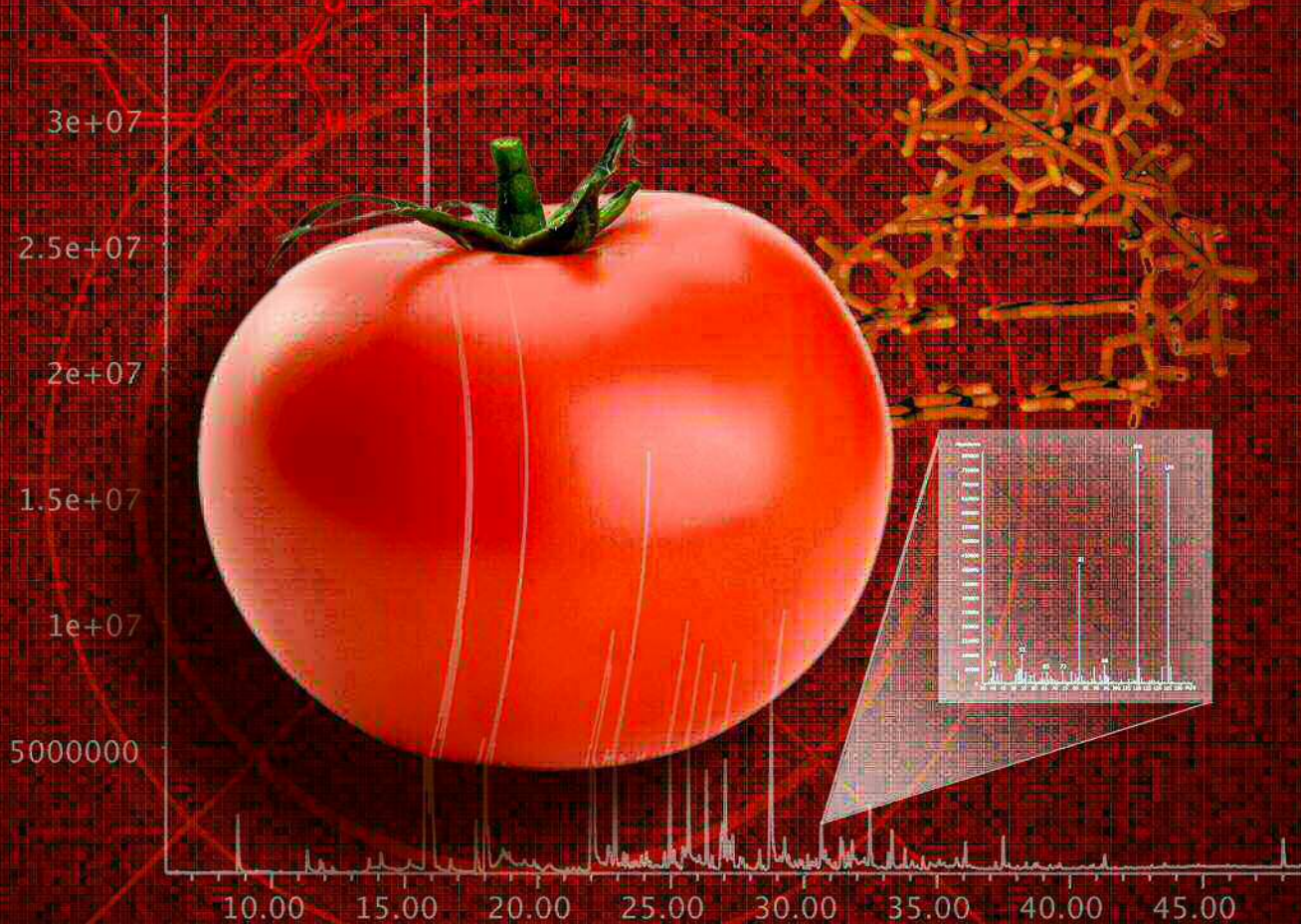


# Genes and genomic regions related to the production of volatile compounds in the tomato fruit

José Luis Rambla Nebot

Supervisor: Antonio Granell Richart

January 2016







**Genes and genomic regions related to  
the production of volatile compounds  
in the tomato fruit**

José Luis Rambla Nebot

Supervisor: Antonio Granell Richart

*January 2016*

Diseño portada y contraportada:

Juan Diego Ingelmo

*Todo tiene su momento,  
y cada cosa su tiempo bajo el cielo.*

*(Qohélet 3, 1)*



*A mis hijos María, José y Francisco,  
que son una bendición de Dios.*





## Agradecimientos

Quiero agradecerle en primer lugar a Antonio Granell la confianza que ha depositado en mí desde el inicio y a lo largo de todo este tiempo, favoreciendo así que recuperara la autoestima que había perdido. Asimismo, le guardo gratitud por las muchas oportunidades que me ha ofrecido de aprender en el campo de la Metabolómica, en el cual me introdujo en un tiempo en el que ni yo, ni nadie de mi entorno, sabíamos aún qué significaba esa palabra.

Quiero manifestar igualmente mi gratitud a Juan Carbonell y a Mari Ángeles Argomániz, *alma mater* del gran grupo de “cuatrolabos”, que me acogieron con gran calidez cuando aún no formaba parte de su equipo. Su cercanía y amistad, fruto de su natural bonhomía, han sido para mí un firme apoyo a lo largo de todos estos años.

Mi profunda gratitud y reconocimiento a Ana Ahuir, profesional ejemplar y mejor persona, sin cuyo apoyo no hubiera podido resistir en los momentos difíciles, y sin el cual, sin duda, no hubiera llegado hasta aquí. Verdaderamente ha sido un lujo, para mí y para el Instituto, el contar con una persona como ella.

Quiero agradecerle a Antonio Monforte cuanto me ha enseñado relativo al análisis de QTLs, en el que mi formación previa era nula, y particularmente por su paciencia y su buena disposición.

También quiero manifestar mi gratitud a todos los compañeros del laboratorio y de los grupos afines donde he desarrollado mi trabajo, y a todos los miembros de otros grupos que de algún modo me han ayudado en este tiempo. Aun no pudiendo nombrar a cada uno por razones de espacio, sí quiero hacer mención particular de Aurora, que es la bondad personificada, de Pako, gran profesional, de Leandro, de quien aprendí a trabajar con *microarrays*, y de María Dolores Gómez, Ana Berbel y Cristina Navarro, por su espontánea ayuda

cuando acababa de embarcarme en la Biología Molecular y andaba aún muy perdido.

Asimismo, quiero expresar mi sincero agradecimiento a todos los miembros de esa gran familia que es el Centro de Ecología Química Agrícola (CEQA): Cristina, Nuria, Vicente, MAPI, Sandra, Juan, Ilde, Javi, Ismael y Pilar, que, lejos de considerarme un intruso, me acogieron desde el primer día y me trataron como a uno más del equipo; y particularmente a su director, Jaime Primo, por su cercanía y calidez humana, y por haber puesto absolutamente todas las facilidades para trabajar con sus equipos (de día y de noche, incluyendo el fin de semana) durante los primeros dos años y medio de este trabajo, cuando en el IBMCP aún no teníamos instrumentación para poder hacer investigación en Metabolómica.

Igualmente agradecido me siento hacia Lourdes Gómez, Ángela, Oussama y todo el equipo de la UCLM en Albacete, con quienes trabajar ha sido siempre muy estimulante, y sin cuyo apoyo clave en algunos momentos críticos tal vez esta tesis nunca hubiera visto la luz.

Debo asimismo mostrar consideración hacia Paola, mi esposa, por haber tenido que soportar mi perseverancia en dedicarme a la investigación, incluso en periodos de gran precariedad en los que se ha visto comprometido el sustento de nuestra familia.

Y, desde luego, sería injusto dejar de agradecer a mis padres por todo cuanto han hecho por mí a lo largo de mi vida. Su amor y su confianza han sido siempre para mí un apoyo firme. Y, en lo relativo a esta tesis, debo agradecerles el apoyo incondicional que, junto con mi hermano, me han brindado en este largo tiempo. En honor a la verdad, bien se puede decir que esta tesis que ahora ve la luz también es suya.





## SUMMARY

Fruits both produce and emit volatile chemical compounds. These are short-chained low polarity molecules involved in many processes, and they are responsible of our perception of fruit aroma and of most of their flavour. This thesis is focused on the study of volatile compounds in the tomato fruit, which is one of the most important horticultural crop worldwide and is a model system for the study of fruit development and ripening.

Some of the analytical methods more frequently used for the analysis of tomato fruit volatiles were systematically compared. Results revealed that the observed volatile profile is highly dependent on the precise analytical method used, both for sample processing and for the technique used for volatile acquisition. It was concluded that the method of election for the comparison of large sets of samples from a multi-omics approach consists on flash freezing the biological material with liquid nitrogen at the selected ripening stage and the use of headspace solid phase microextraction coupled to gas chromatography and mass spectrometry for the subsequent analysis.

This method was implemented and was used for the determination of volatile compounds in selected NILs harbouring QTLs for characters related to flavour and aroma introgressed in different genetic backgrounds. The results allowed the association of several of the organoleptic characters previously identified with modified levels of several volatiles. It was also observed that the genetic background has a major effect on the production of such metabolites. Correlation analysis between the levels of volatiles and primary metabolites led to the conclusion that the production of volatile compounds is generally not determined by the levels of their precursors. Its regulation is most likely to be due to downstream processes such as the availability of either precursors or

intermediate metabolites, the variability in specific processes leading to the conversion of precursors into volatiles, or to other still unknown regulatory mechanisms.

Volatile compounds were also studied in a RIL population derived from a cross between *Solanum pimpinellifolium* accession TO-937, the closest species to cultivated tomato, which produces red fruits, and *S. lycopersicum* cv. 'Moneymaker', a tomato variety for the fresh market. This allowed the identification of 102 QTLs for 39 different volatile compounds, 76 of which had not been previously described. All these QTLs were mapped along the 12 tomato chromosomes by means of the SOLCAP SNPs molecular marker map. Most of the QTLs identified were subsequently evaluated on introgression lines (ILs) generated from the same original genotypes. It was observed that almost half of the QTLs previously identified retained their effect after introgression in the 'Moneymaker' genetic background. Additionally, 16 new QTLs were identified in this IL population. Based on the existing knowledge about the effect of volatile compounds on our perception of flavour and aroma and also on their ability to maintain their effect after introgression in the cultivated tomato, some of the QTLs identified are good candidates to be used in tomato flavour breeding programs.

Eventually, the comparison of the localization in the genome of the QTLs identified in the different populations studied with those already described in the literature revealed a very low degree of co-localization between the different QTLs. This implies that there exists a wide range of variability in the wild species related to tomato available for breeding tomato flavour and aroma.

## RESUMEN

Los frutos producen y emiten compuestos químicos volátiles. Estos son moléculas en general poco polares y de cadena corta que cumplen diversas funciones, y son las responsables de que percibamos el aroma y buena parte del sabor de los frutos. Esta tesis está centrada en el estudio de los volátiles del fruto del tomate, que es uno de los cultivos hortícolas más importantes y un sistema modelo para el estudio del desarrollo y maduración del fruto.

Se compararon de forma sistemática los métodos analíticos más comúnmente utilizados para el análisis de volátiles en fruto de tomate, y se observó que el perfil de volátiles detectado está fuertemente condicionado por el método analítico utilizado, tanto por el proceso de preparación de la muestra como por la técnica de adquisición de los volátiles. Finalmente se concluyó que la técnica más adecuada para la comparación de grandes grupos de muestras desde una aproximación multi-ómica consiste en congelar con nitrógeno líquido el material vegetal una vez alcanzado el momento idóneo de recolección, y su análisis posterior mediante microextracción en fase sólida (SPME) acoplada a cromatografía de gases y espectrometría de masas.

Se puso a punto esta técnica y se utilizó para la determinación de los compuestos volátiles en varias líneas NILs portadoras de QTLs de caracteres relacionados con el sabor y el aroma en distintos fondos genéticos. Los resultados permitieron asociar varios de los caracteres organolépticos identificados con alteraciones en los niveles de algunos volátiles. Igualmente se observó que el fondo genético tiene un efecto importante sobre la producción de estos metabolitos. Los análisis de correlación entre los niveles de volátiles y metabolitos primarios permitieron concluir que la producción de compuestos volátiles, en general, no está determinada por los niveles de sus precursores,

sino que su regulación debe encontrarse más bien en procesos posteriores, tales como la disponibilidad de los precursores o de metabolitos intermedios, variabilidad en procesos específicos relacionados con la conversión de los precursores en volátiles, o algún otro mecanismo regulador aún desconocido.

También se estudiaron los volátiles en una población de RILs derivada de un cruce entre *Solanum pimpinellifolium* entrada TO-937, la especie más próxima al tomate cultivado, la cual produce frutos rojos, y *S. lycopersicum* cv. “Moneymaker”, una variedad de tomate para el mercado en fresco. Esto permitió identificar 102 QTLs para 39 volátiles diferentes, 76 de las cuales no se habían descrito previamente, las cuales se mapearon a lo largo de los 12 cromosomas del tomate utilizando el mapa de marcadores moleculares de SNPs SOLCAP. Posteriormente se evaluaron la mayoría de las QTLs identificadas mediante la determinación de los volátiles en líneas ILs generadas a partir de los mismos materiales. Se observó que casi la mitad de estas QTLs mantuvieron su efecto al ser introgresadas en el fondo genético “Moneymaker”, al tiempo que 16 nuevas QTLs se identificaron en esta población de ILs. Algunas de las QTLs identificadas, en base al conocimiento existente sobre el efecto de los compuestos volátiles en nuestra percepción del sabor y el aroma, y en base a su capacidad para mantener su efecto tras su introgresión en el tomate cultivado, resultan ser candidatos prometedores para su utilización en la mejora genética del sabor del tomate.

Finalmente, el análisis de la localización en el genoma de las QTLs analizadas en las distintas poblaciones objeto de estudio en esta tesis, junto con las descritas en la bibliografía disponible, puso de relieve el bajo grado de co-localización existente entre las distintas QTLs, lo cual implica que en las especies silvestres relacionadas con el tomate existe un amplio rango de variabilidad genética susceptible de ser utilizado para la mejora de su sabor y su aroma.



## RESUM

Els fruits produeixen i emeten compostos químics volàtils. Estos són molècules en general de baixa polaritat i cadena curta que tenen diverses funcions, i són les responsables de la nostra percepció de l'aroma i de bona part del sabor dels fruits. Esta tesi se centra en l'estudi dels volàtils del fruit de la tomata, que és un dels cultius hortícoles més importants i un sistema model per a l'estudi del desenvolupament i la maduració del fruit.

Es van comparar de forma sistemàtica els mètodes analítics més habituals per a l'anàlisi de volàtils en fruits de tomata, i es va observar que el perfil de volàtils detectat està fortament condicionat per el mètode analític utilitzat, tant per el procés de preparació de la mostra com per la tècnica d'adquisició dels volàtils. Es va concloure que la tècnica més adequada per a la comparació de grans grups de mostres des d'una aproximació multiòmica consistix a congelar en nitrògen líquid el material vegetal en el momento idoni de recol.lecció, i analitzar-lo posteriorment per microextracció en fase sòlida (SPME) acoblada a cromatografia de gasos i espectrometria de masses.

Es va posar a punt esta tècnica i es va utilitzar per a la determinació dels compostos volàtils en diverses línies NIL portadores de QTL de caràcters relacionats amb el sabor i l'aroma en fons genètics diversos. Els resultats van permetre associar alguns dels caràcters organolèptics identificats en alteracions en els nivells d'alguns compostos volàtils. També es va observar que el fons genètic té un efecte important sobre la producció d'estos metabolits. Les anàlisis de correlació entre els nivells de volàtils i els metabolits primaris ens van permetre concloure que la producció de compostos volàtils, en general, no està determinada per els nivells dels seus precursors. La regulació dels volàtils és deguda a processos posteriors, com són la disponibilitat dels precursors o de

metabòlits intermediaris, la variabilitat en processos específics relacionats amb la conversió dels precursors en volàtils, o bé algun altre mecanisme regulador encara desconegut.

També es van estudiar els volàtils en una població de RIL derivada d'un creuament entre *Solanum pimpinellifolium* entrada TO-937, l'espècie silvestre més pròxima a la tomatera cultivada i que produeix fruits rojos, i *S. lycopersicum* cv. Moneymaker, una varietat de tomata per al mercat en fresc. Açò va permetre la identificació de 102 QTL de 39 volàtils diferents, 76 dels quals no s'havien descrit prèviament, i que es van mapar al llarg dels 12 cromosomes de la tomatera mitjançant el mapa de marcadors moleculars d'SNPs SOLCAP. Posteriorment es va avaluar la majoria dels QTLs identificades mitjançant la determinació dels volàtils en línies IL generades a partir dels mateixos materials. Es va observar que quasi la meitat d'estos QTL van mantindre el seu efecte quan van ser introgressats en el fons genètic Moneymaker. Addicionalment, 16 nous QTL es van identificar en esta població d'IL. Alguns dels QTL identificats, segons el coneixement que tenim respecte de l'efecte dels volàtils en la nostra percepció del sabor i l'aroma, i tenint en compte la seua capacitat de mantindre el seu efecte quan són introgressats en la tomata cultivada, són candidats prometedors per a ser utilitzats en la millora genètica del sabor de les tomates.

Finalment, l'anàlisi de la localització en el genoma dels QTL analitzats en les distintes poblacions objecte d'aquesta tesi, junt a les descrites en la bibliografia disponible, va evidenciar que hi ha una baixa freqüència de colocalització entre els distintos QTL. Açò implica que existix molta variabilitat genètica en les espècies silvestres relacionades en la tomatera, variabilitat que es pot utilitzar per a la millora del sabor i l'aroma dels seus fruits.





## Table of contents

### INTRODUCTION

<b>The expanded tomato fruit volatile landscape.....</b>	<b>1</b>
Abstract.....	3
The quest for identification of volatiles impacting flavour in tomato.....	5
Variability in volatiles and where to find it and how to harness it.....	10
Genes involved in volatile production in tomato.....	14
Transcriptional regulation of volatile pathways.....	26
Conjugation and volatile management.....	27
Challenges ahead.....	29
References.....	32

<b>OBJECTIVES.....</b>	<b>43</b>
------------------------	-----------

### CHAPTER 1

<b>Tomato fruit volatile profiles are highly dependent on sample processing and capturing methods.....</b>	<b>45</b>
Abstract.....	47
1 Introduction.....	49
2 Materials and methods.....	52
3 Results and discussion.....	57
3.1 <i>Effect of the capturing method on the volatile profile.....</i>	<i>57</i>
3.2 <i>Effect of sample processing on the volatile profile.....</i>	<i>65</i>
3.3 <i>Possible consequences and limitations for metabolomic studies derived from the variability in VOCs introduced by processing and capture methods.....</i>	<i>71</i>
4 Concluding remarks.....	76
References.....	77

## CHAPTER 2

<b>Metabolic characterisation of loci affecting sensory attributes in tomato allows an assessment of the influence of the levels of primary metabolites and volatile organic contents.....</b>	<b>85</b>
Abstract.....	87
Introduction.....	89
Material and methods.....	91
Results.....	96
<i>Elite tomato lines harbour clear metabolic differences.....</i>	<i>96</i>
<i>Analysis of metabolic variation in tomato lines pre-selected for their organoleptic properties.....</i>	<i>100</i>
<i>Variation in volatile organic compound content in tomato lines pre-selected for their organoleptic properties.....</i>	<i>106</i>
<i>Correlation analysis.....</i>	<i>112</i>
Discussion.....	118
Conclusion.....	122
References.....	124

## CHAPTER 3

<b>Identification, validation and introgression of QTL involved in tomato fruit volatile composition from red-fruited wild tomato species <i>Solanum pimpinellifolium</i> L.....</b>	<b>131</b>
Abstract.....	133
Introduction.....	135
Material and methods.....	137
Results and discussion.....	143
<i>The volatile network.....</i>	<i>143</i>
<i>Correlation between metabolites.....</i>	<i>147</i>
<i>Identification of volatile QTLs.....</i>	<i>151</i>
<i>Introgression and validation of the volatile QTLs in a fresh market tomato genetic background.....</i>	<i>158</i>

<i>Volatile QTLs identified in tomato show a low degree of overlapping....</i>	161
Conclusion.....	164
References.....	165
<b>GENERAL DISCUSSION.....</b>	<b>171</b>
<b>CONCLUSIONS.....</b>	<b>175</b>





# **INTRODUCTION**

## **The expanded tomato fruit volatile landscape**

José Luis Rambla, Yury M. Tikunov, Antonio J. Monforte, Arnaud G. Bovy,  
Antonio Granell

*Journal of Experimental Botany*, 65, 4613-4623 (2014)



## **ABSTRACT**

The present review aims to synthesize our present knowledge about the mechanisms implied in the biosynthesis of volatile compounds in the ripe tomato fruit, which have a key role in tomato flavour. The difficulties in identifying not only genes or genomic regions but also individual target compounds for plant breeding are addressed. Ample variability in the levels of almost any volatile compound exists, not only in the populations derived from interspecific crosses but also in heirloom varieties and even in commercial hybrids. Quantitative trait loci (QTLs) for all tomato aroma volatiles have been identified in collections derived from both intraspecific and interspecific crosses with different wild tomato species and they (1) fail to colocalize with structural genes in the volatile biosynthetic pathways and (2) reveal very little coincidence in the genomic regions characterized, indicating that there is ample opportunity to reinforce the levels of the volatiles of interest. Some of the identified genes may be useful as markers or as biotechnological tools to enhance tomato aroma. The current knowledge about the major volatile biosynthetic pathways in the fruit is summarized. Finally, and based on recent reports, it is stressed that conjugation to other metabolites such as sugars seems to play a key role in the modulation of volatile release, at least in some metabolic pathways.

### **Key words**

Aroma, flavour, *Solanum*, fruit, tomato, volatile organic compounds, QTLs, conjugation.



## **The quest for identification of volatiles impacting flavour in tomato**

As with many fruits that are part of the human diet, tomato has been domesticated for a few centuries to satisfy human preferences, but originally evolved to attract seed dispersers. The process of fruit ripening is complex and highly coordinated, and it begins when seeds are fully developed. The variety of physical and chemical changes that occur, such as softening of the fruit, the presence of high levels of organic acids or the conversion of starch into short-chain sugars, make the fruit more attractive to animals. Additionally, the synthesis and accumulation of carotenoids, particularly  $\beta$ -carotene and lycopene, produce a change in fruit colour which may act as a visual cue that the fruit is ripe.

Among the few hundreds of volatile compounds a ripe tomato fruit typically produces (Tikunov *et al.*, 2005) almost all the important volatile compounds related to flavour are derived from essential nutrients such as phenylalanine, leucine, isoleucine or linolenic acid, a fact which has suggested a process of co-evolution between tomato and its predators. Therefore, it has been proposed that volatile compounds produced in the ripe fruit would act as sensory cues for nutritional and health value (Goff and Klee, 2006; Klee and Giovannoni, 2011).

One mechanism for the rapid release of high amounts of selected volatiles in the tomato fruit when physically damaged, for example through chewing by a feeder, relies on their previous accumulation in a conjugated non-volatile form such as glycosides. The accumulation of the appropriate glycosidase in a separate subcellular location would allow the immediate liberation of high amounts of the aglycone when the enzyme and the conjugate glycosylated form would be in contact with each other. It has recently been identified in a subset of tomato varieties that a different pattern of glycosylation

is produced from the breaker stage onwards. Some volatile compounds such as guaiacol, eugenol or methyl salicylate are glycosylated to form a diglycoside during fruit development. Endogenous glycosidases have the ability to cleave this glycosidic bond, therefore liberating the volatile aglycone upon tissue disruption. In some tomato cultivars at the onset of ripening a recently identified glycosyltransferase adds a third sugar to the conjugate, preventing the digestion of the glycoside, which results in a sharp decrease in the release of these volatiles (Tikunov *et al.*, 2010; Tikunov *et al.*, 2013). This mechanism of volatile storage through glycosylation and wound-induced deconjugation and subsequent volatile release suggests that some volatiles may also be part of a protection strategy against predation prior to ripening, discouraging feeding on fruits with immature seeds, in a similar manner as has been reported for other metabolites with a bitter taste such as  $\alpha$ -tomatine, which levels are dramatically reduced upon ripening in most varieties (Rick *et al.*, 1994). According to this view, volatile compounds together with other non-volatile metabolites would have a double effect on seed dispersers: discouraging them to feed on the fruit before the maturation of the seed, and encouraging to feed on them once seed maturation has been achieved.

From a human perspective, a relevant question is which and how many of the volatile compounds produced in the fruit are responsible for our perception of flavour and aroma. There is a great variety of volatile compounds produced in the ripe fruit, and there are differences of many orders of magnitude between their abundance levels, with concentrations ranging from several micrograms per gram of fresh weight for the most abundant such as (*Z*)-3-hexenal or hexanal to the nanogram per gram and even lower levels of  $\beta$ -damascenone or  $\beta$ -ionone (Buttery *et al.*, 1988). A traditional approach for the understanding of which compounds are important in contributing to aroma and

the intensity and odour quality is the application of odour thresholds and odour units. Basically, this approach consists on establishing the lowest concentration of a compound that can be perceived by the human nose. These thresholds were estimated by means of the orthonasal smell perception of decreasing concentrations of volatile compounds in water solution by panels of judges (Guadagni *et al.*, 1963; Buttery *et al.*, 1971; Buttery *et al.*, 1989). According to this approach, the compounds contributing to aroma in the tomato fruit would be those with a concentration higher than the threshold established for that particular compound. The importance of each compound to the resulting aroma perception would be estimated by means of odour units. These are calculated by dividing the concentration of each compound in the tomato samples by its odour threshold, and are usually represented in their logarithmic form. By means of this approach, a list of 16-17 compounds was produced, with the compounds arranged in decreasing order of odour contribution (Buttery *et al.*, 1989).

This approach was widely accepted as a useful tool for a first attempt to identify the volatiles contributing to tomato aroma. However, it has been revealed to be too simplistic to explain the high complexity of our perception of flavour and aroma. First, odour thresholds were calculated by means of the orthonasal perception (sniffing) of volatile compounds, whilst our perception of food aroma is based on the retronasal perception of the volatile compounds released in the mouth. It has been observed that ortho- and retronasal odour thresholds for the same compound are different. Furthermore, particularly for food odours, it has been observed that each of these types of olfaction produces distinct sensory signals (Negoias *et al.*, 2008; Bender *et al.*, 2009). Additionally, odour thresholds were calculated based on the concentration of pure standards in water solution, not in a tomato matrix. It has been described

that there is an important matrix effect on the volatility of aroma compounds, therefore affecting their access to the olfactory receptors. As a consequence, the same amount of compound in a tomato matrix or in a water solution results in a difference of up to an order of magnitude in volatile emission (Bezman *et al.*, 2003).

Another fact which complicates further the making of a list of compounds contributing to aroma is the wide range of variation in their levels between cultivars, so that a particular compound may be at low levels in some cultivars and have no effect on the aroma, but in another variety with high levels could have an important effect. This has apparently been the case of guaiacol, which initially was not considered as participating in tomato aroma, but in some introgression lines derived from *S. lycopersicum* var. *cerasiforme* and some commercial varieties revealed to have a relevant effect as detected by consumer panels (Causse *et al.*, 2002; Zanol *et al.*, 2009; Tikunov *et al.*, 2013).

It has also been observed that the perception of aroma is not due to the additive effect of each individual volatile compound, but to the interaction of different volatile compounds affecting perception in different and sometimes even opposite directions. Furthermore, it has been observed that although taste and olfactory receptors are different and recognize different chemicals, there also exists an interaction in the perception between volatile and non-volatile compounds. So, it has been described that the presence of sugar or organic acids alters the taste panel perception of aromatic descriptors of samples with the same concentration of volatile compounds (Tandon *et al.*, 2003; Baldwin *et al.*, 2008) and, conversely, the perception of taste descriptors such as overall taste, sourness or sweetness can be modified by the addition or naturally occurring levels of some volatiles (Baldwin *et al.*, 2004; Vogel *et al.*, 2010, Tieman *et al.*, 2012).



Taking into consideration this complexity, efforts have been made to generate prediction models for the different descriptors of tomato flavour and consumer preference using regression analysis of both volatile and non-volatile compounds (Tandon *et al.*, 2003), later enhanced by partitioning taste from aromatic flavour notes (Abegaz *et al.*, 2004) and, more recently, by the integration of physicochemical, volatile and sensory parameters in multivariate modelling (Piombino *et al.*, 2013) or by means of targeted metabolomics in order to better understand the interactions between compounds leading to consumer liking (Tieman *et al.*, 2012). According to the latter model, it was revealed that some compounds traditionally considered to be important for tomato aroma based on their odour units, such as phenylacetaldehyde and particularly  $\beta$ -damascenone, would apparently have no contribution to cultivated tomato flavour preference (Tieman *et al.*, 2012).

An additional difficulty for the identification of targets for genetic improvement of tomato flavour is to define what consumers consider a good tomato. It has been concluded that flavour descriptors together with firmness seem to be the most important traits for improving tomato quality (Causse *et al.*, 2010), but a thing such as a perfect tomato which would be considered excellent by all the consumers does not exist. Quite the contrary, consumer preferences are segmented, as has repeatedly been observed in French (Lê and Ledauphin, 2006; Lengard and Kermit, 2006), Italian (Sinesio *et al.*, 2010) and other European consumer studies (Causse *et al.*, 2010). Therefore, diversification of at least flavour and texture in different market varieties would be necessary to satisfy the preferences of all the consumers.

## Variability in volatiles and where to find it and how to harness it

To identify the genetic basis for volatile production is important since a number of surveys reveal a general dissatisfaction of consumers and complaints about the poor organoleptic quality of most commercial tomatoes (Kader *et al.*, 1977; Janse and Schols, 1995). Somewhere along the modern breeding process the aroma of traditional tomatoes has been lost (Klee and Tieman, 2013) and there is an urge to get it back. Although in many cases this loss of organoleptic quality could be due to pre- and postharvest conditioning of the fruit, modern breeding has been focused mainly on biotic resistance, long shelf life and productivity rather than on organoleptic/aroma quality, which in addition is a very complex and difficult trait to breed for (Klee and Tieman, 2013). Important variability in the range of volatile levels, particularly high for some branched-chain and phenolic volatiles, has been found in several experiments (see Table 1 as an example). Moreover, this variability is found actually in heirlooms (Tieman *et al.*, 2012), wild relatives (Tikunov *et al.*, 2013), breeding populations (Causse *et al.*, 2002; Zanon *et al.*, 2009), but also in different commercial hybrids (Tikunov *et al.*, 2005; Ursem *et al.*, 2008). The assessment of such genetic variability opens the opportunity to improve the aroma of modern tomato varieties through breeding.

Flavour is determined by a complex interaction of aroma volatiles, sugars and organic acids. Therefore, flavour and, concomitantly, consumer perception show quantitative variation and are expected to be under complex genetic control. The first systematic attempt to analyse the genetic control of volatiles and aroma in tomato was carried out by Causse *et al.* (2002) in an intraspecific tomato mapping population allowing the identification of some major QTLs for a number of fruit volatiles. Further studies using interspecific populations of *S. habrochaites* and *S. pennellii* with the tomato inbred lines

**Table 1.** Range of volatile variation (fold difference) in fruit from different tomato genotype collections.

Co.1: Tomato breeding lines from company 1 (45 genotypes evaluated); Co.2: Tomato breeding lines from company 2 (22 genotypes); RILs a: Recombinant Introgression Lines originated from a cross *S. lycopersicum* cv. Moneymaker x *S. pimpinellifolium* accession TO-937 (Alba *et al.*, 2009), collected a first season (169 genotypes); RILs b: the same materials collected a different year (169 genotypes); Heirloom: *S. lycopersicum* heirloom varieties (Tieman *et al.*, 2012) (152 genotypes). N/a: Data not available. Amplitude of variation expressed as the fold change in the average values of each given volatile between the genotypes with highest and lowest levels in that population.

<b>Volatile compound</b>	<b>Co.1</b>	<b>Co.2</b>	<b>RILs a</b>	<b>RILs b</b>	<b>Heirloom</b>
3-methylbutanal	13	236	290	75	N/a
2-methylbutanal	36	N/a	14	78	13
3-methylbutanol	N/a	5742	344	2679	58
2-isobutylthiazole	361	185	444	242	174
1-penten-3-one	31	2	8	14	N/a
( <i>Z</i> )-3-hexenal	8	3	12	180	13
Hexanal	33	3	16	52	25
( <i>E</i> )-2-hexenal	11	5	10	16	123
( <i>E</i> )-2-heptenal	12	3	9	12	30
( <i>E,E</i> )-2,4-decadienal	152	3	40	55	211
Phenylacetaldehyde	43	112	30	106	654
Guaiacol	349	73	217	790	290
2-phenylethanol	591	41	118	90	3142
Methyl salicylate	246	273	244	184	3354
1-nitro-2-phenylethane	565	182	794	1920	149
Eugenol	12	36	829	1380	N/a
6-methyl-5-hepten-2-one	150	5	13	16	120
Geraniol	168	6	14	23	N/a
$\beta$ -damascenone	54	5	34	50	86
Geranylacetone	135	8	21	58	195
$\beta$ -ionone	44	4	12	44	47

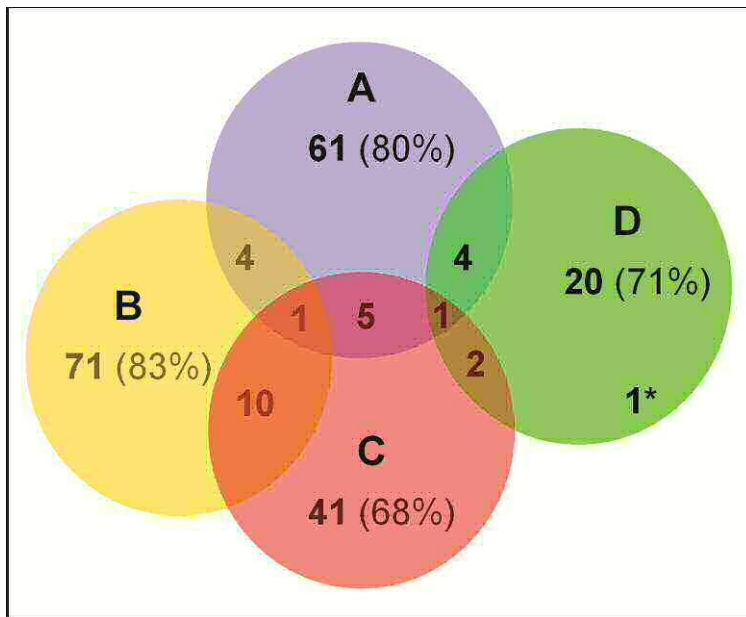
E6203 and M82, respectively, have enlarged the volatile variation range and allowed the identification of new volatile QTLs (Tieman *et al.*, 2006a; Mathieu *et al.*, 2009). Between 25 and 30 loci altered the volatile composition, and in most cases each locus altered several volatiles, most often metabolically related compounds. Interestingly, while all *S. pennellii* alleles increased volatile composition in the M82 background, *S. habrochaites* alleles increased or decreased them in the E6203 background depending on the locus. Although the identification of genes involved in the regulation of biosynthetic pathways of volatile compounds is still in its infancy, volatile QTLs do not co-localize with known structural genes encoding enzymes in any of the described volatile pathways. This makes cloning of these QTLs very attractive as they may underlie important regulatory genes.

Another interesting observation is that only in a few cases the same volatile QTL is conserved among the different mapping populations (Figure 1). This result can be attributed to multiple causes: (i) the volatile profile of the parent genotypes is quite different, indicating an important genetic variability among populations, (ii) fruit volatile composition is strongly influenced by the environment, (iii) differences in sampling, methods of volatile capture and profiling. A standardization of sampling and a large number of studies would be necessary to assess if the lack of co-localization among populations has a strong genetic basis.

Nevertheless, Zanor *et al.* (2009), studying introgression lines developed from a cherry donor into a large fruit tomato background by marker-assisted breeding, demonstrated that single QTL volatiles can be transferred between different genetic backgrounds and a single locus can be sufficient to alter significantly the volatile composition. Therefore, the identification of QTL volatiles in introgression lines could be used to both identify and select for

genomic regions carrying genes associated to accumulation and release of the corresponding volatiles.

