of leucocyanindin and cyaninidin-3-glucoside to catechins. These strong QTL that involve important functional polyphenolic metabolites directly point to *F. vesca* genetic regions controlling (poly) phenol biosynthesis. Further positional cloning studies would lead to fine-mapping those regions and candidate genes characterization.

A previous study with this same NIL collection described 2 QTL for total phenol accumulation in LG2:0–30 cM and LG5:50–76 cM, using the Folin–Ciocalteu method [19]. This technique is non-specific and all anti-oxidant substrates in the sample (including phenolics not identified by LC–MS but also other non-phenolic compounds) can interfere in the quantification results. In addition, not all phenolics have the same molar absorptivity in the assay and therefore do not contribute proportionally to the total quantification [56]. Therefore, these results are not comparable with our total (poly)-phenol quantification, as we sum up the identified phenols by their weights (mg/10 g) not by their antioxidant capacity. However, total (poly)-phenol content calculated by LC–MS showed genetic variation in the NIL collection (Fig. 3, Table 4) and although this was not enough to map stable QTLs, two one-year QTL were mapped at regions LG4:9–20 cM and LG5:50–76 cM. The latter QTL co-localised in position and direction of accumulation with the QTL in LG5:50–76 cM described by Urrutia et al. [19]. The positive QTL described in LG2:0–30 cM by the Folin–Ciocalteu method co-localises with positive and stable QTLs for cinnamoyl-glucose and total hydroxycinnamic acid derivatives.

Introgression of wild species and NIL collections are a simple way to introduce variability in (poly)-phenolic composition and this is the first complete analysis in wild strawberry fruit. The new genetic regions controlling flavonoids accumulation described here open a way for further candidate gene characterization and provide a new powerful tool to increase the antioxidant capacity of fruits and to produce healthier strawberries for consumers.

Acknowledgements

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.plantsci.2015.07.019.

References

[1] J.A. Ross, C.M. Kasum, Dietary flavonoids: bioavailability, metabolic effects, and safety, Annu. Rev. Nutr. 22 (2002) 19–34.

[2] A. Scalbert, C. Manach, C. Morand, C. Remesy, L. Jimenez, Dietary polyphenols and the prevention of diseases, Crit. Rev. Food Sci. Nutr. 45 (2005) 287–306.

[3] C. Ramassamy, Emerging role of polyphenolic compounds in the treatment of neurodegenerative diseases: a review of their intracellular targets, Eur. J. Pharmacol. 545 (2006) 51–64.

[4] N.P. Seeram, Berry fruits: compositional elements, biochemical activities, and the impact of their intake on human health, performance, and disease, J. Agric. Food Chem. 56 (2008) 627–629.

[5] K.S. Gould, C. Lister, Flavonoid functions in plants, in: Ø.M. Andersen, K.R. Markham (Eds.), Flavonoids: Chemistry, Biochemistry and Applications, CRC Press, Boca Raton, 2006, pp. 397–411.

[6] T. Vogt, Phenylpropanoid biosynthesis, Mol. Plant 3 (2010) 2–20.

[7] L. Lepiniec, I. Debeaujon, J.M. Routaboul, A. Baudry, L. Pourcel, N. Nesi, M. Caboche, Genetics and biochemistry of seed flavonoids, Annu. Rev. Plant Biol. 57 (2006) 405–430.

[8] J.R. Almeida, E. D'Amico, A. Preuss, F. Carbone, C.H. de Vos, B. Deiml, F. Mourgues, G. Perrotta, T.C. Fischer, A.G. Bovy, S. Martens, C. Rosati, Characterization of major enzymes and genes involved in flavonoid and proanthocyanidin biosynthesis during fruit development in strawberry (*Fragaria × ananassa*), Arch. Biochem. Biophys. 465 (2007) 61–71.

[9] F. Carbone, A. Preuss, R.C. De Vos, E. D'Amico, G. Perrotta, A.G. Bovy, S. Martens, C. Rosati, Developmental, genetic and environmental factors affect the expression of flavonoid genes, enzymes and metabolites in strawberry fruits, Plant Cell Environ. 32 (2009) 1117–1131.

