Accepted Manuscript

Genetic analysis of the wild strawberry (*Fragaria vesca*) volatile composition

María Urrutia, José L. Rambla, Konstantinos G. Alexiou, Antonio Granell, Amparo Monfort

PII: S0981-9428(17)30342-X

DOI: 10.1016/j.plaphy.2017.10.015

Reference: PLAPHY 5027

To appear in: *Plant Physiology and Biochemistry*

Received Date: 15 August 2017

Revised Date: 13 October 2017

Accepted Date: 17 October 2017

Please cite this article as: Marí. Urrutia, José.L. Rambla, K.G. Alexiou, A. Granell, A. Monfort, Genetic analysis of the wild strawberry (*Fragaria vesca*) volatile composition, *Plant Physiology et Biochemistry* (2017), doi: 10.1016/j.plaphy.2017.10.015.

GENETIC ANALYSIS OF THE WILD STRAWBERRY *(Fragaria vesca)* VOLATILE COMPOSITION

María Urrutia[1], José L. Rambla[2], Konstantinos G. Alexiou[1,3], Antonio Granell[2], Amparo Monfort[1,3*]

[1]Centre for Research in Agricultural Genomics (CRAG) CSIC-IRTA-UAB-UB, Campus UAB, Bellaterra, Barcelona, Spain

[2] Instituto de Biología Molecular y Celular de Plantas (IBMCP), Universidad Politécnica de Valencia (UPV)-Consejo Superior de Investigaciones Científicas (CSIC), Ingeniero Fausto Elio, 46022 Valencia, Spain

[3]Institut de Recerca i Tecnologia Agroalimentàries (IRTA), Barcelona, Spain.

Author for correspondence:

Amparo Monfort, IRTA. Center for Research in Agricultural Genomics (CSIC-IRTA-UAB-UB), Campus UAB, 08193 Bellaterra, Barcelona, Spain

amparo.monfort@irta.es

Abstract:

The volatile composition of wild strawberry (*Fragaria vesca*) fruit differs from that of the cultivated strawberry, having more intense and fruity aromas. Over the last few years, the diploid *F. vesca* has been recognized as a model species for genetic studies of cultivated strawberry (*F. x ananassa*), and here a previously developed *F. vesca/F. bucharica* Near Isogenic Line collection (NIL) was used to explore genetic variability of fruit quality traits. Analysis of fruit volatiles by GC-MS in our NIL collection revealed a complex and highly variable profile. One hundred compounds were unequivocally identified, including esters, aldehydes, ketones, alcohols, terpenoids, furans and lactones. Those in a subset, named key volatile compounds (KVCs), are likely contributors to the special aroma/flavour of wild strawberry. Genetic analysis revealed 50 major quantitative trait loci (QTL) including 14 QTL for KVCs, and one segregating as a dominant monogenetic trait for nerolidol. The most determinant regions affecting QTLs for KVCs, were mapped on LG5 and LG7. New candidate genes for the volatile QTL are proposed, based on differences in gene expression between NILs containing specific fragments of *F. bucharica* and the *F. vesca* recurrent genome. A high percentage of these candidate genes/alleles were colocalized within the boundaries of introgressed regions that contain QTLs, appearing to affect volatile metabolite accumulation acting in *cis.* A NIL collection is a good tool for the genetic dissection of volatile accumulation in wild strawberry fruit and a source of information for genes and alleles which may enhance aroma in cultivated strawberry.

Keywords: *Fragaria vesca*, volatilome, wild aroma, key volatile compounds, QTL, introgression

Introduction

Around the past 30 years, strawberry breeding programs have been directed mainly towards improving agronomical performance, resulting in varieties which produce high yields of large red and firm fruits, but fruit aroma is the quality trait with a major impact in consumers (Bruhn *et al.* 1991; Schwiterman *et al.* 2014). Over 350 volatile compounds have been identified in fruits of *Fragaria* sp., comprising esters, aldehydes, ketones, furanones, alcohols and terpenoids (Latrasse 1991) but only a few have been reported to contribute to the strawberry aroma as perceived by humans (Schieberle & Hofmann 1997; Ulrich *et al.* 1997; Ulrich *et al.* 2007).

As with other fruit crops, the biosynthetic pathways, enzymes and regulation underlying volatile compound accumulation have been partially elucidated in *Fragaria*. Fruit volatile profiles are known to depend on genetic (fruit species and variety), developmental (maturity stage) and postharvest factors, as well as on the analytical technique used. Generally, strawberry fruit volatiles increase with ripening (Goff & Klee 2006) and are classified in three main categories according to their carbon source: fatty acid, amino acid, and carbohydrate derivatives (reviewed by Schwab *et al.* 2008; Granell & Rambla 2013).

Fatty acids are the most important precursors for most fruit aroma volatiles, including straight-chain aldehydes, alcohols, esters, lactones and ketones. These compounds are synthesized mainly through the lipoxygenase (LOX) pathway and α- β-oxidation. In the LOX pathway, linoleic (18:2) and linolenic (18:3) acid are converted to hydroperoxide isomers, which are then cleaved by hydroperoxide lyase (HPL) to form hexanal and (*Z*)-3-hexenal, respectively. The aldehydes are subsequently reduced to the corresponding C_6 alcohols by alcohol dehydrogenase (ADH). Alcohol acyl transferase (AAT) catalyzes the reaction between an acyl moiety and an alcohol to form an ester. It has been proposed that this pathway requires a still-unidentified lipase (Schwab *et al.* 2008; Granell & Rambla 2013). Fatty acids can also be degraded via α- and β-oxidation pathways, although the specific mechanisms in plants are not well understood. In strawberry, alcohol acyl transferases (SAAT) with high sequence similarity but different substrate preferences have been identified: AAT in *F. x ananassa* (SAAT, Aharoni *et al.* 2000) and *F. vesca* (VAAT, Beekwilder *et al.* 2004). Additionally, an omega-6 fatty acid desaturase (*FaFAD*) has been correlated with the presence of γ-decalactone (Chambers *et al.* 2014; Sanchez-Sevilla *et al.* 2014).

Amino acid metabolism is known to be an important source of aroma volatile precursors. This is the case of phenylpropanoid and benzenoid volatiles that derive from phenylalanine. In strawberry, eugenol biosynthesis is mediated by two eugenol synthases (*FaEGS1* and *FaEGS2*) and controlled by one R2R3 MYB transcription factor (*FaEOBII*) (Aragüez *et al.* 2013; Medina-Puche *et al.* 2015). The biosynthetic pathways of other volatile benzenoids have not yet been clearly elucidated. Other branched-chain organic acids and aromatic amino acids are volatile precursors, however their catabolic pathways to form volatile compounds also remain unclear (Granell & Rambla 2013).

Carbohydrates can give rise directly to volatile furanones, without degradation of the carbon skeleton. In *F. x ananassa*, the FaOMT enzyme transforms furaneol to mesifurane (Zorrilla-Fontanesi *et al*. 2012). Volatile terpenoids (mainly mono- and sesqui- terpenoids) are formed from the basic C_5 precursors isopentenyl pyrophosphate (IPP) and its isomer, dimethylallyl pyrophosphate (DMAPP). IPP and DMAPP derive from either the plastidic methylerythriol phosphate or the cytosolic mevalonate pathway. These C_5 units are condensed to pyrophosphate precursors of terpenoids that are converted to final products by terpene synthases (TPS) (Granell & Rambla 2013). In strawberry, the production of the monoterpenoid linalool and the sesquiterpenoid nerolidol, and that of the monoterpene α-pinene, have been shown to be linked to specific alleles of the terpene synthases *FaNES1* and *FvPINS* respectively (Aharoni *et al.* 2004).

Major differences in volatile patterns have been observed among different species within the *Fragaria* genus. The most common volatile compounds contributing to strawberry aroma are esters with methyl butanoate, ethyl butanoate, butyl butanoate, methyl hexanoate, ethyl hexanoate, butyl acetate and hexyl acetate as important contributors to the fruity aroma. Methyl 2-aminobenzoate (also known as methyl anthranilate) has been reported as the single compound which confers the typical “wild strawberry-like” aroma of woodland strawberry (*F. vesca*) accessions, and is only very rarely found in some commercial varieties (Ulrich *et al.* 1997). Methyl cinnamate adds spicy notes and myrtenyl acetate herbaceous notes (Schieberle & Hofmann 1997; Ulrich *et al.* 1997; Jetti *et al.* 2007; Ulrich *et al.* 2007; Olbricht *et al.* 2008; Schwieterman *et al.* 2014). Furans, specifically furaneol and mesifurane, are considered important contributors by adding caramel notes (Schieberle & Hoffmann 1997, Ulrich *et al.* 1997; Ulrich *et al.* 2007; Jetti *et al.* 2007), while the terpenoids linalool and nerolidol, add

flowery notes (Ulrich *et al.* 1997, Olbrich *et al.* 2008, Schwieterman *et al.* 2014), but these compounds have been detected mainly in octoploid cultivars (*F. x ananassa*) and not in diploid wild strawberries (*F. vesca*) (Aharoni *et al.* 2004). The so-called 'green volatile compounds', (*Z*)-3-hexenal, (*E*)-2-hexenal and (*Z*)-3-hexen-1-ol, have been reported to contribute to the aroma characteristics that typically decrease with ripening (Ulrich *et al.* 1997, Schieberle & Hoffman 1997). Another important volatile compound is γ-decalactone, which confers 'peach-like' notes (Ulrich *et al.* 1997, Jetti *et al.* 2007, Olbrich *et al.* 2008).

A distinctive characteristic of volatile composition in *F. vesca* fruit is that it is richer in esters and monoterpenes (α-pinene, β-myrcene, α-terpineol, α-phellandrene) while exhibiting the pleasant and easily identifiable 'wild-strawberry-like' aroma associated with methyl 2-aminobenzoate. These compounds confer more intense and fruity aroma characteristics of this wild species and are not found normally in commercial strawberry fruits (*F. x ananassa*) (Aharoni *et al*. 2004, Ulrich *et al.* 1997; Ulrich *et al.* 2007; Dong *et al.* 2013). It is important to emphasize that large differences have been observed between *F. x ananassa* varieties covering a range of fruit quality phenotypes (Zorrilla-Fontanesi *et al.* 2012; Schwieterman *et al.* 2014).

To date, research has been directed to the characterization of the aroma profile of different octoploid accessions, mapping populations resulting from crosses involving commercial and wild material (Jetti *et al.* 2007; Olbricht *et al.* 2008; Zorrilla-Fontanesi *et al.* 2012; Schwieterman *et al.* 2014), and differences in the aroma profiles between octoploid and diploid strawberries (Aharoni *et al.* 2004; Ulrich *et al.* 2007; Dong *et al.* 2013). It is surprising that, despite the outstanding organoleptic characteristics of *F. vesca*, the genetic basis of its characteristic volatile content have not been yet reported. Given the very high degree of synteny between *F. vesca* and the commercial hybrid *F. x ananassa* (Rousseau-Gueutin *et al.* 2008, Tennessen *et al.* 2014), *F. vesca* is a model for the study of strawberry genetics what facilitates the transfer of information and alleles to modern varieties. In addition, the high quality reference genome sequence available (Shulaev *et al.* 2011), the transcriptomic analysis re-annotation of the especies (Darwish *et al.* 2015) and the recently developed near isogenic line (NIL) mapping collection (Urrutia *et al.* 2015) are powerful tools for the study of genetic traits in strawberry. Specifically, strawberry NIL collection derived from an inter-specific cross between *F. vesca* and *F. bucharica*. The homozygous introgressions of *F. bucharica*, an exotic

147 relative of *F. vesca*, give phenotypic variability that has been used to map QTL for
148 agronomical and metabolic traits (Urrutia *et al*. 2016).
149 This study provides a detailed profiling and QTL mapping of the volatile composition
150 of a *F. vesca* NIL population, as a first step to identifying the genetic basis of the wild
151 strawberry-like aroma. We focused on two genome regions that harbor key aroma
152 volatile QTL, a whole transcriptomic study of the corresponding lines allowed us to
153 select a number of differentially expressed candidate genes as responsible for the
154 differences in volatile accumulation.
155

Materials and Methods

Plant material and sample extraction

The volatilome of diploid strawberry ripe fruits was analyzed using 42 lines from a near isogenic line (NIL) collection in *F. vesca*, its recurrent and donor parents (*F. vesca* var. 'Reine des Vallées' and *F. bucharica* 'FDP 601' respectively) and the yellow-fruited variety of *F. vesca* named 'Yellow Wonder' (YW), which has a very pleasant pineapple-like aroma. Each line was represented by six to eight individuals independently grown from seed in two different years (2012 and 2013) and cultivated in a shaded greenhouse in Caldes de Montbui (latitude: 41º 36'N, longitude: 2º 10' E, altitude 203m above sea level, pre-coastal Mediterranean climate) following the usual agronomical practices for this crop. Pools of berries from each genotype were collected at harvest time and immediately frozen in liquid nitrogen as independent biological replicates. Three to five biological replicates were harvested, ground to fine powder and stored at -80ºC prior to gas chromatography-mass spectrometry (GC-MS) analysis and/or total RNA extraction. The NIL collection is extensively described in Urrutia *et al.* (2015).

Volatile compounds analysis

Volatile compounds were determined in a similar way as described in Rambla *et al.* (2015). Each biological replicate was analyzed as an independent sample. Before the volatile compounds analysis, an aliquot of 500 mg of frozen fruit powder from each sample was weighed in a 7 mL glass vial and thawed at 30ºC for 5 min. Then 500 µL of a saturated NaCl solution were added and the mixture was homogenized gently. Five hundred microliters of the resulting paste were transferred to a 10 mL screw cap headspace vial and analyzed immediately. Volatiles were sampled by HS-SPME (headspace solid phase microextraction) with a 65 µm PDMS/DVB (polydimethylsiloxane/divinyl-benzene) fiber (Supelco, PA, USA). The vials were first tempered at 50ºC for 10 min, then volatiles were extracted by exposing the fiber to the vial headspace for 30 min at 50ºC with agitation at 500 rpm. The extracted volatiles were desorbed in the GC injection port at 250ºC for 1 min in splitless mode. A Combi-PAL autosampler (CTC Analytics, Zwingen, Switzerland) was used for incubation, volatile extraction and desorption. GC-MS was in a 6890N gas chromatograph coupled to a 5975B mass spectrometer (Agilent Technologies, CA, USA). A DB-5ms column (60 m, 0.25 mm, 1 µm) (J&W Scientific, CA, USA) and a constant helium flow of 1.2

mL min^{-1} were used for chromatographic separation. Oven programming conditions were: 40ºC for 2 min, 5ºC min^{-1} ramp to 250ºC, then 5 min at 250ºC. Compounds were monitorized over the mass/charge ratio (m z^{-1}) range of 35-250. Chromatograms and mass spectra were analyzed using the Enhanced ChemStation software (Agilent Technologies, CA, USA). Volatile compounds were unambiguously identified by comparison of both retention time and mass spectra to those of commercial standards (SIGMA-Aldrich, MO, USA) run under the same conditions, except four compounds which were tentatively identified by comparison of their mass spectra to those in the NIST 05 mass spectral library. These compounds are marked with a "T" after the chemical name (Table 1). For quantification, a specific ion was selected for integration of the area of each of the identified compounds. Areas were normalized by comparison with the peak area of the same compound in a reference sample which was injected regularly each five to six samples, in order to correct for variations in sensitivity and fiber aging. This reference sample consisted of a homogeneous mix of all the samples analyzed each year.

Data and mQTL analysis

Volatiles are expressed in relative terms, as a ratio between each sample and a quality control sample (a mix of all studied samples) to correct for technical drift. In order to assess normality for statistical data analysis, ratios were transformed to base 2 logarithm. All the lines that set fruit were processed and analyzed by GC-MS each year (Supplemental Table 1). However, for the exploratory analysis, only those genotypes that produced enough fruits both years were considered (Urrutia et al 2016). For the statistical analysis and graphical representations, the free source software R 2.15.1 (RCoreTeam, 2012) was used, with the Rstudio 0.92.501 interface (Rstudio, 2012) unless otherwise specified. Pearson's correlation was calculated using the *rcorr* function from the Hmisc package (Harrell, 2014). The *Anova* function from the car package (Fox 2011) was used for analysis of variance (ANOVA). Omega squared values (ω^2) were calculated from ANOVA residuals following the formula: $(SS_i - df_i * MS_{err}) * (MS_t + MS_{err})^{-1}$. For Principal Components Analysis (PCA), the *prcomp* function and scaled values were used. The Hierarchical Clustering Analysis (HCA) was calculated considering Euclidean distance and the complete linkage clustering method. The Cluster Network Analysis (CNA) was calculated with the *qgraph* function from the qgraph R package (Epskamp *et al*. 2012). Significance tests were recursively calculated between

each NIL and RV ratio using the *t.test* function and corrected for multi-testing by *p.adjust* (threshold p. adjusted < 0.05) for QTL mapping. QTLs were mapped to a specific genetic region only when all NILs harboring a common *F. bucharica* introgression in this region showed a significant effect and in the same direction over the ratio for the specific metabolite of study. QTL that were mapped to the same region in two harvests were considered stable. Interval mapping analysis with MapQTL v.6 (Van Ooijen 2009) was used to confirm these QTL and estimate their effect. Stable QTL that explained around 20% or more of the variability and had LOD scores >1.8 were considered major QTLs. Non-stable QTL (detected in only one harvest) were considered only if they accounted for more than 20% of the observed variability that year. Graphical representation of the mQTLs was using MapChart 2.2 (Voorrips, 2002).