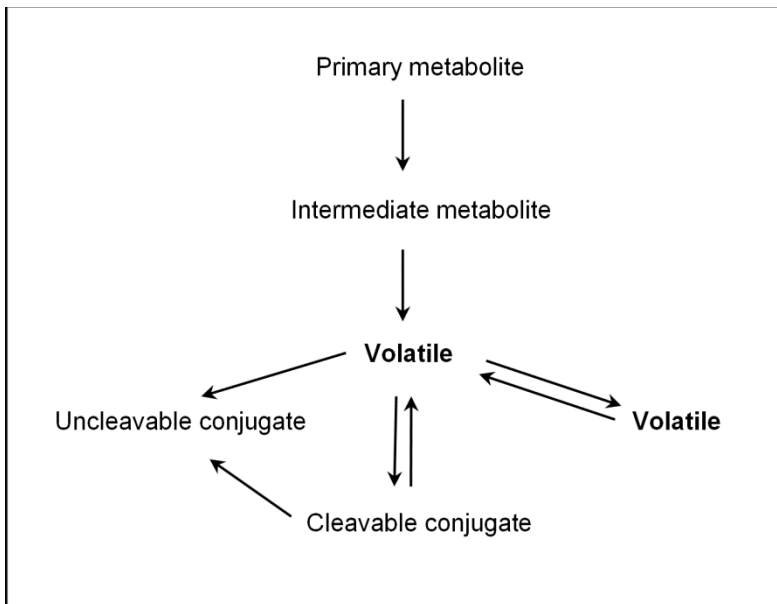
In summary, domesticated tomato still contains large variability for volatile accumulation both in heirloom and commercial varieties, and that variability can be increased by incorporating new loci from wild relatives. Therefore there is ample opportunity to improve volatile composition back into the commercial varieties so far optimized for yield and other traits.



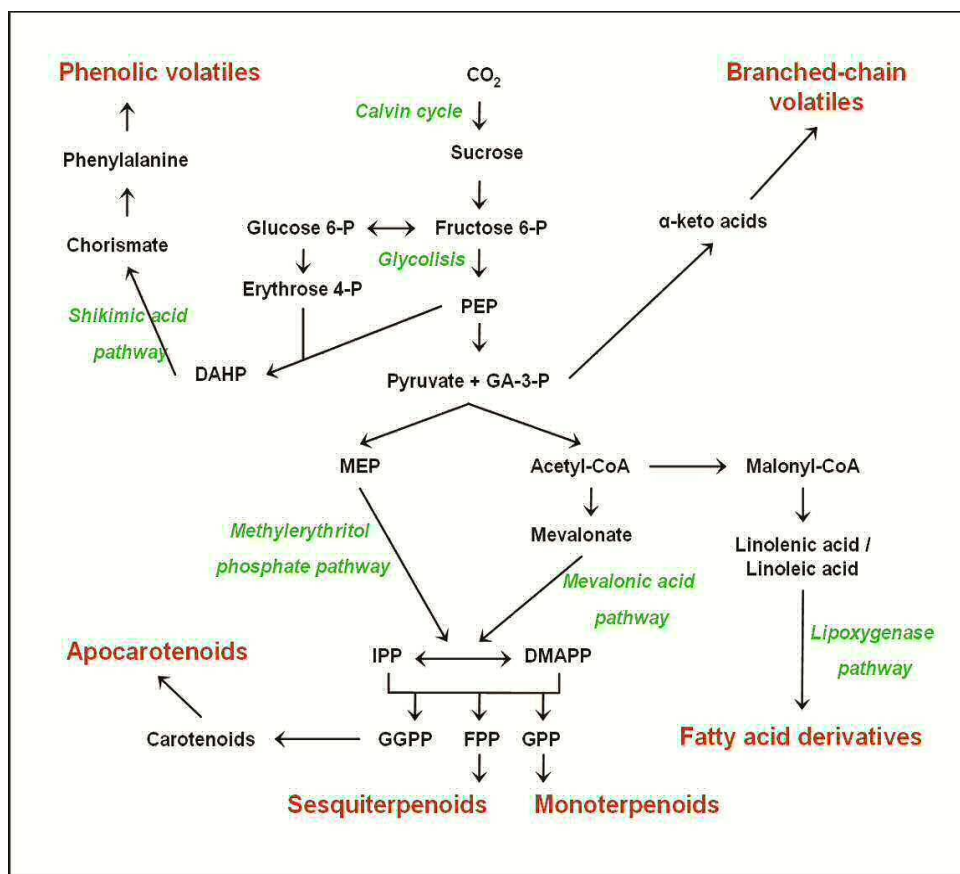
**Figure 1.** Venn diagram showing the degree of overlapping of the QTLs for volatiles in different introgression populations. **A**, RILs from an interspecific cross *S. lycopersicum* x *S. pimpinellifolium* (own data); **B**, ILs from an interspecific cross *S. lycopersicum* x *S. pennellii* (Tieman *et al.*, 2006a); **C**, ILs from an interspecific cross *S. lycopersicum* x *S. habrochaites* (Mathieu *et al.*, 2009); **D**, RILs from an intraspecific cross *S. lycopersicum* x *S. lycopersicum* var. *cerasiforme* (Causse *et al.*, 2002). Numbers in the overlapping areas represent the number of QTLs in common between the populations. Numbers in the non-overlapping areas represent the number of QTLs exclusive of that particular population. In this case data is also expressed as % of the total number of QTLs identified in that population. In D, 1\* refers to 1 QTL in common between populations B and D.

## Genes involved in volatile production in tomato

Volatile compounds are secondary or specialized metabolites which, once synthesized, may undergo different modifications, either reversible or irreversible, e.g. to produce a different volatile compound or a non-volatile conjugate (Figure 2). Although many advances have been achieved in the last decade, many of the genes involved in volatile biosynthesis remain unknown. Figure 3 shows a scheme of the most important metabolic pathways of volatile biosynthesis in tomato fruit, which we will describe later based on the current knowledge in tomato and also supported with the information obtained from other species.



**Figure 2.** General scheme of volatile biosynthesis and modifications.



**Figure 3.** Biosynthetic pathways of the most relevant classes of volatiles in the tomato fruit. Volatile classes are highlighted in brown; metabolic pathways are represented in green. Abbreviations correspond to: DAHP, 3-deoxy-D-arabino-heptulosonate 7-phosphate; DMAPP, Dimethylallyl diphosphate; FPP, Farnesyl diphosphate; GA-3-P, Glyceraldehyde-3-phosphate; GGPP, Geranyl diphosphate; IPP, Isopentenyl diphosphate; MEP, 2-C-methyl-D-erythritol 4-phosphate; PEP, Phosphoenolpyruvate.

### **Fatty acid derivatives**

Volatiles derived from fatty acids constitute a class of compounds which includes the most abundant volatiles produced in the tomato fruit: the C<sub>6</sub> volatiles 1-hexanol, (*Z*)-3-hexenal, (*E*)-2-hexenal or hexenal, and the C<sub>5</sub> volatile

1-penten-3-one. These compounds are classified as green leaf volatiles due to their characteristic 'green', fresh aroma of cut grass, since high amounts of lipid-derived C<sub>6</sub> aldehydes and alcohols are typically released from vegetative tissues when disrupted. In tomato fruit, the production of these compounds is increased at ripening, probably due to the loss of integrity of cellular membranes (Klee, 2010). Despite their abundance in the ripe fruit, their relevance for tomato flavour has been a matter of discussion. Although (*Z*)-3-hexenal and hexanal were originally considered among the most relevant compounds for tomato aroma in studies based on the odour units approach (Buttery *et al.*, 1989), recent studies suggest a reduced impact on tomato flavour and no effect on consumer liking (Chen *et al.*, 2004; Tieman *et al.*, 2012).

The initial step in the biosynthesis of these compounds is still not completely understood. The amount of free fatty acids available in the fruit is very limited, as plants accumulate them as acylglycerides rather than in the toxic free form. Therefore, it is supposed that the catabolism of the acylglycerides by a lipase (or lyolytic acyl hydrolase, LAH), which would liberate the fatty acids, is the initial step in their biosynthesis. This has been observed in *Arabidopsis* leaves, where the production of (*Z*)-3-hexenal was associated with a decrease in the levels of galactolipids, a process which could be repressed by means of a lipase inhibitor (Matsui *et al.*, 2000a).

Free fatty acids are rapidly catabolyzed by means of  $\beta$ -oxidation,  $\alpha$ -oxidation or the lipoxygenase pathway. The latter is the most important for volatile production in the tomato fruit, and includes the sequential activity of lipoxygenase (LOX) and hydroperoxide lyase (HPL) enzymes. Lipoxygenases are non-heme iron-containing fatty acid dioxygenases with the ability to catalyze the regio- and stereospecific dioxygenation of polyunsaturated fatty acids with a (1*Z*,4*Z*)-pentadiene moiety, converting them into fatty acid hydroperoxides



(Liavonchanka and Feussner, 2006). The most important substrates for lipoxygenase activity in tomato fruit are the C<sub>18</sub> fatty acids linolenic acid and, to a lesser extent, linoleic acid.

Lipoxygenases can be divided in two groups: 13-LOX and 9-LOX, depending on the positional specificity of oxygenation, producing 13- or 9-hydroperoxides, respectively. The resulting hydroperoxides are further metabolized by hydroperoxide lyases (HPLs), enzymes of the cytochrome P450 family which produce a volatile aldehyde and an oxoacid. These enzymes are also classified as 13- or 9-HPLs depending on the substrate they act on (Matsui, 2006). In the tomato fruit, there is an important 13-LOX activity, producing (*Z*)-3-hexenal from linolenic acid and hexanal from linoleic acid, particularly when fruits are homogenized. Five isoforms of 13-lipoxygenases have been described in tomato, but apparently only *TomloxC* is expressed in the fruit (Chen *et al.*, 2004), similarly as described in other fruits such as kiwi fruit, where the *AdLox* family is composed by six isoforms, only two of which are responsible for volatile emission in the ripe fruit (Zhang *et al.*, 2006). A 13-fatty acid hydroperoxide lyase has been described in tomato (Howe *et al.*, 2000). Although genes encoding enzymes with 9-LOX activity have been described in the fruits of other species such as cucumber (Matsui *et al.*, 2000b), almond (Mita *et al.*, 2005) or rice (Kuroda *et al.*, 2005), neither *9-LOX* nor *9-HPL* gene expression has been described in tomato fruit, which is in accordance with the low levels of C<sub>9</sub> volatile compounds detected.

Isomeric conversion of (*Z*)-3-hexenal into (*E*)-2-hexenal occurs in the fruit, either non-enzymatically or by means of a 3*Z*,2*E*-enal isomerase (Noordermeer *et al.*, 1999), although at present this enzyme has not been identified. The aldehydes produced from this lipoxygenase pathway, as those produced by other metabolic pathways, can be reduced to alcohols by means of

alcohol dehydrogenases (ADHs), enzymes catalyzing their reversible interconversion. Tomato *ADH2* gene expression was observed to increase during the ripening process, particularly in the last stages and to have an effect on the biosynthesis of hexanol and (*Z*)-3-hexenol (Speirs *et al.*, 1998). Another fruit-ripening-associated alcohol dehydrogenase, SlscADH1, has been described recently in tomato. This enzyme showed *in vitro* activity in the production of hexanol and 1-phenylethanol from hexanal and phenylacetaldehyde respectively, but no *in vivo* effect was observed (Moummou *et al.*, 2012).

Biosynthesis of C<sub>5</sub> lipid volatiles, such as 1-penten-3-one, which is considered as an important contributor to tomato fruit aroma (Baldwin *et al.*, 2000) has not been investigated so far, but LOX could use linolenic acid as a substrate producing 13-oxo radical, which is converted non-enzymatically into 1,3-pentene radical, which could further react with hydroxyl radical yielding C<sub>5</sub> alcohols (Gardner *et al.*, 1996). The activity of this LOX branch could be boosted by a reduction of HPL activity leading to accumulation of hydroperoxides and therefore could be considered as competing with the C<sub>6</sub>-volatile producing LOX pathway (Vancanneyt *et al.*, 2001).

### ***Amino acid derivatives***

A significant number of the volatile compounds considered as relevant for tomato aroma are derived from amino acids. These volatiles can be grouped in two categories: phenolic and branched-chain compounds. Their respective biosynthetic pathways are separately described below.

## Phenolic volatiles

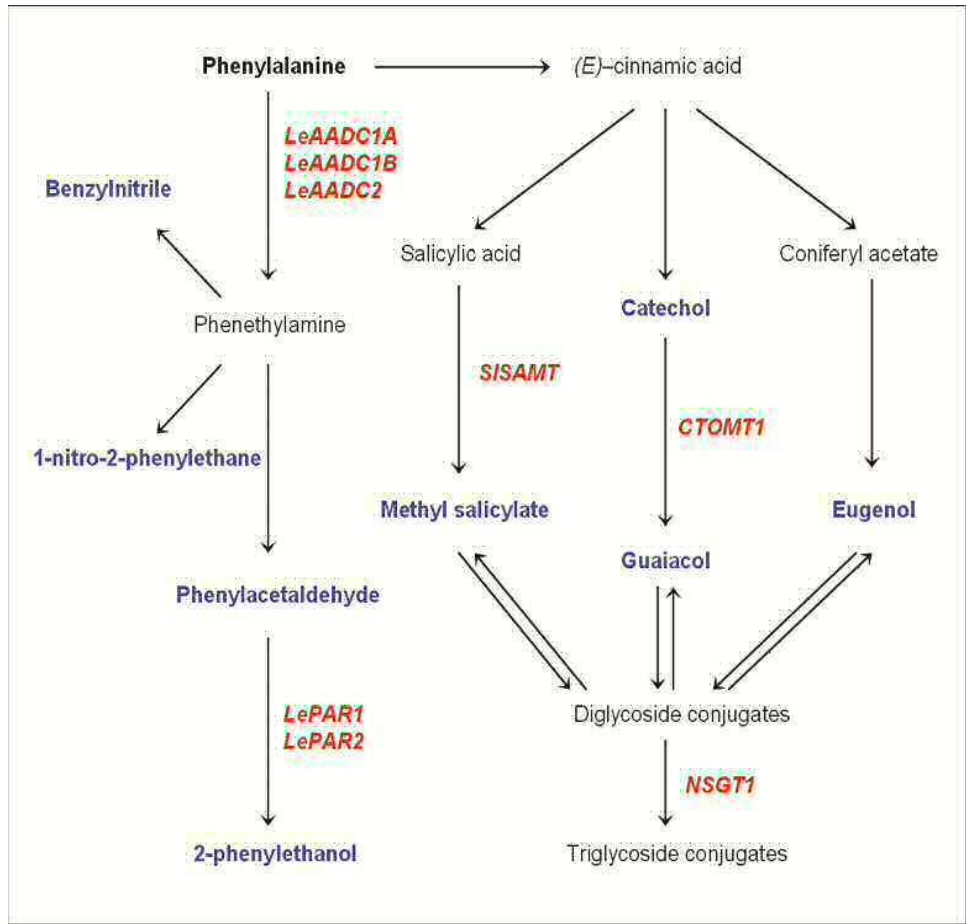
Phenolic volatiles include many compounds that are involved, either positively or negatively, in our perception of tomato flavour and include a variety of compounds derived from the amino acid phenylalanine. In a recent study, transgenic tomatoes with enhanced levels of the phenolic volatiles 2-phenylethanol, phenylacetaldehyde and benzaldehyde had a preferred floral aroma compared to untransformed controls, although they also had diminished levels of some positive aroma apocarotenoids such as  $\beta$ -ionone or geranylacetone (Tzin *et al.*, 2013). 2-phenylethanol had been previously described to have a positive effect on tomato flavour, increasing floral aroma and the perception of sweetness (Baldwin *et al.*, 2008). Nevertheless, introgression lines harbouring the *malodorous* allele from the wild tomato species *S. pennellii*, which produces dramatically increased levels of 2-phenylethanol and its precursor phenylacetaldehyde, showed a strong undesirable flavour, probably due to the very high levels of phenylacetaldehyde produced (Tadmor *et al.*, 2002). This exemplifies the complexity of our perception of flavour based on volatile compounds and the difficulty to predict the effect on flavour and consumer preference when considering to alter a metabolic pathway.

Phenylalanine-derived compounds can be classified in different subfamilies. C<sub>6</sub>-C<sub>2</sub> phenolic volatiles are probably the most important compounds for aroma, and their biosynthesis implies an initial decarboxylation of phenylalanine. A small family of genes (*LeAADC1A*, *LeAADC1B*, *LeAADC2*) has been described in tomato fruit leading to the decarboxylation of phenylalanine into phenethylamine, which would then be de-aminated by means of a yet uncharacterized amine oxidase to produce phenylacetaldehyde. Alternatively, it could be transformed into 1-nitro-2-phenylethane or phenylacetonitrile by

means of other unknown enzymes (Tieman *et al.*, 2006b). 2-phenylethanol, considered to be an important volatile for fruit aroma in many species and also in tomato, is synthesized from phenylacetaldehyde by means of phenylacetaldehyde reductases LePAR1 and LePAR2. These enzymes catalyze the unidirectional reduction of aldehyde into alcohol, and it is supposed that they also use benzaldehyde and cinnamaldehyde as substrates for the synthesis of their respective alcohols (Tieman *et al.*, 2007).

The other group of phenolic compounds originates from the phenylpropanoid branch of Phe catabolism. Biosynthetic pathways of phenylpropanoid compounds have not been completely elucidated in tomato. It is assumed that C<sub>6</sub>-C<sub>3</sub> volatile synthesis would be initiated by means of a phenylalanine ammonia-lyase (PAL) producing (*E*)-cinnamic acid and would go along the pathway of lignin biosynthesis. Some of the compounds in this pathway would be substrates for enzymes producing volatile compounds such as eugenol, as has been reported in other species to be synthesized by means of a eugenol synthase from coniferyl acetate (Koeduka *et al.*, 2006).

Shorter-chain phenylpropanoids also originate from (*E*)-cinnamic acid by the shortening of their side chain and further modifications. The last steps of the biosynthesis of some of these compounds have been recently described in tomato. Methyl salicylate is produced by means of Salicylic Acid Methyl Transferase (SISAMT), an *O*-methyltransferase catalyzing the methylation of salicylic acid (Tieman *et al.*, 2010). The synthesis of guaiacol, another important volatile compound for fruit flavour, would be produced from catechol by means of the catechol-*O*-methyltransferase COMT1 (Mageroy *et al.*, 2012). The biosynthetic pathway of the most important compounds and the identified tomato genes involved in volatile biosynthesis are summarized in Figure 4.



**Figure 4.** Biosynthetic pathway of the phenolic volatile compounds most relevant in tomato fruit. Volatile compounds detected in the fruit are represented in blue; characterized tomato genes involved in the biosynthesis are represented in red. Gene abbreviations are as follows: *CTOMT1*, Catechol-*O*-methyltransferase 1; *LeAADC*, Aromatic Amino Acid Decarboxylases; *LePAR*, Phenylacetaldehyde Reductases; *NSGT1*, Non-smoky Glycosyltransferase 1; *SISAMT*, Salicylic Acid Methyl Transferase.

## Branched-chain volatiles

Another important group of volatiles related to amino acids include branched-chain volatiles, a set of compounds with particularly low molecular weight and high volatility, some of which are considered to participate in tomato aroma, such as 3- and 2-methylbutanal, 3-methylbutanol and 2-isobutylthiazole (Buttery *et al.*, 1989).

Biosynthesis of these compounds has not been elucidated yet in fruits, although their biosynthetic pathway has been described in yeast and bacteria. In these microorganisms, branched-chain amino acids would be the original precursors and would reversibly be converted into  $\alpha$ -ketoacids by means of branched-chain amino acid aminotransferases (BCATs). A set of different volatile compounds can then be formed: (i) an  $\alpha$ -hydroxyacid would be formed by the action of an  $\alpha$ -hydroxyacid dehydrogenase; (ii) an aldehyde through the action of a decarboxylase; (iii) the latter could subsequently be reduced to an alcohol by alcohol dehydrogenases; (iv) an acyl-CoA by means of an  $\alpha$ -ketoacid dehydrogenase; (v) this acyl-CoA can be converted into an acid or else react with an alcohol to form an ester by the action of an alcohol acyltransferase (Marilley and Casey, 2004).

A biosynthetic pathway in fruits similar to that described in yeast and bacteria seemed reasonable after the identification in tomato of a small family of branched-chain amino acid aminotransferases. This family is composed of six members, of which the mitochondria-located SIBCAT1 and SIBCAT2 would be implied in the first step of amino acid catabolism (Maloney *et al.*, 2010). In this case, the amino acids leucine, isoleucine and valine would be the precursors of branched-chain volatile compounds. Nevertheless, it was later revealed that the catabolism of amino acids by means of the BCAT enzymes is unrelated to the

synthesis of volatiles. Therefore, it has been proposed that  $\alpha$ -ketoacids rather than amino acids would be the direct precursors of this family of volatile compounds (Kochevenko *et al.*, 2012).

### ***Esters***

Although very abundant and extremely important for the aroma of fruit in many species such as strawberry (Zorrilla-Fontanesi *et al.*, 2012), peach (Sánchez *et al.*, 2012) or even some citrus species (González-Mas *et al.*, 2011), few esters are found in the volatile fraction of tomato and, with the exception of the previously described phenylpropanoid ester methyl salicylate, they are not relevant for tomato flavour.

On the contrary, new evidence has revealed that this lack of esters in the cultivated species has a positive effect on tomato liking. It has been observed that green-fruited wild tomato species accumulate considerably high levels of acetate esters compared to red-fruited species. The difference is attributed to the insertion of a retrotransposon in a position adjacent to the most enzymatically active tomato esterase, increasing gene expression in all red-fruited species including cultivated tomato. The resulting enhanced esterase activity results in a dramatic reduction in the levels of many esters that are negatively correlated with human preference, which may have provided an adaptive advantage to the ancestor of red-fruited species, such as cultivated tomato (Goulet *et al.*, 2012).

## **Terpenoids**

### Mono- and sesquiterpenoids

This class of volatiles includes an ample variety of structurally complex compounds which are among the most abundant in tomato vegetative tissues and particularly in trichomes, but only a few of them such as limonene, linalool or  $\alpha$ -terpineol are present in the ripe fruit and their impact on tomato aroma is negligible.

Volatile terpenoids can be classified in two groups: monoterpenoids ( $C_{10}$ ) and sesquiterpenoids ( $C_{15}$ ). They are both synthesized from the five carbon precursors isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP). There are two alternate pathways for the biosynthesis of these precursors. The methylerythritol phosphate pathway has been described in the plastids to produce both IPP and DMAPP from pyruvate and glyceraldehyde-3-phosphate. The mevalonic acid pathway has been described in the cytosol to use acetyl-CoA to produce IPP, which can later be converted into DMAPP. Although the different subcellular compartmentation of each of these pathways, some metabolic crosstalk between them has been reported, particularly in the direction from the plastids to the cytosol (Hemmerlin *et al.*, 2003).

Geranyl diphosphate synthase catalyzes the condensation of an IPP and a DMAPP molecule to produce geranyl diphosphate (GPP), the precursor of all monoterpenoids, while farnesyl diphosphate synthase catalyzes the synthesis of farnesyl diphosphate (FPP), the precursor of all sesquiterpenoids, from two IPP and one DMAPP molecules. GPP and FPP are the substrates for the diverse terpene synthases/cyclases, a large family of enzymes, to produce a variety of



monoterpenoids and sesquiterpenoids respectively (Nagegowda, 2010; Granell and Rambla, 2013).

### Carotenoid derived volatiles

Apocarotenoids can be considered as irregular terpenoids, and are synthesized from the oxidative cleavage of double bonds in carotenoids (C<sub>40</sub> terpenoids), compounds which are accumulated at high levels in the ripe fruit. These volatile compounds are produced at low levels in the ripe fruit, but are important in our perception of tomato flavour due to their very low odour thresholds, particularly for some cyclic apocarotenoids such as C<sub>13</sub> ketones  $\beta$ -ionone or  $\beta$ -damascenone, which can be detected orthonasally at concentrations of 0.007 nL/L and 0.002 nL/L respectively (Buttery *et al.*, 1989). Although recent studies have questioned the relevance of individual compounds previously considered important for the flavour of tomato, such as  $\beta$ -damascenone, carotenoid-derived volatiles have proved to have an important role in tomato flavour, as their levels positively correlate with tomato flavour acceptability (Vogel *et al.*, 2010).

In tomato, carotenoid cleavage dioxygenases *LeCCD1A* and *LeCCD1B* have been described as involved in the biosynthesis of at least some of the apocarotenoids produced in the fruit. These genes are highly expressed in the ripening fruit and have proved to cleave multiple carotenoids, both linear and cyclic, producing a C<sub>14</sub> dialdehyde and a variety of C<sub>13</sub> volatiles such as  $\beta$ -ionone, geranylacetone and pseudo-ionone (Simkin *et al.*, 2004).

Proteins of the CCD1 group have been described in other species to have the ability to cleave cyclic carotenoids at the 9,10 position and linear carotenoids at the 5,6 (5',6') or 7,8 (7',8') or 9,10 (9',10') positions, producing

many different compounds. These enzymes are located in the cytosol and show broad substrate specificity, cleaving any carotenoid after  $\zeta$ -carotene in the metabolic pathway. Considering that carotenoids are accumulated in the plastids, it is still unclear how enzymes and substrates come together, although different mechanisms have been proposed (Vogel *et al.*, 2008; Ilg *et al.*, 2009; Floss and Walter, 2009; Walter *et al.*, 2010).

Apocarotenoid levels in the fruit increase dramatically during ripening, although there is significant CCD expression in the fruit during all the stages of fruit development. The coincidence of the conversion of chloroplasts into chromoplasts and the loss of membrane integrity with the increased biosynthesis of apocarotenoids suggests a key role of substrate availability in the regulation of their biosynthesis (Klee, 2010; Vogel *et al.*, 2010).

Similar processes lead to production of another class of carotenoid derived compounds – open-chain carotenoid derived volatiles. Eight-carbon ketone 6-methyl-5-hepten-2-one and C<sub>10</sub> aldehyde  $\alpha$ -citral (geranial) are the most abundant compounds of this class in tomato fruit and contribute to its aroma (Buttery *et al.*, 1989, Baldwin *et al.*, 2000). These two volatiles are derived from open-chain carotenoids – phytoene or phytofluene and lycopene, respectively, and the volatile products correlate strongly to the levels of the carotenoid precursors (Lewinsohn *et al.*, 2005).

### **Transcriptional regulation of volatile pathways**

The production of volatile compounds in the fruit is the result of many interconnected metabolic pathways and a complex regulation network. The ripening of the fruit, which is a highly coordinated process, includes a dramatic change in its volatile profile (Ortiz-Serrano and Gil, 2010), in which

transcriptional regulation seems to be an important aspect. Nevertheless, very little is known about the transcription factors which directly regulate volatile biosynthesis, with few possible exceptions such as the gene encoding the MYB transcription factor SLOD01 (Orzáez *et al.*, 2009).

The levels of most of the volatile compounds are increased by several orders of magnitude during the ripening process, peaking at or shortly before full ripening (Klee and Giovannoni, 2011), whilst a few of them remain constant or are reduced. Therefore, transcription factors which are involved in the regulation of fruit ripening, such as *rin*, *cnr*, *nor* or *nr* have been shown to have pleiotropic effects on biosynthesis and accumulation of aroma related volatile metabolites (Kovács *et al.*, 2009).

### **Conjugation and volatile management**

An effective mechanism to immobilize a volatile for future use once it has been synthesized is by covalent chemical binding to a polar compound, thus producing a non-volatile molecule of higher molecular mass and increased polarity. Conjugation of volatile compounds has been described for a long time in fruits of many species (Marlatt *et al.*, 1992). It usually involves *O*-glycosilation of the volatile compounds (also called aglycones) with one or more sugar moieties. The glycosyltransferase enzyme family is one of the most diverse enzyme families in plant. They lead to the production of a large variation of glycoconjugates with different structures and different biochemical properties. Such conjugation has been reported for different classes of compounds such as linear alcohols, monoterpene alcohols, apocarotenoids and phenylpropanoids (Buttery *et al.*, 1990; Ortiz-Serrano and Gil, 2007; Ortiz-Serrano and Gil, 2010) and the pattern of glycosidically bound volatiles in a particular fruit tends to be

similar to that of the free volatiles produced (Du *et al.*, 2010). These studies show that many volatiles in tomato fruit, e.g. some terpenes, are exclusively present as glycoconjugates. Glycosides of other volatiles can be equally abundant as their corresponding free forms or exceed their concentration. This suggests that conjugated volatiles could have a significant impact on tomato fruit aroma upon release from their glycoconjugates. Potentially, a reversible conjugation would allow the fruit to accumulate significant amounts of a volatile compound, which would otherwise be slowly released and consequently lost, and its deconjugation would enable the massive liberation of high amounts of that volatile when required. On the other hand, an irreversible conjugation could be an effective way to get rid of a compound that is no longer convenient to be released. In any case, the metabolic processes leading to the formation of conjugates and their possible hydrolysis are still poorly understood. In part as a consequence of this, the relevance of conjugation of volatile compounds in the flavour of fruits has probably been underestimated.

A couple of recent papers have shed some light to part of these processes in tomato fruit. It was observed that most of the emission of the phenylpropanoids methyl salicylate, eugenol and guaiacol, some of which are abundant compounds in tomato fruit and have an effect on flavour, relies on their liberation upon tissue disruption from their corresponding accumulated glycoconjugates. Two different patterns of glycoconjugation of these volatiles were observed in a collection of tomato cultivars which were tightly correlated with the emission of the aglycone: fruits accumulating phenylpropanoid volatile diglycosides produced high levels of the volatiles after homogenization, whilst those accumulating triglycosides emitted significantly reduced levels (Tikunov *et al.*, 2010). Subsequent work led to the identification of *NSGT1*, a fruit ripening-induced gene encoding a glycosyltransferase with the ability to transfer an

additional glucose to a set of phenylpropanoid volatile diglycosides. The triglycosides produced cannot be cleaved by tomato glycosidases, whilst diglycosides can be readily hydrolyzed upon fruit disruption. Consequently, NSGT1 activity produces the irreversible immobilization of these volatile phenylpropanoids from the onset of ripening, thereby preventing their emission upon tissue disruption, e.g. through the chewing of the fruit by a predator - or a human being (Tikunov *et al.*, 2013).

Much work has yet to be done in this aspect of volatile biosynthesis, but it seems that conjugation plays an important role in the control of volatile emission in the tomato fruit.

### **Challenges ahead**

Tomato flavour is a very complex trait in which volatile compounds play a key role, but little – if any – attention has been paid to them in plant breeding in the past, due to its high degree of complexity. Our lack of understanding not only concerns the very limited information about the genetic control of volatile levels, but even the definition of which compounds should be selected as targets for breeding.

Much progress has been made in the last years in the identification of genomic regions, genes and enzymes involved in the biosynthesis of volatile compounds in the tomato fruit. As a result, we have been able to initiate the metabolic engineering of volatiles, but still most of the intricate interrelated biological processes leading to volatile emission remain unclear.

The development of the different -omics technologies has provided us a substantial amount of information about biological processes like the ripening

of the fruit. The integration of the different data obtained from them certainly increases the complexity of the dataset, but provides a deeper comprehension and a more complete vision of the biological process studied. Such integrated -omics approaches, which have already been successfully used with non-volatile metabolites (Carrari *et al.* 2006), are expected to provide a better understanding of the high complexity of the whole ripening process, and are also a powerful tool for the identification of new genes responsible for fruit volatile production. Co-localization of volatile QTLs with gene expression QTLs, and the use of modeling networks to integrate the data and predict the most interesting candidates is also a promising approach. In this respect, the availability of the annotated genome sequence of tomato constitutes a highly valuable tool in facilitating the identification of the genes responsible for fruit volatile production in these species.

Once target genes have been selected, tools like the visually reporter system of virus-induced gene silencing in tomato fruit (Orzáez *et al.*, 2009) are useful for a first evaluation of candidate gene function during fruit ripening and allows the selection of most promising candidates from a larger panel. Despite the power of this platform, stable transgenic plants are required to unequivocally confirm the function of a selected candidate gene. In conclusion, we would like to remark that, although until recently not much attention has been paid to the importance of volatiles in tomato fruit flavour, we are quickly increasing our understanding of the metabolic pathways, key enzymes and genes leading to flavour volatile production. This knowledge provides us the ability to start modifying the fruit aroma through both transgenic as well as marker-assisted breeding approaches. Other relevant aspects which require attention are those regarding subcellular compartmentation of substrates, enzymes and products, and the regulatory networks controlling volatile

synthesis, emission and conjugation as a developmentally regulated process coupled to fruit ripening.

The development of new technologies provides the opportunity to obtain important advances in our understanding of the whole metabolic process, which would empower breeders to modify on purpose the fruit aroma in the near future.

### **Acknowledgements**

We want to thank the Metabolomics facility at the IBMCP for technical assistance. AG was supported by grants from MinECO and FECYT. This work was facilitated by the European-funded COST action FA1106 QualityFruit.

## REFERENCES

- Abegaz EG, Tandon KS, Scott JW, Baldwin EA, Shewfelt RL.** 2004. Partitioning taste from aromatic flavour notes of fresh tomato (*Lycopersicon esculentum* Mill.) to develop predictive models as a function of volatile and non-volatile components. *Postharvest Biology and Technology* **34**, 227-235.
- Alba JM, Montserrat M, Fernández-Muñoz R.** 2009. Resistance to the two-spotted spider mite (*Tetranychus urticae*) by acylsucroses of wild tomato (*Solanum pimpinellifolium*) trichomes studied in a recombinant inbred line population. *Experimental and Applied Acarology* **47**, 35-47.
- Baldwin EA, Scott JW, Shewmaker CK, Schuch W.** 2000. Flavor trivia and tomato aroma: Biochemistry and possible mechanisms for control of important aroma components. *Hortscience* **35**, 1013-1022.
- Baldwin EA, Goodner K, Plotto A, Einstein M.** 2004. Effect of Volatiles and their Concentration on Perception of Tomato Descriptors. *Journal of Food Science* **69**, S310-S318.
- Baldwin EA, Goodner K, Plotto A.** 2008. Interactions of Volatiles, Sugars, and Acids on Perception of Tomato Aroma and Flavor Descriptors. *Journal of Food Science* **73**, S294-S307.
- Bender G, Hummel T, Negoias S, Small DM.** 2009. Separate Signals for Orthonasal vs. Retronasal Perception of Food but Not Nonfood Odors. *Behavioral Neuroscience* **123**, 481-489.
- Bezman Y, Mayer F, Takeoka GR, Buttery RG, Ben-Oliel G, Rabinowitch HD, Naim M.** 2003. Differential Effects of Tomato (*Lycopersicon esculentum* Mill) Matrix on the Volatility of Important Aroma Compounds. *Journal of Agricultural and Food Chemistry* **51**, 722-726.
- Buttery RG, Seifert RM, Guadagni DG, Ling LC.** 1971. Characterization of Additional Volatile Components of Tomato. *Journal of Agricultural and Food Chemistry* **19**, 524-529.



- Buttery RG, Teranishi R, Ling LC, Flath RA, Stern DJ.** 1988. Quantitative Studies on Origins of Fresh Tomato Aroma Volatiles. *Journal of Agricultural and Food Chemistry* **36**, 1247-1250.
- Buttery RG, Teranishi R, Flath RA, Ling LC.** 1989. Fresh Tomato Volatiles: Composition and Sensory Studies. In: Buttery RG, Shahidi F, Teranishi R, eds. *Flavor Chemistry: New Trends and Developments*. ACS Symposium series 388, American Chemical Society, Washington DC, 213-222.
- Buttery RG, Takeoka G, Teranishi R, Ling LC.** 1990. Tomato aroma components: Identification of glycoside hydrolysis volatiles. *Journal of Agricultural and Food Chemistry* **38**, 2050-2053.
- Carrari F, Baxter C, Usadel B, et al.** 2006. Integrated Analysis of the Metabolite and Transcript Levels Reveals the Metabolic Shifts that Underlie Tomato Fruit Development and Highlight Regulatory Aspects of Metabolic Network Behaviour. *Plant Physiology* **142**, 1380-1396.
- Causse M, Saliba-Colombani V, Lecomte L, Duffe P, Rousselle P, Buret M.** 2002. QTL analysis of fruit quality in fresh market tomato: a few chromosome regions control the variation of sensory and instrumental traits. *Journal of Experimental Botany* **53**, 2089-2098.
- Causse M, Friguet C, Coiret C, Lépiciér M, Navez B, Lee M, Holthuysen N, Sinesio F, Moneta E, Grandillo S.** 2010. Consumer Preferences for Fresh Tomato at the European Scale: A Common Segmentation on Taste and Firmness. *Journal of Food Science* **75**, S531-S541.
- Chen GP, Hackett R, Walker D, Taylor A, Lin Z, Grierson D.** 2004. Identification of a specific isoform of tomato lipoxygenase (TomloxC) involved in the generation of fatty acid-derived flavor compounds. *Plant Physiology* **136**, 2641-2651.
- Du X, Finn CE, Qian MC.** 2010. Bound Volatile Precursors in Genotypes in the Pedigree of 'Marion' Blackberry (*Rubus* Sp.). *Journal of Agricultural and Food Chemistry* **58**, 3694-3699.