[10] J.A. Tennessen, R. Govindarajulu, T.L. Ashman, A. Liston, Evolutionary origins and dynamics of octoploid strawberry subgenomes revealed by dense targeted capture linkage maps, Genome Biol. Evol. 6 (2014) 3295–3313.

[11] J.J. Ruiz-Rojas, D.J. Sargent, V. Shulaev, A.W. Dickerman, J. Pattison, S.H. Holt, A. Ciordia, R.E. Veilleux, SNP discovery and genetic mapping of T-DNA insertional mutants in *Fragaria vesca* L, Theor. Appl. Genet. 121 (2010) 449–463.

[12] S. Vilanova, D.J. Sargent, P. Arus, A. Monfort, Synteny conservation between two distantly-related Rosaceae genomes: *Prunus* (the stone fruits) and *Fragaria* (the strawberry), BMC Plant Biol. 8 (2008) 67–78.

[13] M. Rousseau-Gueutin, E. Lerceteau-Köhler, L. Barrot, D.J. Sargent, A. Monfort, D. Simpson, P. Arús, G. Guérin, B. Denoyes-Rothan, Comparative genetic mapping between octoploid and diploid *Fragaria* species reveals a high level of colinearity between their genomes and the essentially disomic behavior of the cultivated octoploid strawberry, Genetics 179 (2008) 2045–2060.

[14] E. Illa, D.J. Sargent, E. Lopez Girona, J. Bushakra, A. Cestaro, R. Crowhurst, M. Pindo, A. Cabrera, E. van der Knaap, A. Iezzoni, S. Gardiner, R. Velasco, P. Arus, D. Chagne, M. Troggio, Comparative analysis of rosaceous genomes and the reconstruction of a putative ancestral genome for the family, BMC Evol. Biol. 11 (2011) 9–22.

[15] S. Jung, A. Cestaro, M. Troggio, D. Main, P. Zheng, I. Cho, K.M. Folta, B. Sosinski, A. Abbott, J.M. Celton, P. Arus, V. Shulaev, I. Verde, M. Morgante, D. Rokhsar, R. Velasco, D.J. Sargent, Whole genome comparisons of *Fragaria*, *Prunus* and *Malus* reveal different modes of evolution between Rosaceous subfamilies, BMC Genomics 13 (2012) 129–141.

[16] V. Shulaev, D.J. Sargent, R.N. Crowhurst, T.C. Mockler, O. Folkerts, A.L. Delcher, P. Jaiswal, K. Mockaitis, A. Liston, S.P. Mane, P. Burns, T.M. Davis, J.P. Slovin, N. Bassil, R.P. Hellens, C. Evans, T. Harkins, C. Kodira, B. Desany, O.R. Crasta, R.V. Jensen, A.C. Allan, T.P. Michael, J.C. Setubal, J.-M. Celton, D.J.G. Rees, K.P. Williams, S.H. Holt, J.J.R. Rojas, M. Chatterjee, B. Liu, H. Silva, L. Meisel, A. Adato, S.A. Filichkin, M. Troggio, R. Viola, T.-L. Ashman, H. Wang, P. Dharmawardhana, J. Elser, R. Raja, H.D. Priest, D.W. Bryant, S.E. Fox, S.A. Givan, L.J. Wilhelm, S. Naithani, A. Christoffels, D.Y. Salama, J. Carter, E.L. Girona, A. Zdepski, W. Wang, R.A. Kerstetter, W. Schwab, S.S. Korban, J. Davik, A. Monfort, B. Denoyes-Rothan, P. Arus, R. Mittler, B. Flinn, A. Aharoni, J.L. Bennetzen, S.L. Salzberg, A.W. Dickerman, R. Velasco, M. Borodovsky, R.E. Veilleux, K.M. Folta, The genome of woodland strawberry (*Fragaria vesca*), Nat. Genet. 43 (2011) 109–116.