RNA sequencing and analysis

Total RNA was isolated from three selected NILs (Fb5:0-35 and Fb7:0-10) and the recurrent parental (RV) extracting the nine samples (three biological replicates per line) following the protocol described by Liao *et al.* 2004. A cell lysis step with CTAB buffer, modified with 3% PVP and 4% β-mercaptoethanol, was followed by: 2-3 cleaning steps with chloroform-isoamyl alcohol (24:1 v/v), overnight precipitation with lithium chloride (8 M), 1-2 additional cleaning steps with chloroform-isoamyl alcohol (24:1 v/v) and precipitation with cold absolute ethanol. RNA was quantified and checked for purity and integrity in a Bioanalyzer-2100 (Agilent Technologies, CA, USA). The concentration and quality threshold was set at 150 ng/?L and RNA integrity number (RIN) above eight. Further steps in RNA quality control, library preparation and mRNA paired end (2 x 75bp) sequencing were carried out at the Centro Nacional de Análisis Genómico (CNAG), Spain in a HiSeq2000 sequencer (Illumina, CA, USA). For quality control, trimming of sequencing adapters and removal of low quality and short reads (<40bp), FASTQC v0.10.1 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc) and Trimmomatic v0.32 (Bolger et al 2014) were used respectively. Trimmed reads were mapped against the *F. vesca* reference genome v1.1 using Tophat v2.0.11 with default parameters (Trapnell *et al.* 2009), taking as annotation reference version 2 (a2) (Darwish *et al.* 2015) and version 1 (a1) (https://www.rosaceae.org/species/fragaria/fragaria_vesca). Mapping quality was evaluated with the *bamqc* and *rnaseq* functions from Qualimap v2.1 (García-Alcalde et al 2012).

Differential gene expression analysis and functional annotation

Differential expression analysis was first performed using annotations a2 and then complemented, using the same filters and parameters, with a1. Independent tables of counts per gene were first generated with *HTSeq-count* with mode *union* (Anders et al 2014), considering all annotated genes from the reference annotation a2 and a1 respectively. These tables were provided as input to the DESeq package in R (Anders and Hubers 2010) using the *newCountDataSetFromHTSeqCount* function. DESeq counts all the reads-pairs mapped to a gene and normalizes the number of counts between samples, correcting for the library size. We considered that a gene was expressed in a specific line if at least two of the three biological replicates had >=1 read-counts for the gene. Secondly, 40% of the genes with lowest standard deviation were filtered in order to maximize the discovery rate. Differential expression analyses contrasting each NIL against RV were computed with the *nbinomTest* function (Anders and Hubers 2010). Multi-testing corrected p-values (p-adjust) were calculated using the Benjamini & Hochberg method. The significance threshold for a differentially expressed gene (DEG) was fixed at p-adjust=0.1. Lists of DEGs obtained with a2 (Supplemental Table 6) and a1 were compared for coincidence. DEG lists were inquired for predicted protein similarity with other proteins annotated in plant databases. The mRNA sequence was extracted from predicted exon coordinates. These mRNA sequences were inquired by *blastx* with the *GoAnna* tool from Agbase (McCarthy et al 2006) against the manually annotated protein plant database, with a significance threshold of 0.05. Annotated function and gene ontology terms (GO terms) of best blast hits were assumed as putative functions by mRNA query. In order to obtain a summarized view of the functional annotation results we used *GoSlimViewer* from AgBase (McCarthy et al 2006). In addition, functional enrichment analysis to detect metabolic functions or biological processes that might be over-represented among the DEGs was carried out using the MetGenMAP online platform (Joung et al 2009). Putatively affected metabolic pathways were also explored using MetGenMAP.

Variation calling

SNP and INDEL detection was only carried out for the genomic regions where an introgression of *F. bucharica* was present. Alignment files generated by TopHat for each NIL were indexed and then filtered to contain reads mapping to the respective *F.*

292 *bucharica* introgressed regions, using Samtools (v1.2.0). Further filtering of the
293 alignment files included removal of duplicate reads ("samtools rmdup") and additional
294 steps as described in the "GATK Best Practices workflow for SNP and indel calling on
295 RNAseq data" (GATK-3.1.1;
296 https://www.broadinstitute.org/gatk/guide/article?id=3891). Briefly, after removal of the
297 duplicate reads, sequences overhanging the intronic regions were hard-clipped using
298 '*SplitNCigarReads*', mapping qualities (MAPQs) reassigned using '*PrintReads*' and local
299 INDEL realigned using '*RealignerIndelCreator*' and '*IndelRealigner*'. Clean and
300 reformatted alignment files were used as input for variant calling with Samtools (v1.2.0)
301 using default parameters, except for applying a downgrading of mapping quality for
302 reads containing excessive mismatches (-C 50).

Results

Variability in the profile of fruit volatile compounds in the strawberry NIL collection

In order to detect genetic regions affecting wild strawberry aroma, differences in volatile accumulation were evaluated over two years in ripe fruit of NILs derived from an interspecific *Fragaria* cross (*F. vesca* var. 'Reine des Vallées' (RV) as recurrent parent x *F. bucharica* 'FDP601' (FB), as donor parental; Urrutia *et al.* 2015). Fruits from the RV were used as a reference for the changes in volatiles observed in the population, and fruit from the aromatic white-fruited *F. vesca* var. Yellow Wonder (YW) were used as an external control or out-group. Metabolite profiling by GC-MS analyses and QTL mapping were performed with all the genotypes that set enough fruit each year, but we only considered those that were represented by at least three biological replicates in both years for the statistical analysis (i.e. 25 genotypes, Supplemental Table 1).

We were able to identify 100 volatile compounds, 96 of which were unambiguously identified by comparison of both retention time and mass spectra with those of commercial standards run under the same conditions, whilst the remaining four compounds were tentatively identified based on their mass spectra (these are marked with a T at the end of the chemical name, see Table 1). The unequivocally identified volatile compounds were 11 alcohols, 16 aldehydes, 46 esters, four furans, 14 ketones, eight terpenoids and one lactone, and include most of the compounds described in the literature as contributing to strawberry aroma (Schieberle & Hofmann 1997; Ulrich *et al.* 1997; Ulrich *et al.* 2007). Here we refer to them as 'key volatile compounds' (KVCs), and have marked them with an arrow symbol in Table 1. KVCs that confer specific strawberry aroma are 12 esters butyl acetate, butyl butanoate, (E)-2-hexenyl acetate, ethyl butanoate, ethyl hexanoate, hexyl acetate, methyl-2-aminobenzoate, methyl butanoate, methyl cinnamate, methyl hexanoate, myrtenyl acetate and (Z)-3-hexenyl acetate; two aldehydes (E)-2-hexenal and (Z)-3-hexenal; two furans furaneol and mesifurane; two terpenoids linalool and nerolidol, and one lactone γ-decalactone.

The relative levels (see M&M) for most volatile compounds had mean ratios around one for RV in both harvests (Table 1) consistent with the nearly isogenic nature of the NIL collection, which means lines share much of the common RV genetic background. The variation interval for each volatile (min. and max. ratio) show different ranges of

variation in the NIL indicating that genes involved in accumulation of the volatile compounds segregated in our NIL collection. More extreme values were detected for the lower than for the higher ratios, indicating that, globally, *F. bucharica* alleles decrease volatile accumulation of *Fragaria* berries. Different degrees of variation were detected depending on the volatile, with decanal (4-fold variation from 0.39 to 1.49 in 2012) and γ-decalactone (10,000-fold variation, ranging from 0.01 to 119.96 in 2013) defining the extremes of the variation range. It is also noteworthy to mention that nerolidol segregated as a dominant monogenetic trait in our population, with the *F. bucharica* alleles conferring the ability to produce nerolidol in the otherwise non-nerolidol producer *F. vesca* background (Supplemental Table 1). Dominance of the *F. bucharica* nerolidol allele was determined in the F_1 fruit samples (hybrid *F. vesca* RV x *F. bucharica*), which confirmed their ability to produce nerolidol (assayed in 2013 only).

Relations between volatile compounds and NILs

Each NIL had a characteristic volatile profile according to the *F. bucharica* introgression, and volatile compounds could be clustered according to their levels in the different NILs (Figure 1, Table 1). Volatiles with similar chemical structure or in the same biosynthetic pathways tend to be co-regulated and therefore clustered together. Cluster A (16 volatiles) is enriched in long carboxylesters, particularly in octyl-derived esters. Cluster B (two volatiles) includes (*E*)-2-hexenyl acetate and its free alcohol (*E*)-2-hexen-1-ol. Cluster C (35 volatiles) groups all the aldehydes (except (*E*)-2-decenal), and terpenoids (except α-farnesene) and most C_4 alkyl acetates. Cluster D is divided in two sub-clusters, D1 (7 volatiles) which is enriched in benzenoid-derived volatiles, including two furans (mesifurane and furaneol), and D2 (40 volatiles), enriched in esters derived from butanoic and acetic acids, long chain alcohols and ketones. Compared to *F. vesca* RV, *F. vesca* YW presented quite a different volatile profile which is enriched in esters (clusters A and D2) and with decreased levels of compounds in clusters B, C and D1 (Figure 1). The effect of the *F. bucharica* alleles is obvious in lines with introgressions at the beginning of LG3 (Fb**3**:0-8, Fb**3**:0-15). These lines are characterized by an over-accumulation of the monoterpenoid linalool (96) and the sesquiterpenoid nerolidol (99), which suggests a more active terpene synthase allele from *F. bucharica* associated to this region. Differences were most prevalent in lines with introgressions in LG5, indicating that major QTLs for volatile accumulation are

located in LG5. Lines carrying introgressions of *F. bucharica* in LG7 showed a tendency to over-accumulating esters (cluster A) and under-accumulating of aldehydes and terpenoids (cluster C). Mean ratios for all the samples analyzed each year are provided in Supplemental Table 1.

The patterns of volatile accumulation were quite stable: positive Pearson's pair-wise significant correlations were detected for 82 of the 100 compounds between two years at p-value <0.05 (75 with an adjusted p-value <0.01). This high correlation affected all KVCs except furaneol and butyl acetate (Table 1).

Compounds belonging to the same biosynthetic pathway tended to be highly correlated, as can be seen by cluster network analysis (CNA) in the case of esters and alcohols, fatty acid-derived and phenylalanine-derived compounds and terpenoids (Figure 2). Volatiles whose biosynthetic pathways have not been elucidated, were also highly correlated to other volatile metabolites, which could indicate common regulation. Individual correlation coefficients and significant values are provided in Supplemental Table 2.

Variability in volatile levels across the different NIL, RV and YW fruit samples was also analyzed by principal component analysis (PCA) (Figure 3). PCA suggested that variation of most of the volatiles is continuous, and differences in the aroma pattern between the NILs were restricted to single or small subsets of metabolites. A closer look to the PCA shows that NILs samples spread along PC1 according to their introgressed region (Figure 3A), while PC2 divides the samples again according to their genotype but also according to the harvest year, indicating that a higher proportion of the observed variability between the NILs was due to genotype rather than to environmental factors. This PCA also indicated that volatile accumulation in NIL with introgressions in LG2 and LG3 were especially susceptible to the environmental conditions. According to the corresponding loading plots (Figure 3B), linalool (96), octanal (25) and 6-methyl-5-hepten-2-one (86) together with most esters and alcohols, were mostly responsible for the variability along PC1. Compounds contributing mostly to variability across PC2 were C_6 lipid derivatives (*E*)-2-hexenal (17), (*E*)-2-hexenyl acetate (43), (*E*)-2-hexen-1-ol (9) and (*Z*)-3-hexenal (27), aldehydes (*E*)-2-nonenal (18) and (*E*)-2-

heptenal (16), and the terpenoid myrtenol (97). Among all the samples, YW was the one with the most differentiated volatile profile.

Genotypic and environmental effect on the accumulation of volatile compounds

Genotypic (G) and environmental (E) effect on the volatile accumulation was evaluated by analysis of variance (ANOVA) fitting the model G+E+GxE (years taken as different environments). Several factor combinations influenced variability depending on the given compound. G significantly contributed (p-value<0.05) to variability of 98 out of the 100 studied volatile compounds (Supplemental Table 3). Among them, 33 compounds were significantly influenced by the three factors G, E and GxE. Sixteen volatiles were mostly influenced by G and E but not by the GxE interaction, 33 were influenced by G and GxE but not by E and, most interestingly, 17 volatile compounds were influenced only by G, including some of the KVCs-like methyl 2-aminobenzoate, nerolidol, γ-decalactone, ethyl butanoate and (*Z*)-3-hexenal. Each of the factors also differs in the actual percentage of variability they account for. In general, genotype has a stronger effect on volatile variability than the environment (year) or the GxE interaction (Figure 4; see also Supplemental Table 3). The G factor accounted for >50% of observed variability in 35 compounds (including ten KVCs: 17, 27, 39, 43, 55, 61, 66, 73, 96, 99), but its effect was up to 70% for six volatiles, including four KVCs ((*E*)-2-hexenal, (*Z*)-3-hexenal, (*E*)-2-hexenyl acetate and linalool). The E factor was less important and only surpassed 20% of the observed variability in the case of five compounds (including the KVC mesifurane).

Volatile QTL analysis

Genetic regions controlling ripe-fruit wild strawberry volatile accumulation were detected by QTL mapping. A total of 126 QTL were mapped, 102 of which were stable QTL (detected in two years) and 50 of them were major QTL (stable and explaining >20% of the variability and with LOD>1.8). The QTL corresponded to 81 different compounds (40 esters, 12 aldehydes, 11 alcohols, eight ketones, seven terpenoids and three furans). The effect of the *F. bucharica* alleles on the *F. vesca* RV genetic background was positive (producing an increased volatile accumulation) in 30 of them and negative (reducing their levels) in 96 of them (Table 2).

Considering the major volatile QTL, 25 corresponded to compounds that mapped to a single locus. This included nine KVCs (linalool, nerolidol, mesifurane, methyl hexanoate, methyl cinnamate, (*E*)-2-hexenal, (*E*)-2-hexenyl acetate, (*Z*)-3-hexenal and (*Z*)-3-hexenyl acetate), and three compounds mapped to two major QTL (the KVCs methyl 2-aminobenzoate, nerol and 3-methyl-2-butenyl acetate: Table 2, Figure 5). Genotype had a major effect on most of the volatile compounds for which major QTL were mapped, but the effect of the environment was low (Figure 4). One of the exceptions was mesifurane, which, although clearly influenced by the environment (38%), the effect of the genotype (30%) was enough to map a QTL. There were also some compounds, mainly lipid derivatives including aldehydes (octanal, nonanal, decanal, (*E*)-2-octenal, (*E*)-2-nonenal and (*E*)-2-decenal), alcohols (1-penten-3-ol, 1-hexanol and 2-heptanol) and ketones (1-penten-3-one, 2-pentanone and 2-nonanone), that only resulted in QTLs that could be mapped in a single year and therefore were classified as not stable. Most of these compounds were highly dependent on the environment, with a low correlation between harvests.