- Floss DS, Walter MH.** 2009. Role of carotenoid cleavage dioxygenase 1 (CCD1) in apocarotenoid biogenesis revisited. *Plant signalling and behaviour* **4**, 172-175.
- Gardner HW, Grove MJ, Salch YP.** 1996. Enzymatic pathway of ethyl vinyl 2-pentanal in soybean preparations. *Journal of Agricultural and Food Chemistry* **44**, 882-886.
- Goff SA, Klee HJ.** 2006. Plant Volatile Compounds: Sensory Cues for Health and Nutritional value? *Science* **311**, 815-819.
- González-Mas MC, Rambla JL, Alamar MC, Gutiérrez A, Granell A.** 2011. Comparative analysis of the volatile fraction of fruit juice from different Citrus species. *PLoS ONE* **6**, e22016. doi:10.1371/journal.pone.0022016.
- Goulet C, Mageroy MH, Lam NB, Floystad A, Tieman DM, Klee HJ.** 2012. Role of an esterase in flavour volatile variation within the tomato clade. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 19009-19014.
- Granell A, Rambla JL.** 2013. Biosynthesis of volatile compounds. In: Seymour G, Tucker GA, Poole M, Giovannoni JJ, eds. *The Molecular Biology and Biochemistry of Fruit Ripening*. Wiley-Blackwell, 135-161.
- Guadagni DG, Buttery RG, Okano S.** 1963. Odour thresholds of some organic compounds associated with food flavours. *Journal of the Science of Food and Agriculture* **14**, 761.
- Hemmerlin A, Hoeffler JF, Meyer O, Tritsch D, Kagan IA, Grosdemange-Billiard C, Rohmer M, Bach TJ.** 2003. Cross-talk between the cytosolic mevalonate and the plastidial methylerythritol phosphate pathways in Tobacco Bright Yellow-2 cells. *Journal of Biological Chemistry* **278**, 26666-26676.
- Howe GA, Lee GI, Itoh A, Li L, DeRocher AE.** 2000. Cytochrome P450-Dependent Metabolism of Oxylipins in Tomato. Cloning and Expression of Allene Oxide Synthase and Fatty Acid Hydroperoxide Lyase. *Plant Physiology* **123**, 711-724.

- Ilg A, Beyer P, Al-Babili S.** 2009. Characterization of the rice carotenoid cleavage dioxygenase 1 reveals a novel route for geranyl biosynthesis. *FEBS Journal* **276**, 736-747.
- Janse J, Schols M.** 1995. Une preference pour un gout sucre et non farineux. *Groenten+Fruit* **26**, 16-17.
- Kader AA, Stevens MA, Albright-Holton M, Morris LL, Algazi M.** 1977. Effect of fruit ripeness when picked on flavour and composition in fresh market tomatoes. *Journal of the American Society for Horticultural Science* **102**, 724-731.
- Klee HJ.** 2010. Improving the flavor of fresh fruits: genomics, biochemistry, and biotechnology. *New Phytologist* **187**, 44-56.
- Klee HJ, Giovannoni JJ.** 2011. Genetics and control of Tomato Fruit Ripening and Quality Attributes. *Annual Review Genetics* **45**, 41-59.
- Klee HJ, Tieman DM.** 2013. Genetic challenges of flavor improvement in tomato. *Trends in Genetics* **29**, 257-262.
- Koeduka T, Fridman E, Gang DR et al.** 2006. Eugenol and isoeugenol, characteristic aromatic constituents of spices, are biosynthesized via reduction of a coniferyl alcohol ester. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 10128-10133.
- Kovács K, Fray R, Tikunov Y, Graham N, Bradley G, Seymour GB, Bovy AG, Grierson D.** 2009. Effect of Tomato Pleiotropic Ripening mutations on Flavour Volatile Biosynthesis. *Phytochemistry* **70**, 1003-1008.
- Kochevenko A, Araújo WL, Maloney GS, Tieman, DM, Do PT, Taylor MG, Klee HJ, Fernie AR.** 2012. Catabolism of Branched Chain Amino Acids Supports Respiration but Not Volatile Synthesis in Tomato Fruits. *Molecular Plant* **5**, 366-375.
- Kuroda H, Oshima T, Kaneda H, Takashio M.** 2005. Identification and functional analyses of two cDNAs that encode fatty acid 9-/13-hydroperoxide lyase (CYP74-C) in rice. *Bioscience Biotechnology and Biochemistry* **69**, 1545-1554.

- Lê S, Ledauphin S.** 2006. You like tomato, I like tomato: segmentation of consumers with missing values. *Food Quality and Preference* **17**, 228-233.
- Lengard V, Kermit M.** 2006. 3-Way and 3-block PLS regressions in consumer preference analysis. *Food Quality and Preference* **17**, 234-242.
- Lewinsohn E, Sitrit Y, Bar E, Azulay Y, Ibdah M, Meir A, Yoser E, Zamir D, Tadmor Y.** 2005. Carotenoid Pigmentation Affects the Volatile Composition of Tomato and Watermelon Fruits, as Revealed by Comparative Genetic Analysis. *Journal of Agricultural and Food Chemistry* **53**, 3142-3148.
- Liavonchanka A, Feussner I.** 2006. Lipoxygenases: Occurrence, functions and catalysis. *Journal of Plant Physiology* **163**, 348-357.
- Mageroy MH, Tieman DM, Floystad A, Taylor MG, Klee HJ.** 2012. A *Solanum lycopersicum* catechol-*O*-methyltransferase involved in synthesis of the flavor molecule guaiacol. *Plant Journal* **69**, 1043-1051.
- Maloney GS, Kochevenko A, Tieman DM, Tohge T, Krieger U, Zamir D, Taylor MG, Fernie AR, Klee HJ.** 2010. Characterization of the Branched-Chain Amino Acid Aminotransferase Enzyme Family in Tomato. *Plant Physiology* **153**, 925-936.
- Marilley L, Casey MG.** 2004. Flavours of cheese products: metabolic pathways, analytical tools and identification of producing strains. *International Journal of Food Microbiology* **90**, 139-159.
- Marlatt C, Ho CT, Chien M.** 1992. Studies of Aroma Constituents Bound as Glycosides in Tomato. *Journal of Agricultural and Food Chemistry* **40**, 249-252.
- Mathieu S, Dal Cin V, Fei Z, Li H, Bliss P, Taylor MG, Klee HJ, Tieman DM.** 2009. Flavour compounds in tomato fruits: identification of loci and potential pathways affecting volatile composition. *Journal of Experimental Botany* **60**, 325-337.

- Matsui K, Kurishita S, Hisamitsu A, Kajiwara T.** 2000a. A lipid-hydrolysing activity involved in hexenal formation. *Biochemical Society Transactions* **28**, 857-860.
- Matsui K, Ujita C, Fujimoto SH, Wilkinson J, Hiatt B, Knauf V, Kajiwara T, Feussner I.** 2000b. Fatty acid 9- and 13-hydroperoxide lyases from cucumber. *FEBS Letters* **481**, 183-188.
- Matsui K.** 2006. Green leaf volatiles: hydroperoxyde lyase pathway of oxylipin metabolism. *Current Opinion in Plant Biology* **9**, 274-280.
- Mita G, Quarta A, Fasano P, et al.** 2005. Molecular cloning and characterization of an almond 9-hydroperoxide lyase, a new CYP74 targeted to lipid bodies. *Journal of Experimental Botany* **56**, 2321-2333.
- Moummou H, Tonfack LB, Chervin C, Benichou M, Youmbi E, Ginies C, Latché A, Pech JC, van der Rest B.** 2012. Functional characterization of SlscADH1, a fruit ripening-associated short-chain alcohol dehydrogenase of tomato. *Journal of Plant Physiology* **169**, 1435-1444.
- Nagegowda DA.** 2010. Plant volatile terpenoid metabolism: Biosynthetic genes, transcriptional regulation and subcellular compartmentation. *FEBS Letters* **584**, 2965-2973.
- Negoias S, Visschers R, Boelrijk A, Hummel T.** 2008. New ways to understand aroma perception. *Food Chemistry* **108**, 1247-1254.
- Noordermeer MA, Veldink GA, Vliegenthart JFG.** 1999. Alfalfa contains substantial 9-hydroperoxide lyase activity and a 3Z : 2E-enal isomerase. *FEBS Letters* **443**, 201-204.
- Ortiz-Serrano P, Gil JV.** 2007. Quantitation of Free and Glycosidically Bound Volatiles and Effect of Glycosidase Addition on Three Tomato Varieties (*Solanum lycopersicum* L.). *Journal of Agricultural and Food Chemistry* **55**, 9170-9176.
- Ortiz-Serrano P, Gil JV.** 2010. Quantitative Comparison of Free and Bound Volatiles of Two Commercial Tomato Cultivars (*Solanum lycopersicum* L.) during Ripening. *Journal of Agricultural and Food Chemistry* **58**, 1106-1114.

- Orzáez D, Medina A, Torre S, Fernández-Moreno JP, Rambla JL, Fernández-del-Carmen A, Butelli E, Martin C, Granell A.** 2009. A Visual Reporter System for Virus-Induced Gene Silencing in Tomato fruit Based on Anthocyanin Accumulation. *Plant Physiology* **150**, 1122-1134.
- Piombino P, Sinesio F, Moneta E, et al.** 2013. Investigating physicochemical, volatile and sensory parameters playing a positive or a negative role on tomato liking. *Food Research International* **50**, 409-419.
- Rick CM, Uhlig JW, Jones AD.** 1994. High  $\alpha$ -tomatine content in ripe fruit of Andean *Lycopersicon esculentum* var. *cerasiforme*: Developmental and genetic aspects. *Proceedings of the National Academy of Sciences USA* **91**, 12877-12881.
- Sanchez G, Besada C, Badenes ML, Monforte AJ, Granell A.** 2012. A Non-Targeted Approach Unravels the Volatile Network in Peach Fruit. *PLoS ONE* **7**, e38992. doi: 10.1371/journal.pone.0038992.
- Simkin AJ, Schwartz SH, Auldridge M, Taylor MG, Klee HJ.** 2004. The tomato carotenoid cleavage dioxygenase 1 genes contribute to the formation of the flavor volatiles beta-ionone, pseudoionone, and geranylacetone. *Plant Journal* **40**, 882-892.
- Sinesio F, Cammareri M, Moneta E, Navez B, Peparai M, Causse M, Grandillo S.** 2010. Sensory Quality of Fresh French and Dutch Market Tomatoes: A Preference Mapping Study with Italian Consumers. *Journal of Food Science* **75**, S55-S67.
- Speirs J, Lee E, Holt K, Yong-Duk K, Scott NS, Loveys B, Schuch W.** 1998. Genetic Manipulation of Alcohol Dehydrogenase Levels in Ripening Tomato Fruit Affects the Balance of Some Flavor Aldehydes and Alcohols. *Plant Physiology* **117**, 1047-1058.
- Tadmor Y, Fridman E, Gur A, Larkov O, Lastochkin E, Ravid U, Zamir D, Lewinsohn E.** 2002. Identification of malodorous, a wild species allele affecting tomato aroma that was selected against during domestication. *Journal of Agricultural and Food Chemistry* **50**, 2005-2009.

- Tandon KS, Baldwin EA, Scott JW, Shewfelt RL.** 2003. Linking Sensory descriptors to Volatile and Nonvolatile Components of fresh Tomato Flavor. *Journal of Food Science* **68**, 2366-2371.
- Tieman DM, Zeigler M, Schmelz EA, Taylor MG, Bliss P, Kirst M, Klee HJ.** 2006a. Identification of loci affecting flavour volatile emissions in tomato fruits. *Journal of Experimental Botany* **57**, 887-896.
- Tieman D, Taylor M, Schauer N, Fernie AR, Hanson AD, Klee HJ.** 2006b. Tomato aromatic amino acid decarboxylases participate in synthesis of the flavor volatiles 2-phenylethanol and 2-phenylacetaldehyde. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 8287-8292.
- Tieman DM, Loucas HM, Kim JY, Clark DG, Klee HJ.** 2007. Tomato phenylacetaldehyde reductases catalyze the last step in the synthesis of the aroma volatile 2-phenylethanol. *Phytochemistry* **68**, 2660-2669.
- Tieman D, Zeigler M, Schmelz E, Taylor MG, Rushing S, Jones SB, Klee HJ.** 2010. Functional analysis of a tomato salicylic acid methyl transferase and its role in synthesis of the flavor volatile methyl salicylate. *Plant Journal* **62**, 113-123.
- Tieman D, Bliss P, McIntyre LM, et al.** 2012. The Chemical Interactions Underlying Tomato Flavor Preference. *Current Biology* **22**, 1035-1039.
- Tikunov Y, Lommen A, Ric de Vos CH, Verhoeven HA, Bino RJ, Hall RD, Bovy AG.** 2005. A Novel Approach for Nontargeted Data Analysis for Metabolomics. Large-Scale Profiling of Tomato Fruit Volatiles. *Plant Physiology* **139**, 1125-1137.
- Tikunov YM, de Vos RCH, Gonzalez-Paramas AMG, Hall RD, Bovy AG.** 2010. A Role for Differential Glycoconjugation in the Emission of Phenylpropanoid Volatiles from Tomato Fruit Discovered Using a Metabolic Data Fusion Approach. *Plant Physiology* **152**, 55-70.
- Tikunov YM, Molthoff J, de Vos RC, et al.** 2013. NON-SMOKY GLYCOSYLTRANSFERASE1 Prevents the Release of Smoky Aroma from Tomato Fruit. *Plant Cell* **25**, 3067-3078.

- Tzin V, Rogachev I, Meir S, Ben Zvi MM, Masci T, Vainstein A, Aharoni A, Galili G.** 2013. Tomato fruits expressing a bacterial feedback-insensitive 3-deoxy-D-arabino-heptulosonate 7-phosphate synthase of the shikimate pathway possess enhanced levels of multiple specialized metabolites and upgraded aroma. *Journal of Experimental Botany* **64**, 4441-4452.
- Ursem R, Tikunov Y, Bovy A, van Berloo R, van Eeuwijk, F.** 2008. A correlation network approach to metabolic data analysis for tomato fruits. *Euphytica* **161**, 181-193.
- Vancanneyt G, Sanz C, Farmaki T, Paneque M, Ortego F, Castañera P, Sanchez-Serrano JJ.** 2001. Hydroperoxide lyase depletion in transgenic potato plants leads to an increase in aphid performance. *Proceedings of the National Academy of Science of the United States of America* **98**, 8139-8144.
- Vogel JT, Tan BC, McCarty DR, Klee HJ.** 2008. The carotenoid cleavage dioxygenase 1 enzyme has broad substrate specificity, cleaving multiple carotenoids at two different bond positions. *Journal of Biological Chemistry* **283**, 11364-11373.
- Vogel JT, Tieman DM, Sims CA, Odabasi AZ, Clark DG, Klee HJ.** 2010. Carotenoid content impacts flavour acceptability in tomato (*Solanum lycopersicum*). *Journal of the Science of Food and Agriculture* **90**, 2233-2240.
- Walter MH, Floss DS, Strack D.** 2010. Apocarotenoids: hormones, mycorrhizal metabolites and aroma volatiles. *Planta* **232**, 1-17.
- Zanor MI, Rambla JL, Chaib J, Steppa A, Medina A, Granell A, Fernie AR, Causse M.** 2009. Metabolic characterization of loci affecting sensory attributes in tomato allows an assessment of the influence of the levels of primary metabolites and volatile organic contents. *Journal of Experimental Botany* **60**, 2139-2154.
- Zhang B, Chen KS, Bowen J, Allan A, Espley R, Karunaitretnam S, Ferguson I.** 2006. Differential expression within the LOX gene family in ripening kiwifruit. *Journal of Experimental Botany* **57**, 3825-3836.



**Zorrilla-Fontanesi Y, Rambla JL, Cabeza A, Medina JJ, Sánchez-Sevilla JF, Valpuesta V, Botella MA, Granell A, Amaya, I.** 2012. Genetic Analysis of Strawberry Fruit Aroma and Identification of O-Methyltransferase FaOMT as the Locus Controlling Natural Variation in Mesi-furane Content. *Plant Physiology* **59**, 851-870.



## OBJECTIVES

Considering these precedents, the following objectives were proposed:

1. Establish a robust, high-throughput method for the analysis of volatile metabolites in tomato fruit samples, compatible with a multi-omics approach.
2. Characterize the effect on the volatile profile of quantitative trait loci (QTLs) associated with sensory attributes in tomato fruit in different genetic backgrounds, and evaluate the possible effect of volatile compounds on fruit flavour and aroma.
3. Create a quantitative trait loci (QTLs) map for volatile compounds from a recombinant inbred line (RIL) population derived from red-fruited wild tomato relative *S. pimpinellifolium* accession TO-937.
4. Validate the newly identified QTLs by their introgression in fresh market tomato *S. lycopersicum* cv. 'Moneymaker'.



## **Chapter 1**

### **Tomato fruit volatile profiles are highly dependent on sample processing and capturing methods**

José Luis Rambla, Cristina Alfaro, Aurora Medina, Manuel Zarzo, Jaime Primo,  
Antonio Granell

*Metabolomics*, 11, 1708-1720 (2015)



## **ABSTRACT**

Volatile compounds are together with sugars and organic acids the main determinants of tomato fruit flavour and are therefore important for consumer acceptance. Consequently, in the last years many studies have been performed using different volatile analytical techniques on a large diversity of tomato fruits, aimed mainly at detecting the compounds affecting flavour or at the identification of QTLs and key genes involved in fruit volatile contents. The comparison of three of the analytical methods most commonly applied (headspace, solid phase microextraction, adsorption on Tenax followed by thermal desorption) revealed not only differences in sensitivity, but also dramatic variations in the volatile profile obtained by each of these techniques. The volatile profile was also largely influenced by the way samples were processed before analysis. Four widely used sample processing methods were compared (whole tomato, sliced fruit and two different types of fruit paste), each one producing a characteristic volatile pattern. Therefore, great care should be taken when comparing results available from the literature obtained by means of different methods, or when using the volatile levels obtained in an experiment to predict their influence on tomato flavor or consumer preference, or to assess the success of breeding programs.

## **Keywords**

Tomato fruit, volatile, flavour, solid phase microextraction, headspace, thermal desorption.





## 1 INTRODUCTION

Tomato flavour is mainly the result of the interaction of aroma and taste, together with other sensory inputs such as texture, temperature or mouth feel (Goff and Klee, 2006), and it has been the object of many studies in the last two decades. Volatile organic compounds (VOCs) play an important role in the flavour, as they not only define the aroma but also modify other perceptions related to taste, like sweetness or sourness (Baldwin et al. 1998, 2004; Vogel et al. 2010; Tieman et al. 2012). A large number of volatile compounds have been described in fresh tomato or tomato products (Petró-Turza 1987), but most of them seem to have no effect on our perception of tomato flavour. Thus, based on quantitative results of volatile composition of ripe tomato fruits, extracted from a fresh tomato paste, a list of 16 compounds was proposed (Buttery 1993) to presumably affect fresh tomato flavour and aroma including how relevant each of them should be for flavour perception. The contribution of each individual volatile compound was estimated by determining the odour threshold values and checking whether or not the detected levels in red fruits were above or below these thresholds (Buttery 1993). However, more recent research revealed that some compounds not present in that list were also important for flavour perception at least in some varieties (Causse et al. 2002; Zanor et al. 2009; Tikunov et al. 2010, 2013; Selli et al. 2014), and the relevance of some compounds that had been considered very important for tomato flavour has been questioned (Tieman et al. 2012). The difficulty in defining the volatiles most relevant for flavour is aggravated by the five orders of magnitude variation in the sensitivity to the same compound between individuals due to genetical variation for olfactory receptor genes (Jaeger et al. 2013; McRae et al. 2013). These issues are still under debate and have been discussed in detail in a recent review (Rambla et al. 2014).

Considering its importance on consumer preference, the identification of different factors affecting volatile profiles has been object of many studies. Thus, different studies have evaluated the changes in volatile profiles associated to climatic and cultural conditions (Cebolla-Cornejo et al. 2011), physiological maturity at harvest (Baldwin et al. 1991; Maul et al. 1998) or postharvest handling (MacDonald et al. 1996; Maul et al. 2000; Renard et al. 2013). Considerable efforts have also been carried out to map the genetic regions controlling volatile emission (Causse et al. 2002; Tadmor et al. 2002; Tieman et al. 2006a; Mathieu et al. 2009; Zanor et al. 2009) or to identify the genes involved in the biosynthetic pathways of key volatiles (Chen et al. 2004; Simkin et al. 2004; Tieman et al. 2006b, 2007, 2010; Goulet et al. 2012; Mageroy et al. 2012; Tikunov et al. 2013).

Different methods for the sampling of tomato volatiles have been used in the literature, such as headspace (Baldwin et al. 1991, 1998, 2004; Maul et al. 2000; Tandon et al. 2003), headspace solid phase microextraction (Tikunov et al. 2005, 2013; Zanor et al. 2009; Ortiz-Serrano and Gil 2010), adsorption in a Super Q resin followed by elution in an organic solvent (Tieman et al. 2006a; Mathieu et al. 2009; Goulet et al. 2012; Mageroy et al. 2012), liquid-liquid extraction with an organic solvent (Aubert et al. 2005; Selli et al. 2014), purge-and-trap in a Tenax sorbent followed by thermal desorption (Ruiz et al. 2005), or purified air or dry nitrogen passed over the sample so that the extracted volatiles are retained in a Tenax trap and then extracted with an organic solvent (Buttery et al. 1987, 1988; Buttery 1993; Beltran et al. 2006), just to cite some of them. All these methods are capable of collecting a subset of volatiles from the fruit samples (some of them already present in the whole intact fruit, others produced during the experimental procedure), but no systematic comparative study had been performed to compare them.

Additionally, different alternatives can be found in the literature regarding the way tomato samples are processed before analysis. The method most frequently used is to process the fruit to form a paste, with some variations in the preparation between research groups (Buttery et al. 1987; Baldwin et al. 1991; Ruiz et al. 2005; Tikunov et al. 2005; Ortiz-Serrano and Gil 2010), while some researchers perform the analysis on sliced/chopped fruit (Tieman et al. 2006a; Mathieu et al. 2009; Vogel et al. 2010) or even the whole intact fruit (Carbonell-Barrachina et al. 2006). The results obtained show a wide range of different detected compounds, and a large sample to sample variability for a given compound (Carbonell-Barrachina et al. 2006). This variation is in part, but not completely, due to the high degree of biological variability existing in the levels of volatile compounds between different tomato cultivars (Tieman et al. 2012; Rambla et al. 2014).

The present work unravels the variability in the tomato fruit volatile profile obtained by comparing a number of volatile capturing and sample processing protocols that were chosen among those more frequently reported in the literature. The aim of the present work is to find out to what extent the different methodologies employed affect the volatile profile obtained and also to shed some light on which method is preferable depending on the objective of the study.

## 2 MATERIALS AND METHODS

**2.1 Plant Material.** Tomato fruits (*Solanum lycopersicum* L. cv. Moneymaker) were grown under standard cultural practices in a greenhouse at Experimental Station La Mayora in Málaga (Southern Spain) during the spring season. Red ripe fruits were collected in early June.

**2.2 Chemicals and compound identification.** Compounds in the chromatograms were identified based on the coincidence of both their mass spectra and retention times with those of authentic standards injected under the same analytical conditions. All the standards were injected with each of the techniques used (thermal desorption, headspace and solid phase microextraction) due to small shifts in retention times depending on the technique used. The retention times indicated in tables 1 and S1 correspond to those obtained from thermal desorption. All the compounds used as standards were purchased from Sigma-Aldrich (Madrid, Spain) except for 1-nitro-2-phenylethane that was acquired from Apin Chemicals Ltd. (Abingdon, UK). Dihydrate calcium chloride was purchased from Merck (Darmstadt, Germany). EDTA was obtained from Panreac Química (Barcelona, Spain). Milli-Q water was used throughout the study.

**2.3 Fruit Samples.** Samples were processed following four different methods. In each method four biological replicates were analyzed. The four sample processing methods used were: whole tomato fruit, sliced fruit and two types of tomato paste (A and B). For whole fruit analysis, each replicate consisted of 4 tomato fruits of about 50 g each (total weight about 200 g). For sliced fruit, tomatoes were cut in 16 half-wedges (1 transversal cut and 4 longitudinal cuts through the middle of the fruit). Each piece weighed about 3 g (total weight about 50 g). Tomato paste A was prepared following essentially the procedure

of Buttery et al. (1987). Roughly, fresh tomatoes were homogenized at room temperature ( $25\pm 1^\circ\text{C}$ ) for 30 seconds in a Waring blender. The purée obtained was allowed to stand for 3 min, and then the same volume of a saturated  $\text{CaCl}_2$  solution was added, and the sample was thoroughly mixed, stored in 50 mL aliquots and frozen at  $-20^\circ\text{C}$  until analysis. Immediately before analysis, the sample was thawed at room temperature ( $25\pm 1^\circ\text{C}$ ) and transferred to the corresponding vial. Tomato paste B preparation followed essentially the procedure of Tikunov et al. (2005) with minor modifications. Roughly, fresh tomatoes (two in each replicate) were cut into pieces and flash frozen in liquid nitrogen. The frozen material was homogenized in a cryogenic mill, and stored at  $-80^\circ\text{C}$  until analysis. Immediately before analysis, 1 g of the resulting powder was weighed in a 7 mL vial, incubated for 10 min at  $37^\circ\text{C}$ , and then 2.2 g of  $\text{CaCl}_2\cdot 2\text{H}_2\text{O}$  and 1 mL of a 100 mM EDTA-NaOH pH 7.5 solution were added, gently mixed and sonicated for 5 min. Finally, 2 mL of the resulting paste were transferred to a 22 mL crimp cap vial for analysis.

**2.4 Tenax Adsorption followed by Thermal Desorption (TD) Procedure.** Fifty grams of sample, either fresh fruit or fruit paste (except for the whole fruit, where about 200 g were used), were placed in a glass tray inside a 1.3 L glass cylinder (45 cm length, 6 cm diameter). A zero air flow of 100 mL/min passed through the cylinder with the tray during 30 min at room temperature ( $25\pm 1^\circ\text{C}$ ) and the volatiles emitted by the sample were adsorbed in a Tenax TA stainless steel TD tube (O.D. x L. 1/4 in. x 3 1/2 in.) (Supelco, Pennsylvania, USA) located at the end of the glass cylinder. This acquisition procedure is similar, although not identical, to that described in Tieman et al. (2006a). Desorption was performed thermally in a TurboMatrix TD Thermal Desorber (Perkin Elmer, Wellesley, USA) with the following parameters: primary desorption temperature  $300^\circ\text{C}$ , desorb

flow 100 mL/min, desorption time 10 min, inlet split 3 mL/min, outlet split 20 mL/min, Tenax trap temperature -30/300°C.

**2.5 Static Headspace Trap (HS) Procedure.** 22 mL crimp cap vials (Perkin Elmer) were used, each vial containing 4 mL of tomato paste A. Analyses were performed in a TurboMatrix 40 Trap Headspace Sampler (Perkin Elmer). The analysis conditions were: equilibration time, 80 min; split 1/13; equilibration temperature, 37°C, 50°C or 80°C, as specified in each case.

**2.6 Headspace Solid Phase Microextraction (SPME) Procedure.** For SPME analysis, 2 mL of tomato paste A were introduced in a 22 mL crimp cap vial, and pre-incubated at either 37°C, 50°C or 80°C for 10 min. Then a 65 µm PDMS/DVB fiber (Supelco) was exposed to the headspace for another 10 min at 37°C, 50°C or 80°C respectively. During both pre-incubation and extraction, samples were agitated in a ThermoMixer Comfort agitator (Eppendorf) at 300 rpm. The volatiles trapped on the fiber were desorbed for 1 min at 250°C in the injection port of the GC/MS and then the fiber was cleaned by exposing it for 5 min at 250°C in another injection port to prevent cross-contamination. The mode of injection was splitless. Sampling was performed manually.

**2.7 Gas Chromatography/Mass Spectrometry Conditions.** In all cases, volatile organic compounds were analyzed by GC/MS using a Clarus®500 GC/MS from PerkinElmer, equipped with ZB-5 capillary column (30 m, 0.25 mm, 0.25µm) (Phenomenex, Torrance, CA, USA). Oven programming conditions were 40°C for 3 min, 3°C/min ramp until 60°C, 6°C/min ramp until 160°C, 12°C/min ramp until 260°C, then 5 min at 260°C. Helium was used as the carrier gas: for SPME analysis, with a 1.2 mL/min constant flow; for TD ad HS analysis, at a constant pressure of 19.5 psi. Mass spectra in the electron impact (EI) mode were generated at 70 eV. Chromatograms were recorded in scan mode in the  $m/z$

range 35-300. Chromatograms and spectra were recorded with GC/MS TurboMass software version 5.0 (Perkin Elmer).

To calculate compound abundance, firstly a specific ion was selected for each compound and the resulting peak areas were integrated. Next, the corresponding Total Ion Count (TIC) area of each particular peak was calculated considering the relative abundance of the selected ion in relation to the mass spectrum of that compound. Finally, the Total Ion Count area of all peaks in the chromatogram was integrated. The results are expressed as area of each peak relative to the total area of all compounds in the chromatogram.

**2.8 Statistical analysis.** In order to test the effect of the capturing method, data were arranged in a matrix with 26 rows (4 replicates x 7 experimental conditions: TD, HS<sub>37</sub>, HS<sub>50</sub>, HS<sub>80</sub>, SPME<sub>37</sub>, SPME<sub>50</sub> and SPME<sub>80</sub>) by 49 compounds in columns. For SPME at 50°C and 80°C only 3 replicates were analyzed. In order to study the effect of sample processing experiment, another matrix was arranged containing 16 rows (4 types of sample processing x 4 replicates) by 26 columns (volatile compounds). Each row can be regarded as the volatile profile of the experimental trial. In both cases, data in columns were mean-centered and scaled to unit variance and, next, Principal Component Analysis (PCA) was applied using the software SIMCA-P 10.0 (Umetrics, Sweden) in order to highlight the differences among experimental conditions. Principal components are directions of maximum data variance obtained as linear combinations of the original variables. The contributions of variables (compounds) in the formation of a given component are called loadings,  $p[1]$  being the loadings in the formation of the first principal component (PC1);  $p[2]$ , the loadings of PC2; and so on. The projections of observations (experimental trials in this case) over the directions determined by PC1 and PC2 are called  $t[1]$  scores and  $t[2]$  scores,

respectively. A scatter plot of loadings or scores corresponding to two different components is referred to as loading plot or score plot, respectively.

For Hierarchical Cluster Analysis (HCA), the mean of the data for each capturing method (TD, HS<sub>37</sub>, HS<sub>50</sub>, HS<sub>80</sub>, SPME<sub>37</sub>, SPME<sub>50</sub> and SPME<sub>80</sub>) was used after square root normalisation. HCA was performed by means of Acuity 4.0 software (Axon Instruments; Union City, CA, USA) with the distance metrics based on the Pearson correlation. The normalised data was represented as a heatmap by means of the same software. For both PCA and HCA, the compounds not detected (either below the detection threshold or not present) were assigned a value of 0.



## 3 RESULTS AND DISCUSSION

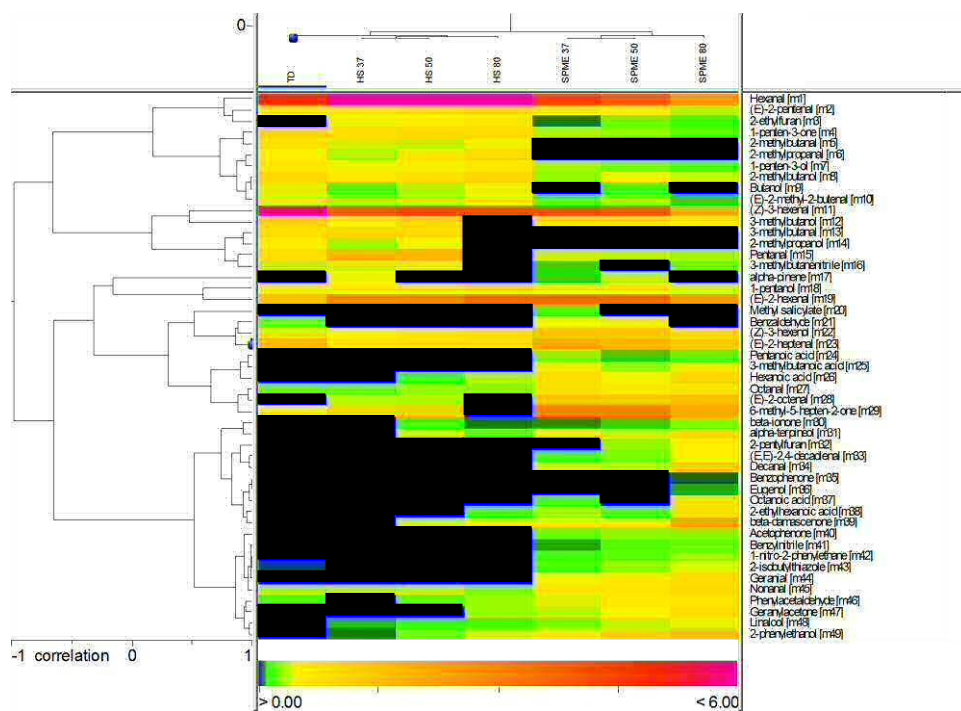
### 3.1 Effect of the capturing method on the volatile profile

A large sample of tomato fruit was processed as described for tomato paste A, and different aliquots of it were subjected to volatile capture/release by means of three different techniques. The first was a Tenax adsorption-Thermal Desorption (TD) method. Volatiles emitted at room temperature (25°C) from 50 mL of tomato paste located on a glass tray inside of a glass cylinder were adsorbed on a Tenax tube by flushing purified air and the trapped volatiles were desorbed thermally in line with the GC/MS. The second was a headspace (HS) method. Volatiles were partitioned between the matrix and the headspace at three different temperatures: 37°C, 50°C and 80°C. According to this method, 4 mL of sample in a 22 mL vial emitted volatile compounds until equilibrium, and all the volatiles in the headspace were then concentrated in a cold trap and thermally desorbed for analysis. The third method was based on headspace solid phase microextraction (SPME). Analyses were also performed at three different temperatures: 37°C, 50°C and 80°C. In this method, the volatiles emitted by 2 mL of sample to the headspace of a 22 mL vial were captured by a PDMS/DVB coated fiber, and the retained volatiles were thermally desorbed in the injection port of the gas chromatograph for analysis.

#### *3.1.1 Capturing by SPME allows the detection of more complex volatile profiles*

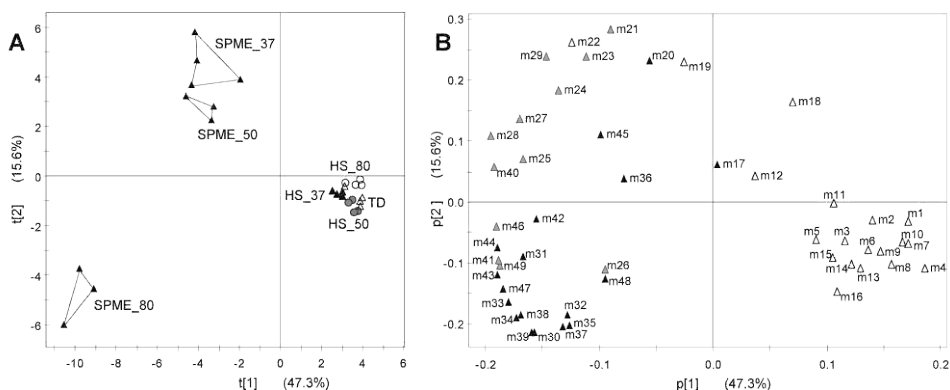
A few hundred compounds were detected in the samples analysed, many of which were present at low levels. A total number of 49 volatile compounds were unequivocally identified. Only this set of compounds was used in the comparative study, although not all were detectable by all trapping

techniques. Headspace Solid Phase Microextraction (SPME) seemed to produce the richest profile of all the methods tested. SPME allowed the detection of about 10-15 compounds more than Headspace-Trap (HS) or Tenax adsorption-Thermal Desorption (TD) methods. In total, 40-41 compounds were unequivocally identified after SPME, depending on the temperature of volatile acquisition, whilst only 26-31 were detected after HS, again depending on temperature, and only 25 compounds were detected following TD (Figure 1, Table S1). Compounds that were only tentatively identified were not considered in our study (data not shown). Figure S1 shows a representative chromatogram obtained by each of the capturing methods.