[17] D.J. Sargent, G. Cipriani, S. Vilanova, D. Gil-Ariza, P. Arús, D.W. Simpson, K.R. Tobutt, A. Monfort, The development of a bin mapping population and the selective mapping of 103 markers in the diploid *Fragaria* reference map, Genome 51 (2008) 120–127.

[18] R.E. Veilleux, K.P. Mills, A.J. Baxter, K.T. Upham, T.J. Ferguson, S.H. Holt, N. Lu, J.J. Ruiz-Rojas, C.J. Pantazis, C.M. Davis, R.C. Lindsay, F.L. Powell, Y. Dan, A.W. Dickerman, T. Oosumi, V. Shulaev, Transposon tagging in diploid strawberry, Plant Biotechnol. J. 10 (2012) 985–994.

[19] M. Urrutia, J. Bonet, P. Arús, A. Monfort, A near-isogenic line (NIL) collection in diploid strawberry and its use in the genetic analysis of morphologic, phenotypic and nutritional characters, Theor. Appl. Genet. (2015) 1–15.

[20] N.V. Bassil, T.M. Davis, H. Zhang, S. Ficklin, M. Mittmann, T. Webster, L. Mahoney, D. Wood, E.S. Alperin, U.R. Rosyara, H. Koehorst-vanc Putten, A. Monfort, D.J. Sargent, I. Amaya, B. Denoyes, L. Bianco, T. van Dijk, A. Pirani, A. Iezzoni, D. Main, C. Peace, Y. Yang, V. Whitaker, S. Verma, L. Bellon, F. Brew, R. Herrera, E. van de Weg, Development and preliminary evaluation of a 90K Axiom® SNP array for the allo-octoploid cultivated strawberry *Fragaria × ananassa*, BMC Genomics 16 (2015) 155–185.

[21] Y. Eshed, D. Zamir, An Introgression line population of lycopersicon pennellii in the cultivated tomato enables the identification and fine mapping of yield-associated Qtl, Genetics 141 (1995) 1147–1162.

[22] A.J. Monforte, S.D. Tanksley, Development of a set of near isogenic and backcross recombinant inbred lines containing most of the *Lycopersicon hirsutum* genome in a L. esculentum genetic background: a tool for gene mapping and gene discovery, Genome 43 (2000) 803–813.

[23] A.J. Monforte, E. Friedman, D. Zamir, S.D. Tanksley, Comparison of a set of allelic QTL-NILs for chromosome 4 of tomato: deductions about natural variation and implications for germplasm utilization, Theor. Appl. Genet. 102 (2001) 572–590.

[24] I. Eduardo, P. Arus, A.J. Monforte, Development of a genomic library of near isogenic lines (NILs) in melon (*Cucumis melo* L.) from the exotic accession PI161375, Theor. Appl. Genet. 112 (2005) 139–148.

[25] E. Moreno, J.M. Obando, N. Dos-Santos, J.P. Fernandez-Trujillo, A.J. Monforte, J. Garcia-Mas, Candidate genes and QTLs for fruit ripening and softening in melon, Theor. Appl. Genet. 116 (2008) 589–602.

[26] J. Vegas, J. Garcia-Mas, A.J. Monforte, Interaction between QTLs induces an advance in ethylene biosynthesis during melon fruit ripening, Theor. Appl. Genet. 126 (2013) 1531–1544.

[27] E.G. Pestsova, A. Borner, M.S. Roder, Development of a set of *Triticum aestivum-Aegilops tauschii* introgression lines, Hereditas 135 (2001) 139–143.

[28] M.J. Jeuken, P. Lindhout, The development of lettuce backcross inbred lines (BILs) for exploitation of the *Lactuca saligna* (wild lettuce) germplasm, Theor. Appl. Genet. 109 (2004) 394–401.

[29] R. Koumproglou, T.M. Wilkes, P. Townson, X.Y. Wang, J. Beynon, H.S. Pooni, H.J. Newbury, M.J. Kearsey, STAIRS: a new genetic resource for functional genomic studies of *Arabidopsis*, Plant J. Cell Mol. Biol. 31 (2002) 355–364.