Co-localized QTL may indicate co-regulated compounds. Two regions in the wild strawberry genome harbor the highest number of major volatile QTL and QTL for KVCs: LG5 and LG7 (Figure 5). The central region of LG5 (LG5:11-35 cM) appears to be very important for the wild strawberry aroma as it had major QTL (negative) for the accumulation of nine esters, five of which were KVCs: methyl 2-aminobenzoate, myrtenyl acetate, methyl butanoate, butyl butanoate and methyl hexanoate. The bottom of LG5 (LG5:50-76 cM) harbors QTLs for fatty acid derived volatiles associated with green-fresh aroma. Positive QTL were mapped for (*Z*)-3-hexenal and (*Z*)-3-hexenyl acetate, and negative QTL for their respective trans-2 isomers (*E*)-2-hexenal and (*E*)-2-hexenyl acetate. This suggests that *F. bucharica* alleles in this region reduce conversion of (*Z*)-3-hexenal (synthesized from linolenic acid) to (*E*)-2-hexenal, that would lead to a higher accumulation of (*Z*)-3- derivatives and a lower accumulation of (*E*)-2-derivatives (Granell & Rambla 2013). In addition, three positive QTL for the terpenoid nerol, the benzenoid eugenol and the aldehyde (*E*)-2-heptenal, and one negative QTL for the alcohol (*E*)-2-hexen-1-ol are localized in the same region. The top region in LG7 also seems to be important for wild strawberry scent as it accumulated 13 major QTL, two of which correspond to key aroma contributors involved in wild strawberry-like aroma (methyl 2-aminobenzoate at LG7:0-10 cM) and sweet-caramel notes (mesifurane

471 at LG7:26-43 cM). Additionally, at the top of LG7:0-10 cM we found four major QTL
472 for the accumulation of long esters and two major QTL for monoterpenoids (limonene
473 and myrtenol). Another interesting genetic region for key aroma volatiles is LG3:0-
474 8 cM where two major QTL for nerolidol and linalool accumulation were mapped.
475 Nerolidol show an absence (RV alleles) - presence (FB alleles) segregating pattern.
476
477 Whole transcriptome analysis of two rich volatile QTL regions
478 The NILs Fb**5**:0-35 and Fb**7**:0-10 (with introgression sizes of 6.51 and 14.20 Mb
479 respectively) carry QTL for key volatile esters in wild strawberry aroma, namely methyl
480 2-aminobenzoate but also myrtenyl acetate, methyl butanoate, butyl butanoate and
481 methyl hexanoate. The transcriptome of ripe berries from these two NILs were analyzed
482 and compared with their recurrent parental (RV) transcriptome in order to identify
483 differences in expression of specific genes that could be linked to the observed
484 phenotypic changes. Transcriptomes were obtained by RNAseq approach using three
485 biological replicates (nine samples in total). A total of 407 million (M) read-pairs were
486 obtained with an average of 45 M read-pairs per sample (min. 33M, max. 58M). The
487 quality of raw read pairs was assessed and sequencing adapters and low quality reads
488 were filtered. A total of 374 M (92%) passed the filter cutoff and were kept for further
489 analysis (average of 41.62 M read-pairs per sample). A high percentage of reads (83-
490 86%) could be mapped to the reference *F. vesca* genome v1.1 (Supplemental Table 4).
491 According to the latest annotation version (Darwish et al. 2015), 73 to 75% of mapped
492 reads were located in exons, 9% in introns and the remaining 16 to 18% in intergenic
493 regions. Differential expression analysis between the selected NILs (Fb5:0-35 and
494 Fb7:0-10) and the recurrent parental (RV), showed that the majority of the 31,778
495 studied genes, 17,906 (56%) were similarly expressed in both NILs and RV.
496 Additionally, 2,847 genes were expressed in at least one of the lines, with 388 detected
497 only inFb5:0-35, 663 in Fb7:0-10, and 437 detected only in RV, while 11,025 (35%)
498 were not expressed in any of the NILs nor in RV (Figure 6).
499
500 Differential expression analysis revealed 257 differentially expressed genes (DEGs)
501 between Fb5:0-35 and RV and 442 DEGs between Fb7:0-10 and RV (DEG significance
502 threshold fixed at p-value=0.1) (Table 3, Supplemental Table 5). The large majority of
503 the DEGs were altered only in one NIL with respect to *F. vesca* RV. This was expected
504 as NILs do not share overlapping introgressions. However, there were also 33 genes

differentially expressed in both NILs when compared with *F. vesca* RV (Figure 6). Analysis of genome position showed that a high percentage of the DEGs in each NIL (54% in Fb**5**:0-35 and 59% in Fb**7**:0-10) were located within the boundaries of their introgressed region, indicating that they are probably acting on *cis*, that is that the differences in expression and their effects are likely to be due to allelic differences of the genes in the region (Figure 7).

Functional annotation of DEGs resulted in significant blast hits for around 83% of them. Gene Ontology (GO) categorization for molecular function and biological process indicated that 48 DEGs were annotated as involved in metabolic activity (Supplemental Table 6). This suggests that *F. bucharica* introgressions are likely to affect fruit metabolism.

In addition, several DEGs were predicted as being involved in known volatile synthetic pathways in *F. vesca* (Table 4), such as the lipoxygenase pathway (13-LOX and 13-HPL pathway) in NIL Fb7:0-10 and terpene synthesis in NIL Fb5:0-35. We carefully selected candidate genes by combining expression data with the metabolic QTL (Table 5). The NILs Fb5:50-76 and Fb7:0-10 contain QTL for fatty-acid derived volatiles. Differentially expressed lipoxygenases (4) and acyltransferases (6) were found in Fb**7**:0-10, and one down-regulated acyl-transferase was detected in Fb**5**:0-35. Selected NILs were also found to harbor several QTL for terpenoids that might be of interest for wild strawberry aroma (Table 2). A differentially expressed sesquiterpene synthase was detected in Fb**5**:0-35 and a terpene synthase in Fb**7**:0-10 (Table 5).

Several transcription factors (TF) were also differentially expressed in NIL Fb**5**:0-35 and Fb**7**:0-10 with respect to RV. As alterations in TF can have wide range effects, all of them were considered candidate genes. A putative MYC2 TF up-regulated in Fb**7**:0-10 (maker-LG7-snap-gene-91.103-mRNA-1) is suspected to be associated with terpenoid biosynthesis as its closest ortholog in *A. thaliana*, (MYC2_ARATH) has also been related to sesquiterpene biosynthesis (Hong *et al.* 2012). Until now, TF were not related to VOC in fruits.

In addition, it should be mentioned that there were 114 differentially expressed genes whose function could not be assigned by sequence similarity. Therefore, we cannot discard these genes may be involved in the volatile phenotypes (Supplemental Table 5).

SNPs between NILs Fb5:0-35 and Fb7:0-10

Although none of the accessions used in this work has been sequenced, the interspecific nature of the NILs is likely to provide a high number of polymorphisms between the introgressed regions (from *F. bucharica* FDP601) and the recurrent parental (*F. vesca* var. 'Reine des Vallées'). The RNAseq results presented here constitute **the first** transcriptome for these accessions and therefore the first global view of the genetic divergence at SNP resolution between them. The transcriptome of the introgressed region of NIL Fb5:0-35 had 6,813 polymorphisms (6,622 SNPs and 191 indels), and Fb7:0-10 10,850 polymorphisms (10,517 SNPs and 333 indels) with respect to RV (Table 6). A detailed list of the SNP polymorphisms and position is given in Supplemental Table 7.

Discussion

Volatile profile particularities of the diploid strawberry

Woodland strawberry (*F. vesca*) aroma is known to have significant qualitative and quantitative differences when compared with commercial varieties (*F. x ananassa*) (Ulrich *et al.* 2007). *F. vesca* fruit produce higher levels of esters and terpenoids and a more intense aroma, besides the production of specific compounds such as methyl 2-aminobenzoate (aka methyl anthranilate) that confers the characteristic 'wild strawberry' aroma (Ulrich *et al.* 2007). In this study we profiled the volatile composition of a NIL collection derived from an inter-specific cross between *F. vesca* and *F. bucharica (*Urrutia *et al*. 2015*).* The genetic background of *F. vesca* confers stability and homogenicity to the collection with outstanding organoleptic quality, but the homozygous introgressions of *F. bucharica*, an exotic relative of *F. vesca*, confer important phenotypic variability that can be used to map QTL for agronomical and metabolic traits (Urrutia *et al*. 2015a; Urrutia *et al*. 2016). The alleles of *F. bucharica* usually had a negative effect on the volatile compounds, as there was a decrease in level of most of the volatiles mapped QTL.

The total number of identified volatile compounds was higher in this *F. vesca* NIL collection (100) than in previous studies with *F. x ananassa* populations (81 in Schwiterman *et al.* (2014) and 87 in Zorrilla-Fontanesi *et al.* (2012)). The *F. vesca* NIL collection volatile profiling revealed a very complex composition. One hundred of the compounds produced were identified, the majority of them being esters (46%), followed by aldehydes (16%), ketones (14%), alcohols (11%), and several terpenoids, furans and lactones (13%). These proportions are in agreement with that described in other studies with octoploid strawberries (Schwiterman *et al.* 2014, Zorrilla-Fontanesi *et al.* 2012). All the compounds identified in the *F. vesca* NIL collection have been previously described in strawberry fruit, and around 20 of them have been reported to be important for its aroma (Latrasse 1991; Schieberle &Hofmann 1997; Ulrich *et al.* 1997; Ulrich *et al.* 2007).

The identified compounds that were not found in octoploid studies correspond to esters such as methyl 2-aminobenzoate, methyl acetate, methyl cinnamate, methyl 3-hydroxyoctanoate, ethyl methylthioacetate and 2,3-butanediol diacetate, and to terpenoids such as α-farnesene and α-pinene (Zorilla-Fontanesi *et al.* 2012) that might

contribute to the special aroma of wild strawberry. We also identified nerolidol and linalool segregating within our collection. These compounds have been reported to be characteristic of octoploid *Fragaria* species and produced by a truncated allele of the *FaNES* gene (Aharoni *et al.* 2004; Chambers *et al.* 2012). However we found a clear QTL at LG3:0-8 cM (Figure 5) for the accumulation of these two compounds that co-locates with the *FaNES* gene. The *F. vesca* RV parental does not produce linalool or nerolidol, but they were both detected in the hybrid (analyzed only in 2013; Supplemental Table 1). This suggests that the *F. bucharica* alleles for the *FaNES* gene produce linalool and nerolidol. Both parentals in the NIL collection (*F. vesca* and *F. bucharica*), the F_1 hybrid and the lines in the collection producing linalool and nerolidol (Fb**3**:0-8 and Fb**3**:0-15), together with a *F. x ananassa* as a positive control, were genotyped for *FaNES* alleles following the method described by Aharoni *et al.* (2004). The conclusion from the observed results is that the truncated *FaNES* allele is absent in our collection (data not shown). This suggests that there may be several alleles producing linalool in strawberry and that some of them may have arisen before octoploidization.

Volatilome comparison between *F. vesca* RV and YW

F. vesca YW is a white fruited strawberry known to have a pleasant, intense fruity aroma with tropical (pineapple-like) notes. Used in this study as an out-group of the NIL collection, it had a different pattern of volatile accumulation, enriched in esters and with higher accumulation ratios than *F. vesca* RV (Figure 1). A recent study with the white fruited octoploid species *F. chiloensis*, also known for its intense, tropical fruity aroma, reported that the characteristic tropical fruit aroma came from a set of six esters, two of which, ethyl hexanoate (49) and hexyl acetate (52), we detected as associated to *F. vesca* YW (Figure 3). The other four compounds (furfuryl acetate, acetyl acetate, 1-methylethyl dodecanoate and ethyl tetradecanoate) were not detected under our experimental conditions. They may be absent in and only detected in other *Fragaria sp*. or failed to be detected by our volatile profiling method (Prat *et al.* 2014).

Volatile QTL in strawberry

Significant year to year correlation was detected for most compounds (82 out of 100) although the correlation index and the significance threshold varied considerably. The correlation values reported here are higher than those reported for volatile compounds in

other studies (Eduardo *et al.* 2013). Differences in the relative volatile accumulation pattern in each NIL in the two studied harvests appear to be mainly associated to their genotypes (Figure 4) and to a lower extent to the environment. This is in contrast to what has been reported in other studies with octoploid strawberry (Forney *et al.* 2000; Zorrilla-Fontanesi *et al.* 2012) and peach (*Prunus persica*) (Eduardo *et al.* 2013; Sanchez *et al.* 2014), where the effect of the environment was more relevant. The special configuration of our mapping population, as near isogenic lines, may be responsible for such stability, avoiding epistatic effects among different QTL. The fact that all lines share a common genetic background, in contrast to other mapping populations where genetic differences between lines is wider, may highlight the effect of the genotype, caused by exotic introgressions, and buffer the effect of the environment over the phenotypic traits, as all lines may respond in a similar way. In fact, stability of the lines has been previously proved with a (poly)-phenolic profiling of the NIL collection (Urrutia *et al*. 2016), and although the correlation between genotypes according to volatile profiling is lower, the median of all genotypes is above 0.70.

QTL mapping revealed 50 major stable QTL that accounted for a high proportion of the variability of 47 compounds, including 14 major QTL identified for 13 KVCs: (E)-2-hexenal, (Z)-3-hexenal, (E)-2-hexenyl acetate, (Z)-3-hexenyl acetate, butyl butanoate, methyl-2-aminobenzoate (2), methyl butanoate, methyl cinnamate, methyl hexanoate, myrtenyl acetate, mesifurane, linalool and nerolidol. Many of the QTL cluster in a few genetic regions, suggesting that the compounds are co-regulated and controlled by a reduced number of loci. LG5 and LG7 seem to be the most determinant regions controlling volatile compounds synthesis as they accumulate the largest number of QTL and harbor nine and two major QTL for KVCs, respectively. Some of the detected QTLs were in agreement with those described by Zorrilla-Fontanesi *et al.* (2012) as they co-locate according to synteny studies (Rousseau-Gueutin *et al.* 2008, Tennessen *et al.* 2014). A QTL for methyl benzoate was located at LG1:26-61 cM in *F. vesca* and at LGI-F.1: 38 cM in *F. x ananassa.* A QTL for benzyl acetate was located at LG7:0-10 cM in *F. vesca* and at LGVII-F.1c: 9 cM in *F. x ananassa.* A QTL for ethyl decanoate was mapped to LG3:8-15 cM for *F. vesca,* and to LGIII-F.1: 4 cM and LGIII-M.1: -8 cM in *F. x ananassa*. A QTL for mesifurane was located at LG7:27-43 cM in *F. vesca,* and to LGVII-F.2: 18 cM and LGVII-M.2:65 cM in *F. x ananassa*. The latter

QTL is associated with the *FaOMT* gene responsible for its accumulation that also co-locates with our QTL (Zorrilla-Fontanesi *et al.* 2012).

There were also QTLs located previously in different regions in *F. x ananassa* and *F. vesca* and volatile compounds that showed significant variability in one population and not in the other, highlighting that different genetic backgrounds and environments can reveal different genetic traits. As an example of this, we found two QTLs controlling the accumulation of methyl 2-aminobenzoate, which is characteristic of *F. vesca* aroma and was not detected in *F. x ananassa*. Previous reports have mapped a QTL for the accumulation of γ-decalactone in the homeolog LGIII-M.2: 50-54 cM (Zorrilla-Fontanesi *et al.* 2012) and a candidate gene *FaFAD1* with an eQTL co-localized (Sanchez-Sevilla *et al.* 2014). However, we found no significant QTL for γ-decalactone in our collection. Although data suggests that there might be an increase in the production of this compound in lines with introgressions at the end of LG5, this increase is not enough to report a significant effect (Supplemental Table 1). However, this suggests there may be other genetic regions controlling γ-decalactone accumulation in *F. vesca.*

C_6 compounds from the lipoxygenase pathway and the corresponding acetate esters ((*E*)-2-hexen-1-ol, (*E*)-2-hexenal, (*E*)-2-hexenyl acetate, (*Z*)-3-hexenal and (*Z)*-3-hexenyl acetate) are commonly described as 'green volatile compounds' and are usually considered too variable within genotypes or varieties to be used as discriminative compounds (Ulrich *et al.* 1997). However, a recent studies in peach (*Prunus persica*) reported stable QTLs for (*E*)-2-hexenyl acetate and (*Z*)-3-hexenyl acetate (Eduardo *et al.* 2012) and in tomato for (*Z*)-3-hexenal and (*E*)-2-hexenal (Rambla *et al*. 2016). Our data revealed a high year to year correlation between these compounds (Table 1) and QTLs that co-localize for all of them at LG5:50-76 cM, suggesting that these compounds were stable and co-regulated under our conditions. By differential expression analysis of the red ripe fruits it was possible to highlight genes differentially expressed between the NILs and the recurrent parental RV, that might contribute to the observed QTL. NILs Fb**5**:0-35 and Fb**7**:0-10 are interesting for further studies in fruity and wild strawberry-like aroma as they harbor QTL for methyl 2-aminobenzoate and several other esters. Differentially-expressed genes include terpene synthases and acyl-transferases, which catalyze the main steps in terpenoid and ester formation, and

688 lipoxygenases, which participate in fatty acid degradation and consequently in FA-
689 derived volatiles.
690
691 In-depth characterization of the volatiles emitted by ripe strawberry fruit in a *F. vesca*
692 NIL mapping collection revealed a complex mixture of 100 compounds, varying in
693 relative abundance across the population presumably because of the effect of *F.*
694 *bucharica* alleles. The high genetic effect on the accumulation of many compounds (35
695 compounds >50% G effect) allowed 50 major QTL to be mapped, including 14 QTL for
696 compounds considered of extreme importance for strawberry aroma. Some, such as
697 methyl 2-aminobenzoate and mesifurane, are only rarely found in commercial varieties
698 (*F. x ananassa*) and are of great interest for breeding programs. Therefore, here we set
699 the ground for further studies on the inheritance of the woodland strawberry aroma that
700 may lead to improved aroma and marketability of new strawberry varieties. Further
701 studies for positional cloning of the QTLs in combination with reverse genetics will
702 shed light on the causal genes of the observed phenotypes.
703

Author contribution statement:

MU analysed the NILs collection, prepared fruit samples, did the statistical analyses, prepared RNAseq samples and evaluated the DEG, in addition to writing the manuscript. AG and JLR did the GC-MS analyses and participated in edition of the manuscript. KA did the SNPs calling analyses. AM lead the project, participated in all steps of phenotyping and in writing the manuscript.

The authors declare that they have no conflict of interest.

Acknowledgements:

This work was funded by grants AGL2010-21414 and RTA2013-00010 from the Spanish Ministry of Economy and Competitiveness and through the "Severo Ochoa Programme for Centres of Excellence in R&D" 2016-2019 (SEV-2015-0533)" and by the CERCA Programme / Generalitat de Catalunya. MU was supported by a FPI fellowship from the Spanish Ministry of Education. AG would like to thank Metabolomic lab and COST action FA1106 for networking activities.