**Figure 1.** Hierarchical Cluster Analysis from the average data of both analytical methods and volatile compounds. Values are represented as a heatmap according to the scale below. Black colour corresponds to compounds not detected; blue colour corresponds to compounds with very low abundance; pink corresponds to the maximum relative abundance values. Data were square root normalized.

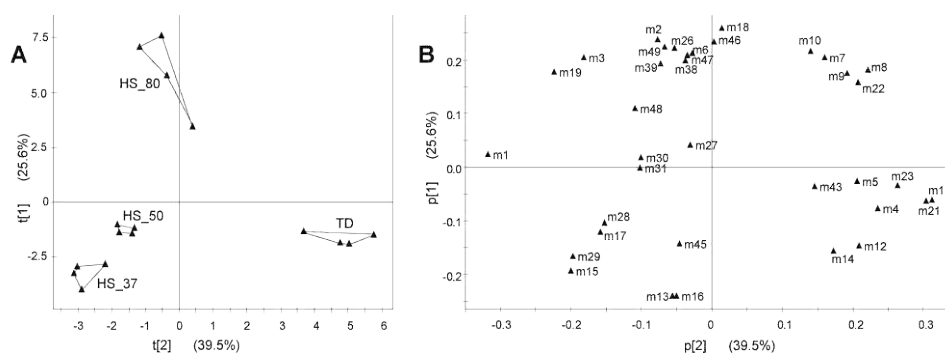
In order to highlight the differences among the resulting volatile compound profiles associated to each capturing method, Principal Component Analysis (PCA) was applied to the dataset matrix. PC1 explained 47.3% of the total data variability, and the score plot (Figure 2A) revealed that SPME produced the most distinctive volatile profiles. In order to better characterize the differences between TD and HS, the PCA was repeated after removing SPME values from the dataset (Figure 3A). From both figures 2A and 3A it is evident that each of the capturing methods resulted in a differentiated volatile profile.



**Figure 2.** Principal component analysis score plot (A) and loading plot (B) of the volatile profiles obtained by different analytical methods. TD, Thermal Desorption; HS, Headspace at 37°C, 50°C or 80°C; SPME, Headspace Solid Phase Microextraction at 37°C, 50°C or 80°C. Observations corresponding to different replicates are joined with solid lines. Compound codes in B are as in Figure 1 and Table S1. Triangles are coloured according to the molecular weight (MW) of each compound: white, MW ≤ 100 Da; gray, 130 Da ≥ MW > 100 Da; black, MW > 130 Da.

The specific compounds that allowed us to discriminate among the different volatile profiles according to the VOC capturing method can be deduced from the PCA loading plots (Figures 2B and 3B). It turned out that the profile obtained by SPME analysis was markedly enriched in higher molecular weight compounds with lower volatility, and particularly in compounds from C<sub>8</sub>

to C<sub>13</sub>. Thirteen compounds were exclusively detected by SPME: relatively long-chain compounds such as decanal, (*E,E*)-2,4-decadienal, geranial, 2-pentylfuran, acetophenone, benzophenone, benzylnitrile, 1-nitro-2-phenylethane, methyl salicylate and eugenol, and also high polarity compounds such as 3-methylbutanoic, pentanoic and octanoic acids. All these compounds apparently fell below the detection limit of the other techniques used. This is consistent with the higher sensitivity of SPME as compared to HS for the analysis of volatile compounds in food products previously reported (Gamero et al., 2013). Additionally, the volatile profiles obtained after SPME were characterized by higher levels of other compounds with low or relatively low volatility such as  $\beta$ -ionone,  $\beta$ -damascenone, geranylacetone, linalool,  $\alpha$ -terpineol, benzaldehyde, phenylacetaldehyde, 2-phenylethanol, 2-isobutythiazole, hexanoic acid, 2-ethylhexanoic acid, nonanal, octanal, (*E*)-2-octenal, (*E*)-2-heptenal, (*Z*)-3-hexenol or 6-methyl-5-hepten-2-one when compared to those profiles obtained by either TD or HS.



**Figure 3.** Principal component analysis score plot (A) and loading plot (B) of the volatile profiles obtained by Thermal Desorption (TD) or Headspace (HS) at 37°C, 50°C or 80°C. Observations corresponding to different replicates are joined with solid lines. Compound codes in B are as in Figure 1 and Table S1.

### 3.1.2 Capturing by TD or HS provides better sensitivity for highly volatile molecules

On the other hand, both TD and HS revealed to be more sensitive for the detection of highly volatile compounds. These techniques allowed the determination of four short-chain compounds that were not detected by SPME: 2-methylpropanal, 2-methylpropanol, 3-methylbutanal and 2-methylbutanal. Additionally, the profiles of both TD and HS were enriched in C<sub>4</sub> and C<sub>5</sub> compounds, and also a few C<sub>6</sub> volatiles, including butanol, 2-methylbutanol, (*E*)-2-methyl-2-butenal, 3-methylbutanenitrile, pentanal, (*E*)-2-pentenal, 1-penten-3-ol, 1-penten-3-one, hexanal, (*Z*)-3-hexenal and 2-ethylfuran. All these compounds show low affinity for the PDMS/DVB fiber coating used, and consequently they were poorly retained.

When the results obtained by HS with different temperatures of incubation and TD were compared, the PCA score plot (Figure 3A) revealed that each of them also produced a characteristic profile. PC1, accounting for 39.5% of the total variability, separated the different variations of HS technique according to the temperature used for the collection of volatiles. PC2 accounted for 25.6% of the variability, and separated the samples acquired by means of TD from those acquired through any variation of HS. The loading plot revealed that acquisition by means of TD produced profiles enriched in (*Z*)-3-hexenal, (*E*)-2-heptenal, benzaldehyde, 1-penten-3-one, 3-methylbutanol, 2-methylbutanal, 2-methylpropanol and 2-isobutylthiazole. On the contrary, when the samples were acquired by means of HS, the profile obtained had lower relative levels of these compounds and higher levels of hexanal, (*E*)-2-hexenal, pentanal, 2-ethylfuran and 6-methyl-5-hepten-2-one (Figure 3B).

### 3.1.3 Effect of the temperature

Three different temperatures were used for capturing volatiles in both HS and SPME: 37°C, 50°C and 80°C. It was decided to use 37°C because this is the temperature at which volatiles are expected to be released in people's mouth when eating tomato. Similarly low temperatures have also been used in headspace analysis of volatiles in other fruit species (Allwood et al. 2014). The temperatures of 50°C and 80°C have been widely used in the literature for volatile analysis of tomato (Maul et al. 2000; Tandon et al. 2003; Tikunov et al. 2005; Baldwin et al. 2008; Zantor et al. 2009). Relatively high temperatures are often used for the analysis of volatile compounds in methods based on compound volatilization in the headspace in order to increase analytical sensitivity (Nongonierma et al. 2006).

Figure 2A reveals that temperature has a major effect on the volatile profile obtained, most notably in SPME, where the volatile profiles at the highest temperature evaluated, 80°C, were very different to those obtained by any of the other analytical conditions evaluated. Results obtained for SPME at 50°C were intermediate to those at 37°C and 80°C, although much more similar to the former. The increase of the sample incubation temperature affected the volatile profile after SPME capture by increasing the relative levels of a number of relatively long-chain semi-volatile compounds, and most remarkably  $\beta$ -ionone,  $\beta$ -damascenone, decanal, 2-ethylhexanoic, hexanoic and octanoic acids, 2-pentylfuran, (*E,E*)-2,4-decadienal, geranylacetone, 2-isobutylthiazole, 2-phenylethanol, phenylacetaldehyde, 1-nitro-2-phenylethane, benzylnitrile, linalool,  $\alpha$ -terpineol, geranial and benzophenone (Figure 2B). The latter compound was only detected with SPME at 80°C.

The temperature of incubation barely affected the number of compounds detected after SPME capture, which yielded 41, 40 and 41 compounds at 37°C, 50°C and 80°C respectively. The main effect was found on the relative abundance of low volatility compounds, all of which increased their levels with the temperature. As a consequence, the maximum levels of organic acids and long chain compounds (C<sub>10</sub> and over) were observed at 80°C.

The effect of temperature on HS analysis can be clearly observed in Figure 3A. PC1 (39.5% of total variability) shows the effect of the temperature on the profiles obtained. As we have previously described for SPME, incubation at 37°C and 80°C produced clearly distinct profiles, whilst incubation at 50°C produced intermediate results closer to those obtained at 37°C. Incubation at high temperatures increased the levels of many semi-volatile compounds we have previously described for SPME such as phenylacetaldehyde, 2-phenylethanol, geranylacetone, hexanoic and 2-ethylhexanoic acids and  $\beta$ -damascenone, but also increased the levels of a number of short-chain compounds including 1-pentanol, (*E*)-2-pentenal, (*E*)-2-methyl-2-butenal, 2-methylpropanol, 1-penten-3-ol and 2-ethylfuran (Figure 3B). Among these compounds, 2-ethylhexanoic acid and geranylacetone were only detected after HS at the highest temperature. On the other hand, several compounds detected by HS after incubation at 37°C and 50°C failed to be detected at 80°C. This is the case of a set of branched-chain small molecular weight compounds such as 3-methylbutanenitrile, 3-methylbutanal, 3-methylbutanol, or 2-methylpropanol, the linear molecules pentanal and (*E*)-2-octenal, and also 6-methyl-5-hepten-2-one. This could be due to degradation of these compounds during the extended time required for gas equilibrium -80 min- at this high temperature.

The total number of compounds detected by HS at 37°C, 50°C and 80°C were 27, 31 and 26 respectively (Figure 1, Table S1). Therefore, for headspace

analysis, an intermediate temperature such as 50°C is apparently the most appropriate for the detection of a higher number of volatile compounds, as an increase in the incubation temperature favours the volatility of the compounds, but too high temperatures would accelerate the processes of degradation of many volatiles.

In conclusion, SPME was revealed to be by far the most sensitive of the three trapping techniques evaluated, yielding the most complex volatile profiles in our tomato samples. Therefore, it seems to be the best acquisition technique for approaches where a high-throughput volatile metabolomics analysis is necessary. Additionally, SPME should be the capturing method of choice when there is an interest on semi-volatile compounds, as in our hands it was the only technique which allowed the detection of most of them. An additional advantage of this technique is the low amount of biological material required (only 1 g of tomato fruit) to obtain a good sensitivity. The main limitation of SPME as performed in our experiments was the low sensitivity for highly volatile compounds, which were not detectable in our assays. This limitation could possibly be overridden by the use of a fiber coating with higher affinity for those compounds, such as divinylbenzene/carboxen/polydimethylsiloxane (DVB/CAR/PDMS). TD and HS, although less sensitive for a wide range of compounds, were very efficient for the detection of short-chain highly volatile compounds and would probably be useful techniques in those cases where this type of volatiles are of particular interest.

Regarding temperature of analysis, a moderately elevated temperature such as 50°C seems to be the most adequate, as it favours the emission and therefore the detection of semi-volatile compounds while minimizing the degradation of the sample.



### 3.2 Effect of sample processing on the volatile profile

In order to assess the effect of the sample processing method on the volatile profile, we compared the results obtained after processing the same biological sample (a pool of red ripe fruits) following four different methods. These were selected among those most commonly reported in the literature: tomato paste A (first processed, then frozen), tomato paste B (first frozen, then processed and pH adjusted), sliced fresh tomato fruit, and the whole intact unprocessed fresh fruit. After having processed the samples according to each method, the emitted volatiles were captured with a Tenax trap system and analyzed by thermal desorption coupled to gas chromatography and mass spectrometry, as described in the corresponding Material and Methods section. This capturing method was used because it allowed the handling of the different types of processed samples analyzed, as the size of sliced and whole fruits did not allow these samples to be introduced inside the vials used for HS or SPME. Representative chromatograms obtained by each of the sample processing methods are shown in Figure S2.

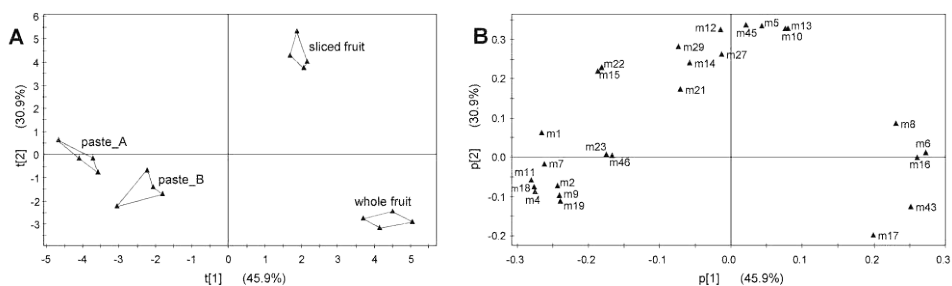
Out of the 49 volatile compounds previously described in this paper (Figure 1, Table S1), only 26 were detected by Thermal Desorption. The effect of sample processing on these compounds was so dramatic that only 3 of the 26 identified compounds (2-methylbutanol, 3-methylbutanenitrile and hexanal) were detected in all four types of samples. Table 1 shows the relative abundance of compounds expressed as the percentage area of each particular compound relative to the area of all peaks in the chromatogram. It turned out that only 6 of these compounds were detected in the whole fruit, whilst 25 were

**Table 1.** Relative abundance of volatile compounds after different sample processing methods. Values represent the average  $\pm$  standard deviation of the contribution of each individual compound related to the sum of the areas of all the peaks in the chromatogram, expressed as a percentage. RT, Retention Time (min); m/z, specific ion used for compound quantitation; *nd*, not detected.

ID	Compound	RT	m/z	Whole fruit	Sliced fruit	Paste A	Paste B
m1	Hexanal	4.62	72	1.37 $\pm$ 0.92	14.82 $\pm$ 1.32	33.04 $\pm$ 0.76	20.33 $\pm$ 7.82
m2	( <i>E</i> )-2-pentenal	3.64	83	nd	nd	0.45 $\pm$ 0.05	0.24 $\pm$ 0.24
m4	1-penten-3-one	2.47	55	nd	nd	1.61 $\pm$ 0.08	1.27 $\pm$ 0.29
m5	2-methylbutanal	2.28	57	nd	4.14 $\pm$ 0.13	1.00 $\pm$ 1.03	0.14 $\pm$ 0.05
m6	2-methylpropanal	1.27	72	4.06 $\pm$ 0.91	2.84 $\pm$ 0.39	0.44 $\pm$ 0.09	nd
m7	1-penten-3-ol	2.43	57	nd	0.24 $\pm$ 0.01	0.66 $\pm$ 0.11	0.83 $\pm$ 0.18
m8	2-methylbutanol	3.34	56	5.02 $\pm$ 2.67	5.00 $\pm$ 0.64	1.46 $\pm$ 0.05	1.60 $\pm$ 0.80
m9	Butanol	2.32	56	nd	nd	0.28 $\pm$ 0.03	0.37 $\pm$ 0.13
m10	( <i>E</i> )-2-methyl-2-butenal	3.40	84	nd	3.20 $\pm$ 0.34	0.28 $\pm$ 0.03	0.19 $\pm$ 0.08
m11	( <i>Z</i> )-3-hexenal	4.58	98	nd	3.63 $\pm$ 0.32	43.05 $\pm$ 1.96	29.01 $\pm$ 5.02
m12	3-methylbutanol	3.23	42	nd	6.33 $\pm$ 0.66	2.11 $\pm$ 0.06	2.45 $\pm$ 1.36
m13	3-methylbutanal	2.17	44	nd	17.53 $\pm$ 1.05	1.11 $\pm$ 0.46	1.11 $\pm$ 0.79
m14	2-methylpropanol	1.99	42	nd	1.45 $\pm$ 0.20	0.49 $\pm$ 0.07	1.13 $\pm$ 0.71
m15	Pentanal	2.61	44	nd	1.05 $\pm$ 0.28	1.09 $\pm$ 0.17	0.83 $\pm$ 0.39
m16	3-methylbutanenitrile	3.15	43	6.30 $\pm$ 2.27	3.91 $\pm$ 1.68	0.19 $\pm$ 0.01	0.23 $\pm$ 0.11
m17	$\alpha$ -pinene	9.40	93	0.16 $\pm$ 0.07	nd	nd	nd
m18	1-pentanol	3.92	42	nd	nd	0.65 $\pm$ 0.02	0.40 $\pm$ 0.06
m19	( <i>E</i> )-2-hexenal	6.34	83	nd	nd	3.60 $\pm$ 0.24	4.49 $\pm$ 2.44
m21	Benzaldehyde	10.69	106	nd	0.08 $\pm$ 0.08	0.07 $\pm$ 0.03	0.04 $\pm$ 0.08
m22	( <i>Z</i> )-3-hexenol	6.56	67	nd	0.93 $\pm$ 0.24	1.15 $\pm$ 0.31	0.49 $\pm$ 0.08
m23	( <i>E</i> )-2-heptenal	10.82	83	nd	nd	3.18 $\pm$ 1.98	nd
m27	Octanal	12.44	84	nd	0.11 $\pm$ 0.10	0.06 $\pm$ 0.03	nd
m29	6-methyl-5-hepten-2-one	11.90	69	nd	0.41 $\pm$ 0.17	0.38 $\pm$ 0.10	nd
m43	2-isobutylthiazole	13.60	99	1.75 $\pm$ 0.47	0.47 $\pm$ 0.04	0.00 $\pm$ 0.00	nd
m45	Nonanal	15.68	57	nd	0.41 $\pm$ 0.11	0.16 $\pm$ 0.04	nd
m46	Phenylacetaldehyde	13.67	91	nd	nd	0.03 $\pm$ 0.03	nd

detected in paste A. Eighteen volatiles were present in both paste B and the sliced fruit, although they are not all the same compounds.

PCA was applied to the data and PC1 and PC2 explained 45.9% and 30.9% of the total variability respectively. The score plot (Figure 4A) revealed the dramatic effect of sample processing on the pattern of volatile compound emission. Three completely separated groups could be easily identified that correspond to each type of processing method: (i) whole fruit, (ii) sliced fruit, and (iii) both tomato pastes. This result indicates how dependable the volatile compound profiles and composition are on the processing method used. The two methods applied to process the pastes (A and B) resulted in the most similar volatile patterns, which were discriminated by PC1 but not so clearly by PC2 (Figure 4A). Nevertheless, paste A produced the richest volatile profile, emitting 7 detectable compounds more than paste B. The most remarkable case is (*E*)-2-heptenal, which is relatively abundant in the first method but not detected in the latter. In any case, results reported here highlight the remarkable effect variations in the protocol used to prepare the sample have on the profiles of volatile compounds obtained.



**Figure 4.** Principal component analysis of volatile profiles obtained with different methods of sample processing: score plot (A) and loading plot (B). Observations corresponding to different replicates are joined with solid lines. Compound codes in B are as in Table 1.

The PCA loading plot (Figure 4B) permits to identify a series of compounds that are more characteristic of each sample cluster observed in the corresponding score plot. It turned out that compounds also tended to group together, which implies that each sampling method tends to produce a distinct, non-overlapping set of volatiles. PC1 discriminates both pastes A and B with respect to the other methods. Both pastes produced considerably higher levels of a set of short-chain fatty acid-derived volatiles: C<sub>6</sub> compounds (*Z*)-3-hexenal, hexanal and (*E*)-2-hexenal; C<sub>5</sub> compounds 1-penten-3-one, 1-penten-3-ol, 1-pentanol, (*E*)-2-pentenal and pentanal; and also C<sub>4</sub> compound butanol. Moreover, (*E*)-2-heptenal was detected exclusively in paste A, but not in paste B. The levels of most of these compounds, although markedly lower, were also induced in the sliced fruit samples when compared to the whole fruit. It has been reported that both biotic and abiotic stresses including physical damage induce the production of a variety of volatile compounds. In plant material containing intact living cells, such as the sliced fruit, the stress associated to cutting the fruit would activate gene responses and the biosynthesis of wound stress-related metabolites, including volatile compounds (Niinemets et al. 2013). Additionally, the intact enzymatic machinery in living cells could modify some of the compounds produced by neighbouring injured cells, as the conversion of wounded cell-produced (*Z*)-3-hexenal into the corresponding alcohol and acetyl ester by neighbouring intact cells (Matsui et al. 2012). In the fruit pastes such response would not take place, because this response requires maintaining homeostasis and this is not happening after homogenization. In the homogenized samples, gene expression is not operational and the only mechanisms altering the volatile composition would be either chemical or enzymatic involving preformed molecules. In fact, it has been described that the homogenization of the fruit would facilitate the contact between enzymes and substrates otherwise localized in different cellular compartments in the living

cells, and this would be responsible for the burst of many volatiles, including many fatty acid derivatives (the so-called *green leaf volatiles*) and several phenylpropanoids (Chen et al. 2004; Granell and Rambla 2013; Shen et al. 2014). The more complete the homogenization (in the paste much more than in the sliced fruit), the higher the production of these compounds. Our results indicate that the latter process rather than the first has a major quantitative effect on the volatile profile obtained (Figure 4, Table 1).

Although the volatile profiles produced by both tomato pastes produced the most similar volatile patterns (Figure 4A), the particular way the tomato paste was produced also had a substantial effect on the volatile profile. Interestingly, paste A (first homogenized and incubated, then frozen) allowed the detection of more compounds than paste B (first frozen, then homogenized and finally thawed and incubated). Octanal, nonanal, (*E*)-2-heptenal, 2-methylpropanal, 6-methyl-5-hepten-2-one, phenylacetaldehyde and 2-isobutylthiazole were detected in tomato paste A but not in paste B. This probably indicates that most volatiles detected in tomato samples are mainly produced when precursors and biosynthetic enzymes of those volatile compounds meet each other after tissue disruption, as it has been documented for several biosynthetic pathways, either by *de novo* biosynthesis (Chen et al. 2004; Shen et al. 2014) or by the release of volatile aglycones accumulated as conjugates (Tikunov et al. 2013). After tissue disruption, volatile analysis protocols usually include some time of incubation so that detectable levels of the volatiles are produced. In the case of tomato paste, a high amount of calcium chloride is usually added after a few minutes of incubation in order to inhibit further reactions and stabilize the volatile profile during the time of analysis. In tomato paste A, incubation took place immediately after homogenization of the fresh tomato. In paste B, the sample was flash frozen

with liquid nitrogen before incubation. This would probably produce a partial or total inactivation of some of the participating enzymes, as has been previously reported (Díaz de León-Sánchez et al. 2009), causing some volatile compounds to either fall below detection levels or even not be present at all.

In accordance to this, the whole unprocessed fruit should produce a very poor volatile profile, which is precisely the case. Under our analytical conditions, only six of the previously identified compounds were present at detectable levels, most of which appear grouped together at the bottom right of the loading plot with the only exception of hexanal (Figure 4B). The whole fruit profile was basically composed of short branched-chain amino acid-related volatiles (2-methylpropanal, 2-methylbutanol, 3-methylbutanenitrile and 2-isobutylthiazole), one fatty acid derivative (hexanal), and also small amounts of  $\alpha$ -pinene, a monoterpene which often accumulates in glandular trichomes, that in tomato are present in both leaves and fruit (Schillmiller et al. 2010). Interestingly, 2-isobutylthiazole is a highly potent odorant and it recalls the smell of tomato leaves. We also have to consider that, in the whole fruit, diffusion of volatiles through the cuticle is extremely slow even for compounds as short as ethane, whilst resistance to diffusion through the sepals/nectary abscission scar is about three orders of magnitude lower (Cameron and Yang 1982). It is therefore likely that a sound tomato fruit needs to drop off the plant in order to release volatiles through the abscission scar.

Finally, the sliced fruit, when compared to the fruit pastes, produced considerably lower levels of many fatty acid derivatives including (*Z*)-3-hexenal and undetectable levels of (*E*)-2-hexenal, (*E*)-2-pentenal, 1-penten-3-one or 1-pentanol, but similar levels of others such as hexanal, pentanal or (*Z*)-3-hexenol. The loading plot (Figure 4B) indicates a group of 11 compounds that yielded the highest values with the sliced samples. Interestingly, higher levels of all

branched-chain amino acid-related volatiles such as 3-methylbutanal, 3-methylbutanol, 2-methylbutanal, 2-methylbutanol, 3-methylbutanenitrile, (*E*)-2-methyl-2-butenal, 2-methylpropanal, 2-methylpropanol, and 2-isobutylthiazole were produced, compared with the pastes. This observation suggests that branched-chain amino acid related compounds are readily produced in the intact fruit, and unlike observed for most fatty acid derivatives, fruit homogenization does not contribute to enhance their levels; on the contrary, their levels seem to be reduced after homogenization.

### **3.3 Possible consequences and limitations for metabolomic studies derived from the variability in VOCs introduced by processing and capture methods**

The comparison of these analytical methods revealed that both the sample processing and the technique used for the capture of volatiles have a dramatic effect on the compounds detected and their abundance in a particular sample. This fact, together with the wide range of biological variation observed among tomato cultivars (Tieman et al. 2012; Rambla et al. 2014), explains the high degree of variability in the abundance of volatile compounds in tomato fruit reported in the literature.

There is no single methodology that could be claimed as 'the best', but the comparative results here described can be used as a guide to select the most suitable approach for a particular experiment, depending on its objective or technical limitations. Regarding sampling processing, paste A yielded the highest number of detected compounds. Paste B produced a lower number of detectable volatiles, but it has the advantage that the same flash frozen material used for the volatile analysis can be stored at -80C until further use or shipped to other labs for determination of other metabolites, proteomic

analysis or even gene expression analysis, and therefore it is compatible with a multi-omics approach of a given biological sample. Sliced and whole fruit procedures have the disadvantage that the analysis requires to be performed on the fresh sample, and a complex set up is needed when parallel simultaneous acquisition of volatiles from a high number of samples is required, such as when profiling breeding collections which are highly dependent on harvest time (Tieman et al. 2012). Nevertheless, sliced fruit samples produce a reasonably rich volatile profile, with particularly high levels of short branched-chain compounds, and therefore this sample processing method has been useful for a large number of studies (Tieman et al. 2006a, 2006b, 2007, 2010, 2012; Mathieu et al. 2009; Vogel et al. 2010; Goulet et al. 2012; Mageroy et al. 2012; Shen et al. 2014). In fact, this sample processing technique is particularly useful when these volatiles are of particular interest, since it allows the detection of a higher number of such compounds. The whole fruit produced a very poor profile, but it could be useful in assays to determine how odour-attractive intact fruits are for seed dispersers. Regarding the technique used for volatile acquisition, each of the evaluated techniques has its own advantages, although SPME was the one that provided the largest number of compounds with a very low requirement of sample (only 1 g).

Considering that the high degree of variation in the volatile profiles depends largely on the method of analysis used, a concern arises about the limitations of each analytical study. The quantitative results obtained by different research groups using a range of methodological conditions cannot be directly compared. For example, the quantitative amount of volatiles of a particular variety analyzed, *e.g.* from a paste, cannot be directly compared to results obtained from another cultivar analyzed in a different laboratory from *e.g.* sliced fruit, because a very relevant part of the differences observed would



not be due to the effect of the variety, but to that of the sampling procedure. Obviously, this does not mean that the analyses of volatile compounds in tomatoes are unreliable. Far from that, the comparison of different sets of samples using the same procedure has produced fruitful results, including the effect of different treatments to the volatile profile (Baldwin et al. 1991; MacDonald et al. 1996; Maul et al. 1998, 2000; Cebolla-Cornejo et al. 2011; Renard et al. 2013), the identification of genomic regions responsible of particular traits (Causse et al. 2002; Tadmor et al. 2002; Tieman et al. 2006a; Mathieu et al. 2009; Zanor et al. 2009) or the identification of genes implied in biosynthetic pathways (Chen et al. 2004; Simkin et al. 2004; Tieman et al. 2006b, 2007, 2010; Goulet et al. 2012; Mageroy et al. 2012; Tikunov et al. 2013). Thus, the identification of the QTL and gene underlying the accumulation of smoky flavour in tomatoes resulted in the cloning of NSGT1; this hallmark in the way volatiles are kept / released by higher level glycosilation could have been more difficult to be unveiled should the authors had used whole or sliced fruits, since incubation of the extracts and activity of the glycosidases to release the volatile was consubstantial with its discovery. Therefore, a sample treatment which involved homogenization of the tissue and allowed the glycosidases in the extract to act on the non-volatile glycoside form so as to liberate or not the volatile was necessary (Tikunov et al. 2010, 2013). In summary, the identification of gene/gene products depends on the volatile profile obtained during the screening methods and this, as we have demonstrated, is highly influenced by the capture and processing method. Therefore, the identification of genes involved in volatile biosynthesis (Chen et al. 2004; Simkin et al. 2004; Tieman et al. 2006b, 2007, 2010; Goulet et al. 2012; Mageroy et al. 2012; Tikunov et al. 2013) has been so far highly dependent on the conditions the volatiles were released and analyzed. Our results obtained here indicate that each methodological approach has its *pros* and *cons*, and great care should be

taken when trying to compare different data from the literature, as they have been obtained under different methodological conditions.

The present study reveals that the 'quantitative' results obtained for the fruit of a given variety of tomato when analyzed by means of a particular technique may in some way be more representative of the sampling procedure utilized than on the variety itself, as a consequence of the strong effect of the method used on the volatile profile obtained. This issue has important implications when trying to translate the quantitative results (profile of volatile compounds) obtained for each volatile compound in a cultivar in terms of flavour and aroma. The most widely accepted approach to characterize tomato aroma is based on odour units, although it is a simplistic approach with rather limitations, as discussed recently (Tieman et al. 2012; Rambla et al. 2014). Basically, a threshold of human perception is determined for each volatile compound dissolved in water, and then compared to the quantitative results obtained from a tomato sample. In theory, only those compounds produced at levels above the threshold would participate in our perception of flavour and aroma. The odour units currently reported for tomato are based on red ripe tomato fruit samples processed exactly as paste A. Buttery (1993) quantified 16 compounds over the respective threshold levels, and they were considered thereafter as responsible of tomato flavour and aroma. Our results suggest that a different list of compounds would be obtained if a different sampling procedure (such as paste B or the sliced fruit) was used. Therefore, depending on the precise way how the tomato fruit is manipulated, different amounts of each volatile compound will be released, either above or below our detection threshold. An important difficulty to relate the volatile content with consumer liking is that none of the sampling procedures used up to date resembles much the 'procedure' that takes place in the consumer's mouth (chewing for a short

time, insalivating, heating at mouth temperature...). In this sense, new sampling procedures have been recently developed (Farneti et al. 2013) with the objective to obtain less artifactual analytical results, which would be more easily translatable to our perception of tomato flavour.

## 4 CONCLUDING REMARKS

Our results clearly show that both the process of sample preparation and the technique used for capturing the volatiles have a dramatic effect on the volatile profiles obtained, and they contribute to the wide range of variability in volatile profiles reported in the literature for tomato fruit. Different protocols provide different views of the volatile content which are not readily comparable, therefore suggesting that although each method can be suitable for a specific purpose, great care should be taken when comparing results between experiments using different volatile technologies, or when using the resulting volatile levels to predict the influence of particular compounds on our perception of tomato flavour or on consumer preference. Although each technique has its own *pros* and *cons*, a sample processing method starting from frozen material such as paste B would be the most adequate from an *-omics* perspective, as the same sample can be used for the different types of analysis. Concerning the capture method, headspace solid phase microextraction yields the highest number of volatile compounds even from a small amount of sample. The combination of these two techniques would probably be the most adequate for a multi-*omics* approach.

## Acknowledgements

We thank Rafael Fernández for providing excellent tomato fruits for this study. Funding to AG was provided through CALITOM and ESPSOL from FECYT and EUSOL (EU FP7 program) and Mineco.

## REFERENCES

Allwood, J.W., Cheung, W., Xu, Y., et al. (2014) Metabolomics in melon: a new opportunity for aroma analysis. *Phytochemistry*, *99*, 61-72.

Aubert, C., Baumann, S. & Arguel, H. (2005) Optimization of the analysis of flavor volatile compounds by liquid-liquid microextraction (LLME). Application to the aroma analysis of melons, peaches, grapes, strawberries, and tomatoes. *Journal of Agricultural and Food Chemistry*, *53*, 8881-8895.

Baldwin, E.A., Nisperos-Carriedo, M.O. & Moshonas, M.G. (1991) Quantitative analysis of flavor and other volatiles and for certain constituents of 2 tomato cultivars during ripening. *Journal of the American Society for Horticultural Science*, *116*, 265-269.

Baldwin, E.A., Scott, J.W., Einstein, M.A., et al. (1998) Relationship between sensory and instrumental analysis for tomato flavor. *Journal of the American Society for Horticultural Science*, *123*, 906-915.

Baldwin, E.A., Goodner, K., Plotto, A. & Einstein, M. (2004) Effect of volatiles and their concentration on perception of tomato descriptors. *Journal of Food Science*, *69*, S310-S318.

Baldwin, E.A., Goodner, K. & Plotto, A. (2008) Interaction of Volatiles, Sugars, and Acids on Perception of Tomato Aroma and Flavor Descriptors. *Journal of Food Science*, *73*, S294-S307.

Beltran, J., Serrano, E., Lopez, F.J., Peruga, A., Valcarcel, M. & Rosello, S. (2006) Comparison of two quantitative GC-MS methods for analysis of tomato aroma based on purge-and-trap and on solid-phase microextraction. *Analytical and Bioanalytical Chemistry*, *385*, 1255-1264.

Buttery, R.G., Teranishi, R. & Ling, L.C. (1987) Fresh tomato aroma volatiles: a quantitative study. *Journal of Agricultural and Food Chemistry*, *35*, 540-544.

Buttery, R.G., Teranishi, R., Ling, L.C., Flath, R.A. & Stern, D.J. (1988) Quantitative Studies on Origins of Fresh Tomato Aroma Volatiles. *Journal of Agricultural and Food Chemistry*, *36*, 1247-1250.

Buttery, R.G. (1993) Quantitative and sensory aspects of flavor of tomato and other vegetables and fruits. In: Acree T.E., Teranishi, R. (Eds), *Flavor Science: Sensible Principles and Techniques*. American Chemical Society: Washington, DC, pp. 259-286.

Cameron, A.C. & Yang, S.F. (1982) A simple method for the determination of resistance to gas diffusion in plant organs. *Plant Physiology*, *70*, 21-23.

Carbonell-Barrachina, A.A., Agustí, A. & Ruiz, J.J. (2006) Analysis of flavor volatile compounds by dynamic headspace in traditional and hybrid cultivars of Spanish tomatoes. *European Food Research and Technology*, *222*, 536-542.

Causse, M., Saliba-Colombani, V., Lecomte, L., Duffe, P., Rousselle, P. & Buret, M. (2002) QTL analysis of fruit quality in freshmarket tomato: a few chromosome regions control the variation of sensory and instrumental traits. *Journal of Experimental Botany*, *53*, 2089-2098.

Cebolla-Cornejo, J., Roselló, S., Valcárcel, M., Serrano, E., Beltrán, J. & Nuez, F. (2011) Evaluation of genotype and environment effects on taste and aroma flavor components of Spanish fresh tomato varieties. *Journal of Agricultural and Food Chemistry*, *59*, 2440-2450.

Chen, G.P., Hackett, R., Walker, D., Taylor, A., Lin, Z. & Grierson, D. (2004) Identification of a specific isoform of tomato lipoxygenase (TomloxC) involved in the generation of fatty acid-derived flavor compounds. *Plant Physiology*, *136*, 2641-2651.

Díaz de León-Sánchez, F., Pelayo-Zaldívar, C., Rivera-Cabrera, F., et al. (2009) Effect of refrigerated storage on aroma and alcohol dehydrogenase activity in tomato fruit. *Postharvest Biology and Technology*, *54*, 93-100.

Farneti, B., Alarcon, A.A., Cristescu, S.M., et al. (2013) Aroma volatile release kinetics of tomato genotypes measured by PTR-MS following artificial chewing. *Food Research International*, *54*, 1579-1588.

Gamero, A., Wesselink, W. & de Jong, C. (2013) Comparison of the sensitivity of different aroma extraction techniques in combination with gas

chromatography-mass spectrometry to detect minor aroma compounds in wine. *Journal of Chromatography A*, 1272, 1-7.

Goff, S.A. & Klee, H.J. (2006) Plant volatile compounds: Sensory cues for health and nutritional value? *Science*, 311, 815-819.

Goulet, C., Mageroy, M.H., Lam, N.B., Floystad, A. Tieman, D.M. & Klee, H.J. (2012) Role of an esterase in flavour volatile variation within the tomato clade. *Proceedings of the National Academy of Sciences, USA*, 109, 19009-19014.