[30] R. Fletcher, J. Mullen, S. Yoder, W. Bauerle, G. Reuning, S. Sen, E. Meyer, T. Juenger, J. McKay, Development of a next-generation NIL library in *Arabidopsis thaliana* for dissecting complex traits, BMC Genomics 14 (2013) 1–13.

[31] B. Buendia, M.I. Gil, J.A. Tudela, A.L. Gady, J.J. Medina, C. Soria, J.M. Lopez, F.A. Tomas-Barberan, HPLC–MS analysis of proanthocyanidin oligomers and other phenolics in 15 strawberry cultivars, J. Agric. Food Chem. 58 (2010) 3916–3926.

[32] C. Munoz, J.F. Sanchez-Sevilla, M.A. Botella, T. Hoffmann, W. Schwab, V. Valpuesta, Polyphenol composition in the ripe fruits of *Fragaria* species and transcriptional analyses of key genes in the pathway, J. Agric. Food Chem. 59 (2011) 12598–12604.

[33] K. Aaby, S. Mazur, A. Nes, G. Skrede, Phenolic compounds in strawberry (*Fragaria × ananassa* Duch.) fruits: composition in 27 cultivars and changes during ripening, Food Chem. 132 (2012) 86–97.

[34] RCoreTeam, R, A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, 2012.

[35] RStudio, R Studio, Integrated Development Environment for R, RStudio, Boston, MA, USA, 2012.

[36] F.E., Harrell, Hmisc: Harrell miscellaneous R package version 3. 14-3, in, 2014.

[37] J.W., Fox, S., An R companion to applied regression. R package version 2. 0-19 in, Thousand Oaks, CA, USA, 2011.

[38] S.C., Epskamp, G., Cramer, A.O.J., Waldorp, L.J., Schmittmann, VD., Borsboom, D., Network representations of relationships in data. R package version 1.2.4, in, 2012.

[39] J.M., Van Ooijen, MapQTL version 6.0: Software for the mapping of quantitative trait loci in experimental populations of diploid species, in Kazyma, Wageningen, The Netherlands, 2009.

[40] R.E. Voorrips, MapChart: software for the graphical presentation of linkage maps and QTLs, J. Heredity 93 (2002) 77–78.

[41] K.R. Määttä-Riihinen, A. Kamal-Eldin, A.R. Törrönen, Identification and quantification of phenolic compounds in berries of *Fragaria* and *Rubus* Species (Family Rosaceae), J. Agric. Food Chem. 52 (2004) 6178–6187.

[42] A. Fait, K. Hanhineva, R. Beleggia, N. Dai, I. Rogachev, V.J. Nikiforova, A.R. Fernie, A. Aharoni, Reconfiguration of the achene and receptacle metabolic networks during strawberry fruit development, Plant Physiol. 148 (2008) 730–750.

[43] L. Ring, S.Y. Yeh, S. Hucherig, T. Hoffmann, R. Blanco-Portales, M. Fouche, C. Villatoro, B. Denoyes, A. Monfort, J.L. Caballero, J. Munoz-Blanco, J. Gershenson, W. Schwab, Metabolic interaction between anthocyanin and lignin biosynthesis is associated with peroxidase FaPRX27 in strawberry fruit, Plant Physiol. 163 (2013) 43–60.

[44] L. Medina-Puche, G. Cumplido-Laso, F. Amil-Ruiz, T. Hoffmann, L. Ring, A. Rodriguez-Franco, J.L. Caballero, W. Schwab, J. Munoz-Blanco, R. Blanco-Portales, MYB10 plays a major role in the regulation of flavonoid/phenylpropanoid metabolism during ripening of *Fragaria × ananassa* fruits, J. Exp. Bot. 65 (2014) 401–417.

[45] C. Muñoz, T. Hoffmann, N.M. Escobar, F. Ludemann, M.A. Botella, V. Valpuesta, W. Schwab, The Strawberry fruit fra a allergen functions in flavonoid biosynthesis, Mol. Plant 3 (2010) 113–124.