720 *Figure legends*

721 **Figure 1 Hierarchical clustering (HCA) and heatmap of volatile compounds levels**.
722 Ratio values of all studied volatile compounds per genotype are shown in the heatmap
723 on a blue (negative) to red (positive) scale. Compounds are numerically codified as
724 specified in Table 1. Genotypes include the NILs that were analyzed both years, RV and
725 YW. Top bar identifies the sample harvest year: 2012 (blue) and 2013 (green). The
726 HCA and dendrogram of volatile compounds was according to metabolite ratio
727 distances (Euclidean distance, complete linkage). Clusters are indicated with capital
728 letters in both the dendrogram and Table 1.

729 **Figure 2 Cluster network analysis (CNA).**
730 Metabolites are represented as nodes colored according to their biosynthetic pathway (if
731 known) or chemical structure as specified by the legend. Positive (green) and negative
732 (red) correlations with absolute values >|0.5| are shown as links between the nodes.
733 Links representing absolute correlations >|0.8| are wider the stronger they are and have
734 the maximum color saturation. Absolute correlations <|0.8| are vaguer the weaker they
735 are and have the least width.

736 **Figure 3 PCA (PC1 and PC2) scores and loading plot**.
737 **A:** PCA scores plot. NIL, RV and YW are colored according to the harvest year as
738 specified in legend. **B:** loading plot Compounds are coded as specified in Table 1 and
739 colored according to their chemical family as specified in legend.

740 **Figure 4 ω2 values**. Percentage of the observed variability attributable to each of the
741 factors: genotype (G), environment (E), their interaction (GxE) or to error.

742 **Figure 5 Volatile QTL**. Graphical representation of the major QTL mapped. QTL
743 shown were found to be significantly different (corrected p-value < 0.05) from the
744 recurrent parental (*F. vesca* RV), in the same direction, in both harvests for all the NILs
745 harboring the introgressed region, and explained around 20% of the variability
746 regarding the NIL collection. QTL names correspond to the volatile compound affected.
747 Colored bars indicate the biosynthetic pathway (if known) or the chemical structure of
748 the compound as in Figure 2. The positive or negative effect of the QTL over the ratio
749 regarding *F. vesca* RV is represented by the full or empty color bars respectively. For
750 locating the QTL, the LG and position (in cM) of the microsatellites (SSRs) used for
751 genotyping are given.

752 **Figure 6 Venn diagrams.** Venn diagram **A** depicts the number of annotated genes (a2)
753 expressed by each line. Colored ellipses represent analyzed lines (Fb**5**:0-35, Fb**7**:0-10

and RV). Venn diagram **B** depicts the number of differentially expressed genes detected between each NIL and the recurrent parental (RV). Colored ellipses represent comparisons (NIL vs. RV). Numbers in intersecting areas indicate that the genes are shared between the lines/comparisons meeting in the area. Non-intersecting areas indicate the number of genes that are specifically expressed/differentially expressed in a line.

Figure 7 Differentially expressed gene distribution (Manhattan plot). Graphical representation of all genes in their physical position (x-axis), and their associated -log10 (p-value) from the differential expression analysis (y-axis).

Table legends

Table 1. Volatile compounds summary, and between harvests correlations.
All identified compounds and their assigned number codes and clusters are presented. Tentatively identified compounds are indicated with a T after the chemical name. A selected set of important compounds contributing to strawberry aroma are indicated with an arrow. Data are expressed as the ratio between samples and a reference. Mean ratios and standard deviation (sd) were calculated for each compound in the recurrent parental *F.vesca* (RV) and in the average NIL collection for 2012 and 2013 harvests. The range of the ratios (min and max values) was calculated for the NIL collection in both harvests. Correlation between harvests was calculated using average genotype values in both years. Asterisk after the values indicate significance at different thresholds: p-value <0.05 '*' , p-value <0.01 '**', p-value <0.001 '***'. No significant correlations are indicated by 'ns'.

Table 2. QTL for volatile compounds detected in a *F. vesca* NIL collection.
Detected QTL listed by compound's alphabetical order. The position of the QTL (LG number followed by the start and end position in cM), the positive (up) or negative (down) effect of the QTL over the metabolite's ratio compared with *F. vesca* RV, the NIL harboring the shorter *F. bucharica* introgression (in cM) that includes the QTL, the results of the t-test (corrected p-value) and interval mapping analysis (LOD score), the percentage of variance explained by the QTL regarding the NIL collection and the stability of the QTL (detected in 1 or 2 harvests) are provided.

Table 3 Differentially expressed genes (DEG) summary
Number total, up- and down-regulated DEG obtained with annotation version 2 (a2) for both contrasting hypothesis (NIL vs. RV).

Table 4 Metabolic pathways affected
List of known metabolic pathways related to DEG detected in each NIL using MetGenMAP software.

Table 5 Selected candidate genes

List of selected DEG between genotypes (NIL vs RV) for each metabolic QTL

Table 6 Polymorphism summary

Electronic Suplemental Material Legend

Supplemental table1: Volatile compounds average values per genotype per year. Average values (per genotype per harvests) of all detected volatile compounds are provided. Data are expressed as the ratio between the samples and a reference sample.

Supplemental table 2. Pearson correlation values between volatile compounds for 2012 and 2013 independent harvests. Asterisk after the values indicate significance at different thresholds: p-value <0.05 '*' , p-value <0.01 '**', p-value <0.001 '***'

Supplemental Table 3. Analysis of variance (ANOVA) fitting the model G+E+GxE and w2 values. ANOVA was calculated for all volatile compounds independently considering two factors, genotype (G) and environment (E), and their interaction (GxE). The resulting parameters of the ANOVA test Sum of squares (SS), degrees of freedom (df) and p-values are provided. Omega squared values (w2) were calculated from the ANOVA parameters for G, E and GxE and reflect the percentage of variability accounted by each one of them. The error is 1 minus the percentual variability accounted by G, E and GxE.

Supplemental Table 4 RNAseq reads quality

Supplemental table 5 List of DEG for each contrasting hypothesis (NIL vs. RV). DEG for each NIL are presented in ascending order of log2(fold change).
a, gene id is according to F. vesca annotation 2 nomenclature
b, log2(fold change) values use as reference RV, so negative values indicate down-regulation in NIL vs. RV and positive values up-regulation in NIL vs. RV
c, best blast hit found for the DEG predicted proteins. Codes are according to UniProtUK entries

Supplemental Figure 6 Go terms summary. Molecular function and Biological process GO terms summary of the Differentially Expressed Genes.

Supplemental table 7 Polymorphisms. SNPs and Indels detected between transcriptomes of NILs (Fb5:0-35 and Fb7:0-10) and RV.

851

852

853 **References**

854 Aharoni A, Giri AP, Verstappen FWA, Bertea CM, Sevenier R, Sun ZK, Jongsma MA,
855 Schwab W & Bouwmeester HJ Gain and loss of fruit flavor compounds produced by
856 wild and cultivated strawberry species. *Plant Cell* (2004) **16**:3110-31.

857 Aharoni A, Keizer LCP, Bouwmeester HJ, Sun ZK, Alvarez-Huerta M, Verhoeven HA,
858 Blaas J, van Houwelingen AMML, De Vos RCH, van der Voet H, Jansen RC, Guis M,
859 Mol J, Davis RW, Schena M, van Tunen AJ & O'Connell AP Identification of the
860 SAAT gene involved in strawberry flavor biogenesis by use of DNA microarrays. *Plant*
861 *Cell* (2000) **12**:647-61.

862 Anders S & Huber W Differential expression analysis for sequence count data. *Genome*
863 *Biology* (2010) **11**:ppR106.

864 Anders S, Pyl PT & Huber W HTSeq—a Python framework to work with high-
865 throughput sequencing data. *Bioinformatics* (2015) **31**:166-9.

866 Aragüez I, Osorio S, Hoffmann T, Rambla JL, Medina-Escobar N, Granell A, Botella
867 MÁ, Schwab W & Valpuesta V Eugenol Production in Achenes and Receptacles of
868 Strawberry Fruits Is Catalyzed by Synthases Exhibiting Distinct Kinetics. *Plant*
869 *Physiology* (2013) **163**:946-58.

870 Beekwilder J, Alvarez-Huerta M, Neef E, Verstappen FWA, Bouwmeester HJ. &
871 Aharoni A Functional characterization of enzymes forming volatile esters from
872 strawberry and banana. *Plant Physiology* (2004) **135**:1865-78.

873 Bolger AM, Lohse M. & Usadel B Trimmomatic: a flexible trimmer for Illumina
874 sequence data. *Bioinformatics* (2014) **30**:2114-20.

875 Bruhn CM, Feldmann N, Garlitz C, Harwood J, Ivans E, Marshall M, Riley A, Thurber
876 D & Williamson E Consumer perceptions of quality: apricots, cantaloupes, peaches,
877 pears, strawberries, and tomatoes. *Journal of Food Quality* (1991) **14**:187-95.

878 Chambers A, Whitaker VM, Gibbs B, Plotto A & Folta KM Detection of the linalool-
879 producing NES1 variant across diverse strawberry (Fragaria spp.) accessions. *Plant*
880 *Breeding* (2012) **131**:437-43.

881 Chambers AH, Pillet J, Plotto A, Bai JH, Whitaker VM & Folta KM Identification of a
882 strawberry flavor gene candidate using an integrated genetic-genomic-analytical
883 chemistry approach. *BMC Genomics* (2014) **15**:217.

884 Darwish O, Shahan R, Liu ZC, Slovin JP & Alkharouf NW Re-annotation of the
885 woodland strawberry (Fragaria vesca) genome. *BMC Genomics* (2015) **16**:29.

886 Dong J, Zhang YT, Tang XW, Jin WM & Han ZH Differences in volatile ester
887 composition between Fragaria x ananassa and F. vesca and implications for strawberry
888 aroma patterns. *Scientia Horticulturae* (2013) **150**:47-53.

889 Eduardo I, Chietera G, Pirona R, Pacheco I, Troggio M, Banchi E, Bassi D, Rossini L,
890 Vecchietti A & Pozzi C Genetic dissection of aroma volatile compounds from the

essential oil of peach fruit: QTL analysis and identification of candidate genes using dense SNP maps. *Tree Genetics & Genomes* (2013) **9**:189-204.

Epskamp SC, G; Cramer, AOJ: Waldorp, LJ; Schmittmann, VD; Borsboom, D Network representations of relationships in data. R package version 1.2.4. (2012)

Forney CF, Kalt W & Jordan MA The composition of strawberry aroma is influenced by cultivar, maturity, and storage. *Hortscience* (2000) **35**:1022-6.

García-Alcalde F, Okonechnikov K, Carbonell J, Cruz LM, Götz S, Tarazona S, Dopazo J, Meyer TF & Conesa A. Qualimap: evaluating next-generationsequencing alignment data. *Bioinformatics* (2012) **28**:2678-9.

Goff SA & Klee HJ Plant Volatile Compounds: Sensory Cues for Health and Nutritional Value? *Science* (2006) **311**:815-9.

Granell A & Rambla JL Biosynthesis of volatile compounds. In: The Molecular Biology and Biochemistry of Fruit Ripening (eds. by Seymour GB, Poole M, Giovannoni JJ & Tucker GA) Blackwell Publishing Ltd, Oxford, UK. (2013) pp. 135-161.

Hong GJ, Xue XY, Mao YB, Wang LJ & Chen XY Arabidopsis MYC2 Interacts with DELLA Proteins in Regulating Sesquiterpene Synthase Gene Expression. *Plant Cell* (2012) **24**:2635-48.

Jetti RR, Yang E, Kurnianta A, Finn C & Qian MC Quantification of selected aroma-active compounds in strawberries by headspace solid-phase microextraction gas chromatography and correlation with sensory descriptive analysis. *Journal of Food Science* (2007) **72**:S487-S96.

Joung JG, Corbett AM, Fellman SM, Tieman DM, Klee HJ, Giovannoni JJ & Fei ZJ Plant MetGenMAP: An Integrative Analysis System for Plant Systems Biology. *Plant Physiology* (2009) **151**:1758-68.

Latrasse A. Fruits III. In: Volatile Compounds in Fruits and Beverages (ed. by Maarse H), Dekker, New York, USA. (1991) pp. 333-87.

Liao ZH, Chen M, Guo L, Gong YF, Tang F, Sun XF & Tang KX Rapid isolation of high-quality total RNA from Taxus and Ginkgo. *Preparative Biochemistry & Biotechnology* (2004) **34**:209-14.

McCarthy F, Wang N, Magee GB, Nanduri B, Lawrence M, Camon E, Barrell D, Hill D, Dolan M, Williams WP, Luthe D, Bridges S & Burgess S AgBase: a functional genomics resource for agriculture. *BMC Genomics* (2006) **7**:229.

Medina-Puche L, Cumplido-Laso G, Amil-Ruiz F, Hoffmann T, Ring L, Rodriguez-Franco A, Caballero JL, Schwab W, Munoz-Blanco J & Blanco-Portales R MYB10 plays a major role in the regulation of flavonoid/phenylpropanoid metabolism during ripening of Fragaria x ananassa fruits. *J Exp Bot* (2014) **65**:401-17.

Olbricht K, Grafe C, Weiss K & Ulrich D Inheritance of aroma compounds in a model population of Fragaria x ananassa Duch. *Plant Breeding* (2008) **127**:87-93.

Prat L, Espinoza MI, Agosin E & Silva H Identification of volatile compounds associated with the aroma of white strawberries (Fragaria chiloensis). *Journal of the Science of Food and Agriculture* (2014) **94**:752-9.

Rambla JL, López-Gresa MP, Bellés JM, Granell A Metabolomic profiling of plant tissues. In: Plant Functional Genomics (Methods in Molecular Biology 1284 series) (eds. Alonso J.M. & Stepanova A.N.). Springer, New York, USA. doi: 10.1007/978-1-4939-2444-8_11. (2015) pp. 221-235

Rambla JL, Medina A, Fernández-del-Carmen A, Barrantes W, Grandillo S, Cammareri M, López-Casado G, Rodrigo G, Alonso A, García-Martínez S, Primo J, Ruiz JJ, Fernández-Muñoz R, Monforte AJ, Granell A Identification, introgression, and validation of fruit volatile QTLs from a red-fruited wild tomato species. *J Exp Bot* (2016) **68**: 429-442.

RCoreTeam R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. (2012)

Rousseau-Gueutin M, Lerceteau-Köhler E, Barrot L, Sargent DJ, Monfort A, Simpson D, Arús P, Guérin G & Denoyes-Rothan B 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* (2008) **179**:2045-60.

RStudio RStudio: Integrated development environment for R RStudio, Boston, MA, USA. (2012)

Sanchez-Sevilla JF, Cruz-Rus E, Valpuesta V, Botella MA & Amaya I Deciphering gamma-decalactone biosynthesis in strawberry fruit using a combination of genetic mapping, RNA-Seq and eQTL analyses. *BMC Genomics* (2014) **15**:218.

Sanchez G, Martinez J, Romeu J, Garcia J, Monforte AJ, Badenes ML & Granell A The peach volatilome modularity is reflected at the genetic and environmental response levels in a QTL mapping population. *BMC Plant Biology* (2014) **14**:137.

Schieberle P & Hofmann T Evaluation of the character impact odorants in fresh strawberry juice by quantitative measurements and sensory studies on model mixtures. *J Agric Food Chem* (1997) **45**:227-32.

Schwab W, Davidovich-Rikanati R & Lewinsohn E Biosynthesis of plant-derived flavor compounds. *The Plant Journal* (2008) **54**:712-32.

Schwieterman ML, Colquhoun TA, Jaworski EA, Bartoshuk LM, Gilbert JL, Tieman DM, Odabasi AZ, Moskowitz HR, Folta KM, Klee HJ, Sims CA, Whitaker VM & Clark DG Strawberry Flavor: Diverse Chemical Compositions, a Seasonal Influence, and Effects on Sensory Perception. *PLoS ONE* (2014) **9**:e88446.

Shulaev V, Sargent DJ, Crowhurst RN, Mockler TC, Folkerts O, Delcher AL, Jaiswal P, Mockaitis K, Liston A, Mane SP, Burns P, Davis TM, Slovin JP, Bassil N, Hellens RP, Evans C, Harkins T, Kodira C, Desany B, Crasta OR, Jensen RV, Allan AC, Michael TP, Setubal JC, Celton J-M, Rees DJG, Williams KP, Holt SH, Rojas JJR, Chatterjee M, Liu B, Silva H, Meisel L, Adato A, Filichkin SA, Troggio M, Viola R, Ashman T-L, Wang H, Dharmawardhana P, Elser J, Raja R, Priest HD, Bryant DW, Fox SE, Givan

SA, Wilhelm LJ, Naithani S, Christoffels A, Salama DY, Carter J, Girona EL, Zdepski A, Wang W, Kerstetter RA, Schwab W, Korban SS, Davik J, Monfort A, Denoyes-Rothan B, Arus P, Mittler R, Flinn B, Aharoni A, Bennetzen JL, Salzberg SL, Dickerman AW, Velasco R, Borodovsky M, Veilleux RE & Folta KM The genome of woodland strawberry (Fragaria vesca). *Nat Genet* (2011) **43**:109-16.