Granell, A. & Rambla, J.L. (2013) Biosynthesis of volatile compounds. In: Seymour, G., Tucker, G.A., Poole, M. & Giovannoni, J.J. (Eds), *The Molecular Biology and Biochemistry of Fruit Ripening*, Wiley-Blackwell, Oxford, UK, pp. 135-161 doi: 10.1002/9781118593714.

Jaeger, S.R., McRae, J.F., Bava, C.M. et al. (2013) A Mendelian trait for olfactory sensitivity affects odor experience and food selection. *Current biology*, 23, 1601-1605.

McDonald, R.E., McCollum, T.G. & Baldwin, E.A. (1996) Prestorage heat treatments influence free sterols and flavor volatiles stored at chilling temperature. *Journal of the American Society for Horticultural Science*, 121, 531-536.

Mageroy, M.H., Tieman, D.M., Floystad, A., Taylor, M.G. & Klee, H.J. (2012) A *Solanum lycopersicum* catechol-O-methyltransferase involved in synthesis of the flavor molecule guaiacol. *Plant Journal*, 69, 1043-1051.

Mathieu, S., Dal Cin, V., Fei, Z., et al. (2009) Flavour compounds in tomato fruits: identification of loci and potential pathways affecting volatile composition. *Journal of Experimental Botany*, 60, 325-337.

Matsui, K., Sugimoto, K., Mano, J., Ozawa, R. & Takabayashi, J. (2012) Differential metabolisms of green leaf volatiles in injured and intact parts of a wounded leaf meet distinct ecophysiological requirements. *PLOS ONE*, 7, doi: 10.1371/journal.pone.0036433.

Maul, F., Sargent, S.A., Balaban, M.O., Baldwin, E.A., Huber, D.J. & Sims, C.A. (1998) Aroma volatile profiles from ripe tomatoes are influenced by

physiological maturity at harvest: an application for electronic nose technology. *Journal of the American Society for Horticultural Science*, 123, 1094-1101.

Maul, F., Sargent, S.A., Sims, C.A., Baldwin, E.A., Balaban, M.O. & Huber, D.J. (2000) Tomato flavor and aroma quality as affected by storage temperature. *Journal of Food Science*, 65, 1228-1237.

McRae, J.F., Jaeger, S.R., Bava, C.M. et al. (2013) Identification of regions associated with variation in sensitivity to food-related odors in the human genome. *Current biology*, 23, 1596-1600.

Niinemets, U., Kannaste, A. & Copolovici, L. (2013) Quantitative patterns between plant volatile emissions induced by biotic stresses and the degree of damage. *Frontiers in Plant Science*, 4, doi: 10.3389/fpls.2013.00262.

Nongonierma, A., Cayot, P., Le Quere, J.L., Springett, M. & Voilley, A. (2006) Mechanisms of extraction of aroma compounds from foods, using adsorbents. Effect of various parameters. *Food Reviews International*, 22, 51-94.

Ortiz-Serrano, P. & Gil, J.V. (2010) Quantitative comparison of free and bound volatiles of two commercial tomato cultivars (*Solanum lycopersicum* L.) during ripening. *Journal of Agricultural and Food Chemistry*, 58, 1106-1114

Petró-Turza, M. (1987) Flavor of tomato and tomato products. *Food Reviews International*, 2, 309-351.

Rambla, J.L., Tikunov, Y.M., Monforte, A.J., Bovy, A.G. & Granel A. (2014) The expanded tomato fruit volatile landscape. *Journal of Experimental Botany*, 65, 4613-4623.

Renard, C.M.G.C., Ginies, C., Gouble, B., Bureau, S. & Causse, M. (2013) Home conservation strategies for tomato (*Solanum lycopersicum*): Storage temperature vs. duration – Is there a compromise for better aroma preservation? *Food Chemistry*, 139, 825-836.

Ruiz, J.J., Alonso, A., García-Martínez, S., Valero, M., Blasco, P. & Ruiz-Bevia, F. (2005) Quantitative analysis of flavour volatiles detects differences among closely related traditional cultivars of tomato. *Journal of the Science of Food and Agriculture*, 85, 54-60.



Shen, J.Y., Tieman, D., Jones, J. B., et al. (2014) A 13-lipoxygenase, TomloxC, is essential for synthesis of C5 flavour volatiles in tomato. *Journal of Experimental Botany*, 65, 419-428.

Selli, S., Kelebek, H., Ayseli, M.T. & Tokbas, H. (2014) Characterization of the most aroma-active compounds in cherry tomato by application of the aroma extract dilution analysis. *Food Chemistry*, 165, 540-546.

Schillmiller A., Shi, F., Kim, J., et al. (2010) Mass spectrometry screening reveals widespread diversity in trichome specialized metabolites of tomato chromosomal substitution lines. *Plant Journal*, 62, 391-403.

Simkin, A.J., Schwartz, S.H., Aldridge, M., Taylor, M.G. & Klee, H.J. (2004) The tomato carotenoid cleavage dioxygenase 1 genes contribute to the formation of the flavor volatiles beta-ionone, pseudoionone, and geranylacetone. *Plant Journal*, 40, 882-892.

Tadmor, Y., Fridman, E., Gur, A., et al. (2002) Identification of malodorous, a wild species allele affecting tomato aroma that was selected against during domestication. *Journal of Agricultural and Food Chemistry*, 50, 2005-2009.

Tandon, K.S., Baldwin, E.A., Scott, J.W. & Shewfelt, R.L. (2003) Linking sensory descriptors to volatile and nonvolatile components of fresh tomato flavor. *Journal of Food Science*, 68, 2366-2371.

Tieman, D.M., Zeigler, M., Schmelz, E.A., et al. (2006a) Identification of loci affecting flavour volatile emissions in tomato fruits. *Journal of Experimental Botany*, 57, 887-896.

Tieman, D., Taylor, M., Schauer, N., Fernie, A.R., Hanson, A.D. & Klee, H.J. (2006b) Tomato aromatic amino acid decarboxylases participate in synthesis of the flavour volatiles 2-phenylethanol and 2-phenylacetaldehyde. *Proceedings of the National Academy of Sciences, USA*, 103, 8287-8292.

Tieman, D.M., Loucas, H.M., Kim, J.Y., Clark, D.G. & Klee, H.J. (2007) Tomato phenylacetaldehyde reductases catalyze the last step in the synthesis of the aroma volatile 2-phenylethanol. *Phytochemistry*, 68, 2660-2669.

Tieman, D., Zeigler, M., Schmelz, E., et al. (2010) Functional analysis of a tomato salicylic acid methyl transferase and its role in synthesis of the flavor volatile methyl salicylate. *Plant Journal*, 62, 113-123.

Tieman, D., Bliss, P. & McIntyre, L.M. (2012) The chemical interactions underlying tomato flavour preference. *Current Biology*, 22, 1035-1039.

Tikunov, Y., Lommen, A., De Vos, C.H.R., et al. (2005) A novel approach for nontargeted data analysis for metabolomics. Large-scale profiling of tomato fruit volatiles. *Plant Physiology*, 139, 1125-1137.

Tikunov, Y.M., de Vos, R.C.H., Gonzalez-Paramas, A.M.G., Hall, R.D. & Bovy, A.G. (2010) A role for differential glycoconjugation in the emission of phenylpropanoid Volatiles from tomato fruit discovered using a metabolic data fusion approach. *Physiology*, 152, 55-70.

Tikunov, Y.M., Molthoff, J., de Vos, R.C., et al. (2013) NON-SMOKY GLYCOSYLTRANSFERASE1 prevents the release of smoky aroma from tomato fruit. *Plant Cell*, 25, 3067-3078.

Vogel, J.T, Tieman, D.M., Sims, C.A., Odabasi, A.Z., Clark, D.G. & Klee, H.J. (2010) Carotenoid content impacts flavour acceptability in tomato (*Solanum lycopersicum*). *Journal of the Science of Food and Agriculture*, 90, 2233-2240.

Zanor, M.I., Rambla, J.L., Chaib, J., et al. (2009) Metabolic characterization of loci affecting sensory attributes in tomato allows an assessment of the influence of the levels of primary metabolites and volatile organic contents. *Journal of Experimental Botany*, 60, 2139-2154.





## **Chapter 2**

### **Metabolic characterisation of loci affecting sensory attributes in tomato allows an assessment of the influence of the levels of primary metabolites and volatile organic contents**

María Inés Zanon, José Luis Rambla, Jamila Chaïb, Agnes Steppa, Aurora Medina, Antonio Granell, Alisdair R. Fernie, Mathilde Causse



## **ABSTRACT**

Numerous studies have revealed the extent of genetic and phenotypic variation between both species and cultivars of tomato. Using a series of tomato lines resulting from crosses between a cherry tomato and three independent large fruit cultivar (Levovil, VilB, and VILD), extensive profiling of both central primary metabolism and volatile organic components of the fruit was performed. In this study, it was possible to define a number of quantitative trait loci (QTLs) which determined the levels of primary metabolites and/or volatile organic components and to evaluate their co-location with previously defined organoleptic QTLs. Correlation analyses between either the primary metabolites or the volatile organic compounds and organoleptic properties revealed a number of interesting associations, including pharmaceutical aroma–guaiacol and sourness–alanine, across the data set. Considerable correlation within the levels of primary metabolites or volatile organic compounds, respectively, were also observed. However, there was relatively little association between the levels of primary metabolites and volatile organic compounds, implying that they are not tightly linked to one another. A notable exception to this was the strong association between the levels of sucrose and those of a number of volatile organic compounds. The combined data presented here are thus discussed both with respect to those obtained recently from wide interspecific crosses of tomato and within the framework of current understanding of the chemical basis of fruit taste.

### **Key words**

Metabolite profiling, QTL sensory profiling, Tomato, Volatile profiling.





## INTRODUCTION

Human perception of flavour involves the integration of multiple signals emanating from taste and olfactory receptors. In tomato, as in most fruits, flavour is largely dependent on sugar and acid contents, but also on the sugar/acid ratio (Dennison *et al.*, 1953; Stevens 1972; Saliba-Colombani *et al.*, 2001). However, whilst taste receptors clearly respond to relatively few cues, olfactory receptors respond to thousands of chemicals and as such are thought to be responsible for the vast diversity of unique food flavours (Goff and Klee, 2006; Tieman *et al.*, 2006a). In the case of tomato fruits, 400 volatile organic compounds have been identified (Petro-Turza, 1987), between 15 and 20 of which are thought to constitute the flavour of fresh tomatoes (Buttery *et al.*, 1971; Baldwin *et al.*, 2000). These volatile compounds are generally derived from various precursors including fatty acids, carotenoids, and amino acids. However, the exact definition of the biosynthetic pathways of many of them remains elusive (Tieman *et al.*, 2006a). In addition to the chemical components of fruit quality, physical components related to texture are of crucial importance to the consumer (Causse *et al.*, 2003; Serrano-Megias and Lopez-Nicolas, 2006; Chaïb *et al.*, 2007). Fruit texture is composed of many traits including flesh firmness, mealiness, meltiness, juiciness, and crispness (Harker *et al.*, 1997; Redgwell and Fischer, 2002; Szczesniak, 2002). During fruit ripening, major changes in texture occur. Fruit softening has a major impact on many aspects of post-harvest physiology, including transport, shelf life, and disease resistance (Brummell and Harpster, 2001; Saladie *et al.*, 2006).

Given that consumers have complained about tomato flavour for >10 years in Europe (Decoene, 1995; Janse and Schols, 1995), the USA (De Giglio, 2003), and Australia (Ratanachinakorn *et al.*, 1997), much research attention has

focused on ways to improve it. As a first step in this process a number of surveys of natural variation in the chemical composition of tomatoes have been carried out either on the cultivar/species basis (Schauer *et al.*, 2005b; Spencer *et al.*, 2005; Tikunov *et al.*, 2005; Fernie *et al.*, 2006), or utilizing either recombinant inbred or introgression lines (Chaib *et al.*, 2006, 2007; Schauer *et al.*, 2006, 2008; Tieman *et al.*, 2006b; Hovav *et al.*, 2007). Several of these studies have identified genomic loci controlling the levels either of sugars and organic acids or of volatiles (Saliba-Colombani *et al.*, 2001; Causse *et al.*, 2002; Tieman *et al.*, 2006b; Schauer *et al.*, 2006, 2008), whilst other studies have concentrated on more physical aspects of organoleptic quality (Lecomte *et al.*, 2004; Chaïb *et al.*, 2007). In the current study, the metabolite composition of quantitative trait loci near isogenic lines (QTL-NILs) that had previously been demonstrated, by use of a trained tasting panel, to possess characteristic organoleptic properties (Chaïb *et al.*, 2006) were evaluated. For this purpose, both polar primary metabolites and volatile organic compounds in the lines were evaluated using well-established GC-MS-based profiling methods for each type of compound. In total, the levels of 100 metabolites were determined and it was possible to evaluate co-localization and correlation of changes in these metabolic traits with changes in the previously determined organoleptic traits. Data are discussed with respect to current models of determinants of fruit organoleptic quality and its underlying molecular basis.

## MATERIALS AND METHODS

### *Plant material*

The experiments were performed on parental lines and two types of introgressed lines in different genetic backgrounds: genotypes combining five regions of interest for fruit quality and QTL-NILs carrying one introgressed region of chromosome 1, 2, 4, and 9 (two regions 9A and 9B). The five regions carried several QTLs involved in fruit quality (see Fig. 3, Causse *et al.*, 2002). The initial QTL analysis was performed on a population of recombinant inbred lines (RILs) developed from an intraspecific cross between Cervil (a cherry tomato, *Solanum lycopersicum*, var. *cerasiforme*) with 7 g fruits, a good taste, and a high aroma intensity, and Levovil (a *S. lycopersicum* line) with 125 g fruits and a common taste (Causse *et al.*, 2002). Based on the QTL map, five regions (located on chromosomes 1, 2, 4, and 9, respectively) were introgressed in the Levovil genetic background. A QTL for titratable acidity was detected in region 1, QTLs for sweetness, tomato aroma intensity, mealiness, and meltiness were detected in region 2, a QTL for mealiness and several QTLs for volatiles were detected in region 4, QTLs for sourness, tomato aroma intensity, mealiness, meltiness, and flesh firmness were detected in region 9A, and a QTL for pharmaceutical aroma was detected in region 9B. QTLs for physical and chemical traits were also detected in these regions. The introgressed lines were produced as described in Chaib *et al.* (2006). Briefly, as the favourable alleles for fruit quality were conferred by the C parent in most of the cases, the cherry tomato alleles at the five regions were introgressed into large fruit genotypes in order to obtain QTL-NILs. A single RIL with C alleles at the five regions was used as the donor parent of the breeding programme. The same marker-assisted backcross programme was performed with three different recipient lines, kindly provided by Vilmorin: Levovil, VilB, and VilD, hereafter L, B and D, respectively. As the donor parent

contained 47% of recipient genome L, the first cross with each recipient line was considered as a BC1. The BC1 progeny was genetically homogenous; it was thus backcrossed without any selection to the recipient line to produce a BC2 population. Almost 300 plants were grown for each background, and, after a marker-assisted selection step, one BC2 individual was selected and backcrossed again to produce a BC3 population. Similarly, one BC3 individual was selected and three selfing generations were performed. In each BC3S1 population, the segregation of markers in the five regions of interest was comparable with that of an F<sub>2</sub> population. Then, BC3S3 lines with homozygous alleles at the five regions were selected and BC3S3 lines carrying C alleles at a single introgressed region were evaluated. These lines were nearly isogenic to their recipient line and were thus called QTL-NILs (Van Berloo *et al.*, 2001). The QTL-NILs were named with a letter corresponding to their genetic background and a number for the QTL region carried. For example, the line carrying the C allele at the region of interest on chromosome 2 with a genetic background L was denoted L2. In each genetic background, a line was obtained for each QTL region, with the exception of NIL-B9A that contained a C fragment introgressed on chromosome 1. The lines combining the five regions in the Levovil and VilB genetic background were named Lx and Bx, respectively.

### ***Plant growth conditions trials***

Three trials were performed during spring 2004, 2005, and 2006 in a heated glasshouse in Avignon (France, 43°55'N; 4°52'E). Planting took place on February at a density of 3.2 plants m<sup>-2</sup>, and the day–night temperature set-point was 24–16 °C. Plant nutrition and chemical pest and disease control followed commercial practices and plants were grown on a single vine. From anthesis of the first truss, flowers were pollinated with an electrical shaker every 2–3 d. In each trial, the parental lines, the lines combining the five regions, and the QTL-

NILs in the three genetic backgrounds were grown. Each line was represented by six plants grown in a fully randomized design. Several types of analyses were performed on red ripe tomatoes: physical measurements, sensory profiling, metabolic profiling, and volatile profiling.

### ***Physical and physiological measurements***

Red ripe fruits were harvested on the six plants of each line twice a week for 6 weeks. For metabolic profiles, six fruits per line were peeled and pericarp maintained frozen at  $-80^{\circ}\text{C}$ . For volatiles another six fruits per line were used and sections of the fruit were stored at  $-80^{\circ}\text{C}$  until further use.

### ***Sensory profiling***

Sensory profiles were obtained in 2004. Red ripe tomatoes were harvested in the morning of the day of the tasting, and homogeneous fruit samples were selected and stored at  $20^{\circ}\text{C}$  in an air-conditioned room. The sensory panel was composed of 15 judges, who had previously been trained in the quantitative description of tomato attributes according to selection trials based on French norms (ISO8586-1, AFNOR V09-003). For each line, fruits were tasted twice by each judge, giving 30 scores per genotype. Fifteen sessions took place in a sensory analysis laboratory (AFNOR norm V09-105), on 2 d per week, and eight fruits were tasted by each judge on each occasion. The attributes chosen were colour intensity and heterogeneity, ribbed and translucent fruit intensity, to describe aspect, typical odour, sourness and sweetness, metal aroma, global aroma intensity, typical tomato aroma, pharmaceutical aroma, and firmness, juiciness, fleshiness, mealiness, and embarrassing skin to describe fruit texture. Each descriptor was scored on a 10-point scale.

### **Primary metabolite analysis**

The relative levels of metabolites were determined using the GC-MS protocol exactly as described in Liseč *et al.* (2006) with the exceptions that the method was optimized for tomato fruit (Schauer *et al.*, 2006) and the mass spectra were cross-referenced with those in the Golm Metabolome Database (Kopka *et al.*, 2005; Schauer *et al.*, 2005a). The absolute concentrations of several metabolites were determined by comparison with calibration standard curve response ratios of various concentrations of standard substance solutions, including the internal standard ribitol (Roessner-Tunali *et al.*, 2003).

### **Volatile analysis**

Fruit volatile analysis was performed essentially as described in Tikunov *et al.* (2005), with minor variations. Frozen tomato samples were milled in liquid nitrogen. A 1 g aliquot of the frozen fruit powder was weighed in a 7 ml vial, and the vial was sealed, and incubated at 37 °C for 10 min. An EDTA-NaOH water solution was prepared by adjusting 100 mM EDTA to a pH of 7.5 with NaOH. Then 1 ml of the EDTA-NaOH solution was added to the sample to a final EDTA concentration of 50 mM. A 2.2 g aliquot of solid CaCl<sub>2</sub>·2H<sub>2</sub>O was then immediately added. The closed vials were agitated and sonicated for 5 min. A 1 ml aliquot of the pulp was transferred into a 22 ml crimp cap vial (Perkin-Elmer), capped, and used for HS-SPME-GC-MS analysis. The vials were tempered at 50 °C for 10 min. The volatiles were then extracted by exposing a 65 μm polydimethylsiloxane-divinylbenzene SPME fibre (Supelco) to the vial headspace for 20 min under continuous agitation and heating at 50 °C. The fibre was manually inserted into a Clarus 500 (Perkin-Elmer) injection port and volatiles were desorbed for 1 min at 250 °C. Chromatography was performed on a ZB-5 (30 m×0.25 mm×0.25 μm) column with helium as carrier gas, at a constant flow

of 1.2 ml min<sup>-1</sup>. The GC interface and MS source temperatures were 260 °C and 180 °C, respectively. The oven programming conditions were 40 °C for 2 min, 5 °C min<sup>-1</sup> ramp until 180 °C, then a 15 °C min<sup>-1</sup> ramp until 250 °C, and a final hold at 250 °C for 4 min. The total run time, including oven cooling, was 60 min. Mass spectra in the 35–250 *m/z* range were recorded by a Clarus 500 electron impact MS (Perkin-Elmer) at a scanning speed of five scans s<sup>-1</sup> and an ionization energy of 70 eV. The chromatography and spectral data were evaluated using TurboMass software version 5.0 (Perkin-Elmer).

### ***Data analysis***

Statistical analyses were performed using either R statistical software or Microsoft Excel 7.0 (Microsoft, 2000). If two observations are described as different this means that their difference was determined to be statistically significant ( $P < 0.05$ ) by the performance of Student's *t*-tests. The QTLs were evaluated by using Student's *t*-tests at a significance threshold of 0.05 to compare statistically each trait of each introgression line with its respective reference control. Principal component analysis was performed by means of SIMCA-P 11 software (Umetrics). Pearson correlation coefficients were calculated using the embedded CORREL function in Microsoft Excel 7.0 (Microsoft, 2000).

### ***Heat map***

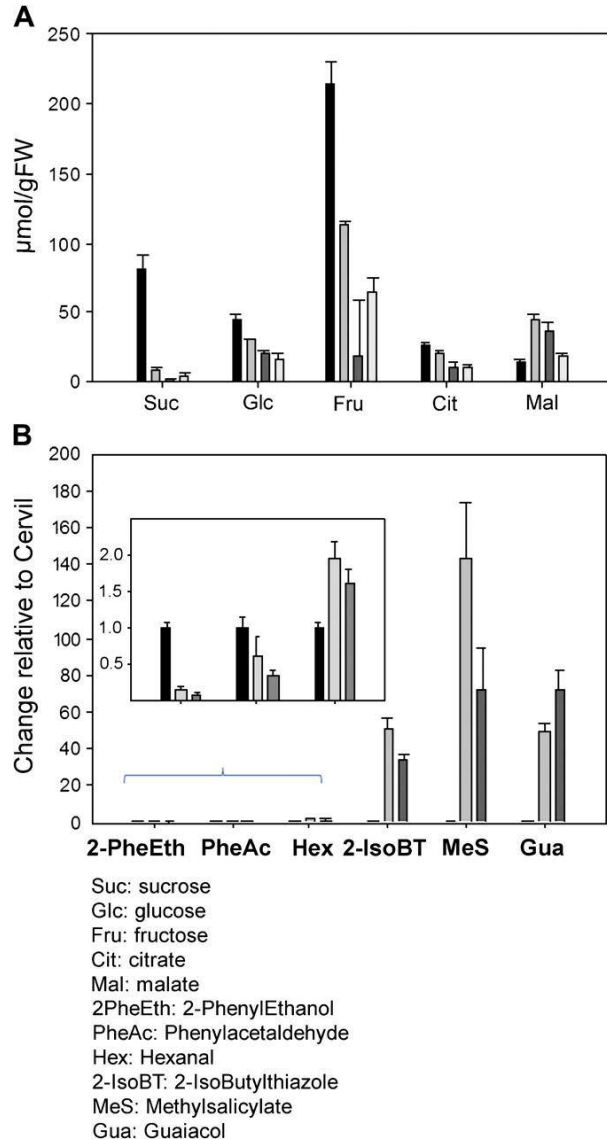
Heat maps were calculated using the 'heatmap' module of the statistical software environment R (<http://www.r-project.org>) version 1.9. False colour imaging was performed on the log<sub>2</sub>-transformed data. Regions of red and blue indicate negative or positive correlation between traits as depicted in the reference colour bar.

## RESULTS

### *Elite tomato lines harbour clear metabolic differences*

Given that both previous sensory profiling results (Saliba-Colombani *et al.*, 2001; Causse *et al.*, 2003; Lecomte *et al.*, 2004) and common perception suggest that the cherry tomatoes are tastier than the large-fruited tomatoes, it was decided to analyse the basis of these differences at the metabolic level. For this purpose, an established GC-MS-based metabolite profiling method (Fernie *et al.*, 2004; Lisec *et al.*, 2006) was applied to the four parental lines used in this study [the cherry tomato line Cervil (C) and the large-fruited lines Levovil (L), ViLB (B), and ViLD (D)]. This analysis revealed profound differences between the lines in the levels of several metabolites. The initial focus was on the major sugar and acid contents (Fig. 1A). As could be anticipated, there were huge differences in sugar and acid levels between the three elite lines and the cherry tomato line, with the latter displaying greater levels of the major soluble sugars (sucrose, glucose, and fructose) whilst the larger fruited tomatoes had higher levels of malate and lower levels of citrate. In line with this observation, the sugar/acid ratio of the parental lines (calculated as  $\mu\text{mol gFW}^{-1}$  of sucrose, glucose, and fructose versus  $\mu\text{mol gFW}^{-1}$  of citrate and malate) was highest in the cherry variety (8.5) and lowest in the L variety (L=0.9; B=2.4; D=3.2). A more detailed analysis of the metabolite profiles of the parental lines revealed that many other metabolites were present at significantly different levels between the lines. One-way analysis of variance (ANOVA) tests revealed additional significant differences in the abundance of maltose, trehalose, arabinose, xylose, rhamnase, ribose, isocitrate, citramalate, malate,  $\alpha$ -ketoglutarate, proline, valine, alanine,  $\beta$ -alanine, glutamate, serine, threonine, and phenylalanine between the parental lines. These data are presented in Table 1A which shows the fold changes observed in the levels of primary metabolites between each of the large-fruited





**Figure 1.** Metabolic analysis of the parental lines. (A) Quantitative determination of the concentration of selected primary metabolites: sucrose, glucose, fructose, malate, and citrate in samples harvested in 2004. Cervil (black bars), ViIB (light grey), Levovil (dark grey bars), and ViID (very light grey bars). Each bar represents the mean  $\pm$ SE of six independent biological determinations. (B) Relative changes of 2-phenylethanol, phenylacetaldehyde, hexanal, 2-isobutylthiazole, methyl salicylate, and guaiacol in samples harvested in 2005. Cervil (black bars), ViIB (light grey), and Levovil (dark grey). Data are normalized to the mean response for Cervil. Each bar represents the mean  $\pm$ SE of five biologically independent replicates.

**Table 1A.** Fold changes in the primary metabolites relative to Cervil in the parental lines.

Data are normalized to the mean response calculated for the Cervil line. Values are presented as the mean of six biologically independent determinations. Those metabolites which were significantly different from the Cervil line ( $P < 0.05$ ) by the performance of Student's *t*-tests are marked in bold. nd indicates that metabolites were not detected.

	<b>VilB</b>	<b>Levovil</b>	<b>VilD</b>		<b>VilB</b>	<b>Levovil</b>	<b>VilD</b>
Alanine	<b>0.17</b>	<b>0.25</b>	0.28	Trehalose	<b>0.53</b>	<b>0.21</b>	<b>0.27</b>
β-Alanine	<b>2.65</b>	0.33	1.83	Xylose	<b>0.69</b>	<b>2.54</b>	<b>0.38</b>
Arginine/ornithine	0.86	<b>0.31</b>	0.71	Glycerol	0.73	<b>0.12</b>	<b>0.47</b>
Asparagine	<b>0.50</b>	<b>0.18</b>	<b>0.26</b>	Myo-inositol	0.88	<b>0.33</b>	<b>0.35</b>
Aspartate	1.52	<b>0.32</b>	<b>0.26</b>	α-Ketoglutarate	<b>0.06</b>	<b>0.01</b>	<b>0.04</b>
γ-Aminobutyrate	<b>2.91</b>	0.84	1.34	Benzoate	0.87	0.58	<b>0.55</b>
Glutamate	<b>0.34</b>	<b>0.15</b>	<b>0.20</b>	Citramalate	<b>0.70</b>	<b>0.48</b>	<b>0.48</b>
Glutamine	0.40	0.46	<b>0.10</b>	Citrate	<b>0.73</b>	<b>0.39</b>	<b>0.35</b>
Homoserine	1.01	0.47	0.64	Dehydroescorbate	<b>0.33</b>	<b>0.38</b>	<b>0.20</b>
Isoleucine	1.63	0.51	0.66	Fumarate	1.53	<b>1.76</b>	0.88
Lysine	<b>0.05</b>	0.68	<b>0.05</b>	Galacturonate	<b>0.18</b>	1.14	<b>0.26</b>
Phenylalanine	0.89	<b>0.44</b>	<b>0.29</b>	Gluconate	<b>1.97</b>	1.15	1.13
Proline	<b>0.08</b>	<b>0.00</b>	<b>0.06</b>	Glycerate	<b>1.54</b>	<b>2.78</b>	<b>2.02</b>
Serine	<b>2.24</b>	1.68	<b>2.96</b>	Glycolate	0.91	<b>0.63</b>	<b>0.53</b>
Threonine	<b>2.12</b>	0.72	1.24	Isocitrate	<b>0.37</b>	<b>0.19</b>	<b>0.29</b>
Valine	<b>3.06</b>	0.72	1.37	Malate	<b>3.26</b>	<b>2.63</b>	1.31
Arabinose	<b>0.58</b>	<b>0.58</b>	<b>0.34</b>	Nicotinate	<b>17.41</b>	nd	<b>7.47</b>
Fructose	<b>0.52</b>	<b>0.08</b>	<b>0.30</b>	Phosphorate	0.86	<b>0.49</b>	<b>0.54</b>
Fructose-6-P	0.83	<b>0.37</b>	<b>0.44</b>	Pyroglutamate	<b>0.44</b>	<b>0.28</b>	<b>0.28</b>
Gentiobiose	1.00	<b>0.06</b>	0.71	Quinate	<b>0.34</b>	<b>0.40</b>	<b>0.27</b>
Glucose	<b>0.65</b>	<b>0.46</b>	<b>0.38</b>	Saccharate	<b>3.83</b>	<b>2.34</b>	<b>2.23</b>
Glucose-6-P	0.82	<b>0.02</b>	<b>0.39</b>	Shikimate	0.56	2.78	0.06
Maltose	1.09	<b>0.26</b>	<b>0.30</b>	Succinate	<b>0.46</b>	<b>0.40</b>	<b>0.31</b>
Rhamnose	<b>0.61</b>	<b>0.71</b>	<b>0.37</b>	Threonate	<b>0.37</b>	<b>0.68</b>	<b>0.38</b>
Ribose	<b>0.15</b>	<b>0.35</b>	<b>0.11</b>	FA 16:00	<b>2.11</b>	0.98	1.12
Sucrose	<b>0.10</b>	<b>0.02</b>	<b>0.05</b>	FA 18:00	<b>2.07</b>	0.97	1.14

**Table 1B.** Fold changes in the volatiles relative to Cervil in the parental lines.

Data are normalized to the mean response calculated for the Cervil line. Values are presented as the mean of five biologically independent determinations. Those metabolites which were significantly different from the Cervil line ( $P < 0.05$ ) by the performance of Student's *t*-tests are marked in bold. Data for eugenol and ethyl salicylate should be considered as higher than the value present in the respective parental lines since these compounds were not present in the Cervil parental line.

	VilB	Levovil		VilB	Levovil
2-Methylpropanol	<b>0.31</b>	<b>0.27</b>	Hexanoic acid	1.79	0.76
3-Methylbutanal	0.81	1.25	Octanal	2.15	1.58
Butanol	0.42	0.82	Benzylalcohol	<b>15.94</b>	<b>25.59</b>
1-Penten-3-ol	<b>0.65</b>	<b>0.60</b>	2-Isobutylthiazole	<b>51.02</b>	<b>33.96</b>
1-Penten-3-one	<b>0.69</b>	0.69	Phenylacetaldehyde	0.63	<b>0.34</b>
Pentanal	0.76	<b>0.55</b>	(E)-2-Octenal	<b>0.42</b>	<b>0.55</b>
2-Ethylfuran	<b>0.60</b>	0.74	Acetophenone	2.47	2.54
3-Methylbutanenitrile	0.98	<b>2.02</b>	<i>p</i> -Tolualdehyde	2.23	1.31
3-Methylbutanol	<b>0.28</b>	0.58	Guaiacol	<b>48.19</b>	<b>64.76</b>
2-Methylbutanol	<b>0.40</b>	<b>0.31</b>	Linalool	2.39	<b>9.91</b>
(E)-2-Methyl-2-butenal	<b>0.20</b>	<b>0.18</b>	Nonanal	2.21	1.79
(E)-2-Pentenal	<b>0.53</b>	<b>0.47</b>	2-Phenylethanol	<b>0.17</b>	<b>0.10</b>
1-Pentanol	<b>0.62</b>	<b>0.68</b>	2-Ethylhexanoic acid	1.24	0.84
(Z)-3-Hexenal	0.87	1.01	Benzyl nitrile	0.87	0.87
Hexanal	<b>1.97</b>	<b>1.59</b>	Octanoic acid	1.64	1.08
3-Methylbutanoic acid	<b>0.44</b>	0.52	Terpineol	<b>3.56</b>	<b>9.27</b>
(E)-2-Hexenal	<b>0.41</b>	<b>0.52</b>	Methyl salicylate	<b>142.88</b>	<b>69.15</b>
(Z)-3-Hexen-1-ol	1.18	<b>2.06</b>	Geranial	1.95	<b>3.33</b>
Pentanoic acid	<b>1.89</b>	1.05	Ethyl salicylate	<b>&gt;29.55</b>	<b>&gt;120.93</b>
(E,E)-2,4-Hexadienal	0.83	1.01	1-Nitro-2-phenylethane	2.06	1.53
$\alpha$ -Pinene	0.82	<b>0.33</b>	(E,E)-2,4-Decadienal	<b>0.24</b>	<b>0.41</b>
(E)-2-Heptenal	<b>0.72</b>	<b>0.64</b>	Eugenol	<b>&gt;53.48</b>	<b>&gt;45.62</b>
Benzaldehyde	<b>1.94</b>	<b>2.51</b>	$\beta$ -Damascenone	<b>1.63</b>	0.90
6-Methyl-5-hepten-2-one	1.39	2.71	Geranylacetone	0.66	0.97
2-Pentylfuran	0.92	0.86	$\beta$ -Ionone	0.63	1.39

cultivars and the cherry cultivar. It is well known that aroma makes a major contribution to the human perception of flavour (Goff *et al.*, 2006); therefore, analysis of volatile organic compounds was also conducted on the lines C, L, and B. This analysis revealed huge differences between the cherry variety C and the large-fruited varieties L and B, including changes in the levels of volatiles thought to be relevant for the definition of tomato aroma. The most prominent differences were in 2-phenylethanol, which was present at 6- to 13-fold higher levels in the C variety, and for a group of phenolic derivatives: eugenol, methyl salicylate, ethyl salicylate and guaiacol, found at levels 20- to 100-fold lower than those observed in the large-fruited lines (Fig. 1B and Table 1B). Many other volatiles showed statistically significant different levels between C and the other lines, such as terpineol, linalool, (E)-2-octenal, hexanal, (E)-2-pentenal, 1-penten-3-ol, 2-methylbutanol, (E)-2-methyl-2-butenal, 2-methylpropanal, benzaldehyde, phenylacetaldehyde, and 2-isobutylthiazole.

### ***Analysis of metabolic variation in tomato lines pre-selected for their organoleptic properties***

Having established that the elite lines displayed considerable metabolic variation, the primary metabolite content of a subset of tomato lines resulting from their crossings which had been selected on the basis of their organoleptic properties (Lecomte *et al.*, 2004) were next evaluated. These lines consisted of marker-defined introgressions of five regions, controlling fruit quality variation, from the cherry tomato into each of the large-fruited lines. Lines in all three genetic backgrounds were evaluated in the first year but, due to the relatively low metabolic variation of the lines in the D background (see Supplementary Table S1) subsequent studies were focused only on lines carrying the L and B backgrounds. The lack of phenotypic variation in the D background lines is largely in accordance with results of previous studies in suggesting unfavourable

interaction on introgression of genome regions of C into the D variety (Lecomte *et al.*, 2004). A total of 45 primary metabolites were accurately quantified in every chromatogram. These compounds included most plant amino and organic acids, sugars, sugar alcohols, and fatty acids. The range of content of specific metabolites in the introgression lines was generally within that observed between the parental controls. In B background lines, only a relatively small number of metabolites exhibited transgressive behaviour in both harvests. These included glucose (which exhibited a range of relative levels of 0.54–1.49 in comparison with the recipient genotype control), aspartate (0.55–1.13), gluconate (0.00–1.33),  $\beta$ -alanine (0.79–4.06) and myo-inositol (0.52–1.07). All other metabolites only displayed transgressive behaviour either in a single harvest or not at all (see Table 2 for details). The occurrence of transgressive behaviour was even rarer in the L background and only reproducible in the case of alanine (which exhibited a range of relative levels of 0.58–9.58 in comparison with the recipient genotype control; Table 3).