[46] J. Thill, S. Miosic, T.P. Gotame, M. Mikulic-Petkovsek, C. Gosch, R. Veberic, A. Preuss, W. Schwab, F. Stampar, K. Stich, H. Halbwirth, Differential expression of flavonoid 3′-hydroxylase during fruit development establishes the different B-ring hydroxylation patterns of flavonoids in *Fragaria × ananassa* and *Fragaria vesca*, Plant Physiol. Biochem. 72 (2013) 72–78.

[47] S. Lunkenbein, M. Bellido, A. Aharoni, E.M. Salentijn, R. Kaldenhoff, H.A. Coiner, J. Munoz-Blanco, W. Schwab, Cinnamate metabolism in ripening fruit. Characterization of a UDP-glucose:cinnamate glucosyltransferase from strawberry, Plant Physiol. 140 (2006) 1047–1058.

[48] M. Griesser, B. Vitzthum, B. Fink, M.L. Bellido, C. Raasch, J. Munoz-Blanco, W. Schwab, Multi-substrate flavonol O-glucosyltransferases from strawberry (*Fragaria × ananassa*) achene and receptacle, J. Exp. Bot. 59 (2008) 2611–2625.

[49] K. Aaby, R.E. Wrolstad, D. Ekeberg, G. Skrede, Polyphenol composition and antioxidant activity in strawberry purees; impact of achene level and storage, J. Agric. Food Chem. 55 (2007) 5156–5166.

[50] M.C. Rousseaux, C.M. Jones, D. Adams, R. Chetelat, A. Bennett, A. Powell, QTL analysis of fruit antioxidants in tomato using Lycopersicon pennellii introgression lines, Theor. Appl. Genet. 111 (2005) 1396–1408.

[51] N. Schauer, Y. Semel, U. Roessner, A. Gur, I. Balbo, F. Carrari, T. Pleban, A. Perez-Melis, C. Bruedigam, J. Kopka, L. Willmitzer, D. Zamir, A.R. Fernie, Comprehensive metabolic profiling and phenotyping of interspecific introgression lines for tomato improvement, Nat. Biotechnol. 24 (2006) 447–454.

[52] S. Alseekh, T. Tohge, R. Wendenberg, F. Scossa, N. Omranian, J. Li, S. Kleessen, P. Giavalisco, T. Pleban, B. Mueller-Roeber, D. Zamir, Z. Nikoloski, A.R. Fernie, Identification and mode of inheritance of quantitative trait loci for secondary

metabolite abundance in tomato, Plant Cell 27 (2015) 485–512.

[53] J.M. Obando-Ulloa, I. Eduardo, A.J. Monforte, J.P. Fernandez-Trujillo, Identification of QTLs related to sugar and organic acid composition in melon using near-isogenic lines, Sci. Hortic. Amsterdam 121 (2009) 425–433.

[54] J. Lisec, R.C. Meyer, M. Steinfath, H. Redestig, M. Becher, H. Witucka-Wall, O. Fiehn, O. Torjek, J. Selbig, T. Altmann, L. Willmitzer, Identification of metabolic and biomass QTL in *Arabidopsis thaliana* in a parallel analysis of RIL and IL populations, Plant J. 53 (2008) 960–972.

[55] A. Di Matteo, V. Ruggieri, A. Sacco, M.M. Rigano, F. Carriero, A. Bolger, A.R. Fernie, L. Frusciante, A. Barone, Identification of candidate genes for phenolics accumulation in tomato fruit, Plant Sci. Int. J. Exp. Plant Biol. 205–206 (2013) 87–96.

[56] V.L. Singleton, R. Orthofer, R.M. Lamuela-Raventós, Analysis of total phenols and other oxidation substrates and antioxidants by means of folin-ciocalteu reagent, in: P. Lester (Ed.), Methods in Enzymology, Academic Press, 1999, pp. 152–178.