Tennessen JA, Govindarajulu R, Ashman TL & Liston A Evolutionary origins and dynamics of octoploid strawberry subgenomes revealed by dense targeted capture linkage maps. *Genome Biol Evol* (2014) **6**:3295-313.

Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ & Pachter L Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature biotechnology* (2010) **28**:511-518.

Ulrich D, Hoberg E, Rapp A & Kecke S Analysis of strawberry flavour - discrimination of aroma types by quantification of volatile compounds. *Zeitschrift Fur Lebensmittel-Untersuchung Und-Forschung a-Food Research and Technology* (1997) **205**:218-23.

Ulrich D, Komes D, Olbricht K & Hoberg E Diversity of aroma patterns in wild and cultivated Fragaria accessions. *Genetic Resources and Crop Evolution* (2007) **54**:1185-96.

Urrutia M, Bonet J, Arús P & Monfort A A near-isogenic line (NIL) collection in diploid strawberry and its use in the genetic analysis of morphologic, phenotypic and nutritional characters. *Theoretical and Applied Genetics* (2015) **128**:1261-75.

Urrutia M, Schwab W, Hoffmann T & Monfort A Genetic dissection of the (poly)phenol profile of diploid strawberry (Fragaria vesca) fruits using a NIL collection. *Plant Science* (2016) **242**:151-168.

Zorrilla-Fontanesi Y, Rambla JL, Cabeza A, Medina JJ, Sánchez-Sevilla JF, Valpuesta V, Botella MA, Granell A & Amaya I Genetic Analysis of Strawberry Fruit Aroma and Identification of O-Methyltransferase FaOMT as the Locus Controlling Natural Variation in Mesifurane Content. *Plant Physiology* (2012) **159**:851-70.

1000

1001

1002

1003 **Table 1** Volatile compounds summary and between harvests correlations.

						Correlation		Recurrent parental (RV)				NIL collection						
								2012		2013		2012				2013		
KVC	Code	Compound	Pathway	Cluster	Family	corr.	sig.	mean	sd	mean	sd	mean	sd	range		mean	sd	range
	1	1-decanol		A	alcohol	0.9	***	1,24	0,80	0,93	0,51	0,60	1,11	0,01	5,43	0,60	0,78	0,01 3,61
	2	1-hexanol	Fatty Acid Deriv.	C	alcohol	-0,22	ns	1,05	0,28	1,05	0,67	1,29	0,95	0,40	6,87	1,24	0,68	0,34 3,92
	3	1-octanol		D2	alcohol	0,78	***	1,73	0,94	1,01	0,48	0,72	0,71	0,09	3,51	0,67	0,52	0,04 2,19
	4	1-penten-3-ol		C	alcohol	0,33	ns	1,26	0,60	0,89	0,46	1,30	0,85	0,24	6,15	1,15	0,45	0,26 2,45
	5	2-heptanol		D2	alcohol	0,72	***	1,42	1,35	1,00	0,48	0,71	1,02	0,01	4,53	0,46	0,56	0,01 4,13
	6	2-nonanol		D2	alcohol	0,83	***	1,34	1,10	1,08	0,47	0,66	0,84	0,00	4,03	0,51	0,48	0,01 1,89
	7	2-tridecanol		D2	alcohol	0,62	***	1,36	1,11	1,16	1,06	0,95	1,18	0,05	7,57	0,63	0,77	0,02 4,04
	8	2-undecanol		D2	alcohol	0,78	***	1,25	0,76	1,18	0,87	0,77	0,92	0,04	5,21	0,53	0,48	0,03 1,87
	9	(E)-2-hexen-1-ol	Fatty Acid Deriv.	B	alcohol	0,82	***	1,08	0,35	0,82	0,48	1,36	1,56	0,02	7,41	1,08	0,90	0,01 4,35
	10	Ethanol		A	alcohol	0,41	*	0,80	0,65	0,39	0,21	1,09	1,70	0,01	7,26	0,68	1,07	0,02 6,00
	11	Eugenol	Benzoid	D2	alcohol	0,81	***	0,42	0,14	0,94	0,94	0,88	2,45	0,05	19,29	1,08	2,93	0,04 19,19
	12	3,4-dimethylbenzaldehyde	Benzoid	C	aldehyde	0,42	*	1,02	0,24	0,97	0,10	1,01	0,49	0,42	3,03	0,97	0,20	0,52 2,35
	13	Benzaldehyde	Benzoid	C	aldehyde	0,78	***	1,05	0,26	1,01	0,28	1,35	0,85	0,31	5,58	1,17	0,54	0,31 2,73
	14	Decanal		C	aldehyde	-0,01	ns	1,00	0,24	0,99	0,36	0,88	0,23	0,39	1,49	1,02	0,35	0,47 1,85
	15	(E)-2-decenal		D2	aldehyde	0,57	**	1,16	0,42	0,74	0,27	1,16	0,46	0,26	2,58	0,95	0,75	0,13 6,22
	16	(E)-2-heptenal		C	aldehyde	0,90	***	1,30	0,40	1,17	0,68	3,14	4,85	0,26	24,08	3,23	4,23	0,19 19,28
→	17	(E)-2-hexenal	Fatty Acid Deriv.	C	aldehyde	0,88	***	1,32	0,18	1,03	0,30	1,11	0,40	0,24	1,85	1,07	0,40	0,31 1,83
	18	(E)-2-nonenal		C	aldehyde	0,47	*	0,63	0,15	0,96	0,29	0,68	0,21	0,24	1,26	0,99	0,42	0,27 2,29

	19	(E)-2-octenal		C	aldehyde	0,52	**	1,73	0,84	1,10	0,34	1,40	0,67	0,35	3,61	1,18	0,48	0,25	2,80
	20	(E)-2-pentenal		C	aldehyde	0,58	**	1,88	1,10	1,75	1,60	2,32	1,60	0,15	8,63	2,19	1,33	0,30	7,13
	21	(E,Z)-2,4-heptadienal		C	aldehyde	0,34	ns	1,27	0,45	1,00	0,46	1,13	0,56	0,52	3,89	1,12	0,50	0,37	3,87
	22	Heptanal		C	aldehyde	0,54	**	1,28	0,55	1,29	0,74	1,20	0,57	0,51	3,51	1,51	0,91	0,42	4,88
	23	Hexanal	Fatty Acid Deriv.	C	aldehyde	0,62	**	1,18	0,21	1,02	0,40	1,27	0,36	0,45	2,35	1,27	0,36	0,46	2,34
	24	Nonanal		C	aldehyde	0,09	ns	1,30	0,61	1,60	0,74	1,30	0,70	0,39	4,92	1,28	0,79	0,43	4,69
	25	Octanal		C	aldehyde	0,41	*	2,32	1,45	3,38	3,30	2,81	2,58	0,27	17,03	2,91	3,07	0,31	14,83
	26	Pentanal		C	aldehyde	0,12	ns	1,24	0,49	1,55	0,89	1,41	0,70	0,25	3,27	1,57	0,81	0,31	4,49
→	27	(Z)-3-hexenal	Fatty Acid Deriv.	C	aldehyde	0,94	***	1,23	0,31	1,40	0,86	2,67	3,17	0,50	15,03	4,08	5,62	0,52	25,13
	28	1-methylbutyl butanoate		D2	ester	0,76	***	1,91	3,33	1,79	1,19	0,72	1,21	0,07	7,41	0,74	1,61	0,13	12,17
	29	1-methylethyl butanoate		D2	ester	0,16	ns	2,11	1,94	1,50	0,83	1,24	0,91	0,25	4,23	0,68	0,66	0,04	3,58
	30	1-methylethyl acetate		C	ester	0,00	ns	1,27	0,64	0,58	0,28	1,48	0,73	0,27	3,56	0,74	0,48	0,11	2,57
	31	1-methylhexyl acetate		D2	ester	0,31	ns	1,24	1,16	0,93	0,53	1,02	1,51	0,00	8,69	0,47	0,57	0,01	3,73
	32	1-methyloctyl butanoate		D2	ester	0,68	***	0,78	0,54	1,33	0,61	0,56	0,84	0,06	4,76	0,63	0,96	0,05	6,13
	33	2,3-butanedioldiacetate T		A	ester	0,75	***	0,70	0,65	0,33	0,29	1,24	2,36	0,04	12,91	0,74	1,53	0,01	9,07
	34	2 -methylbutyl acetate		C	ester	0,60	**	1,25	0,51	0,76	0,35	1,65	0,93	0,42	5,28	0,88	0,59	0,19	3,16
	35	3-methyl-2-butenyl acetate		C	ester	0,62	**	1,33	0,51	0,87	0,43	1,57	1,27	0,18	7,26	1,20	1,05	0,17	5,58
	36	3-methylbutyl acetate		C	ester	0,27	ns	1,44	0,46	0,32	0,10	1,26	0,88	0,02	6,92	0,69	0,54	0,10	3,16
	37	Benzyl acetate	Benzoid	D1	ester	0,75	***	2,51	1,54	1,54	1,04	1,31	1,09	0,13	4,82	1,16	0,91	0,08	3,97
→	38	Butyl acetate		D2	ester	0,25	ns	1,08	0,43	0,93	0,28	1,38	1,14	0,13	4,96	0,98	0,79	0,08	4,04
→	39	Butyl butanoate		D2	ester	0,63	***	1,13	0,62	1,74	0,96	0,88	1,31	0,02	6,87	1,06	1,47	0,01	7,66
	40	Butyl hexanoate		A	ester	0,77	***	1,12	0,73	1,21	0,53	0,71	1,01	0,02	4,99	1,11	1,71	0,01	9,21
	41	Cinnamyl acetate	Benzoid	D1	ester	0,67	***	1,13	0,49	1,90	2,14	1,40	2,79	0,02	19,84	0,61	0,81	0,01	3,75
	42	Decyl acetate		A	ester	0,88	***	1,03	0,50	0,85	0,45	0,52	0,85	0,01	4,53	0,57	0,68	0,01	2,85
→	43	(E)-2-hexenyl acetate	Fatty Acid Deriv.	B	ester	0,92	***	3,13	1,06	0,75	0,44	1,25	0,85	0,00	4,08	1,01	1,06	0,01	5,46
	44	Ethyl 2-hexenoate		A	ester	0,65	***	0,59	0,63	0,69	0,32	1,15	2,20	0,01	11,63	0,67	1,04	0,02	6,22

	45	Ethyl acetate		A	ester	0,54	**	1,27	1,49	0,43	0,22	1,61	3,18	0,01	17,75	0,58	0,69	0,01	2,96
→	46	Ethyl butanoate		D2	ester	0,56	**	1,59	1,20	1,15	0,72	0,96	1,04	0,01	4,08	0,82	0,64	0,01	3,11
	47	Ethyl decanoate		A	ester	0,77	***	0,67	0,61	0,37	0,23	0,74	1,65	0,00	7,41	0,58	1,83	0,01	12,59
	48	Ethyl dodecanoate		A	ester	0,79	***	0,80	1,10	0,28	0,18	1,26	3,24	0,01	17,03	0,80	3,38	0,01	23,06
→	49	Ethyl hexanoate		A	ester	0,68	***	1,22	0,94	1,17	0,61	0,61	0,87	0,00	3,18	0,59	0,57	0,01	2,39
	50	Ethyl methylthioacetate		D2	ester	0,35	ns	2,76	2,35	1,25	0,64	1,18	1,54	0,03	5,78	1,39	1,18	0,02	5,19
	51	Ethyl octanoate		A	ester	0,76	***	1,13	1,08	0,75	0,43	0,69	1,29	0,00	4,72	0,54	1,10	0,01	6,84
→	52	Hexyl acetate		D2	ester	0,52	**	1,47	0,51	1,26	0,35	1,02	0,58	0,16	2,64	1,00	0,61	0,12	3,34
	53	Hexyl butanoate		D2	ester	0,72	***	1,02	0,79	1,26	0,65	0,80	1,09	0,03	4,63	1,15	1,75	0,01	10,46
	54	Hexyl hexanoate		D2	ester	0,63	***	0,91	0,47	1,45	1,03	0,67	0,84	0,02	3,89	1,03	1,29	0,01	7,17
→	55	Methyl 2-aminobenzoate		D2	ester	0,84	***	0,77	0,27	1,85	1,72	1,37	1,25	0,01	5,90	1,26	1,96	0,01	13,03
	56	Methyl 2-hexenoate		D2	ester	0,44	*	2,07	1,23	1,18	0,74	1,60	2,54	0,12	15,56	0,85	0,97	0,08	4,92
	57	Methyl 3-hydroxyoctanoate T		D2	ester	0,54	**	1,89	0,86	2,41	2,09	0,89	1,07	0,00	6,15	0,90	1,04	0,01	8,18
	58	Methyl acetate T		D2	ester	-0,01	ns	1,78	1,65	0,72	0,37	1,88	1,63	0,23	11,16	0,90	0,67	0,07	4,40
	59	Methyl benzoate	Benzoid	D1	ester	0,74	***	1,67	1,11	0,75	0,44	1,17	1,19	0,10	7,01	1,15	1,32	0,05	7,10
→	60	Methyl butanoate		D2	ester	0,49	*	2,09	1,75	1,29	0,88	1,01	1,12	0,04	6,96	0,82	0,81	0,01	3,67
→	61	Methyl cinnamate T	Benzoid	D1	ester	0,71	***	0,75	0,50	0,45	0,28	1,65	1,82	0,01	11,16	1,34	2,61	0,02	22,49
	62	Methyl decanoate		D2	ester	0,86	***	1,07	0,49	0,80	0,51	0,78	0,75	0,01	3,07	0,70	0,84	0,01	4,73
	63	Methyl dodecanoate		A	ester	0,87	***	1,25	1,20	0,72	0,53	0,88	1,49	0,09	6,59	0,65	1,49	0,04	9,76
→	64	Methyl hexanoate		D2	ester	0,81	***	1,27	0,92	1,21	0,73	0,69	0,80	0,01	3,20	0,82	0,80	0,02	3,46
	65	Methyl octanoate		D2	ester	0,81	***	1,29	0,52	0,78	0,47	0,88	0,70	0,05	2,81	0,69	0,65	0,01	3,09
→	66	Myrtenyl acetate		D2	ester	0,74	***	1,94	0,69	1,51	0,70	1,12	0,57	0,23	3,25	0,96	0,69	0,17	4,73
	67	Nonyl acetate		D2	ester	0,59	**	1,11	0,39	1,27	0,88	0,96	0,67	0,20	2,95	0,92	0,95	0,20	4,88
	68	Octyl acetate		A	ester	0,79	***	1,41	0,69	1,08	0,54	0,61	0,79	0,01	3,29	0,58	0,54	0,01	2,48
	69	Octyl butanoate		A	ester	0,68	***	1,04	0,81	1,03	0,54	0,50	1,09	0,03	6,19	0,84	1,39	0,02	9,86