Comparison of individual changes in primary metabolite content between the two harvests revealed that the data sets are generally in very high accordance, indicating that the observed changes are probably due to quantitative genetic factors. For subsequent analysis, the mean change between the two harvests was used since this allows a greater confidence that the changes reported are due to genetic rather than environmental factors. Whilst it is clearly difficult to display such a large data set in a truly quantitative manner, it can be stated that the mean difference in the content of any given metabolite ranged between 0.4 and 38.1 times the value observed in the L line for the L genotypes and between 0.3 and 9.7 times the value observed in the B line for the B genotypes. The metabolic changes observed in the hybrids, LxC and BxC, were similar in trend, but of more moderate magnitude, to the changes

**Table 2.** Metabolic analysis of the lines derived from the cross between ViIB and Cervil parents.

Values are presented as the mean of six biologically independent determinations. The fold changes are relative to the ViIB parent. In bold are those values which were significantly different with  $P < 0.05$  by the performance of Student's  $t$ -tests.

	Harvest 2004						Harvest 2005					
	B1	B2	B4	B9a	B9b	CxB	B1	B2	B4	B9a	B9b	Bx
β-Alanine	0.95	1.04	0.68	<b>1.64</b>	1.18	<b>0.51</b>	1.34	<b>4.06</b>	0.79	1.34	1.39	<b>2.21</b>
Alanine	2.96	<b>2.37</b>	0.79	<b>2.32</b>	1.59	<b>2.41</b>	1.24	1.16	1.09	<b>2.17</b>	<b>1.84</b>	0.69
Asparagine	0.83	1.51	0.71	0.97	1.10	1.46	3.19	1.63	1.16	0.73	0.92	1.27
Aspartate	0.76	0.80	0.85	1.03	0.70	0.61	0.87	0.55	0.88	1.13	0.94	0.74
Cysteine	0.94	1.33	0.78	1.57	nd	0.29	0.77	0.90	0.77	0.69	0.61	0.58
γ-Aminobutyrate	0.88	0.92	0.92	1.15	0.79	<b>0.61</b>	<b>0.61</b>	0.91	<b>0.57</b>	<b>0.61</b>	0.69	<b>0.44</b>
Glutamate	1.34	0.97	0.92	1.08	0.67	1.51	1.04	1.00	0.96	1.12	1.20	1.08
Glutamine	0.92	1.09	0.67	1.04	0.95	nd	1.09	1.02	0.95	0.95	1.01	0.95
Glycine	0.78	1.22	0.66	1.20	1.49	0.23	0.82	1.56	0.91	1.64	2.24	0.65
Homoserine	0.66	0.77	0.53	0.66	0.65	0.57	0.40	0.35	0.87	0.41	0.35	0.39
Isoleucine	<b>0.55</b>	0.92	0.73	0.70	0.79	<b>0.59</b>	<b>0.76</b>	1.18	0.96	0.94	1.14	0.76
Lysine	0.85	0.89	1.00	0.71	1.78	27.82	0.70	0.53	1.28	0.44	<b>0.50</b>	0.65
Phenylalanine	0.79	1.16	0.61	0.72	0.73	<b>0.60</b>	0.88	1.13	0.94	0.92	1.24	1.02
Proline	<b>3.86</b>	<b>1.80</b>	0.63	<b>3.61</b>	1.10	<b>4.46</b>	<b>5.64</b>	<b>1.69</b>	1.07	<b>2.52</b>	<b>1.72</b>	<b>7.54</b>
Putrescine	<b>1.69</b>	1.15	1.83	<b>2.25</b>	1.57	<b>2.56</b>	1.15	1.06	<b>1.60</b>	<b>1.64</b>	1.19	1.75
Pyroglutamate	1.13	1.00	0.92	1.04	0.88	0.55	1.04	0.97	1.03	1.04	1.03	1.03
Serine	0.77	1.07	0.90	0.99	0.99	<b>0.37</b>	<b>0.79</b>	1.19	0.82	0.91	1.19	<b>0.59</b>
Threonine	0.83	0.96	0.56	1.05	0.80	<b>0.44</b>	0.46	0.58	<b>0.40</b>	<b>0.87</b>	<b>0.71</b>	0.33
Valine	<b>0.62</b>	1.08	0.74	1.12	1.15	<b>0.32</b>	<b>0.72</b>	<b>1.43</b>	0.91	1.37	<b>1.71</b>	0.70
Fructose	<b>1.20</b>	1.03	1.04	1.08	<b>0.82</b>	0.93	1.09	0.92	0.94	1.02	1.03	1.09
Fructose-6-P	1.22	1.14	0.61	0.88	nd	0.77	<b>1.27</b>	1.01	1.06	1.16	<b>0.69</b>	<b>1.93</b>
Gentiobiose	<b>1.97</b>	1.31	1.03	1.20	1.18	<b>0.70</b>	1.16	0.96	0.76	<b>1.21</b>	1.04	1.05
Glucose	<b>1.14</b>	0.96	0.96	<b>1.09</b>	<b>0.79</b>	1.02	1.28	0.90	<b>0.54</b>	1.49	<b>0.74</b>	0.82
Glucose-6-P	1.20	1.11	0.65	1.01	1.03	0.68	<b>1.45</b>	<b>1.31</b>	0.99	<b>1.30</b>	1.12	<b>1.42</b>
Isomaltose	<b>1.95</b>	1.32	1.31	1.27	1.18	1.07	1.81	0.85	0.62	1.41	<b>2.93</b>	1.16
Maltose	1.49	1.00	0.53	0.61	5.37	0.73	1.27	0.93	1.37	0.96	1.25	<b>1.50</b>
Sucrose	1.63	1.82	0.68	<b>2.44</b>	<b>1.82</b>	<b>2.25</b>	<b>1.73</b>	<b>1.24</b>	0.79	<b>1.68</b>	<b>1.38</b>	<b>1.77</b>
Trehalose	<b>1.46</b>	1.02	0.97	<b>1.39</b>	2.11	0.83	1.93	1.01	1.16	<b>2.01</b>	<b>6.39</b>	1.33
Xylose	0.97	0.94	1.43	0.95	0.69	0.79	<b>0.63</b>	<b>0.65</b>	<b>0.73</b>	<b>0.56</b>	<b>0.61</b>	0.79
Glycerol	1.25	1.15	1.09	1.00	0.82	1.38	1.24	1.47	0.65	0.86	0.58	0.67
Myo-ino-1-P	1.31	0.96	0.81	1.01	0.94	nd	1.29	1.16	0.94	0.93	1.09	0.97
Myo-ino	<b>0.70</b>	<b>0.69</b>	0.56	0.99	1.06	<b>0.38</b>	0.99	0.84	<b>0.72</b>	1.03	1.07	0.95

	Harvest 2004						Harvest 2005					
	B1	B2	B4	B9a	B9b	CxB	B1	B2	B4	B9a	B9b	Bx
Benzoate	<b>1.87</b>	1.13	1.38	<b>2.26</b>	1.81	0.90	1.15	1.43	1.34	1.01	1.02	1.14
Citramalate	0.90	<b>0.73</b>	1.18	0.85	0.72	0.86	1.29	1.08	1.10	1.30	1.27	1.26
Citrate	<b>1.26</b>	1.04	0.89	<b>1.15</b>	<b>0.89</b>	<b>0.73</b>	1.08	0.93	0.88	0.98	0.95	1.04
Gluconate	1.17	1.07	1.17	<b>1.19</b>	1.33	<b>0.66</b>	<b>0.00</b>	<b>0.00</b>	<b>0.00</b>	<b>0.00</b>	<b>0.00</b>	<b>0.29</b>
Glycerate	<b>0.79</b>	<b>0.75</b>	0.78	<b>0.80</b>	0.74	<b>0.43</b>	0.90	<b>0.69</b>	1.40	0.98	1.09	1.03
Malate	<b>0.67</b>	0.96	0.67	<b>1.66</b>	1.23	<b>0.27</b>	1.09	0.92	0.97	<b>1.50</b>	1.12	0.73
Nicotinate	1.30	1.03	0.99	1.05	0.90	1.63	0.92	0.81	<b>0.74</b>	1.06	0.88	0.91
Phosphorate	<b>1.27</b>	0.96	0.97	1.01	0.77	<b>0.68</b>	0.96	0.90	1.10	0.92	1.12	0.90
Saccharate	<b>0.71</b>	<b>0.14</b>	0.47	<b>0.43</b>	0.21	<b>0.26</b>	0.95	0.78	0.79	0.90	0.89	1.07
Succinate	1.16	<b>1.16</b>	0.83	0.94	1.00	1.68	<b>1.71</b>	1.20	0.94	0.81	1.27	<b>1.91</b>
16:00	<b>1.20</b>	1.03	1.04	0.99	1.69	0.38	<b>0.59</b>	0.71	1.41	1.23	0.87	0.76
18:00	<b>1.19</b>	0.95	0.94	0.91	0.98	0.36	1.04	0.90	1.02	1.18	0.92	1.09

observed between the parental lines (Tables 2, 3). QTLs were determined by using Student's *t*-tests at a significance threshold of 0.05 in order to compare statistically every trait of each introgression line with its respective recipient genotype. Using this criterion, 35 single-trait metabolite QTLs were identified in the L background and 16 in the B background (see Fig. 2, although those for the introgression of chromosome 2 into the L background should be regarded as putative, since they only represent a single year analysis). Although most of the QTLs presented here were previously unknown, several, including those for sucrose and malate, have already been documented in the literature either in studies using the population described here or in studies reliant on the *S. pennellii* introgression line populations (Causse *et al.*, 2004; Schauer *et al.*, 2006, 2008). The number of QTLs was similar irrespective of the background into which the C genome segments were introgressed. Moreover, the F<sub>1</sub> hybrids between C and both L and B were largely equivalent with respect to the degree of metabolic changes observed [displaying changes in 50% of traits (52% for L and 54% for B)].

**Table 3.** Metabolic analysis of the lines derived from the cross between Levovil and Cervil parents.

Values are presented as the mean of six biologically independent determinations. The fold changes are relative to the Levovil parent. In bold are those values which were significantly different with  $P < 0.05$  by the performance of Student's *t*-tests.

	Harvest 2004							Harvest 2005					
	L1	L2	L4	L9a	L9b	Lx	CxL	L1	L4	L9a	L9b	Lx	CxL
β-Alanine	2.26	2.49	1.31	2.24	1.44	0.75	<b>5.88</b>	1.53	<b>0.63</b>	1.56	0.85	0.84	0.86
Alanine	<b>9.58</b>	<b>7.52</b>	<b>3.22</b>	1.97	1.28	<b>4.42</b>	2.20	<b>2.58</b>	0.63	0.77	0.58	<b>2.52</b>	1.52
Asparagine	<b>3.31</b>	<b>2.78</b>	1.77	2.14	1.37	0.89	<b>3.00</b>	1.19	0.62	0.58	0.60	0.58	0.51
Aspartate	<b>2.25</b>	1.61	1.78	2.06	1.12	0.74	<b>2.68</b>	<b>1.97</b>	1.12	1.80	1.23	0.88	1.35
γ-Aminobutyrate	<b>2.02</b>	<b>2.29</b>	1.42	1.90	6.71	0.95	2.26	0.89	<b>0.48</b>	0.77	<b>0.60</b>	<b>0.51</b>	<b>0.40</b>
Glutamate	10.80	<b>14.83</b>	6.78	3.35	3.68	1.21	<b>3.80</b>	1.27	1.17	1.44	1.38	0.91	<b>2.07</b>
Glutamine	3.39	1.33	0.60	1.25	0.62	0.72	2.74	1.68	0.73	6.45	.63	0.67	0.62
Glycine	0.50	0.65	0.45	0.45	0.45	0.20	0.81	1.30	0.68	1.53	0.78	0.79	0.61
Homoserine	1.92	1.52	0.81	1.21	1.26	0.77	1.67	0.62	0.53	0.68	0.60	0.61	0.69
Isoleucine	<b>2.38</b>	<b>2.11</b>	1.17	1.58	1.64	0.88	<b>2.08</b>	<b>1.39</b>	0.83	1.18	1.30	0.88	1.12
Leucine	<b>3.59</b>	<b>2.76</b>	<b>1.91</b>	1.80	<b>1.97</b>	1.36	0.03	<b>1.67</b>	0.99	<b>1.34</b>	1.38	1.11	<b>1.56</b>
Lysine	4.49	4.77	3.01	0.92	1.38	0.73	2.02	1.35	0.69	1.58	0.83	1.41	<b>2.32</b>
Phenylalanine	<b>2.02</b>	1.92	1.36	1.39	<b>1.92</b>	0.84	2.08	1.24	0.61	0.62	0.79	0.64	0.64
Proline	<b>23.25</b>	<b>13.66</b>	7.15	<b>15.86</b>	1.30	13.05	nd	<b>2.75</b>	0.85	1.50	0.89	<b>1.66</b>	<b>6.01</b>
Putrescine	0.88	1.40	1.49	0.90	1.41	1.12	1.22	1.49	0.61	0.72	0.61	0.67	0.62
Pyroglutamate	<b>1.80</b>	<b>1.68</b>	1.35	<b>1.49</b>	1.31	0.99	<b>2.31</b>	1.08	0.68	0.89	0.65	0.66	0.64
Serine	1.71	1.56	1.14	1.73	1.29	0.63	1.01	<b>2.25</b>	0.86	<b>2.04</b>	1.57	0.85	0.81
Threonine	1.91	1.51	1.15	1.09	1.17	0.64	1.98	1.41	0.59	0.67	1.30	0.55	0.62
Tyrosine	2.54	2.92	1.53	0.56	1.45	0.22	1.40	<b>2.18</b>	<b>1.43</b>	<b>2.06</b>	<b>1.60</b>	<b>2.34</b>	<b>3.30</b>
Valine	1.54	1.60	1.64	1.44	1.13	0.46	<b>2.33</b>	1.37	0.88	0.74	0.75	0.66	0.73
Arabinose	<b>3.47</b>	<b>2.58</b>	1.16	1.60	1.24	1.31	1.23	1.68	0.73	6.45	0.63	0.67	0.62
Fructose	<b>1.61</b>	1.26	1.05	<b>1.48</b>	1.20	<b>1.71</b>	1.45	0.94	0.89	1.07	0.92	1.11	1.11
Fructose-6-P	<b>3.48</b>	<b>2.09</b>	1.56	<b>1.80</b>	1.45	<b>2.92</b>	<b>2.02</b>	<b>1.86</b>	1.30	<b>2.11</b>	1.36	<b>3.86</b>	<b>4.66</b>
Gentiobiose	<b>3.54</b>	2.00	1.28	1.67	2.61	<b>2.46</b>	<b>13.46</b>	1.16	0.90	<b>1.41</b>	1.04	1.27	<b>1.56</b>
Glucose	<b>1.35</b>	<b>1.27</b>	1.23	<b>1.32</b>	1.18	<b>1.34</b>	<b>1.49</b>	<b>0.79</b>	0.86	0.85	0.95	0.85	<b>0.66</b>
Glucose-6-P	<b>3.08</b>	1.80	1.39	<b>1.86</b>	1.27	<b>2.72</b>	<b>2.24</b>	<b>2.81</b>	<b>1.71</b>	<b>2.46</b>	<b>1.72</b>	<b>2.64</b>	<b>3.83</b>
Isomaltose	<b>4.72</b>	2.89	2.20	<b>4.43</b>	0.89	<b>4.00</b>	<b>2.43</b>	<b>3.32</b>	1.22	<b>1.95</b>	<b>1.80</b>	<b>2.49</b>	<b>5.54</b>
Maltose	<b>4.67</b>	2.11	1.67	<b>4.26</b>	nd	<b>2.91</b>	<b>2.33</b>	1.24	1.08	<b>2.04</b>	<b>1.24</b>	<b>2.46</b>	<b>3.97</b>
Rhamnose	<b>3.15</b>	<b>2.51</b>	0.93	1.17	1.11	0.94	1.06	<b>1.49</b>	0.88	1.15	1.07	1.38	<b>1.55</b>
Sucrose	<b>6.53</b>	<b>4.38</b>	2.34	<b>5.80</b>	<b>2.48</b>	<b>9.12</b>	<b>15.97</b>	<b>2.38</b>	<b>0.54</b>	<b>1.60</b>	0.73	<b>6.00</b>	<b>4.68</b>
Trehalose	<b>2.63</b>	<b>2.08</b>	1.72	<b>2.88</b>	1.17	<b>2.51</b>	<b>2.73</b>	<b>2.13</b>	<b>1.26</b>	<b>2.32</b>	<b>1.48</b>	<b>3.50</b>	<b>4.78</b>
Xylose	1.31	1.19	1.21	1.32	1.08	<b>1.46</b>	<b>0.25</b>	0.88	1.28	<b>1.49</b>	1.06	1.16	1.12



	Harvest 2004							Harvest 2005					
	L1	L2	L4	L9a	L9b	Lx	CxL	L1	L4	L9a	L9b	Lx	CxL
Glycerol	<b>1.48</b>	<b>1.52</b>	1.37	<b>1.32</b>	<b>1.44</b>	1.24	<b>13.73</b>	1.21	1.21	1.81	1.39	1.29	1.49
Myo-ino-1-P	<b>2.17</b>	<b>1.75</b>	<b>1.71</b>	1.21	1.33	<b>1.59</b>	1.15	<b>2.33</b>	<b>1.54</b>	1.28	1.15	0.86	0.87
Myo-ino	<b>1.56</b>	1.28	1.36	<b>1.82</b>	1.11	1.12	1.35	1.51	1.04	1.36	1.35	1.99	<b>2.62</b>
Benzoate	1.43	1.46	1.39	1.36	1.28	1.26	1.87	1.17	0.92	<b>1.37</b>	1.08	1.07	1.39
Citramalate	<b>2.34</b>	1.47	1.23	1.04	1.32	<b>2.05</b>	<b>2.04</b>	<b>2.52</b>	1.56	<b>1.55</b>	<b>2.17</b>	<b>3.26</b>	<b>3.61</b>
Citrate	<b>1.60</b>	<b>1.44</b>	1.18	<b>1.45</b>	1.16	<b>1.37</b>	<b>1.68</b>	<b>0.76</b>	0.93	0.76	1.07	<b>0.54</b>	<b>0.33</b>
Gluconate	<b>1.88</b>	1.50	1.14	1.42	0.89	<b>1.67</b>	<b>1.79</b>	1.38	1.20	<b>1.61</b>	<b>1.76</b>	<b>2.13</b>	<b>2.89</b>
Glycerate	0.96	1.06	1.02	0.89	1.48	1.18	<b>0.59</b>	<b>2.22</b>	1.98	1.19	1.34	1.22	<b>0.52</b>
Malate	0.76	0.79	0.83	1.05	1.02	<b>0.49</b>	<b>0.56</b>	<b>0.71</b>	0.89	<b>0.83</b>	0.83	<b>0.72</b>	<b>0.65</b>
Nicotinate	<b>2.40</b>	<b>2.01</b>	<b>1.75</b>	<b>1.74</b>	<b>1.43</b>	1.56	<b>35.60</b>	<b>1.59</b>	1.18	1.57	0.86	<b>1.59</b>	<b>1.98</b>
Phosphorate	<b>1.78</b>	<b>1.67</b>	1.14	<b>1.31</b>	1.23	1.11	1.50	<b>2.39</b>	1.30	<b>1.95</b>	1.23	1.20	1.09
Saccharate	<b>0.52</b>	0.79	<b>0.59</b>	1.16	<b>0.75</b>	<b>0.67</b>	<b>0.42</b>	1.44	0.86	1.37	1.42	1.66	<b>1.90</b>
Succinate	<b>2.58</b>	1.85	1.23	<b>1.56</b>	<b>1.38</b>	<b>3.94</b>	1.44	1.41	0.59	0.67	1.30	0.55	0.62
16:00	1.16	1.10	0.98	1.38	1.32	1.03	1.12	1.36	0.84	1.32	0.88	1.12	1.37
18:00	1.28	1.17	nd	1.36	1.32	1.06	1.08	<b>2.56</b>	<b>1.70</b>	<b>2.09</b>	<b>1.73</b>	<b>2.58</b>	<b>3.47</b>

The lines carrying the five introgressed segments simultaneously and hence the highest proportion of the parental cherry Cervil genome (Lx and Bx) showed a similar percentage of overall changes ( 36% for Lx and 32% for Bx). Figure 2 shows the full list of QTLs (and, in the case of the Levovil introgression of chromosome 2, for which replicate data were not obtained, putative QTLs) for metabolite content, volatile content, and organoleptic properties analysed in the NILs. These QTLs were compared with the QTLs detected in a recombinant inbred population derived from the cross of Cervil and Levovil (Causse *et al.*, 2002). QTLs for sucrose were found in L1 and L2, which have previously been documented to display fruit sweetness QTLs. When the co-localization behaviour of the metabolites themselves is assessed, clustering of QTLs of metabolites of similar chemical structure is clearly visible, as would be expected both from previous studies of other traits in tomato (Causse *et al.*, 2002) and from studies of metabolic traits in both tomato and *Arabidopsis* (Schauer and Fernie, 2006; Liseć *et al.*, 2008; Rowe *et al.*, 2008).

### ***Variation in volatile organic compound content in tomato lines pre-selected for their organoleptic properties***

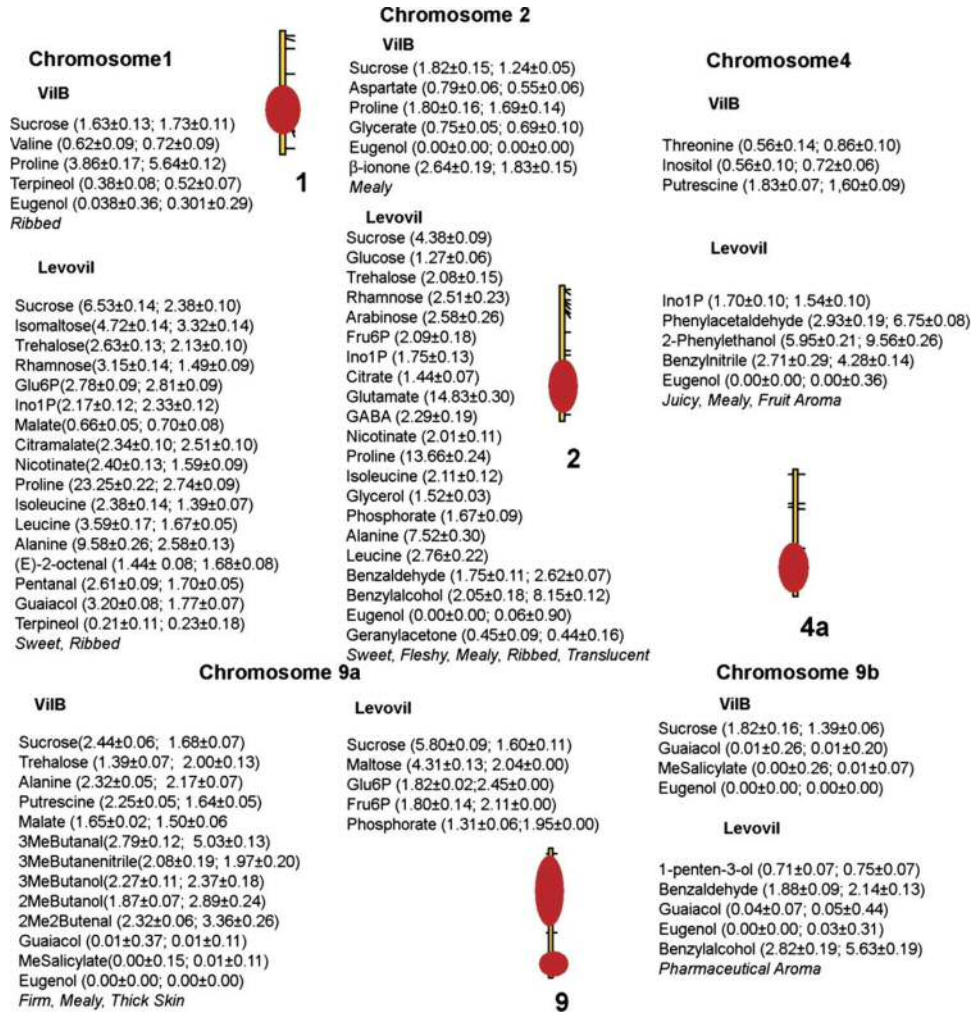
Having assessed the level of variation of primary metabolites in these lines, attention was next focused on the levels of volatile organic compounds. For this purpose, only L and B lines were studied. As for the primary metabolites, these compounds were measured in two different harvests—those of the 2005 and 2006 seasons (due to logistical difficulties it was not possible to perform these experiments in the exact same harvests; however, the close agreement of the primary metabolite results in the two harvests described above render this unproblematic). Fifty volatile organic compounds were accurately quantified by means of a HS-SPME-GC-MS method. In contrast to the observations for primary metabolites, many of the volatiles exhibited a transgressive behaviour. Guaiacol, (E)-2-pentenal, 1-pentanol, (Z)-3-hexenal, *p*-tolualdehyde, 3-methylbutanoic acid, and 2-pentylfuran showed transgressive behaviour in both genetic backgrounds analysed (Tables 4, 5). Additionally, 3-methylbutanal, 1-penten-3-one, 3-methylbutanenitrile, 3-methylbutanol, 2-methyl-1-butanol, (E)-2-methyl-2-butenal, hexanal, (E)-2-heptenal, hexanoic acid, and acetophenone displayed transgressive behaviour in the B lines, whilst 1-penten-3-ol, pentanal, 2-ethylfuran,  $\alpha$ -pinene, benzaldehyde, 1-nitro-2-phenylethane,  $\beta$ -damascenone, and geranylacetone exhibited such behaviour in the L lines. A total of 18 volatiles were transgressive in the B-derived lines, with a range of variation of 0.01–5.03 (ratio of relative abundance of the most extreme compounds compared with the parental line). Similarly, 15 volatiles were transgressive in the L lines, with a relative range of variation of between 0.01 and 12.8. Unlike the situation observed for primary metabolites, there is no a clear increase in the overall volatile content in the introgression lines. Indeed, the most remarkable differences are the dramatic decrease in a group of

phenylpropanoid derivatives: eugenol, methyl salicylate, ethyl salicylate, and guaiacol, to barely detectable levels in the lines harbouring a fragment of chromosome 9. The differences in the volatile patterns between the introgression lines and the varieties from which they are derived should thus be attributed more to the differences in levels of individual volatiles (or families thereof) rather than to differences in the overall volatile content.

Comparison of the levels of volatiles in the independent harvests (see Tables 4, 5) revealed that in contrast to the primary metabolite content, the data sets displayed large variation, indicating an important influence of environmental factors. The mean difference across the two harvests in the content of any given metabolite ranged between 0.00 and 75.18 times the value observed in the L line for L recipient genotypes and between 0.00 and 79.12 times the value observed in the B line for B recipient genotypes.

QTLs were determined for these traits, revealing a total of 17 QTLs in the L background and 15 in the B background (see Fig. 2 and Tables 4, 5). Whilst many of the QTLs presented here were previously uncharacterized, several, including those for pentanal, (E)-2-methyl-2-butenal, guaiacol, and eugenol, have already been documented within this population (Saliba-Colombani *et al.*, 2001), whereas others, including 3-methylbutanal, 3-methylbutanenitrile, 3-methylbutanol, 2-methyl-1-butanol, and  $\beta$ -ionone, have also been previously described in the *S. pennellii* introgression lines (Tieman *et al.*, 2006a). The number of QTLs for volatiles was similar irrespective to the background into which the C genome segments were introgressed, with both L and B displaying approximately similar numbers of QTLs. Principal component analysis illustrates how many of the introgression lines are clearly distinguishable on the basis of their volatile profile. Variance in the levels of a group of phenolic derivatives (1-nitro-2-phenylethane, 2-phenylethanol, phenylacetaldehyde, and benzylnitrile)

are responsible for the discrimination of the introgression line which harbours chromosome 4 fragments, whilst other NILs are segregated by their relative levels of other volatile compounds (Supplementary Fig. S1).



**Figure 2.** Quantitative trait loci controlling the content of the primary metabolites, volatiles, and sensory properties (in italics) in ViiB- and Levovil-derived lines.

**Table 4.** List of volatiles measured on fruits harvested from ViLB-derived lines. Values are presented as the mean of six biologically independent determinations. In bold are those values which were significantly different with  $P < 0.05$  by the performance of Student's *t*-tests.

	Harvest 2005						Harvest 2006					
	B1	B2	B4	B9a	B9b	Bx	B1	B2	B4	B9a	B9b	Bx
2-Methylpropanol	0.62	0.47	0.81	0.71	0.78	0.69	1.18	0.73	0.84	1.39	1.78	0.72
3-Methylbutanal	1.14	1.35	<b>0.20</b>	<b>2.80</b>	1.18	1.03	<b>2.55</b>	1.00	0.53	<b>5.03</b>	1.12	0.92
Butanol	1.07	0.89	0.81	1.60	1.20	1.27	0.94	1.48	0.58	1.98	3.75	1.56
1-Penten-3-ol	1.24	<b>1.25</b>	0.94	1.11	0.99	1.42	0.97	0.78	0.88	0.98	0.86	1.20
1-Penten-3-one	1.09	1.26	0.88	0.96	1.06	1.20	0.89	0.84	0.89	1.07	0.94	1.26
Pentanal	1.20	1.17	0.81	<b>0.64</b>	0.99	0.95	1.26	1.18	0.92	0.89	3.73	1.24
2-Ethylfuran	<b>1.78</b>	1.57	<b>1.68</b>	<b>2.12</b>	<b>1.72</b>	<b>2.24</b>	0.90	0.84	0.97	0.94	0.78	0.79
3-Methylbutanenitrile	0.94	1.65	<b>0.30</b>	<b>2.09</b>	2.55	1.51	<b>1.74</b>	1.73	0.46	<b>1.97</b>	1.57	0.93
3-Methylbutanol	1.38	1.18	<b>0.20</b>	<b>2.27</b>	1.32	1.08	<b>1.87</b>	0.80	0.71	<b>2.37</b>	<b>1.52</b>	0.97
2-Methylbutanol	0.81	1.06	<b>0.40</b>	<b>1.87</b>	1.25	1.11	1.30	0.84	0.88	<b>2.89</b>	<b>2.29</b>	1.39
(E)-2-Methyl-2-butenal	1.25	0.88	0.65	<b>2.32</b>	1.32	0.92	1.14	0.51	0.66	<b>3.36</b>	2.65	0.96
(E)-2-Pentenal	1.05	<b>1.30</b>	0.87	0.91	1.11	5.02	0.80	0.74	0.74	0.92	0.80	1.20
1-Pentanol	<b>1.34</b>	1.01	0.91	0.94	1.09	0.99	1.16	0.90	0.90	0.86	0.89	1.16
(Z)-3-Hexenal	1.32	1.36	1.37	<b>1.63</b>	1.35	1.30	0.81	1.08	1.16	1.10	1.08	1.14
Hexanal	1.02	1.04	0.92	0.92	0.90	0.93	<b>1.38</b>	<b>1.30</b>	1.18	1.05	<b>1.27</b>	1.05
3-Methylbutanoic acid	2.70	1.79	0.51	<b>2.34</b>	1.73	1.35	<b>2.45</b>	0.49	0.49	1.71	1.52	0.70
(E)-2-Hexenal	1.30	1.38	1.20	<b>1.45</b>	1.04	<b>1.71</b>	0.85	0.83	0.96	0.92	0.84	0.94
(Z)-3-Hexen-1-ol	1.84	1.05	1.07	1.12	0.75	1.23	<b>1.48</b>	0.85	1.05	1.06	0.73	1.08
Pentanoic acid	0.82	0.80	0.62	<b>0.63</b>	<b>0.65</b>	<b>0.65</b>	1.26	0.44	0.94	0.92	1.11	0.93
(E,E)-2,4-Hexadienal	1.03	0.98	1.37	1.32	1.21	1.49	0.84	1.01	1.04	0.99	1.02	1.07
$\alpha$ -Pinene	0.68	1.00	1.44	0.74	3.42	3.55	1.16	1.00	1.09	0.77	1.16	0.93
(E)-2-Heptenal	<b>1.36</b>	2.32	2.16	2.97	1.17	2.28	0.91	1.08	1.12	1.16	1.03	1.22
Benzaldehyde	0.62	1.42	0.98	0.51	1.05	<b>2.48</b>	0.73	1.02	0.71	1.01	0.89	<b>2.64</b>
6-Methyl-5-hepten-2-one	0.83	1.04	1.15	0.98	1.19	1.12	<b>1.73</b>	1.42	1.08	1.12	1.50	<b>1.67</b>
2-Pentylfuran	0.87	1.29	0.87	1.22	1.28	1.30	3.95	1.46	1.45	1.27	1.53	1.32
Hexanoic acid	0.60	0.70	0.58	0.62	1.02	0.74	1.08	1.41	1.29	4.96	0.70	1.10
Octanal	0.61	0.60	0.96	0.58	0.73	0.69	1.17	0.67	1.14	0.88	1.46	0.67
Benzylalcohol	<b>0.35</b>	1.23	0.92	<b>0.45</b>	1.07	<b>2.22</b>	0.95	1.04	0.79	0.63	1.50	<b>1.71</b>
2-Isobutylthiazole	0.78	0.81	0.27	0.67	0.83	0.61	<b>1.76</b>	1.24	0.41	0.86	0.87	0.77
Phenylacetaldehyde	0.48	1.30	0.64	0.73	0.72	1.10	0.70	1.46	<b>0.59</b>	1.23	0.76	1.43
(E)-2-Octenal	1.44	1.31	1.15	1.05	1.28	<b>1.66</b>	<b>1.81</b>	<b>1.37</b>	1.01	1.25	1.11	<b>2.36</b>
Acetophenone	0.60	0.85	1.23	0.65	0.57	0.81	0.79	1.26	1.04	0.86	0.81	1.01
<i>p</i> -Tolualdehyde	0.32	0.43	1.45	0.62	0.92	0.45	0.80	1.31	0.90	0.81	1.28	0.99

	Harvest 2005						Harvest 2006					
	B1	B2	B4	B9a	B9b	Bx	B1	B2	B4	B9a	B9b	Bx
Guaiacol	<b>1.32</b>	<b>2.25</b>	<b>2.46</b>	<b>0.01</b>	<b>0.01</b>	<b>0.03</b>	0.74	0.68	1.54	<b>0.01</b>	<b>0.02</b>	<b>0.03</b>
Linalool	0.28	1.11	1.35	1.20	1.14	0.25	0.62	1.69	1.24	0.97	1.42	<b>0.43</b>
Nonanal	0.66	0.69	1.23	0.64	0.83	0.62	1.26	1.15	1.31	1.20	<b>1.57</b>	0.92
2-Phenylethanol	<b>0.65</b>	1.21	0.82	0.89	1.19	1.11	0.79	1.16	0.75	1.20	1.06	<b>1.60</b>
2-Ethylhexanoic acid	0.67	0.84	1.37	0.89	1.61	1.19	0.48	1.27	0.86	0.98	0.28	0.70
Benzyl nitrile	0.52	1.31	0.97	0.79	1.47	1.13	1.19	<b>2.38</b>	0.83	1.98	1.77	2.38
Octanoic acid	0.65	0.77	1.42	0.85	1.34	1.01	0.08	0.52	0.28	0.34	0.04	0.27
Terpineol	<b>0.38</b>	0.97	<b>1.62</b>	0.95	1.03	<b>0.43</b>	<b>0.52</b>	<b>1.63</b>	1.21	0.90	<b>1.39</b>	<b>0.45</b>
Methyl salicylate	<b>2.18</b>	2.08	2.18	<b>0.00</b>	<b>0.00</b>	<b>0.01</b>	0.92	<b>0.39</b>	0.91	<b>0.01</b>	<b>0.01</b>	<b>0.01</b>
Geraniol	0.75	0.85	1.55	0.92	1.21	1.16	2.14	2.05	2.08	2.23	1.74	1.67
Ethyl salicylate	0.49	<b>0.13</b>	1.62	<b>0.00</b>	<b>0.01</b>	0.01	2.97	2.06	3.43	0.00	0.74	1.34
1-Nitro-2-phenylethane	0.52	0.91	0.49	0.80	1.42	1.14	1.66	<b>3.19</b>	0.58	2.68	1.57	<b>5.11</b>
(E,E)-2,4-Decadienal	1.28	1.74	1.28	1.42	1.63	<b>2.66</b>	<b>2.32</b>	<b>1.40</b>	0.97	1.36	1.54	<b>2.90</b>
Eugenol	<b>0.04</b>	<b>0.00</b>	0.56	<b>0.00</b>	<b>0.00</b>	<b>0.00</b>	<b>0.30</b>	<b>0.00</b>	1.30	<b>0.00</b>	<b>0.00</b>	<b>0.00</b>
$\beta$ -Damascenone	0.69	1.10	1.04	0.69	0.96	0.52	1.31	1.26	0.80	0.92	1.45	0.67
Geranylacetone	1.00	1.65	1.40	1.21	1.27	1.97	1.96	1.53	1.53	1.69	1.70	1.94
$\beta$ -Ionone	1.99	<b>2.64</b>	1.82	<b>1.76</b>	<b>1.47</b>	<b>3.48</b>	<b>2.05</b>	<b>1.83</b>	1.50	2.16	1.69	<b>2.45</b>

There are many co-localizations of volatile and organoleptic QTLs. The fruit aroma QTL co-localized with the QTL for 2-phenylethanol, benzyl nitrile, and phenylacetaldehyde (chromosome 4), all of them phenolic derivatives with increased contents in the lines containing C alleles at this QTL. 2-Phenylethanol, the volatile which showed the highest increase, has been described to provide a sweet and fruity aroma (Togari *et al.*, 1995), and could be responsible for this fruit aroma perception. Pharmaceutical aroma QTL co-localized on chromosome 9 with the QTL of guaiacol and methyl salicylate, with both phenylpropanoid derivatives levels being 20-fold lower in the lines containing the C alleles at this QTL. As previously stated, guaiacol and eugenol provide a medicinal-like aroma. Thus, these compounds could conceivably be responsible for the pharmaceutical aroma perception.

**Table 5.** List of volatiles measured on fruits harvested from Levovil-derived lines. Values are presented as the mean of six biologically independent determinations. In bold are those values which were significantly different with  $P < 0.05$  by the performance of Student's *t*-tests.

	Harvest 2005							Harvest 2006						
	L1	L2	L4	L9a	L9b	Lx	CxL	L1	L2	L4	L9a	L9b	Lx	CxL
2-Methylpropanol	1.33	0.68	2.35	0.92	0.77	0.78	1.26	0.77	<b>0.46</b>	1.48	0.56	0.86	<b>0.23</b>	1.25
3-Methylbutanal	0.82	0.86	1.04	0.98	1.46	0.85	0.79	1.90	0.78	0.17	0.95	1.04	0.31	0.10
Butanol	1.80	<b>0.28</b>	1.31	0.70	1.30	0.53	1.26	2.15	0.71	1.34	1.01	2.13	0.58	4.04
1-Penten-3-ol	1.12	0.88	0.77	1.01	<b>0.71</b>	1.05	1.18	0.90	0.87	<b>0.77</b>	0.94	<b>0.75</b>	1.03	1.14
1-Penten-3-one	<b>1.60</b>	0.86	1.15	0.97	0.91	1.17	1.24	1.18	0.99	1.08	1.01	0.91	1.08	1.03
Pentanal	<b>2.61</b>	0.93	1.83	1.28	0.68	1.48	1.56	1.70	1.25	1.03	<b>0.77</b>	<b>0.70</b>	<b>1.36</b>	1.24
2-Ethylfuran	0.71	<b>1.61</b>	<b>0.57</b>	0.86	0.82	1.30	<b>1.45</b>	0.52	0.92	0.73	0.76	0.74	0.74	1.43
3-Methylbutanenitrile	<b>0.00</b>	<b>0.36</b>	0.60	0.75	<b>0.64</b>	<b>0.64</b>	<b>0.62</b>	1.29	1.22	0.57	1.32	0.69	0.78	0.50
3-Methylbutanol	0.81	0.94	0.91	1.00	1.48	0.86	1.34	0.62	0.70	0.52	0.99	1.24	0.57	1.11
2-Methylbutanol	2.15	0.73	4.40	1.15	<b>2.81</b>	0.71	<b>1.78</b>	0.79	<b>0.48</b>	1.76	0.90	1.57	0.84	1.33
(E)-2-Methyl-2-butenal	0.99	0.52	12.8	0.80	1.16	0.56	<b>1.76</b>	0.28	0.90	<b>3.46</b>	0.75	0.78	0.83	1.50
(E)-2-Pentenal	<b>1.82</b>	0.87	1.27	1.11	0.91	1.36	<b>1.59</b>	1.08	1.28	1.19	1.03	0.78	1.15	1.40
1-Pentanol	1.36	0.84	1.19	1.24	0.74	1.19	1.39	1.47	1.16	0.97	0.77	0.72	1.30	<b>1.36</b>
(Z)-3-Hexenal	0.85	1.04	0.90	1.01	0.97	1.30	<b>1.47</b>	0.65	0.94	0.96	1.21	1.22	1.07	1.13
Hexanal	0.75	0.75	0.76	0.80	0.84	0.98	0.73	1.26	1.14	0.97	<b>0.82</b>	0.81	<b>1.25</b>	0.84
3-Methylbutanoic acid	0.68	1.36	0.77	1.23	2.23	1.53	1.09	0.62	2.31	0.82	2.23	1.55	1.53	1.04
(E)-2-Hexenal	<b>1.57</b>	<b>1.56</b>	1.00	1.02	0.99	<b>1.29</b>	1.79	1.12	1.30	0.92	0.96	1.00	1.12	0.99
(Z)-3-Hexen-1-ol	<b>0.65</b>	0.75	0.73	1.11	<b>0.48</b>	<b>0.69</b>	0.79	0.71	1.16	1.16	1.33	0.98	1.29	<b>1.84</b>
Pentanoic acid	0.93	0.98	1.45	1.59	1.58	1.05	0.84	0.63	1.51	1.17	0.97	0.55	0.69	1.54
(E,E)-2,4-Hexadienal	0.91	0.83	0.79	1.00	0.78	1.16	1.24	0.85	1.00	0.94	0.99	1.01	1.01	1.00
$\alpha$ -Pinene	<b>3.00</b>	<b>2.75</b>	23.6	9.39	1.04	1.85	1.80	0.65	0.89	1.01	0.86	13.18	<b>0.55</b>	0.76
(E)-2-Heptenal	<b>1.47</b>	1.27	1.40	<b>1.28</b>	0.85	2.47	1.15	0.67	0.89	0.84	1.15	1.04	1.07	1.32
Benzaldehyde	0.74	<b>1.75</b>	1.28	1.29	<b>1.88</b>	<b>1.99</b>	1.14	1.87	<b>2.62</b>	<b>3.78</b>	<b>3.19</b>	<b>2.15</b>	<b>4.07</b>	<b>2.29</b>
6-Methyl-5-hepten-2-one	0.35	<b>0.25</b>	0.55	0.43	0.58	0.44	0.35	0.90	0.66	0.63	<b>0.55</b>	0.56	0.69	<b>0.42</b>
2-Pentylfuran	0.79	0.79	0.96	0.97	0.99	1.32	0.75	0.96	0.92	0.76	0.79	0.60	1.09	0.85
Hexanoic acid	2.29	1.32	1.71	1.67	1.99	1.41	0.91	0.88	2.13	2.22	2.77	1.97	1.57	2.66
Octanal	1.03	0.96	1.01	1.11	0.93	0.77	0.59	1.08	0.98	0.74	0.67	0.78	0.61	0.40
Benzylalcohol	0.95	<b>2.05</b>	1.09	1.32	<b>2.82</b>	<b>2.68</b>	1.02	8.15	<b>7.00</b>	<b>9.84</b>	9.15	<b>5.63</b>	<b>16.6</b>	<b>4.78</b>
2-Isobutylthiazole	<b>0.15</b>	<b>0.29</b>	<b>0.36</b>	<b>0.08</b>	<b>0.28</b>	<b>0.08</b>	<b>0.10</b>	0.88	1.22	0.51	0.12	0.50	<b>0.19</b>	<b>0.20</b>
Phenylacetaldehyde	0.99	<b>0.54</b>	<b>2.93</b>	0.62	0.75	0.74	1.43	1.77	1.73	<b>6.76</b>	1.02	1.52	1.38	<b>4.77</b>
(E)-2-Octenal	<b>1.44</b>	1.00	0.97	1.25	0.71	<b>1.63</b>	<b>1.47</b>	1.68	1.21	0.85	<b>0.91</b>	<b>0.71</b>	<b>2.10</b>	<b>2.36</b>
Acetophenone	0.80	0.58	0.62	0.75	0.56	0.71	0.56	0.83	0.83	0.93	0.76	1.10	0.72	1.06
<i>p</i> -Tolualdehyde	0.79	1.36	2.14	1.73	1.56	<b>0.56</b>	0.93	0.95	0.55	0.45	0.60	1.00	0.68	2.37

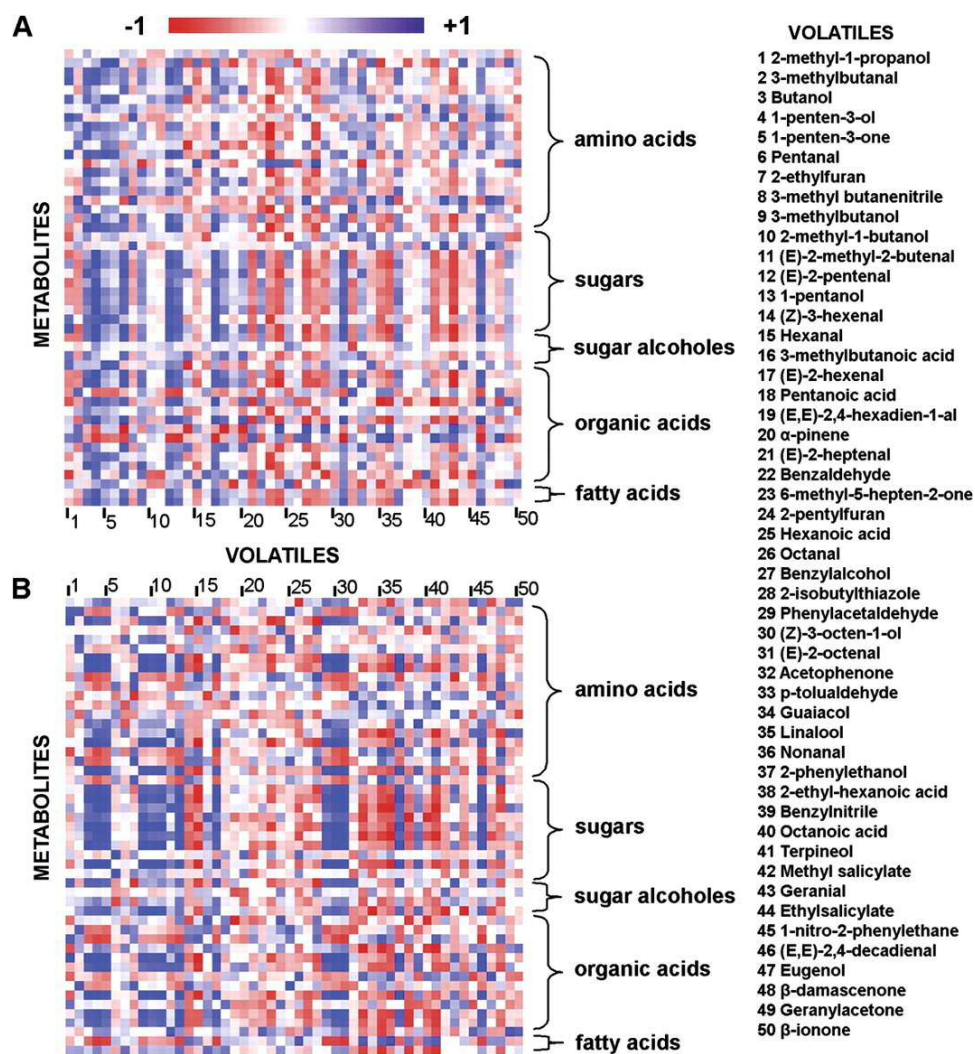
	Harvest 2005								Harvest 2006							
	L1	L2	L4	L9a	L9b	Lx	CxL	L1	L2	L4	L9a	L9b	Lx	CxL		
Guaiacol	<b>3.20</b>	1.18	0.82	1.29	<b>0.04</b>	<b>0.02</b>	<b>0.01</b>	1.77	0.96	0.75	<b>0.83</b>	<b>0.05</b>	<b>0.02</b>	<b>0.01</b>		
Linalool	<b>0.10</b>	0.55	0.83	0.57	0.77	<b>0.16</b>	<b>0.32</b>	0.23	1.05	0.83	1.35	1.09	0.24	0.45		
Nonanal	1.11	0.96	1.14	1.06	1.11	0.65	0.68	0.85	0.84	0.75	0.97	0.83	<b>0.49</b>	0.60		
2-Phenylethanol	0.67	0.73	<b>5.95</b>	<b>0.51</b>	0.88	0.59	<b>2.35</b>	1.90	1.18	<b>9.56</b>	0.93	1.46	<b>2.13</b>	7.56		
2-Ethylhexanoic acid	1.74	1.36	1.97	1.28	1.82	1.46	0.77	1.53	3.08	3.57	5.21	2.14	1.89	3.28		
Benzyl nitrile	0.46	<b>0.34</b>	<b>2.71</b>	0.50	<b>0.42</b>	<b>0.43</b>	0.81	1.59	1.48	<b>4.38</b>	0.87	1.16	0.90	<b>2.31</b>		
Octanoic acid	0.95	1.15	0.93	1.12	1.55	1.35	0.79	3.82	6.65	5.62	32.3	8.57	9.39	<b>16.66</b>		
Terpineol	<b>0.21</b>	<b>0.55</b>	1.08	0.61	0.60	<b>0.16</b>	<b>0.34</b>	0.23	0.80	0.78	0.92	0.82	<b>0.18</b>	<b>0.40</b>		
Methyl salicylate	<b>2.59</b>	<b>5.76</b>	0.68	3.37	<b>0.02</b>	<b>0.02</b>	<b>0.01</b>	1.36	1.64	1.17	1.26	0.09	0.07	0.08		
Geraniol	<b>0.37</b>	<b>0.35</b>	0.74	0.50	0.61	<b>0.47</b>	<b>0.33</b>	0.74	0.57	0.51	<b>0.43</b>	0.61	0.57	<b>0.39</b>		
Ethyl salicylate	3.21	<b>0.19</b>	0.53	1.35	<b>0.00</b>	<b>0.00</b>	<b>0.00</b>	2.43	2.64	0.53	1.09	0.00	0.00	0.00		
1-Nitro-2-phenylethane	0.43	<b>0.20</b>	5.84	0.69	0.47	<b>0.46</b>	0.65	1.89	1.08	<b>7.53</b>	0.59	0.70	1.27	<b>2.66</b>		
(E,E)-2,4-Decadienal	1.07	0.88	1.06	1.60	0.53	1.65	<b>2.33</b>	1.29	0.80	0.69	0.54	<b>0.52</b>	1.72	1.51		
Eugenol	1.13	<b>0.00</b>	<b>0.00</b>	0.70	<b>0.00</b>	<b>0.00</b>	<b>0.00</b>	0.62	<b>0.06</b>	<b>0.00</b>	0.60	<b>0.03</b>	<b>0.00</b>	<b>0.00</b>		
$\beta$ -Damascenone	1.13	1.29	1.42	1.12	0.73	0.44	1.11	0.70	0.77	0.50	1.20	1.50	0.56	0.53		
Geranylacetone	0.72	<b>0.45</b>	0.77	0.64	0.66	0.81	0.76	0.72	<b>0.44</b>	<b>0.37</b>	<b>0.35</b>	0.53	0.87	<b>0.38</b>		
$\beta$ -Ionone	0.65	0.66	0.63	0.94	0.59	<b>1.78</b>	0.85	0.71	0.56	<b>0.45</b>	<b>0.51</b>	0.53	1.24	0.53		

### Correlation analysis

For a fuller characterization of the associations between traits, a correlation-based approach was adopted in which the mean values determined above for each metabolite were compared with those determined for each volatile. For this purpose, a combinatorial analysis of all metabolites (both primary and volatile) was carried out, by running the data points through pairwise correlation analysis. Of the 4560 possible pairs analysed, 806 and 750 resulted in significant correlations ( $P \leq 0.05$ ) for L and B lines, respectively. Of these pairs, 609 and 466 showed positive ( $r > 0.65$ ) and 197 and 284 showed negative ( $r$  less than  $-0.65$ ) correlation coefficients for L- and B-derived lines, respectively. The heat map of Fig. 3 (and Supplementary Tables S2, S3) shows the correlations between primary metabolites and volatiles (to simplify interpretation, metabolites are grouped on the basis of their compound class).



**Figure 3.** Heat map showing the correlation analysis between traits in tomato NILs. (A) Mean of metabolites and volatiles during the two years for Levovil-derived NILs. (B) Mean of metabolites and volatiles during the two years for VilB-derived NILs. Regions in red and blue indicate negative or positive correlations between traits, respectively (the complete data set is also available in Supplementary Tables S2 and S3).

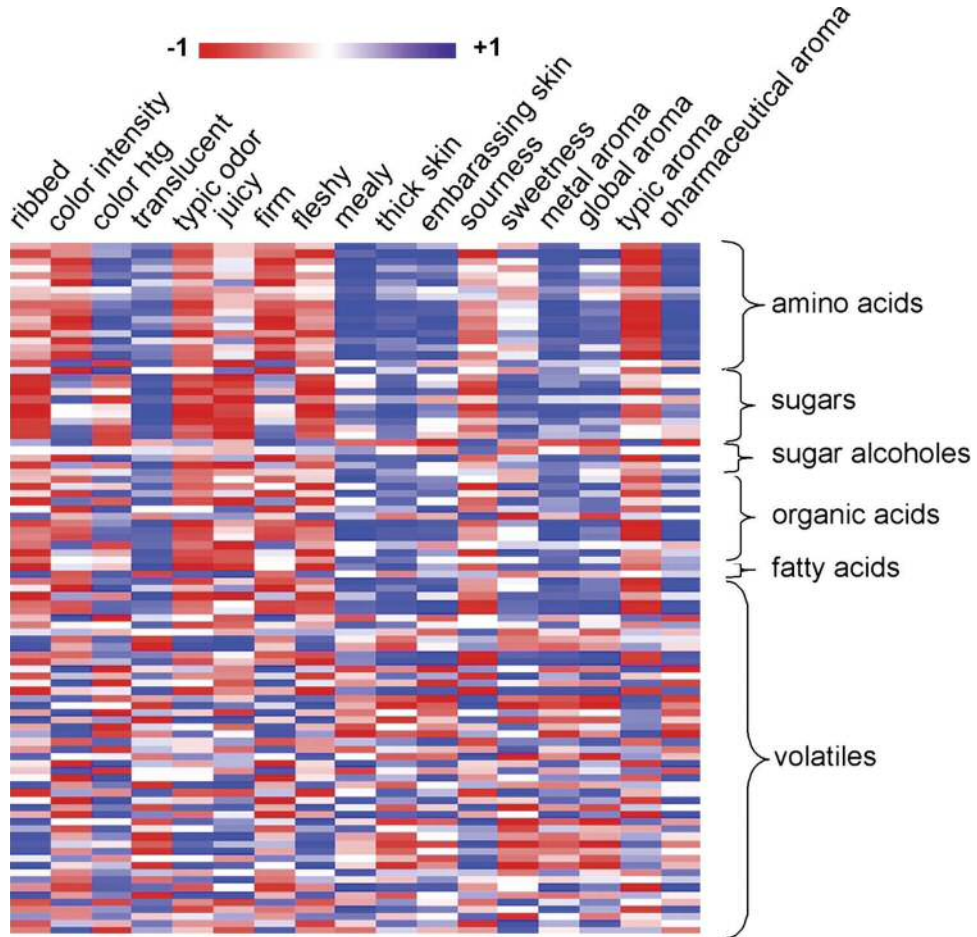


Negative correlations were significant between the sugars and sugar derivatives fructose, fructose-6P, glucose, glucose-6P, isomaltose, and sucrose, and the volatiles linalool, terpineol, and nonanal in both genetic backgrounds, whilst geranial was also strongly negatively correlated with sugars in the L background but not in the B background. In contrast, positive correlations were observed between 1-penten-3-ol, (E)-2-hexenal, (E)-2-octenal, and (E,E)-2,4-decadienal and the above-mentioned sugars. There is little correlation between the levels of the volatile organic compounds and their direct precursors from primary metabolism. Correlations within primary metabolites and volatiles were also analysed. The full data set of correlation coefficients is presented in Supplementary Tables S2 –S7. Among primary metabolites (Supplementary Tables S4, S5), correlations were qualitatively similar to those reported previously in data sets wherein metabolite contents varied either across a developmental time course (Carrari *et al.*, 2006) or across the *S. pennellii* introgression lines (Schauer *et al.*, 2006, 2008). As observed previously in Carrari *et al.* (2006), phosphorylated intermediates displayed the greatest number of significant correlations to other primary metabolites. Among the different classes of primary metabolites, the sugars displayed the highest number of correlations irrespective to the genotype analysed. For example, sucrose, fructose, and glucose exhibited 20, 20, and 15 significant correlations in L-derived lines and 17, 21, and 19 in B-derived lines, respectively. Other compounds displayed a different number of correlations when the two genotypes were considered. Aspartate and asparagine displayed 23 and 21 significant correlations, respectively, in the L-derived lines but no significant correlations in the B-derived lines. Additionally the number of correlations for glutamate in the L-derived lines was lower compared with those observed in B-derived lines (10 and 15, respectively).  $\gamma$ -Aminobutyric acid (GABA) and saccharate displayed a low number of correlations in L-derived lines (0 and 3,

respectively) but a high number in B-derived lines (12 and 15, respectively). Similarly, the volatile–volatile correlations (Supplementary Tables S6, S7) observed across the lines were largely in accordance with those described by Tikunov *et al.* (2005) across a panel of 94 tomato cultivars. The results were consistent with most of the previously described correlations such as those of eugenol, guaiacol, methyl salicylate, and ethyl salicylate. Some novel correlations were also uncovered in the present study such as those between 1-nitro-2-phenylethane and benzylnitrile or other phenylpropanoid derivatives, or the tight correlations between (E)-2-octenal and (E,E)-2,4-decadienal, or 1-penten-3-ol and other lipid derivatives. A strong correlation was additionally observed between linalool and terpineol, and also between 2-methyl-1-propanol, 2-phenylethanol, and butanol. As described for the primary metabolites, many of the correlations were observed in both genetic backgrounds (L and B), whilst others were significant only in one of them (Supplementary Tables S6, S7).

As a final analysis, correlations between all chemical traits measured in L-derived lines with organoleptic properties assessed on the same harvest were studied by sensory profiling (Fig. 4 and Supplementary Tables S8, S9). Of 1615 pairs of traits, 181 showed significant correlations ( $P \leq 0.05$ ), among which 101 exhibited positive correlations ( $r > 0.65$ ) and 80 displayed negative correlations ( $r$  less than  $-0.65$ ). Some of the chemical traits showed opposite behaviour with respect to different sensory properties. For example, xylose correlated positively with firmness but negatively with juiciness, whilst malate correlated positively with sourness and negatively with sweetness. However, there were other cases, such as those of sweetness and global aroma, in which sensory traits displayed highly similar correlative behaviour with the same metabolites. When analysed specifically from the perspective of the organoleptic traits, some

**Figure 4.** Heat map showing the correlation analysis between primary metabolites, volatiles, and sensory properties in Levovil-derived tomato NILs. Regions in red and blue indicate negative or positive correlation between traits, respectively (for details see Supplementary Table S9).



strong correlations were observed, such as colour intensity–glutamic acid ( $r=0.98$ ), pharmaceutical aroma–guaiacol ( $r=0.97$ ), typical tomato aroma–phenylalanine ( $r=-0.97$ ), global aroma–2-ethyl-hexanoic acid [ $r=-0.98$ ; global aroma corresponded to the general impression of aroma before swallowing (Causse *et al.*, 2001)], sweetness–citramalic acid ( $r=0.99$ ), sourness–alanine ( $r=-$

0.97), juiciness–trehalose ( $r=-0.99$ ), firmness–glutamic acid ( $r=0.99$ ), embarrassing skin–xylose [ $r=-0.97$ ; embarrassing skin is a sensory attribute which describes how difficult it is to swallow fruit skin and therefore it has a higher tendency to remain in the mouth (Causse *et al.*, 2001)]. Some of these correlations could probably be predicted on the basis of the chemical properties of the metabolites, such as, for example, the volatile guaiacol (which correlated positively with pharmaceutical aroma), is described as having a smoke-like or medicinal odour, and 2-ethyl-hexanoic acid (which negatively correlated with global aroma) which exhibited a wine-like odour. A more in-depth analysis of the organoleptic traits revealed complex interactions among many metabolites. Global aroma, for instance, significantly correlated to many volatiles, both positively [1-pentanol ( $r=1.00$ ), (E)-2-hexenal ( $r=0.97$ ), (E)-2-pentenal ( $r=0.93$ ), 1-penten-3-one ( $r=0.91$ )] and negatively [2-ethylhexanoic acid ( $r=-0.98$ ), pentanoic acid ( $r=-0.96$ ), linalool ( $r=-0.95$ )], and also to non-volatile compounds [alanine ( $r=0.98$ )]. Typical tomato aroma displayed significant positive correlation only with the volatile benzaldehyde ( $r=0.91$ ), but exhibits negative correlation with 12 metabolites, most of them being non-volatile.

## DISCUSSION

Fruit flavour is known to be considerably influenced by several factors. For example, the contents of primary metabolites such as organic acids and sugars are known to be important, but the sugar/acid ratio is also an important determinant of taste. In practical terms, this can be summarized as follows: both high sugar content and acidity result in a good flavour, low acidity and high sugar content gives a bland flavour, high acidity and low sugar content give a tart flavour, and finally low acidity and sugar content results in an essentially tasteless flavour. On the other hand, volatile components which build fruit aroma greatly influence human perception of flavour. Here the metabolomic approach was used to describe the phenotypic variation of a broad range of primary and volatile metabolites across diverse genetic backgrounds. The results of the most highly abundant primary metabolite analysis of cherry and large-fruited tomatoes lines were largely in accordance with those obtained from previous studies (Causse *et al.*, 2002). The low sugar and high malate content of the L parent and the corresponding very low sugar/acid ratio could explain the lower acceptance of the fruit by the food panel tasters, especially given that malate is perceived as sourer tasting than citrate (Marsh *et al.*, 2003).

Other less abundant primary metabolites were also found at different levels in the parental lines. A recent survey of metabolite content in the fruits of a range of wild tomato species revealed that whilst these displayed large variations in sugar and amino acid content they were essentially unaltered in the content of tricarboxylic acid (TCA) cycle intermediates (Schauer *et al.*, 2005b). This suggests that the variation observed here is probably the result of breeding-based selection. One metabolite of particular interest is glutamate, known to be sensed as the fifth basic taste (umami), which evokes a savoury feeling. In addition to the changes observed in sugars and acids in cherry

tomatoes, the glutamate level was found to be considerably higher in the C variety than in the large-fruited varieties. This finding is additionally in accordance with the fact that cherry tomatoes were found to be tastier than the other parental lines used in this study.

Within the aroma components, 2-phenylethanol is known to provide a sweet and fruity perception (Togari *et al.*, 1995). It is thus expected that the increased levels of 2-phenylethanol in line C would synergistically interact with sugars to produce an even sweeter flavour. Moreover, guaiacol has been described as an undesirable compound in many fruits, as it provides a medicinal-like aroma (Zierler *et al.*, 2004).

The evaluation of the primary metabolite content of a subset of tomato lines containing marker-defined introgressions, of five regions controlling fruit quality variation from the cherry tomato into large-fruited genetic backgrounds, revealed only a relatively small number of metabolites which exhibited transgressive behaviours across both harvests. This contrasted with the situation observed in interspecific introgression lines in which segments of the *S. pennellii* genome were inserted into the background of the M82 cultivar of *S. lycopersicum*, in which transgressive behaviour was observed for the majority of metabolic traits (Schauer *et al.*, 2006). Irrespective of whether they were transgressive or not, the changes in metabolites showed a strong bias toward an increase in metabolite contents in the introgressed lines relative to either recipient background. This could have been anticipated since the cherry tomato line was characterized as generally displaying a higher metabolite content than the large-fruited cultivars, but this is not true for all metabolites since increases were also found in the metabolite valine that was present at lower levels within the cherry tomato than in the large-fruited species.

As stated before, unlike the situation observed in primary metabolites, there is no clear increase in the overall volatile content of the introgression lines. Thus, the differences in the volatile pattern between parental and introgression lines are due to the differences in individual volatiles (or families of them), their modified levels depending on the introgressed chromosome fragment.

Few clear patterns emerged when co-localization between metabolite and volatile traits was examined. Co-localizations of QTLs for two metabolites could be due either to physiological relationships or to the action of two genes genetically linked and introgressed in the same region, as the size of introgressed regions is still large ( 10–40 cM). For example, the negative association between sucrose and eugenol content must be due to genetic linkage rather than to a common physiological origin since there are other examples of these traits varying independently of one another and, moreover, the molecular mechanism underlying this association cannot be formally resolved in the current study. Evaluation of the *S. pennellii* introgression lines revealed that increased levels of 2-phenylethanol and 2-phenylacetaldehyde were independent of changes in the level of phenylalanine (Tieman *et al.*, 2006*b*). In this study, the content in volatiles correlated more with the levels of soluble sugars than with their direct precursors. The most likely explanation is that sink strength regulates part of the production of secondary metabolites. Nevertheless, it is also possible to speculate that these changes could be due to sugar-mediated changes in gene expression of enzymes involved in their biosynthetic pathways or that they merely resulted from spurious associations resulting from gene linkages within the large introgressions of the C genome. Considerably more experimental evidence is, however, required in order to provide mechanistic insight into these phenomena. This is indeed the case for



any of the associations presented here since the data provided only indicate linkages between the various traits and do not provide any information concerning the causality underlying their association. Whilst some of the correlations found in the present work could probably be predicted on the basis of the chemical properties of the metabolites, the vast majority are novel, and as such could provide valuable information in helping to unravel the complex basis of sensory fruit traits. It seems likely that considerable research effort is still needed in order to identify the causality, if any, underlying these relationships.

## CONCLUSION

A comprehensive profiling of both small molecule primary metabolites and the important volatile organic compounds of tomato was performed in independent cultivars of tomato containing equivalent introgression regions from a cherry tomato variety. The results confirmed and extended earlier studies (Causse *et al.*, 2001, 2002, 2004), suggesting that chemical composition QTLs were identifiable and hence probably tractable from these crosses. In addition, they revealed that the expression of the QTLs is highly dependent on the genetic background, D-derived lines displaying far fewer QTLs for primary metabolites than L- and B-derived lines (a fact exacerbated when it is taken into account that the QTLs for the D genotype could only be regarded as putative). The current study utilized a broad level profiling of primary metabolites and volatiles to facilitate the evaluation of possible links between them. The lack of correlation between the levels of specific volatile organic compounds and the levels of their precursor metabolites is perhaps at first sight surprising. However, this is not without precedent since the levels of 2-phenylacetaldehyde and 2-phenylethanol have previously been shown to vary greatly independently of the levels of phenylalanine (Tieman *et al.*, 2006). This finding suggests that the rate of volatile production is generally not governed by precursor supply but rather at the transcriptional or post-transcriptional level. Although more studies will be required to understand the complex factors underlying consumer preference in tomato, the results provide several candidate molecules that may be useful leads for this purpose.

## **Acknowledgments**

The help of Nicolas Schauer (metabolite profiling), Aaron Fait (bioinformatic analysis), Emmanuel Botton and Yolanda Carretero (fruit harvests and sample preparations), Karine Robini (sensory data management), Cristina Alfaro and Jaime Primo (volatile analysis) is gratefully acknowledged. Furthermore, we acknowledge the financial support of the trilateral project GENMETFRUQUAL (BMBF FKZ 0313151).

## REFERENCES

- Baldwin EA, Scott JW, Shewmaker CK, Schuch W.** 2000. Flavor trivia and tomato aroma: Biochemistry and possible mechanisms for control of important aroma components. *Hortscience* **35**: 1013-1022.
- Brummell DA, Harpster MH.** 2001. Cell wall metabolism in fruit softening and quality and its manipulation in transgenic plants. *Plant Mol Biol.* **47**: 311-340.
- Buttery RG, Seifert RM, Guadagni DG, Ling LC.** 1971. Characterization of additional volatile components of tomato. *Journal of Agricultural and Food Chemistry.* **19**: 524-529.
- Causse M, Saliba-Colombani V, Lecomte L, Duffé P, Rousselle P, Buret M.** 2002. QTL analysis of fruit quality in fresh market tomato: a few chromosome regions control the variation of sensory and instrumental traits. *J Exp Bot.* **53**: 2089-2098.
- Causse M, Buret M, Robini K, Verschave P.** 2003. Inheritance of nutritional and sensory quality traits in fresh market tomato and relation to consumer preferences. *J Food Sci* **68**, **7**, 2342-2350.
- Chaïb J, Lecomte L, Buret M, Causse M.** 2006. Stability over genetic backgrounds, generations and years of quantitative trait locus (QTLs) for organoleptic quality in tomato. *Theor Appl Genet.* **112**: 934-944.
- Chaïb J, Devaux MF, Grotte MG, Robini K, Causse M, Lahaye M, Marty I.** 2007. Physiological relationships among physical, sensory, and morphological attributes of texture in tomato fruits. *J Exp Bot.* **58**: 1915-1925.
- Decoene C.** 1995. Tomates, qu'en pensent les consommateurs? *Infos-Ctifl* **112**: 8-11.
- Dennison RA, Hall CB, Nettles VF.** 1953. Factors influencing tomato quality. *Proc Fla State Hort Soc.* **65**: 108-111.
- Fernie AR, Trethewey RN, Krotzky AJ, Willmitzer L.** 2004. Metabolite profiling: from diagnostics to systems biology. *Nat. Rev. Mol. Cell. Biol.* **5**:763-769.

**Fernie AR, Tadmor Y, Zamir D.** 2006. Natural genetic variation for improving crop quality. *Curr Opin Plant Biol.* **9**: 196-202.

**Goff SA, Klee HJ.** 2006. Plant volatile compounds: Sensory cues for health and nutritional value? *Science* **311**: 815-819.

**Harker FR, Stec MGH, Hallett IC, Bennet CL.** 1997. Texture of parenchymatous plant tissue: A comparison between tensile and other instrumental and sensory measurements of tissue strength and juiciness. *Postharvest biology and technology.* **11**: 63-72.

**Hovav R, Chehanovsky N, Moy M, Jetter R, Schaffer AA.** 2007. The identification of a gene (*Cwp1*), silenced during *Solanum* evolution, which causes cuticle microfissuring and dehydration when expressed in tomato fruit. *Plant J* **52**: 627-639.

**Janse J, Schols M.** 1995. Une préférence pour un goût sucré et non farineux. *Groenten+Fruit* **26**: 16-17.

**Kader AA, Stevens MA, Albright-Holton M, Morris LL, Algazi M.** 1977. Effect of fruit ripeness when picked on flavour and composition in fresh market tomatoes. *Journal of the American Society for Horticultural Scienc.* **102**: 724-731.

**Kopka J, Schauer N, Krueger S, Birkemeyer C, Usadel B, Bergmuller E, Dormann P, Weckwerth W, Gibon Y, Stitt M, Willmitzer L, Fernie AR, Steinhauser D.** 2005. GMD@CSB.DB: the Golm Metabolome Database. *Bioinformatics.* **21**: 1635-1638.

**Lecomte L, Duffé P, Buret M, Servin B, Hospital F, Causse M.** 2004. Marker-assisted introgression of five QTLs controlling fruit quality traits into three tomato lines revealed interactions between QTLs and genetic backgrounds. *Theor Appl Genet.* **109**: 658-668.

**Lisec J, Meyer RC, Steinfath M, Redestig H, Becher M, Witucka-Wall H, Fiehn O, Törjek O, Selbig J, Altmann T, Willmitzer L** 2006. Identification of metabolic and biomass QTL in *Arabidopsis thaliana* in a parallel analysis of RIL and IL populations. *Plant J* **53**:960-972.

**Lisec J, Schauer N, Kopka J, Willmitzer L, Fernie AR.** 2006. Gas chromatography mass spectrometry-based metabolite profiling in plants. *Nat Protoc.* **1**: 387-396.

**Marsh KB, Rossiter K, Lau K, Walker S, Gunson A, MacRae E.** 2003. Using fruit pulps to explore flavour in kiwifruit. *Acta Hort.* **610**: 229–238.

**Petro-Turza M.** 1987. Flavour of tomato and tomato products. *Food Review International* **2**: 309–351.

**Ratanachinakorn B, Klieber A, Simons DH.** 1997. Effect of short-term controlled atmospheres and maturity on ripening and eating quality of tomatoes. *Postharvest Biology and Technology*. **11**: 149-154.

**Redgwell RJ, Fischer M.** 2002. Fruit texture, cell wall metabolism and consumer perceptions. *Fruit quality and its biological basis*. Editor(s): Knee M. Series: Sheffield Biological Sciences. Publisher: Sheffield Academic Press, Mansion House 19 Kingfield. PP: 46-88.