	70	Octyl hexanoate		A	ester	0,77	***	1,71	1,16	1,49	0,95	0,51	1,27	0,03	7,89	0,74	1,25	0,02	8,75
	71	Pentyl acetate		C	ester	0,30	ns	1,46	0,41	0,78	0,15	1,25	0,56	0,47	3,03	0,93	0,59	0,13	3,87
	72	Propyl butanoate		D2	ester	0,48	*	1,21	0,52	0,98	0,46	1,12	1,16	0,03	5,24	0,91	0,85	0,02	3,77
→	73	(Z)-3-hexenyl acetate	Fatty Acid Deriv.	C	ester	0,80	***	1,31	0,82	0,71	0,43	1,38	1,79	0,17	10,20	1,85	2,83	0,09	22,01
	74	2,1-pentenylfuran		C	furan	0,72	***	1,64	0,56	1,05	0,42	1,40	0,79	0,08	4,08	1,26	0,63	0,13	2,79
	75	2-pentylfuran		C	furan	0,61	**	1,14	0,23	0,88	0,21	1,31	0,59	0,22	2,89	1,14	0,40	0,47	2,49
→	76	Furaneol		D1	furan	-0,27	ns	1,73	1,22	0,81	1,02	1,58	1,78	0,02	9,65	3,20	13,07	0,01	120,64
→	77	Mesifurane		D1	furan	0,69	***	1,59	0,79	0,28	0,20	1,48	1,12	0,05	5,43	0,31	0,28	0,01	1,23
	78	1-penten-3-one		C	ketone	0,54	**	1,74	0,56	1,63	0,30	1,84	0,97	0,31	4,47	1,36	0,60	0,36	3,20
	79	2-heptanone		D2	ketone	0,67	***	1,39	0,68	1,45	0,65	0,80	0,71	0,02	3,66	0,86	0,52	0,01	2,55
	80	2-nonanone		D2	ketone	0,83	***	1,24	0,53	1,42	0,68	0,77	0,69	0,00	3,39	0,79	0,51	0,01	2,01
	81	2-pentadecanone		A	ketone	0,71	***	2,03	2,39	1,46	1,45	0,99	1,44	0,02	6,32	0,72	1,06	0,01	5,06
	82	2-pentanone		D2	ketone	0,76	***	1,55	0,98	1,63	0,75	0,82	1,18	0,01	6,73	0,83	0,72	0,01	3,62
	83	2-tridecanone		D2	ketone	0,64	***	1,27	0,77	1,05	0,84	0,73	0,76	0,01	4,06	0,57	0,55	0,01	2,62
	84	2-undecanone		D2	ketone	0,66	***	1,26	0,58	1,32	0,97	0,91	0,83	0,02	4,32	0,71	0,51	0,02	2,35
	85	4-tridecanone		D2	ketone	0,76	***	1,26	1,10	1,37	1,08	1,17	1,33	0,48	9,58	1,10	0,95	0,45	5,27
	86	6-methyl-5-hepten-2-one		C	ketone	0,63	***	0,85	0,31	1,00	0,38	1,06	0,46	0,29	2,46	1,37	0,44	0,35	2,82
	87	Acetone		C	ketone	0,74	***	1,46	0,60	0,94	0,57	1,08	0,66	0,13	3,43	0,79	0,44	0,08	2,02
	88	Acetophenone	Benzoid	D1	ketone	0,76	***	1,68	0,65	1,35	0,70	1,11	0,77	0,07	3,34	1,04	1,21	0,09	10,73
	89	α-ionone		D2	ketone	0,51	**	1,67	0,58	1,63	1,19	1,39	1,17	0,16	5,31	1,19	0,90	0,21	4,91
	90	β-ionone		D2	ketone	0,55	**	1,48	0,64	0,88	0,40	1,26	0,68	0,23	3,46	0,98	0,43	0,13	1,98
	91	(Z)-geranyl acetone		D2	ketone	0,32	ns	1,08	0,56	0,98	0,40	1,09	0,57	0,19	3,10	1,09	0,43	0,29	2,57
→	92	γ-decalactone		D2	lactone	0,75	***	2,05	1,25	14,31	37,55	0,96	1,59	0,01	10,34	2,36	11,72	0,01	119,96
	93	α-farnesene	Terpenoids	D2	terpenoid	0,55	**	1,72	1,84	1,60	1,06	1,27	1,36	0,04	6,96	1,01	1,01	0,14	6,68
	94	α-pinene	Terpenoids	C	terpenoid	0,76	***	1,10	0,41	0,61	0,23	1,45	0,86	0,32	4,14	1,11	1,03	0,14	6,00
	95	Limonene	Terpenoids	C	terpenoi	0,31	ns	1,25	0,52	0,79	0,34	1,05	0,78	0,36	5,54	0,86	0,32	0,34	1,89

→	96	Linalool	Terpenoids	C	d terpenoid	0,78	***	1,06	0,33	0,72	0,35	1,32	1,44	0,25	9,92	1,70	1,98	0,31	11,65
	97	Myrtenol	Terpenoids	C	terpenoid	0,64	***	1,88	0,48	0,78	0,29	1,10	0,64	0,17	3,18	0,98	1,06	0,17	7,86
	98	Nerol	Terpenoids	C	terpenoid	0,84	***	1,40	0,36	0,77	0,21	1,08	0,71	0,05	3,61	1,21	1,12	0,01	5,88
→	99	Nerolidol	Terpenoids	C	terpenoid	0,95	***	1,00	0,00	1,00	0,00	1,24	0,94	1,00	6,73	1,20	0,77	1,00	5,83
	100	Terpineol	Terpenoids	C	terpenoid	0,25	ns	1,25	0,36	1,04	0,31	1,18	0,40	0,43	2,33	1,20	0,68	0,30	3,80

Table 2 QTL for volatile compounds detected in a F. vesca NIL collection.

	Compound	direction	qtl (cM)	shorter NIL	t-test (corrected p.value)	LOD	% explained variance	stable
	(E)-2-decenal	down	LG7:0-26	Fb7:0-27	<0,05	4.64	46%	1
	(E)-2-decenal	down	LG7:27-59	Fb7:0-59	<0,05	6.14	56%	1
	(E)-2-heptenal	down	LG7:0-10	Fb7:0-10	<0,05	2.10	25%	1
	(E)-2-heptenal	up	LG5:50-76	Fb5:50-76	<0,05	15.17	46-87%	2
	(E)-2-hexen-1-ol	down	LG5:50-76	Fb5:50-76	<0,05	15.66	63-88%	2
→	(E)-2-hexenal	down	LG5:50-76	Fb5:50-76	<0,05	16.04	74-88%	2
→	(E)-2-hexenyl acetate	down	LG5:50-76	Fb5:50-76	<0,05	16.75	82-89%	2
	(E)-2-nonenal	down	LG5:50-76	Fb5:50-76	<0,05	4.59	46%	1

	(E)-2-octenal	down	LG7:52-59	Fb7:52-59	<0,05	2.61	30%	1
	(E)-2-pentenal	down	LG7:0-10	Fb7:0-10	<0,05	2.91	30-32%	2
→	(Z)-3-hexenal	up	LG5:50-76	Fb5:50-76	<0,05	14.76	58-86%	2
→	(Z)-3-hexenyl acetate	up	LG5:50-76	Fb5:50-76	<0,05	5.68	44-53%	2
	1-decanol	down	LG5:0-11	Fb5:0-11	<0,05	1.86	16-22%	2
	1-decanol	down	LG3:8-15	Fb3:0-15	<0,05	<1,80	2-5%	2
	1-decanol	down	LG4:20-44	Fb4:0-44	<0,05	<1,80	1-5%	2
	1-hexanol	up	LG5:50-76	Fb5:50-76	<0,05	1.99	23%	1
	1-methylbutyl butanoate	down	LG5:11-35	Fb5:0-35	<0,05	<1,80	10-11%	2
	1-methylbutyl butanoate	down	LG7:0-10	Fb7:0-10	<0,05	<1,80	2-3%	2
	1-methylhexyl acetate	down	LG4:9-44	Fb4:0-44	<0,05	2.68	30%	1
	1-methyloctyl butanoate	down	LG2:0-30	Fb2:0-30	<0,05	<1,80	7-13%	2
	1-methyloctyl butanoate	down	LG5:11-35	Fb5:0-35	<0,05	<1,80	17-21%	2
	1-octanol	down	LG1:26-61	Fb1:26-61	<0,05	<1,80	1-3%	2
	1-octanol	down	LG2:0-30	Fb2:0-30	<0,05	<1,80	10-16%	2
	1-octanol	down	LG5:11-35	Fb5:0-35	<0,05	<1,80	5-17%	2
	1-penten-3-ol	down	LG7:0-10	Fb7:0-10	<0,05	5.25	51%	1
	1-penten-3-one	down	LG7:0-10	Fb7:0-10	<0,05	3.97	42%	1
	2,1-pentenyl furan	down	LG7:0-10	Fb7:0-10	<0,05	4.45	36-45%	2
	2,3-butanedioldiacetate T	up	LG7:0-10	Fb7:0-10	<0,05	2.49	6-28%	2
	2-heptanol	down	LG4:9-44	Fb4:0-44	<0,05	1.94	23%	1
	2-methylbutyl acetate	down	LG7:43-59	Fb7:43-59	<0,05	<1,80	7-8%	2
	2-nonanol	down	LG1:26-61	Fb1:26-61	<0,05	<1,80	1%	2
	2-nonanol	down	LG5:11-35	Fb5:0-35	<0,05	<1,80	13-15%	2
	2-nonanol	down	LG4:9-44	Fb4:0-44	<0,05	4.95	48%	1
	2-nonanone	down	LG4:9-44	Fb4:0-44	<0,05	6.48	58%	1
	2-pentanone	down	LG4:9-44	Fb4:0-44	<0,05	2.09	25%	1

	2-pentylfuran	down	LG7:0-10	Fb7:0-10	<0,05	3.16	35%	1
	2-pentylfuran	up	LG2:0-30	Fb2:0-30	<0,05	3.16	21-35%	2
	2-tridecanol T	down	LG3:8-15	Fb3:0-15	<0,05	<1,80	3-17%	2
	2-undecanol T	down	LG4:20-44	Fb4:0-44	<0,05	<1,80	1-15%	2
	2-undecanol T	down	LG5:11-35	Fb5:0-35	<0,05	<1,80	18-20%	2
	2-undecanone T	down	LG4:20-44	Fb4:0-44	<0,05	<1,80	2-31%	2
	3-methyl-2-butenyl acetate	up	LG3:54-94	Fb3:54-94	<0,05	2.07	11-24%	2
	3-methyl-2-butenyl acetate	up	LG2:39-45	Fb2:39-47	<0,05	5.54	5-49%	2
	3-methylbutyl acetate	up	LG3:54-94	Fb3:54-94	<0,05	3.29	36%	1
	acetone	down	LG4:9-44	Fb4:0-44	<0,05	2.24	26%	1
	acetone	up	LG5:50-76	Fb5:50-76	<0,05	2.99	33%	1
	acetophenone	down	LG3:54-94	Fb3:54-94	<0,05	1.80	14-21%	2
	acetophenone	down	LG4:0-20	Fb4:0-20	<0,05	<1,80	6-15%	2
	acetophenone	down	LG6:101-101	Fb6:101-101	<0,05	<1,80	14-20%	2
	acetophenone	down	LG7:0-10	Fb7:0-10	<0,05	<1,80	8-19%	2
	a-farnesene	down	LG4:20-44	Fb4:0-44	<0,05	2.29	10-26%	2
	a-farnesene	down	LG3:8-15	Fb3:0-15	<0,05	<1,80	16%	2
	a-ionone	down	LG1:26-61	Fb1:26-61	<0,05	3.35	16-36%	2
	a-pinene	up	LG5:0-11	Fb5:0-11	<0,05	4.20	35-42%	2
	benzaldehyde	up	LG2:0-30	Fb2:0-30	<0,05	<1,80	13-19%	2
	benzyl acetate	down	LG7:0-10	Fb7:0-10	<0,05	2.26	15-26%	2
	benzyl acetate	down	LG6:101-101	Fb6:101-101	<0,05	<1,80	18-21%	2
	b-ionone	down	LG1:26-61	Fb1:26-61	<0,05	2.58	18-30%	2
→	butyl acetate	up	LG1:26-61	Fb1:26-61	<0,05	<1,80	6-15%	2
→	butyl butanoate	down	LG5:11-35	Fb5:0-35	<0,05	3.51	30-38%	2
→	butyl butanoate	down	LG7:0-10	Fb7:0-10	<0,05	<1,80	1-2%	2
	butyl hexanoate	down	LG5:0-11	Fb5:0-11	<0,05	3.26	30-35%	2

	butyl hexanoate	down	LG2:0-30	Fb2:0-30	<0,05	<1,80	19-20%	2
	butyl hexanoate	up	LG7:0-10	Fb7:0-10	<0,05	<1,80	11-14%	2
	cinnamyl acetate	down	LG3:54-94	Fb3:54-94	<0,05	<1,80	5-8%	2
	cinnamyl acetate	down	LG7:0-10	Fb7:0-10	<0,05	<1,80	4-6%	2
	decanal	up	LG4:0-20	Fb4:0-20	<0,05	2.51	29%	1
	decyl acetate	down	LG5:11-35	Fb5:0-35	<0,05	1.80	20-22%	2
	decyl acetate	down	LG4:20-44	Fb4:0-44	<0,05	<1,80	1-16%	2
	ethanol	up	LG7:0-10	Fb7:0-10	<0,05	3.36	36%	1
	ethyl 2-hexenoate	down	LG5:11-35	Fb5:0-35	<0,05	<1,80	15-18%	2
→	ethyl butanoate	down	LG7:0-10	Fb7:0-10	<0,05	<1,80	3-10%	2
	ethyl decanoate	down	LG1:26-61	Fb1:26-61	<0,05	<1,80	2-4%	2
	ethyl decanoate	up	LG7:0-10	Fb7:0-10	<0,05	3.59	10-38%	2
	ethyl dodecanoate	down	LG2:0-30	Fb2:0-30	<0,05	<1,80	4-15%	2
	ethyl dodecanoate	up	LG7:0-10	Fb7:0-10	<0,05	3.57	5-38%	2
→	ethyl hexanoate	down	LG1:26-61	Fb1:26-61	<0,05	<1,80	1-3%	2
	ethyl methylthioacetate T	down	LG7:0-10	Fb7:0-10	<0,05	1.84	1-22%	2
	ethyl octanoate	down	LG1:26-61	Fb1:26-61	<0,05	<1,80	1-4%	2
	ethyl octanoate	up	LG7:0-10	Fb7:0-10	<0,05	3.19	10-35%	2
	eugenol	up	LG5:50-76	Fb5:50-76	<0,05	4.44	33-45%	2
	hexanal	up	LG3:54-94	Fb3:54-94	<0,05	<1,80	6-7%	2
	hexyl butanoate	down	LG5:11-35	Fb5:0-35	<0,05	4.55	34-46%	2
	hexyl butanoate	down	LG7:0-10	Fb7:0-10	<0,05	<1,80	1%	2
	hexyl hexanoate	down	LG2:0-30	Fb2:0-30	<0,05	3.53	27-38%	2
	limonene	down	LG7:0-10	Fb7:0-10	<0,05	2.03	21-24%	2
→	linalool	up	LG3:0-8	Fb3:0-8	<0,05	6.64	54-59%	2
→	mesifurane	down	LG7:26-43	Fb7:26-45	<0,05	7.27	16-62%	2
→	methyl 2-aminobenzoate T	down	LG7:0-10	Fb7:0-10	<0,05	2.54	8-29%	2

→	methyl 2-aminobenzoate T	down	LG5:11-35	Fb5:0-35	<0,05	6.79	33-60%	2
	methyl 2-hexenoate	down	LG5:11-35	Fb5:0-35	<0,05	3.26	14-35%	2
	methyl 3-hydroxyoctanoate T	down	LG3:8-15	Fb3:0-15	<0,05	<1,80	4-6%	2
	methyl benzoate	down	LG3:54-94	Fb3:54-94	<0,05	<1,80	10-11%	2
	methyl benzoate	down	LG6:101-101	Fb6:101-101	<0,05	<1,80	13-15%	2
	methyl benzoate	down	LG7:0-10	Fb7:0-10	<0,05	<1,80	11-21%	2
	methyl benzoate	up	LG1:26-61	Fb1:26-61	<0,05	<1,80	14-18%	2
→	methyl butanoate	down	LG5:11-35	Fb5:0-35	<0,05	2.79	16-31%	2
→	methyl butanoate	down	LG7:0-10	Fb7:0-10	<0,05	<1,80	1-19%	2
→	methyl cinnamate T	up	LG2:0-30	Fb2:0-30	<0,05	2.81	18-32%	2
	methyl decanoate	down	LG5:0-11	Fb5:0-11	<0,05	2.49	24-28%	2
	methyl decanoate	down	LG4:9-44	Fb4:0-44	<0,05	2.71	31%	1
	methyl dodecanoate	down	LG1:26-61	Fb1:26-61	<0,05	<1,80	1%	2
	methyl dodecanoate	down	LG2:0-30	Fb2:0-30	<0,05	<1,80	8-18%	2
	methyl dodecanoate	down	LG5:11-35	Fb5:0-35	<0,05	<1,80	13-14%	2
	methyl dodecanoate	up	LG7:0-10	Fb7:0-10	<0,05	2.80	31%	1
→	methyl hexanoate	down	LG5:11-35	Fb5:0-35	<0,05	5.54	35-52%	2
	methyl octanoate	down	LG5:0-11	Fb5:0-11	<0,05	4.83	43-48%	2
	myrtenol	down	LG7:0-10	Fb7:0-10	<0,05	4.32	8-44%	2
	myrtenol	down	LG3:8-15	Fb3:0-15	<0,05	<1,80	2-7%	2
	myrtenol	up	LG5:50-76	Fb5:50-76	<0,05	6.71	60%	1
→	myrtenyl acetate	down	LG5:11-35	Fb5:0-35	<0,05	4.67	45-47%	2
→	myrtenyl acetate	down	LG6:101-101	Fb6:101-101	<0,05	<1,80	1%	2
	nerol	down	LG4:9-20	Fb4:0-20	<0,05	4.40	17-45%	2
	nerol	down	LG7:43-59	Fb7:43-59	<0,05	<1,80	7-11%	2
	nerol	up	LG5:50-76	Fb5:50-76	<0,05	4.33	38-44%	2
→	nerolidol	up	LG3:0-8	Fb3:0-8	<0,05	22.38	76-95%	2

	nonanal	down	LG7:43-59	Fb7:43-59	<0,05	5.16	50%	1
	octanal	down	LG7:43-59	Fb7:43-59	<0,05	2.85	32%	1
	octyl acetate	down	LG5:11-35	Fb5:0-35	<0,05	2.40	25-27%	2
	octyl butanoate	down	LG2:0-30	Fb2:0-30	<0,05	2.11	21-25%	2
	octyl hexanoate	down	LG2:0-30	Fb2:0-30	<0,05	1.95	21-23%	2
	octyl hexanoate	down	LG1:26-61	Fb1:26-61	<0,05	<1,80	1-2%	2
	octyl hexanoate	down	LG5:0-11	Fb5:0-11	<0,05	<1,80	13-20%	2
	pentyl acetate	up	LG1:26-61	Fb1:26-61	<0,05	1.99	8-23%	2
	propyl butanoate	down	LG7:0-10	Fb7:0-10	<0,05	2.81	5-31%	2
	propyl butanoate	up	LG1:26-61	Fb1:26-61	<0,05	<1,80	7-13%	2

Table 3 Differentially expressed genes (DEG) summary

		a2			
NIL vs. RV	**Introgression size (Mb)**	**DEG**	**blast homologies**	**Up regulated**	**Down regulated**
Fb5:0-35	6.51	257	218	106	151
Fb7:0-10	14.20	442	367	204	234

1018

1019

Table 4 Metabolic pathways affected

NIL vs. RV	pathway	p-value
Fb5:0-35	glutathione biosynthesis	1.51E-02
Fb5:0-35	ß-alanine biosynthesis I	1.58E-02
Fb5:0-35	farnesene biosynthesis	2.22E-02
Fb5:0-35	cis-zeatin biosynthesis	2.26E-02
Fb5:0-35	linalool biosynthesis	3.41E-02
Fb5:0-35	g-glutamyl cycle	4.48E-02
Fb7:0-10	valine degradation I	7.67E-03
Fb7:0-10	divinyl ether biosynthesis II (13-LOX)	1.89E-02
Fb7:0-10	13-LOX and 13-HPL pathway	1.89E-02
Fb7:0-10	asparagine degradation I	3.38E-02
Fb7:0-10	homogalacturonan degradation	4.82E-02

1030 **Table 5** Selected candidate genes

1031

comparison vs. RV	gene id[a]	log_2(fold change)[b]	p-value	p-adjusted	blast hit[c]	blast hit protein description	predicted function in reference annotation (a1)
Fb5:0-35	maker-LG4-augustus-gene-138.110-mRNA-1	-Inf	1.76E-10	6.05E-08	PAT1_ARATH	Scarecrow-like transcription factor PAT1	
Fb5:0-35	mrna09934.1-v1.0-hybrid	-11.82	4.56E-44	8.49E-40	F4JBC7_ARATH	HXXXD-type acyl-transferase-like protein	Vinorine synthase (probable)
Fb5:0-35	maker-LG4-snap-gene-135.249-mRNA-1	-2.09	2.79E-05	3.74E-03	STPS1_SANAL	Sesquiterpene synthase	(+)-delta-cadinene synthase isozyme A (D-cadinene synthase A) (probable)
Fb5:0-35	augustus_masked-LG6-processed-gene-175.2-mRNA-1	1.62	5.65E-04	4.81E-02	EIF3C_ARATH	Eukaryotic translation initiation factor 3 subunit C	
Fb5:0-35	mrna32494.1-v1.0-hybrid	1.68	6.21E-04	5.20E-02	GL3_ARATH	Transcription factor GLABRA 3	Transcription factor GLABRA 3 (bHLH 1) (putative)
Fb5:0-35	maker-LG4-augustus-gene-136.257-mRNA-1	3.97	2.44E-12	9.39E-10	MFS_MENPI	(+)-menthofuran synthase	
Fb7:0-10	maker-LG7-snap-gene-1.135-mRNA-1	-Inf	5.34E-04	2.88E-02	F4KGA3_ARATH	Putative PHD finger transcription factor	
Fb7:0-10	maker-LG7-snap-gene-129.164-mRNA-1	-Inf	7.03E-08	8.53E-06	ZDH22_ARATH	Protein S-acyltransferase 24	
Fb7:0-10	snap_masked-LG7-processed-gene-42.93-mRNA-1	-7.15	2.77E-09	4.07E-07	VRN1_ARATH	B3 domain-containing transcription factor VRN1	
Fb7:0-10	mrna23606.1-v1.0-hybrid	-6.85	1.24E-14	3.45E-12	LOX2_ORYSJ	Linoleate 9S-lipoxygenase 2	3-deoxy-manno-octulosonate cytidylyltransferase (CKS) (similar to)
Fb7:0-10	augustus_masked-LG7-processed-gene-126.10-mRNA-1	-6.69	2.99E-05	2.20E-03	TA12B_ARATH	Transcription initiation factor TFIID subunit 12b	

Fb7:0-10	mrna34011.1-v1.0-hybrid	-6.02	1.38E-25	1.52E-22	F4JBC7_ARATH	HXXXD-type acyl-transferase-like protein	BAHD acyltransferase At5g47980 (probable)
Fb7:0-10	maker-LG3-augustus-gene-99.141-mRNA-1	-5.22	2.93E-10	4.85E-08	HIBC1_ARATH	3-hydroxyisobutyryl-CoA hydrolase 1	3-hydroxyisobutyryl-CoA hydrolase, mitochondrial (HIB-CoA hydrolase), Precursor (probable)
Fb7:0-10	augustus_masked-LG7-processed-gene-56.12-mRNA-1	-4.02	1.56E-10	2.72E-08	LOXC2_ORYSJ	Probable lipoxygenase 8, chloroplastic	Probable lipoxygenase 8, chloroplastic, Precursor (similar to)
Fb7:0-10	maker-LG7-augustus-gene-88.90-mRNA-1	-3.90	4.37E-13	9.30E-11	HIBC1_ARATH	3-hydroxyisobutyryl-CoA hydrolase 1	3-hydroxyisobutyryl-CoA hydrolase, mitochondrial (HIB-CoA hydrolase), Precursor (probable)
Fb7:0-10	augustus_masked-LG7-processed-gene-56.13-mRNA-1	-3.51	6.02E-09	8.65E-07	LOXC2_ORYSJ	Probable lipoxygenase 8, chloroplastic	Probable lipoxygenase 8, chloroplastic, Precursor (similar to)
Fb7:0-10	augustus_masked-LG3-processed-gene-102.20-mRNA-1	-2.30	6.74E-06	5.72E-04	ERF61_ARATH	Ethylene-responsive transcription factor ERF061	
Fb7:0-10	maker-LG5-snap-gene-206.105-mRNA-1	-1.97	3.70E-05	2.70E-03	GLYC7_ARATH	Serine hydroxymethyltransferase 7	Serine hydroxymethyltransferase 2 (SHMT 2) (probable)
Fb7:0-10	maker-LG7-augustus-gene-95.135-mRNA-1	-1.79	2.12E-04	1.29E-02	VRN1_ARATH	B3 domain-containing transcription factor VRN1	
Fb7:0-10	genemark-LG7-processed-gene-22.65-mRNA-1	-1.75	1.91E-04	1.19E-02	F4JW79_ARATH	Kow domain-containing transcription factor 1	
Fb7:0-10	augustus_masked-LG6-processed-gene-175.2-mRNA-1	-1.56	1.31E-03	6.03E-02	EIF3C_ARATH	Eukaryotic translation initiation factor 3 subunit C	
Fb7:0-10	maker-LG3-augustus-gene-10.249-mRNA-1	0.44	3.61E-01	1.00E+00	ASAT1_ARATH	Acyl-CoA--sterol O-acyltransferase 1	Probable long-chain-alcohol O-fatty-acyltransferase 5
Fb7:0-10	maker-LG6-augustus-gene-341.179-mRNA-1	1.86	1.53E-03	6.87E-02	TPS10_RICCO	Terpene synthase 10	Myrcene synthase, chloroplastic, Precursor (probable)
Fb7:0-10	augustus_masked-	2.83	1.59E-	7.05E-02	ZDH14_ARATH	Probable protein S-	Probable S-acyltransferase

	LG7-processed-gene-50.27-mRNA-1		03			acyltransferase 14	At3g60800 (putative)
Fb7:0-10	maker-LG5-augustus-gene-34.145-mRNA-1	3.14	5.88E-04	3.12E-02	LOX2_ARATH	Lipoxygenase 2, chloroplastic	Lipoxygenase 2, chloroplastic (AtLOX2), Precursor (similar to)
Fb7:0-10	maker-LG7-snap-gene-91.103-mRNA-1	3.35	3.42E-05	2.50E-03	MYC2_ARATH	Transcription factor MYC2	
Fb7:0-10	maker-LG7-augustus-gene-8.110-mRNA-1	6.44	4.87E-22	2.37E-19	NAC86_ARATH	NAC domain-containing protein 86	
Fb7:0-10	mrna23453.1-v1.0-hybrid	11.01	2.67E-20	2.02E-17	O23392_ARATH	HXXXD-type acyl-transferase family protein	Vinorine synthase (probable)
Fb7:0-10	mrna34009.1-v1.0-hybrid	13.29	3.28E-39	6.22E-35	F4JBC7_ARATH	HXXXD-type acyl-transferase-like protein	Vinorine synthase (probable)
Fb7:0-10	augustus_masked-LG7-processed-gene-21.17-mRNA-1	Inf	3.17E-05	2.33E-03	HIBC1_ARATH	3-hydroxyisobutyryl-CoA hydrolase 1	

1032 [a] gene id is according to F. vesca annotation version 2 nomenclature

1033 [b] log2(fold change) values use as reference RV, so negative values indicate down-regulation in NIL and positive values up-regulation in NIL

1034 [c] best blast hit found for the DEG predicted proteins. Codes are according to UniProtUK entries

1035

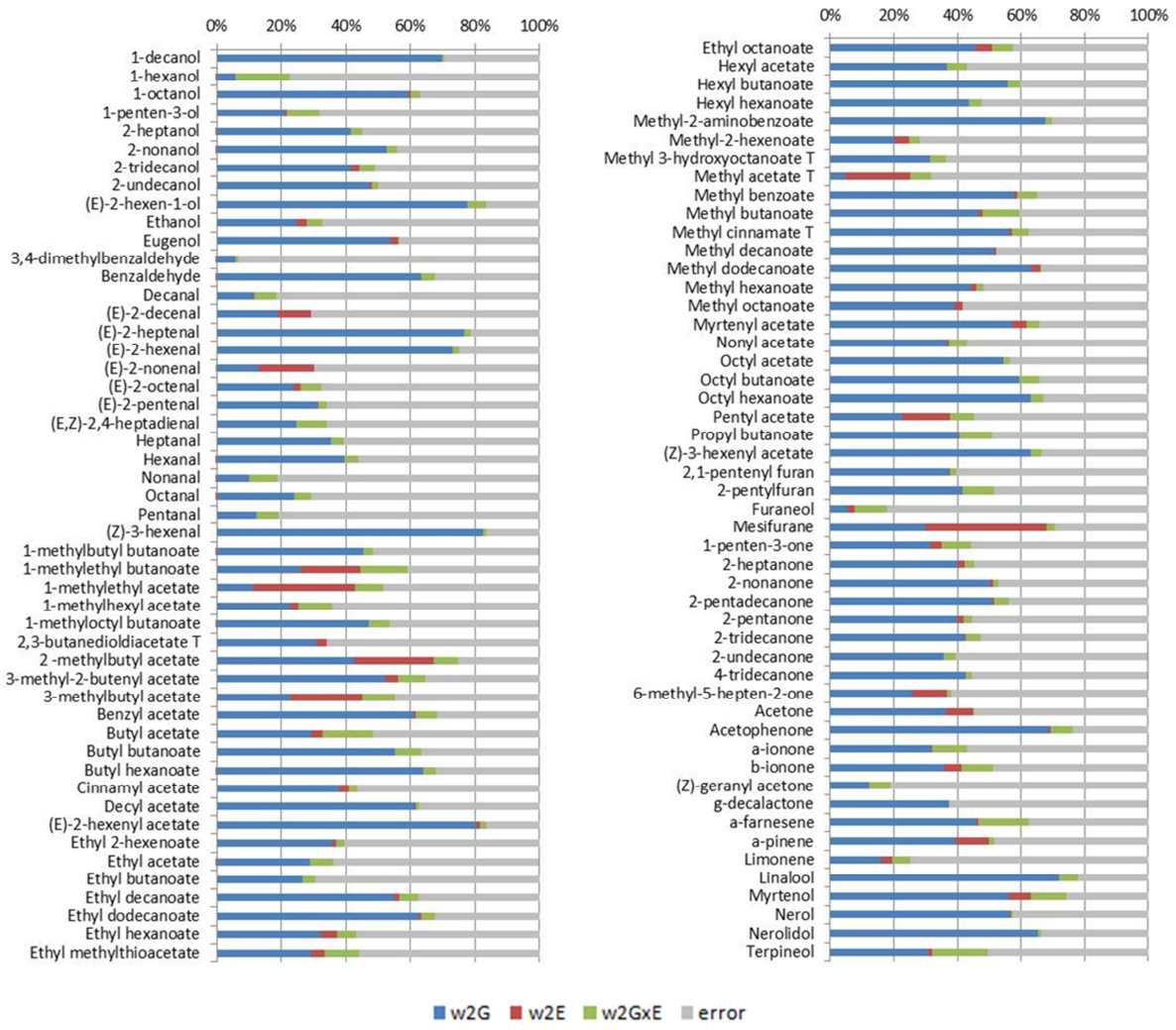
1036
1037

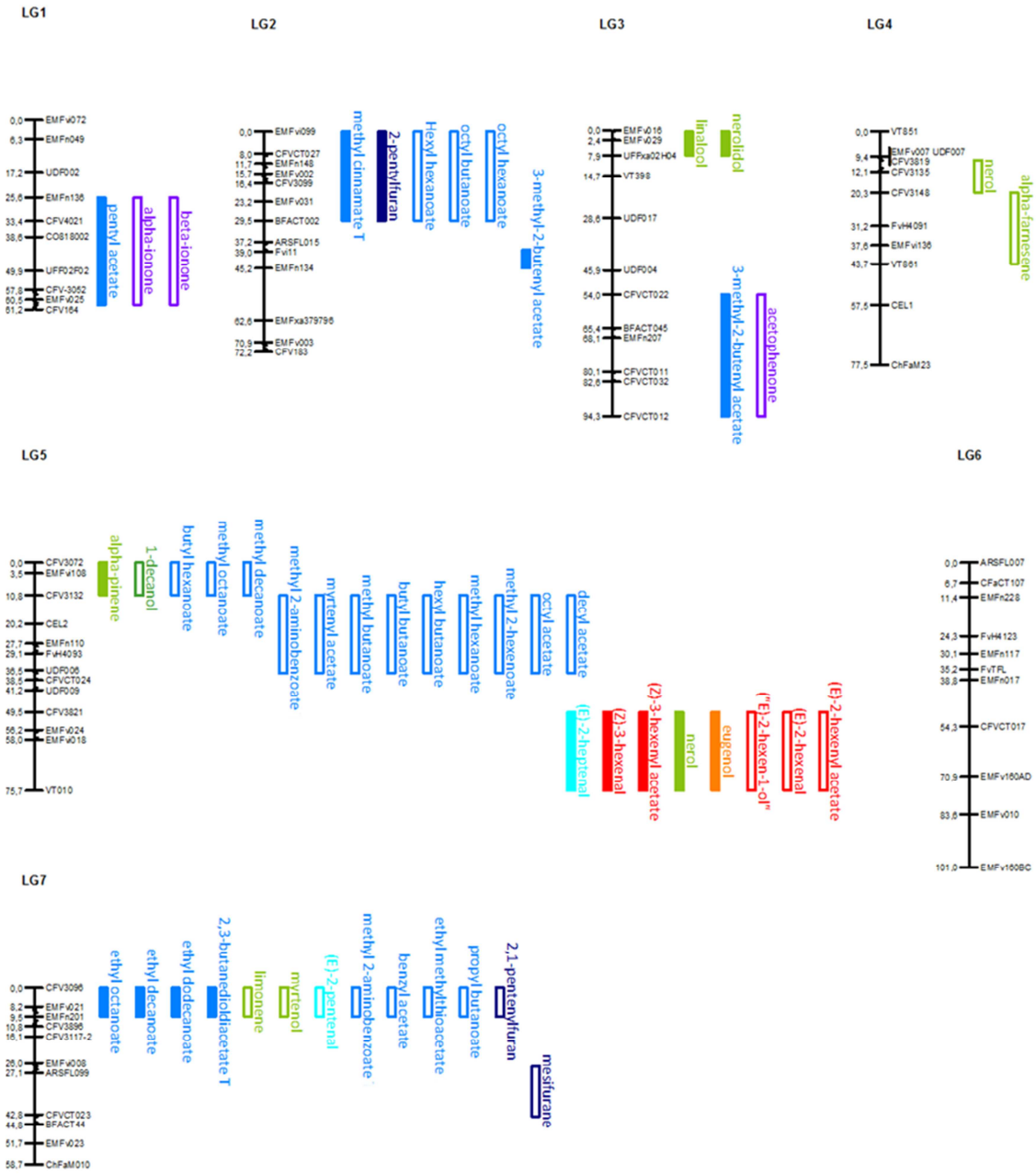
1038 **Table 6** Polymorphism summary

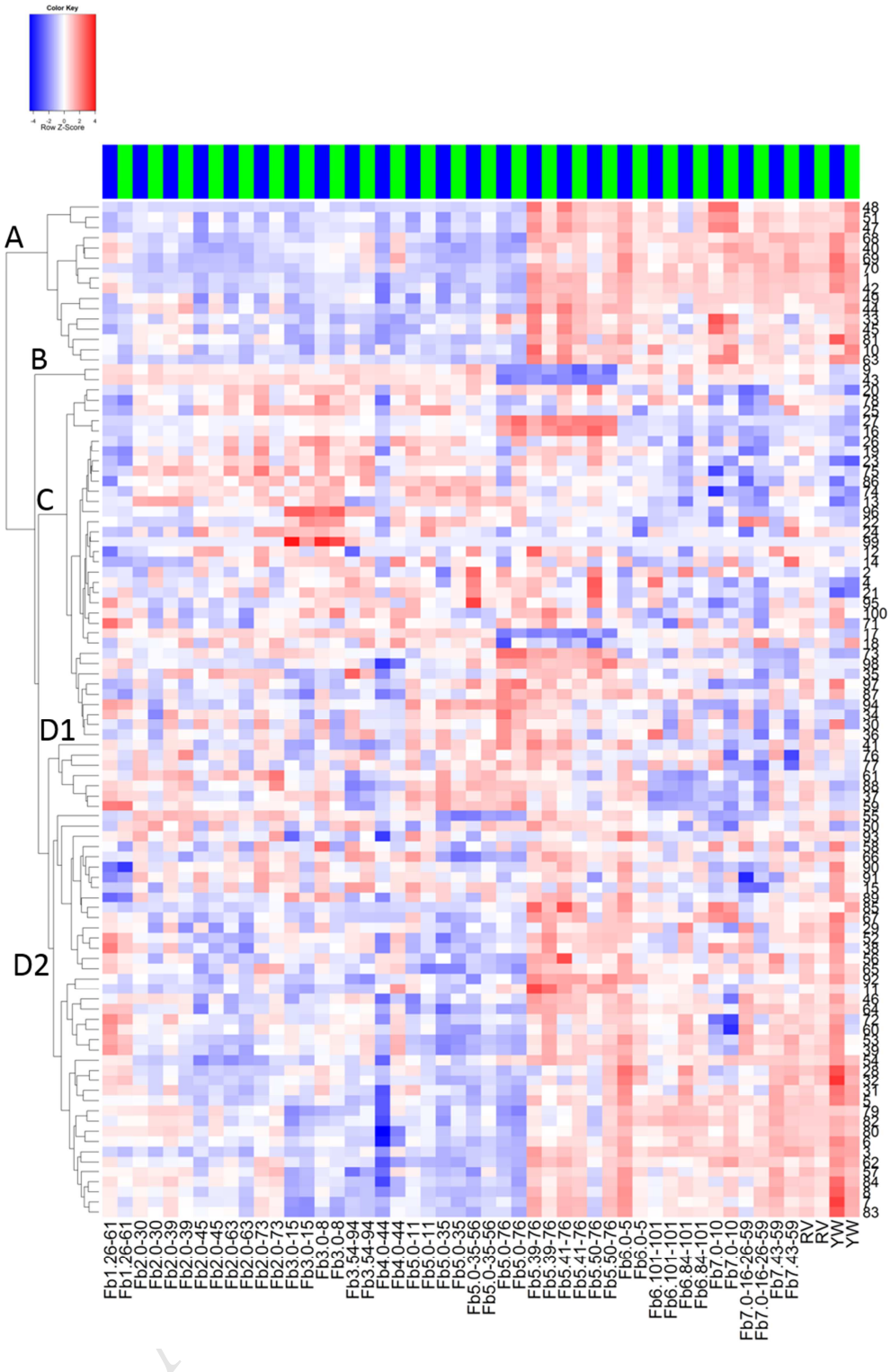
	Fb5:0-35	Fb7:0-10	total
Introgression cM	35	10	
Introgression bp	5.593.948	15.652.556	
SNPs vs. RV	6622	10517	17139
Indels vs. RV	191	333	524
total polymorphisms	6813	10850	17663

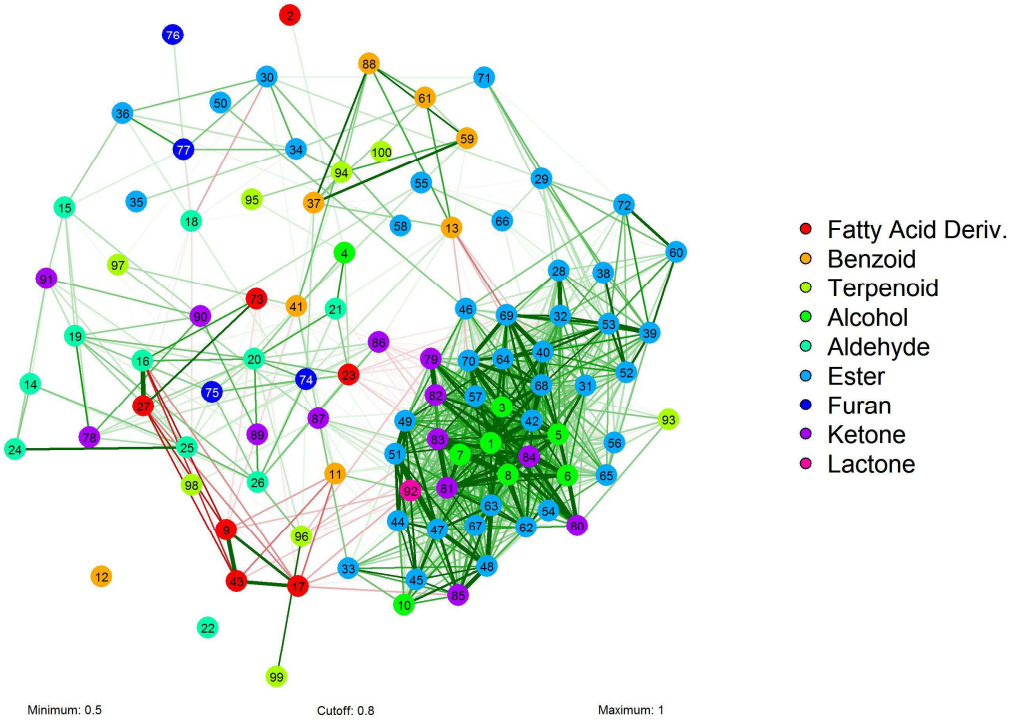
1039

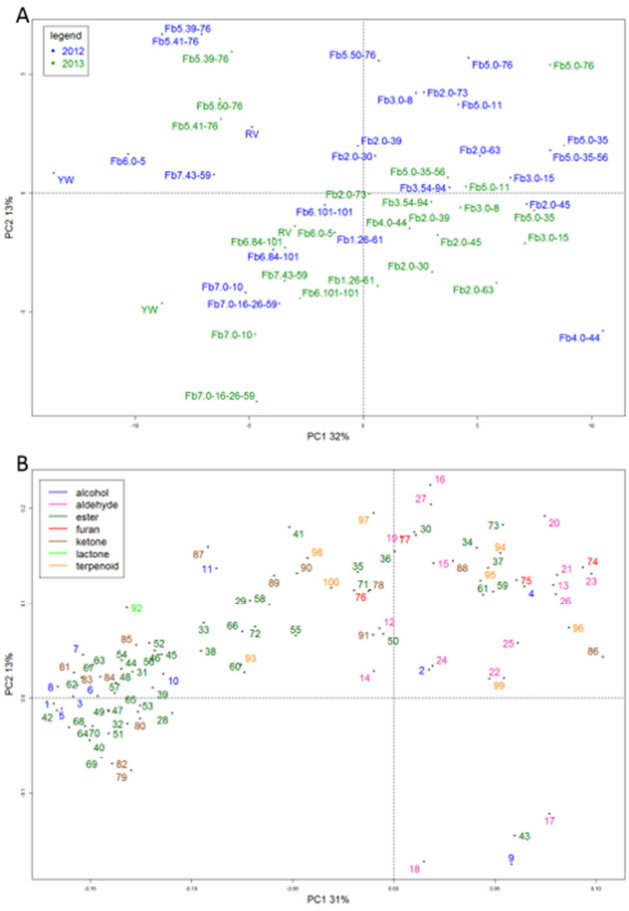
1040

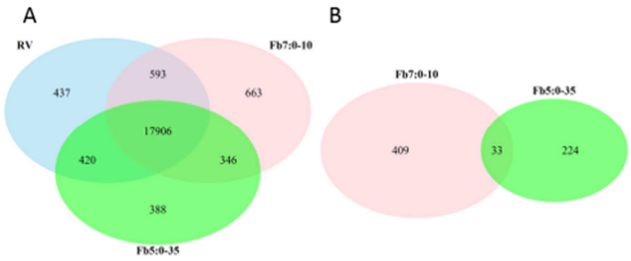

0%
20%
40%
60%
80%
100%
1-decanol
1-hexanol
1-octanol
1-penten-3-ol
2-heptanol
2-nonanol
2-tridecanol
2-undecanol
(E)-2-hexen-1-ol
Ethanol
Eugenol
3,4-dimethylbenzaldehyde
Benzaldehyde
Decanal
(E)-2-decenal
(E)-2-heptenal
(E)-2-hexenal
(E)-2-nonenal
(E)-2-octenal
(E)-2-pentenal
(E,Z)-2,4-heptadienal
Heptanal
Hexanal
Nonanal
Octanal
Pentanal
(Z)-3-hexenal
1-methylbutyl butanoate
1-methylethyl butanoate
1-methylethyl acetate
1-methylhexyl acetate
1-methyloctyl butanoate
2,3-butanedioldiacetate T
2 -methylbutyl acetate
3-methyl-2-butenyl acetate
3-methylbutyl acetate
Benzyl acetate
Butyl acetate
Butyl butanoate
Butyl hexanoate
Cinnamyl acetate
Decyl acetate
(E)-2-hexenyl acetate
Ethyl 2-hexenoate
Ethyl acetate
Ethyl butanoate
Ethyl decanoate
Ethyl dodecanoate
Ethyl hexanoate
Ethyl methylthioacetate
0%
20%
40%
60%
80%
100%
Ethyl octanoate
Hexyl acetate
Hexyl butanoate
Hexyl hexanoate
Methyl-2-aminobenzoate
Methyl-2-hexenoate
Methyl 3-hydroxyoctanoate T
Methyl acetate T
Methyl benzoate
Methyl butanoate
Methyl cinnamate T
Methyl decanoate
Methyl dodecanoate
Methyl hexanoate
Methyl octanoate
Myrtenyl acetate
Nonyl acetate
Octyl acetate
Octyl butanoate
Octyl hexanoate
Pentyl acetate
Propyl butanoate
(Z)-3-hexenyl acetate
2,1-pentenyl furan
2-pentylfuran
Furaneol
Mesifurane
1-penten-3-one
2-heptanone
2-nonanone
2-pentadecanone
2-pentanone
2-tridecanone
2-undecanone
4-tridecanone
6-methyl-5-hepten-2-one
Acetone
Acetophenone
a-ionone
b-ionone
(Z)-geranyl acetone
g-decalactone
a-farnesene
a-pinene
Limonene
Linalool
Myrtenol
Nerol
Nerolidol
Terpineol
w2G
w2E
w2GxE
error


LG1
0,0 EMFv072
6,3 EMFn049
17,2 UDF002
25,6 EMFn136
33,4 CFV4021
38,6 COS18002
49,9 UFR02F02
57,8 CFV-3052
60,5 EMFv025
61,2 CFV164
pentyl acetate
alpha-ionone
beta-ionone
LG2
0,0 EMFv099
8,0 CFVCT027
11,7 EMFn148
15,7 EMFv002
16,4 CFV3099
23,2 EMFv031
29,5 BFACT002
37,2 ARSFL015
39,0 Fvi11
45,2 EMFn134
62,6 EMFxa379796
70,9 EMFv003
72,2 CFV183
2-pentylfuran
methyl cinnamate T
Hexyl hexanoate
octyl butanoate
octyl hexanoate
3-methyl-2-butenyl acetate
LG3
0,0 EMFv016
2,4 EMFv029
7,9 UFFxa02H04
14,7 VT398
28,6 UDF017
45,9 UDF004
54,0 CFVCT022
65,4 BFACT045
68,1 EMFn207
80,1 CFVCT011
82,6 CFVCT032
94,3 CFVCT012
linalool
nerolidol
3-methyl-2-butenyl acetate
acetophenone
LG4
0,0 VT851
9,4 EMFv007 UDF007
12,1 CFV3819
20,3 CFV3135
31,2 CFV3148
37,6 FvH4091
43,7 EMFvi136
57,5 VT861
77,5 CEL1
ChFaM23
nerol
alpha-farnesene
LG5
0,0 CFV3072
3,5 EMFvi108
10,8 CFV3132
20,2 CEL2
27,7 EMFn110
29,1 FvH4093
36,5 UDF006
38,5 CFVCT024
41,2 UDF009
49,5 CFV3821
56,2 EMFv024
58,0 EMFv018
75,7 VT010
alpha-pinene
1-decanol
butyl hexanoate
methyl octanoate
methyl decanoate
methyl 2-aminobenzoate
myrtenyl acetate
methyl butanoate
butyl butanoate
hexyl butanoate
methyl hexanoate
methyl 2-hexenoate
octyl acetate
decyl acetate
(E)-2-heptenal
(Z)-3-hexenal
(Z)-3-hexenyl acetate
nerol
eugenol
("E)-2-hexen-1-ol"
(E)-2-hexenal
(E)-2-hexenyl acetate
LG6
0,0 ARSFL007
6,7 CFaCT107
11,4 EMFn228
24,3 FvH4123
30,1 EMFn117
35,2 FvTFL
38,8 EMFn017
54,3 CFVCT017
70,9 EMFv160AD
83,6 EMFv010
101,0 EMFv160BC
LG7
0,0 CFV3096
8,2 EMFv021
9,5 EMFn201
10,8 CFV3896
16,1 CFV3117-2
26,0 EMFv008
27,1 ARSFL099
42,8 CFVCT023
44,8 BFACT44
51,7 EMFv023
58,7 ChFaM010
ethyl octanoate
ethyl decanoate
ethyl dodecanoate
2,3-butanediol diacetate T
limonene
myrtenol
(E)-2-pentenal
methyl 2-aminobenzoate
benzyl acetate
ethyl methylthioacetate
propyl butanoate
2,1-pentenylfuran
mesifurane

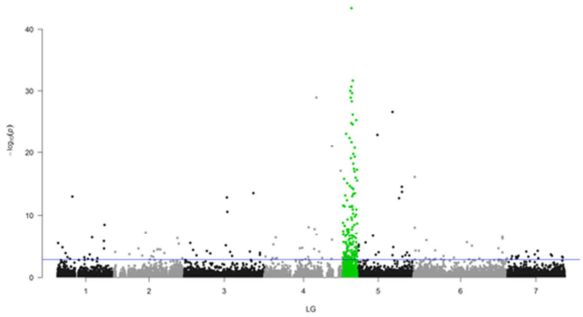

Color Key
Row Z-Score
A
B
C
D1
D2
Fb1.26-61
Fb1.26-61
Fb2.0-30
Fb2.0-30
Fb2.0-39
Fb2.0-39
Fb2.0-45
Fb2.0-45
Fb2.0-63
Fb2.0-63
Fb2.0-73
Fb2.0-73
Fb3.0-15
Fb3.0-15
Fb3.0-8
Fb3.0-8
Fb3.54-94
Fb3.54-94
Fb4.0-44
Fb4.0-44
Fb5.0-11
Fb5.0-11
Fb5.0-35
Fb5.0-35
Fb5.0-35-56
Fb5.0-35-56
Fb5.0-76
Fb5.0-76
Fb5.39-76
Fb5.39-76
Fb5.41-76
Fb5.41-76
Fb5.50-76
Fb5.50-76
Fb6.0-5
Fb6.0-5
Fb6.101-101
Fb6.101-101
Fb6.84-101
Fb6.84-101
Fb7.0-10
Fb7.0-10
Fb7.0-16-26-59
Fb7.0-16-26-59
Fb7.43-59
Fb7.43-59
RV
RV
YW
YW



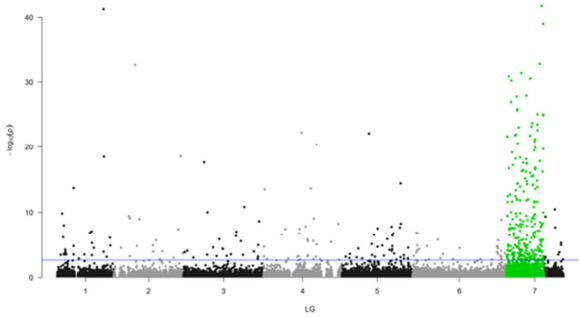




Fb5:0-35



Fb7:0-10



- Volatile composition of wild strawberry as model of octoploide cultivated fruit.
- NIL collection a tool to explore genetic variability of fruit quality traits and aroma volatiles
- 50 major QTLs controlling volatile accumulation to increase wild strawberry flavour
- Two wild strawberry genome regions harbor key aroma volatile QTL
- Differences in gene expression between NILs show possible genes important to enhance aroma.

Author contribution statement:

MU analysed the NILs collection, prepared fruit samples, did the statistical analyses, prepared RNAseq samples and evaluated the DEG, in addition to writing the manuscript. AG and JLR did the GC-MS analyses and participated in edition of the manuscript. KA did the SNPs calling analyses. AM lead the project, participated in all steps of phenotyping and in writing the manuscript.

The authors declare that they have no conflict of interest.