**Roessner-Tunali U, Hegeman B, Lytovchenko A, Carrari F, Bruedigam C, Granot D, Fernie AR.** 2003 Metabolic Profiling of transgenic tomato plants overexpressing hexokinase reveals that the influence of hexose phosphorylation diminishes during fruit development. *Plant Physiol.* **133**:84-89.

**Rowe HC, Hansen BG, Halkier BA, Kliebenstein DJ.** 2008. Biochemical networks and epistasis shape the *Arabidopsis thaliana* metabolome. *Plant Cell* doi10.1105/tpc.108.058131.

**Saladie M, Matas AJ, Isaacson T, Jenks MA, Goodwin SM, Niklas KJ, Ren XL, Labavitch JM, Shackel KA, Fernie AR, Lytovchenko A, O'Neill MA, Watkins CB, Rose JKC.** 2007. A reevaluation of the key factors that influence tomato fruit softening and integrity. *Plant Physiol* **144**: 1012-1028.

**Saliba-Colombani V, Causse M, Langlois D, Philouze J, Buret M.** 2001. Genetic analysis of organoleptic quality in fresh market tomato. 1. Mapping QTLs for physical and chemical traits. *Theoretical and Applied Genetics* **102**, 259–272.

**Schaffer AA, Petreikov M, Miron D, Fogelman M, Spiegelman M, Bnei-Moshe Z, Shen S, Granot D, Hadas R, Dai N, Levin I, Bar M, Friedman M, Pilowsky M, Gilboa N, Chen L.** 1999. Modification of carbohydrate content in developing tomato fruit. *Hortscience*. **34**: 1024-1027.

**Schauer N, Steinhäuser D, Strelkov S, Schomburg D, Allison G, Moritz T, Lundgren K, Roessner-Tunali U, Forbes MG, Willmitzer L, Fernie AR, Kopka J.** 2005a. GC-MS libraries for the rapid identification of metabolites in complex biological samples. *FEBS Letters*. **579**: 1332-1337.

**Schauer N, Zamir D, Fernie AR.** 2005b. Metabolic profiling of leaves and fruit of wild species tomato: a survey of the *Solanum lycopersicum* complex. *J Exp Bot*. **56**: 297-307.

**Schauer N, Fernie AR.** 2006. Plant metabolomics: towards biological function and mechanism *Trends in Plant Sci*. **11**: 508-516.

**Schauer N, Semel Y, Roessner U, Gur A, Balbo I, Carrari F, Pleban T, Perez-Melis A, Bruedigam C, Kopka J, Willmitzer L, Zamir D, Fernie AR.** 2006. Comprehensive metabolic profiling and phenotyping of interspecific introgression lines for tomato improvement. *Nat Biotech*. **24**: 447-454.

**Schauer N, Semel Y, Balbo I, Steinfath M, Repsilber D, Selbig J, Pleban T, Zamir D, Fernie AR.** 2008. Mode of inheritance of primary metabolic traits in tomato. *Plant Cell*. **20**:509-523.

**Serrano-Megias M, Lopez-Nicolas, JM.** 2006. Application of agglomerative hierarchical clustering to identify consumer tomato preferences: influence of physicochemical and sensory characteristics on consumer response. *Journal of the Science of Food and Agriculture*. **86**: 493-499.

**Spencer JPE, Kuhnle GGC, Hajirezaei MR, Mock HP, Sonnewald U, Rice-Evans C.** 2005. The genotypic variation of the antioxidant potential of different tomato varieties. *Free Radical Research* **39**: 1005-1016.

**Stevens MA.** 1972 Relationships between components contributing to quality variation among tomato lines. *J Am Soc Hort Sci*. **97**: 70-73.

**Szczesniak AS.** 2002. Texture is a sensory property. *Food Quality and preference*. **13**: 215-225. Publisher: Elsevier Sci Ltd, The Boulevard, Langford Lane, Kidlington, Oxford OX5 1GB, Oxon, England.

**Telef N, Stammitti-Bert L, Mortain-Bertrand A, Marcourt M, Carde JP, Rolin D, Gallusci P.** 2006. Sucrose deficiency delays lycopene accumulation in tomato fruit pericarp discs. *Plant Mol Biol* **62**: 453-469.

**Tieman DM, Zeigler M, Schmelz EA, Taylor MG, Bliss P, Kirst M, Klee HJ.** 2006a. Identification of loci affecting flavour volatile emissions in tomato fruits. *J Exp Bot* **57**: 887-896.

**Tieman D, Taylor M, Schauer N, Fernie AR, Hanson AD, Klee HJ.** 2006b. Tomato aromatic amino acid decarboxylases participate in synthesis of the flavor volatiles 2-phenylethanol and 2-phenylacetaldehyde. *PNAS*. **103**: 8287-8292.

**Tikunov Y, Lommen A, de Vos CHR, Verhoeven HA, Bino RJ, Hall RD, Bovy AG.** 2005. A novel approach for nontargeted data analysis for metabolomics. Large-scale profiling of tomato fruit volatiles *Plant Physiol* **139**: 1125-1137.

**Togari N, Kobayashi A, Aishima T.** 1995. Relating sensory properties of tea aroma to gas chromatographic data by chemometric calibration methods. *Food Research International* **28**: 485-493.

**Zierler B, Siegmund B, Pfannhauser W.** 2004. Determination of off-flavour compounds in apple juice caused by microorganisms using headspace solid phase microextraction-gas chromatography-mass spectrometry. *Analytica Chimica Acta* **520**: 3-11.







## **Chapter 3**

**Identification, validation and introgression of QTL involved in tomato fruit volatile composition from red-fruited wild tomato species *Solanum pimpinellifolium* L.**



## ABSTRACT

Volatile organic compounds are major determinants of fruit flavor, which has become a major objective for tomato breeding. A recombinant inbred line (RIL) population consisting of 169 lines derived from a cross between *Solanum lycopersicum* and a red-fruited wild tomato species *S. pimpinellifolium* accession was characterized in three different seasons. Over one hundred quantitative trait loci (QTLs) corresponding to 39 different fruit volatile compounds were identified. Most of these volatile QTLs resulted to be new, whilst only one quarter of them had been previously described in other collections, indicating that there exists ample variability/ opportunity in the tomato clade available for breeding fruit volatile content. Correlation and hierarchical cluster analyses were performed on the 52 volatile compounds identified, providing a tool for the putative assignation of individual compounds to metabolic pathways. Most of the QTLs for apocarotenoid volatiles, which are regarded as desirable for tomato liking, have a positive effect indicating that alleles inherited from *S. pimpinellifolium* are a valuable resource for breeding tomato flavour. An introgression line (IL) population was developed from the same parental genotypes, and 12 of these lines carrying single introgression spanning the regions of QTLs for volatile compounds were characterized over three different locations. The results concluded that almost half of the QTLs previously identified in the RILs maintained their effect on the volatile levels when introgressed in different *S. lycopersicum* genetic backgrounds, reinforcing the value of these QTLs for flavour/aroma breeding in cultivated tomato.



## INTRODUCTION

Tomato is cultivated worldwide, being one of the most important vegetable fruit crops in the world, and also a model for the study of the processes involved in fruit ripening. During the XX century, important breeding efforts have been carried out in order to generate mainly high yielding and disease resistance varieties. However, fruit quality and consumer preferences usually have not been taken into account in breeding programs, leading to a reduction of the taste appreciated by consumers. Tomato taste is a complex trait, composed by primary characteristics such as sugar content and acidity and also a large number of volatile compounds that have an important impact on the flavour (Klee and Tieman, 2013).

Breeding for improved flavour by controlling the production of volatile compounds in the fruit has been seldom attempted due to many factors. First, our perception of flavour is influenced by the interaction of a relatively large number of both volatile (around 16-20 or more), (Buttery, 1993, (Tieman et al., 2012) and non-volatile compounds (Abegaz et al., 2004, Baldwin et al., 2008, Tieman et al., 2012) which makes the task of identifying which compounds would be desirable to enhance or reduce a difficult one. Second, volatile composition is under polygenic genetic control, and the genetic basis of the mechanisms controlling their levels is largely unknown. Third, volatile composition also depends on agronomic management, environment and genetic background. Therefore, the identification of the regions in the genome modifying the levels of volatiles in the ripe fruit is a necessary tool to understand the genetic control and for breeding tomato flavour.

Information on the genetic control is limited. QTL have been identified in experimental populations obtained from crosses between tomato cultivars

with different germplasm sources such as cherry tomato (Saliba-Colombani et al., 2001, Zanor et al., 2009) or green-fruited distantly related wild tomato species *Solanum pennellii* (Tadmor et al., 2002, Tieman et al., 2006b) and *S. habrochaites* (Mathieu et al., 2009). In the case of the latter two species, wild chromosomal regions were introgressed in processing cultivars, whereas in the first example, two fresh-market varieties were used.

In the last decade, a number of genes involved in different metabolic pathways in the tomato fruit have been identified, shedding some light into the biosynthesis of volatile compounds. Several genes have been characterized involved in the biosynthesis of fatty acid-derived volatiles (Matsui et al., 2000, Chen et al., 2004, Shen et al., 2014, Matsui et al., 2007, Speirs et al., 1998), apocarotenoids (Simkin et al., 2004), esters (Goulet et al., 2012, Goulet et al., 2015), phenylpropanoids (Tieman et al., 2010, Mageroy et al., 2012), other phenylalanine-derived volatile compounds (Tieman et al., 2006a, Tieman et al., 2007) and the role of conjugation in the accumulation and emission of volatiles (Louveau et al., 2011, Tikunov et al., 2013). Nevertheless, we have only achieved a partial knowledge of the metabolic pathways and the genes involved in volatile biosynthesis and regulation.

In the present work, QTLs for a large set of volatile compounds were identified from *S. pimpinellifolium*, a red-fruited close relative to cultivated tomato. Most of the genetic regions containing these loci were introgressed in a fresh market tomato variety, confirming their value to be used in breeding programs for the tomato fresh market. Also, genes and associated functions located within the QTL regions are provided, which will facilitate the identification of candidates for those genes involved in volatile biosynthesis / regulation.



## MATERIAL AND METHODS

### *Plant material*

The Recombinant Inbred Line (RIL) population originated from an interspecific cross between *S. lycopersicum* cv. 'MoneyMaker' and *S. pimpinellifolium* accession TO-937, thus producing F<sub>1</sub> seeds. Selfing of a single F<sub>1</sub> plant generated an F<sub>2</sub> segregating population from which a RIL population consisting of 169 F<sub>7:8</sub> lines was generated by the *single seed descendant* (SSD) method (Alba et al., 2009). Five plants of each RIL were grown during winter-spring cycle in a plastic greenhouse under standard commercial growing conditions and three biological replicates were analysed for each line, each replicate consisting on a mixture of 3 to 5 different red ripe fruits obtained from trusses 2 to 4. This experimental design was repeated for 3 independent years.

An Introgression Line (IL) collection was generated previously from the same original genotypes than the RILs (Barrantes et al., 2014). Twelve ILs covering different regions harbouring selected volatiles and the recurrent parent 'MoneyMaker' were cultivated under plastic greenhouse during the summer of 2013 in three Spanish locations: Alginet (Coagri cooperative), Orihuela (Miguel Hernández University) and Málaga (Institute for Mediterranean and Subtropical Horticulture "La Mayora", IHSM-UMA-CSIC). The experimental design followed a randomized block with 8 blocks, with one single plant replicate per IL and 6 replicates of 'MoneyMaker' in each block. Eight independent biological replicates were analysed per genotype, each consisting of between 3 to 5 fruits of a single plant, which were harvested and pooled before processing.

### ***Analysis of volatile compounds***

For the analysis of volatile compounds, tomato fruits were collected at the red ripe stage, carefully washed with water and dried with a paper cloth to avoid external contamination from trichome secretions from stem / leaves. A slice of pericarp (avoiding the septa) was excised from each fruit taking care to remove all the locular tissue, and immediately frozen in liquid nitrogen.

Volatile compounds were captured by means of headspace solid phase microextraction (HS-SPME) and separated and detected by means of gas chromatography coupled to mass spectrometry (GC/MS). Samples were processed similarly as described in Rambla et al. (2015). Roughly, frozen tomato pericarp was ground in a cryogenic mill and stored at -80°C until analysis. Five hundred milligrams of the resulting powder were introduced in a 7 mL glass vial and incubated at 37°C for 10 min in a water bath. Five hundred mL of an EDTA 100 mM, pH 7.5 solution and 1.1 g of CaCl<sub>2</sub>·2H<sub>2</sub>O were added, mixed gently and sonicated for 5 min. One mL of the resulting paste was transferred to a 10 mL screw cap headspace vial with silicon/PTFE septum and analysed within 12 hours. Volatile compounds were extracted from the headspace by means of a 65 µm PDMS/DVB solid phase microextraction fiber (SUPELCO). In the analysis of the RIL population in the first season, HS-SPME extraction was performed manually as follows. Vials were first incubated at 50°C for 10 min with agitation at 500 rpm in a Thermomixer Comfort (Eppendorf), then the fiber exposed to the headspace of the vial for 20 min under the same conditions of temperature and agitation. Volatiles adsorbed in the fiber were desorbed at 250°C for 1 min in the injection port of a Clarus 500 gas chromatograph (Perkin Elmer) in splitless mode. After desorption, the fiber was cleaned in another injection port at 250°C for 5 min under a nitrogen flow. Chromatography was performed on a ZB-5 (30 m, 0.25 mm, 0.25 µm) column with helium as carrier gas, at a constant

flow of 1.2 mL/min. The GC interface and MS source temperatures were 260°C and 180°C respectively. Oven programming conditions were 40°C for 2 min, 5°C/min ramp until 180°C, then a 15°C/min ramp until 250°C, and a final hold at 250°C for 4 min. Data was recorded in a Clarus 500 mass spectrometer (Perkin Elmer) in the 35-250 m/z range at 5 scans/s, with electronic impact ionization at 70 eV. Chromatograms were processed by means of the TurboMass software version 5.0 (Perkin-Elmer). For the rest of the analyses, volatile extraction was performed automatically by means of a CombiPAL autosampler (CTC Analytics) with identical settings of agitation, time and temperature. Desorption was performed at 250°C during 1 min in splitless mode in the injection port of a 6890N gas chromatograph (Agilent Technologies). After desorption, the fiber was cleaned in an SPME fiber conditioning station (CTC Analytics) at 250°C for 5 min under a helium flow. Chromatography was performed on a DB-5ms (60 m, 0.25 mm, 1.00 µm) column with helium as carrier gas, at a constant flow of 1.2 mL/min. The GC interface and MS source temperatures were 260°C and 230°C respectively. Oven programming conditions were 40°C for 2 min, 5°C/min ramp until 250°C, and a final hold at 250°C for 5 min. Data was recorded in a 5975B mass spectrometer (Agilent Technologies) in the 35-250 m/z range at 7 scans/s, with electronic impact ionization at 70 eV. Chromatograms were processed by means of the Enhanced ChemStation E.02.02 software (Agilent Technologies).

Identification of compounds was performed by the comparison of both retention time and mass spectrum with those of pure standards. All the standards were purchased from Sigma-Aldrich, except 1-nitro-2-phenylethane which was supplied by Apin Chemicals.

For quantitation, one specific ion was selected for each compound, and the corresponding peak from the extracted ion chromatogram was integrated. The criteria for ion selection were the highest signal-to-noise ratio and being

specific enough in order to provide good peak integration in that particular region of the chromatogram. An admixture reference sample was prepared for each season by mixing thoroughly equal amounts of each sample. A 500 mg aliquot of the admixture was analysed regularly (one admixture for every six to seven samples) and processed as a regular sample as part of the injection series. This admixture contained all the compounds identified in any of the samples at an intermediate concentration between those in the individual samples, and it was used as a reference to normalize for temporal variation and fiber aging. Finally, the normalized results (corrected for temporal variation and fiber aging) for a sample were expressed as the ratio of the abundance of each compound in that particular sample to those present in the reference admixture.

Thirty-two compounds were identified and quantified in the samples from the RIL population in the first season, whilst the number in all the remaining analyses was fifty-two.

### ***Genotyping and map construction***

RILs and parental lines were genotyped with the 8K SolCap Illumina Infinium SNP tomato array (Sim et al., 2012) at the INCLIVA Genotyping and Genetic Diagnosis Unit (Valencia, Spain). Monomorphic markers, as well as markers with high number of missing values (>30%) or unknown parental genotypes were removed before further analysis. Linkage analysis and map construction were performed using the mapping software JoinMap 4 (Kyazma B.V, Netherlands) (Van Ooijen, 2006) using the maximum likelihood mapping algorithm. Markers were assigned to linkage groups at a minimum LOD value of 5.0.

### ***QTL analysis in RILs***

In order to facilitate computational analysis, the genetic map was condensed to 297 markers. QTL analysis was performed with the log 2 transformed data in each season independently by Composite Interval Mapping with WinQTLCart (Wang S., C. J. Basten, and Z.-B. Zeng (2012). Windows QTL Cartographer 2.5. Department of Statistics, North Carolina State University, Raleigh, NC. (<http://statgen.ncsu.edu/qtlcart/WQTLCart.htm>). The confidence interval of QTL position was established to 2-LOD units from the maximum LOD. QTLs detected in at least two seasons were retained for further discussions, discarding QTLs detected in a single season. Genetic and QTL map were drawn with Mapchart 2.2 (Voorrips, 2002).

### ***Other statistical analysis***

All the data obtained from volatile analyses of the samples was log 2 transformed before all the statistical analyses in order to achieve a normal distribution.

Correlation matrices between metabolites were performed from the data of individual samples of seasons 2 and 3 separately, and also with the around one thousand samples of both seasons together. Data from season 1 were not used, as it contained a lower number of compounds. Pearson correlations were calculated by means of the SPSS 16.0 software.

Two-way ANOVA were performed for the determination of the effects of genotype, environment and the interaction genotype x environment on the levels of each of the volatiles determined. For the RIL population, data from

seasons 2 and 3, in which the same number of compounds were analysed, was used. For the IL population, data from the three different locations was utilized. Analysis of variance was performed by means of the software Statgraphics Centurion XVI.II 64 bits.

To evaluate statistical significance between the average of the volatiles in each of the ILs and those in the parental 'Moneymaker', Dunnett's test ( $p < 0.05$ ) was used. We considered that a QTL previously identified in the RILs was confirmed in the respective IL when the level of that particular compound was significantly different in at least one of the three locations evaluated. Our criterium for considering a new volatile QTL in was that it was observed in fruits of an IL in at least two of the three locations. Dunnett's test was performed by means of the JMP 12 software.

### ***Hierarchical Cluster Analysis and cluster similarity***

A hierarchical clustering based on correlations (Eisen et al., 1998) was performed over the combined dataset obtained from the RILs of seasons 2 and 3. It was also performed over each dataset individually. To assess the similarity between clusterings with statistical significance, the metric proposed in (Fowlkes and Mallows, 1983) was used and random datasets were generated by permutation (bootstrapping). Data were analyzed with Matlab (MathWorks).

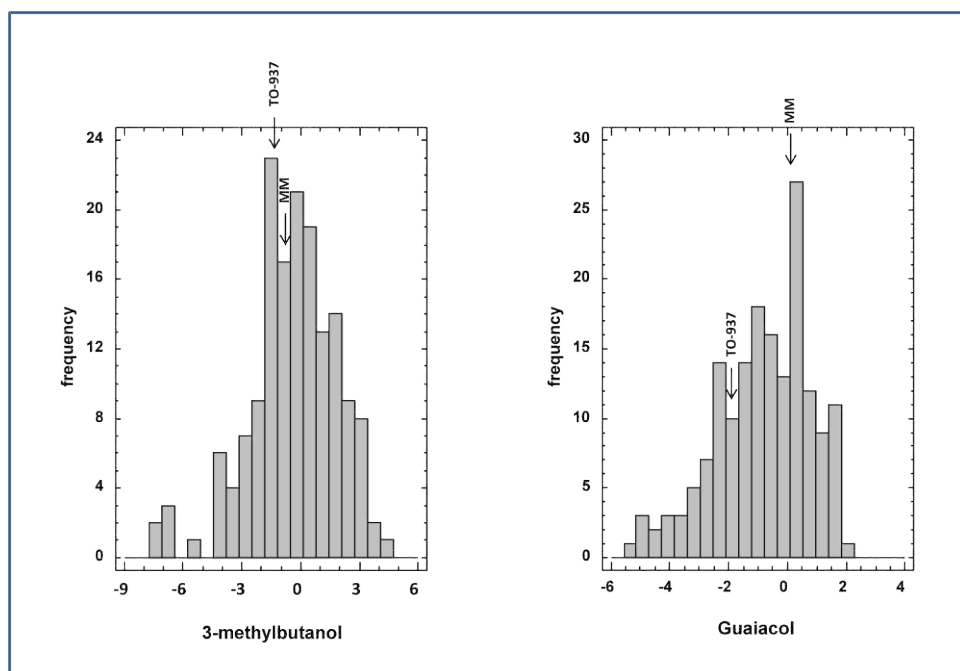
## RESULTS AND DISCUSSION

A *S. lycopersicum* x *S. pimpinellifolium* recombinant inbred line (RIL) population has been used in this work which had previously been employed for the localization of QTL involved in pest resistance (Salinas et al., 2013) and other fruit-quality components (Capel et al., 2015). Line TO-937 was derived from a *S. pimpinellifolium* accession originally collected in Lambayeque, Peru (Manuel Alba et al., 2009), which produces the species' typical small, red fruits containing higher levels of flavour and a number of healthy compounds such as sugars, organic acids, vitamin C or carotenoids (Capel et al., 2005). 'Moneymaker' is a medium-size fruit cultivar for fresh market with indeterminate growth habit that is commonly used in genetic, physiological, phytopathological and developmental studies. The two species are close relatives and, contrarily to wild green-fruited tomato species that have been used for fruit quality traits genetic analyses, they share common characteristics for fruit ripening, metabolite profiling, etc. Additionally, no major cross-compatibility is present that may cause strong segregation distortions or the elimination of introgressed genomic regions in the mapping populations.

### The volatile network

Fifty-two different volatile compounds were unequivocally identified in the ripe fruit samples obtained from the RIL population. A continuous variation was observed, with transgressive segregation in both directions for all the compounds (Figure S1), showing a large degree of variation for most of the volatiles, with up to 15 of them displaying over 100-fold difference in the levels between the RILs with extreme values. Compounds displaying the highest degree of variation included phenylpropanoids such as guaiacol, eugenol or

methyl salicylate, other phenylalanine derivatives such as 1-nitro-2-phenylethane, benzylnitrile or 2-phenylethanol, and branched-chain amino acid related compounds such as 2-isobutylthiazole, 3-methylbutanol, 2-methylbutanol, 3-methylbutanal, 3-methylbutanoic acid or 3-methylbutanenitrile. It must be noted that broad variability was detected not only for volatiles that were differentially accumulated among the parents of the population (i.e. the case of guaiacol) but also for other ones that showed very similar levels in both of them (i.e. 3-methylbutanol) as illustrated in Figure 1. On the other hand, a minority of volatiles showed a relatively low degree of variation, such as fatty acid derivatives octanal, nonanal and decanal, with a range below 5-fold between the RILs with extreme values (Table S1).

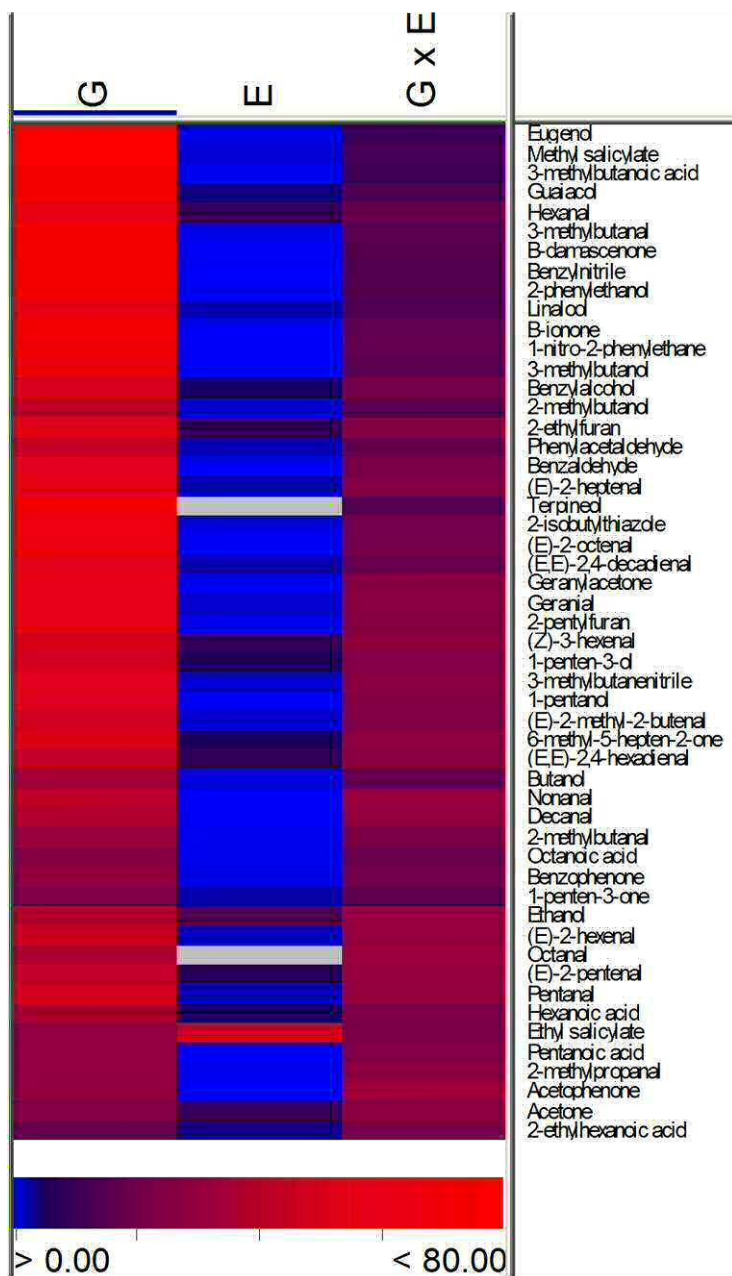


**Figure 1.** Distribution histogram for 3-methylbutanol and guaiacol in the RIL population. The position of the parents *S. lycopersicum* cv. 'Moneymaker' (MM) and *S. pimpinellifolium* accession TO-937 are indicated with an arrow.



This observation indicates that by combining the genomes of the two parents in the RIL population we generated new unpredictable variability in fruit volatile content.

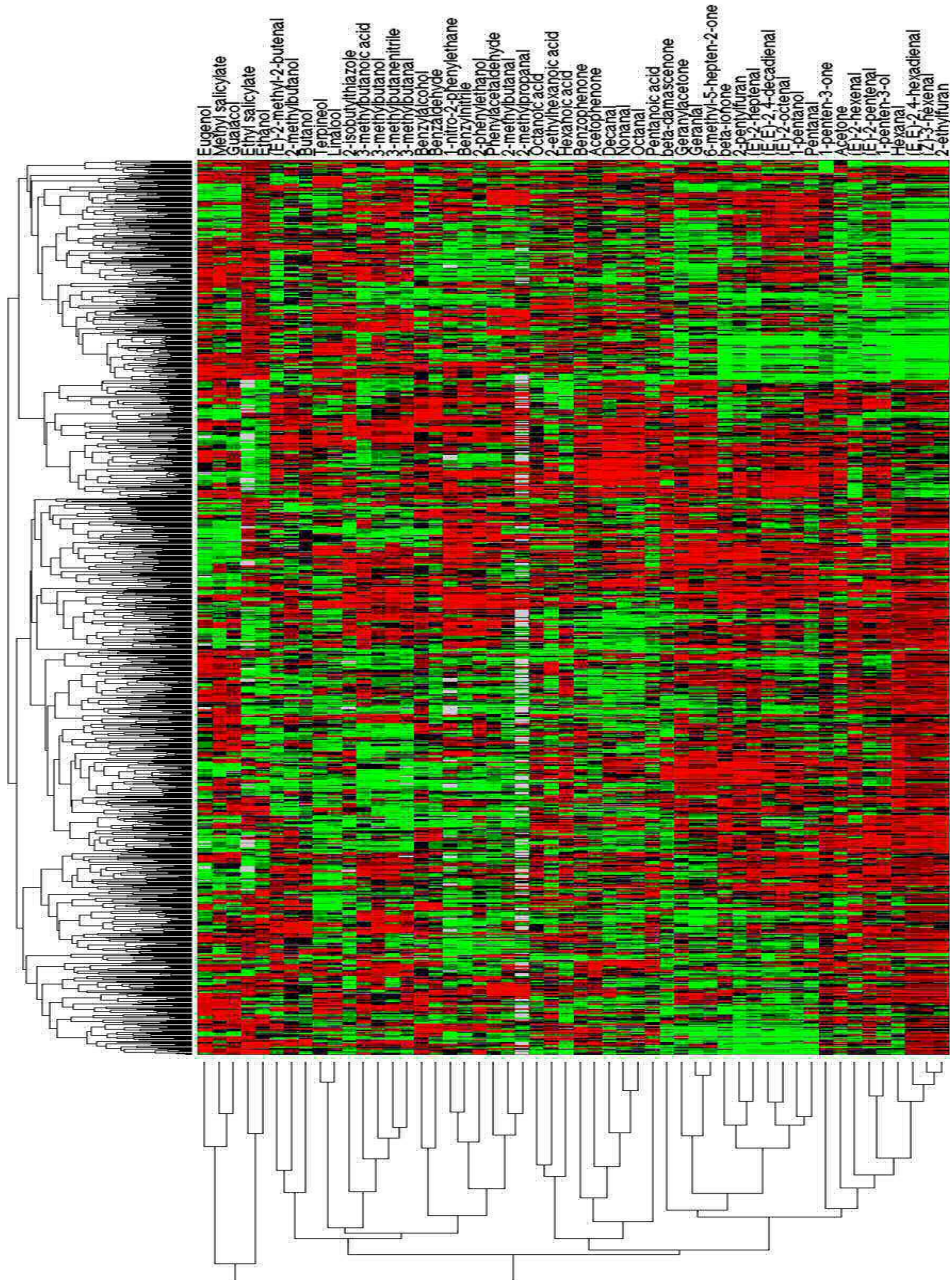
Statistical analysis on the volatile levels revealed that the genotype had a highly significant effect on all the volatiles profiled, being the main factor explaining the existing variability (Table S2). The genotypic effect was particularly high in the case of some apocarotenoid volatiles such as  $\beta$ -ionone and  $\beta$ -damascenone, several branched-chain amino acid related volatiles, phenylalanine derivatives and particularly phenylpropanoids such as eugenol and methyl salicylate, for which the genotype explained over 80% of the total variance. Interestingly, most of the traits with higher heritability were also among those with a larger degree of variation between the RILs. For compounds with low heritability, the genotypically explained variability scored about 20%, and included volatiles from different metabolic pathways such as fatty acid derivatives 2-ethylhexanoic acid and 1-penten-3-one. It should be noted that the interaction between genotype and environment was also highly significant for almost the totality of the volatiles, and accounted for about 15-30% of the total variability for most of the compounds. The effect of the environment was significant but remarkably low for most of the volatiles with just a few exceptions, such as in the case of ethyl salicylate (Figure 2, Table S2). Therefore, this suggests that the environment – at least under the conditions of our experiments – did not have a general effect on the variability in volatile biosynthesis, which was modulated mostly by the particular combination of genes in each RIL.



**Figure 2.** Heat map with the effect of genotype (G), environment (E) and the genotype-environment (GxE) interaction for each volatile in the RILs population. Data corresponds to the percentage of variability explained, according to the scale below. Red colour corresponds to high values (up to 80%); blue corresponds to low values. The exact values are detailed in Table S2.

## Correlation between metabolites

A correlation matrix was created with the results obtained from the analysis of volatiles for each of the two seasons separately and together in order to unravel the relationship between all the 52 volatile compounds identified in the ripe fruit (Table S3). Statistical analyses were performed from over 1000 biological samples corresponding to 169 independent recombinant inbred lines derived from the cross between cultivated tomato and red-fruited wild relative *S. pimpinellifolium*, each line harbouring about one half of the randomly inherited genome of each parental. Therefore, correlation obtained from this set of data can be considered as a reliable, largely unbiased description of the relationship between volatile compounds in tomato. Provided that *S. pimpinellifolium* is the closest wild relative of tomato, it could be assumed that each RIL represents different alterations of the tomato volatile network, but not as drastic as those imposed by combining more distant relatives. Correlation results of both seasons matched largely, as revealed by hierarchical clustering (HCA) based on correlations performed on the dataset of one season or on the other. The similarity between clusterings was assessed following the metric proposed by Fowlkes and Mallows (1983), which consists in cutting the two hierarchical trees to then count the number of matching entries in a number of clusters in each tree. We obtained a clustering comparison metric of 0.45 for 10 clusters (note that this metric ranges from 0 to 1; the higher the metric, the more similarity). To obtain the corresponding statistical significance, we kept the dataset of year 2011 and generated 1,000 random datasets by permutation of the values in the dataset of year 2006 (bootstrapping). We got a  $P$ -value = 0, and a mean clustering comparison metric of 0.13. Note that for 5 clusters, we also got a  $P$ -value = 0, with 0.53 the clustering comparison metric between seasons, and 0.27 the metric for random comparison (Figure S2).



**Figure 3.** Heat map of volatiles in the RILs population in seasons 2 and 3 and hierarchical cluster of both volatiles and RILs. Red colour corresponds to higher values; green colour corresponds to lower values.

Hierarchical cluster analysis was also performed using the results of the volatile levels in each of the samples of both seasons together, producing clusters for both volatile compounds and the samples, thus facilitating the classification of both the compounds and the RILs according to their respective correlations (Figure 3, Figure S2).

Paralelism in the levels of different compounds in the RIL population such as those shown in both the correlation matrices and the HCA may be due to different factors. There exists the possibility that two or more genes controlling the metabolism of different compounds are physically linked very close one to another, so that recombination between them is rare and the lines with a particular allele of one of the genes also inherited the same allele of the other(s). Nevertheless, the most suitable explanation in most of the cases is that significant correlations between volatiles are due to the fact that there is some common process in their biosynthesis, such as sharing the same precursor, or an enzyme in the pathway, or there is an independent process favouring both pathways (i.e. cell wall degradation which would facilitate the availability of different substrates). Therefore, cluster analysis has already been used to propose an assignment of volatile compounds to metabolic pathways (Tikunov et al., 2005, Mathieu et al., 2009), consisting on a valuable tool to unravel the biosynthesis of those compounds for which a metabolic pathway has not been well established. Thus, the most relevant results obtained from both hierarchical cluster and correlation analyses will be discussed thereafter in these terms.

According to its chemical structure, geranial is a monoterpenic aldehyde, and could be expected to cluster with the monoterpenoids linalool and terpineol. Nevertheless, geranial is located in the apocarotenoid cluster, as previously noticed (Tikunov et al., 2005) and shows a very tight positive

correlation with 6-methyl-5-hepten-2-one (0.94), thus suggesting a common substrate in the biosynthesis of both compounds, probably the linear carotenoid lycopene, and maybe also a common carotenoid cleavage dioxygenase enzyme with both 5,6 and 7,8 cleavage activities.

In the case of  $\beta$ -damascenone, it has been proposed that  $\beta$ -carotene could be its precursor, based on this compound clustering with  $\beta$ -ionone in fruit volatile datasets obtained from introgression lines derived from other wild tomato species (Mathieu et al., 2009). Nevertheless, in our RIL population, levels of  $\beta$ -damascenone did not tightly correlate with  $\beta$ -ionone. On the contrary, it was located in other cluster, with linear apocarotenoids 6-methyl-5-hepten-2-one, geranylacetone and geranial, to which it showed even slightly higher correlation (0.30-0.33) than with  $\beta$ -ionone (0.26). Therefore, this reinforces the hypothesis that precursors other than  $\beta$ -carotene might be reconsidered, such as the previously proposed neoxanthin (Skouroumounis et al., 1993).

It has been proposed that the disruption or alteration in permeability of the thylakoid membranes during the chloroplast to chromoplast transition would facilitate the access of both lipoxygenases and carotenoid cleavage dioxygenases to their substrates, non-esterified fatty acids and carotenoids respectively (Mathieu et al., 2009). Our data may support this hypothesis, considering that  $\beta$ -ionone showed the highest correlations with fatty acid derivatives such as (*E*)-2-heptenal and hexanal (0.74 and 0.71 respectively).  $\beta$ -ionone also shows notably higher correlation with geranylacetone than with any other apocarotenoid, which is consistent with both compounds resulting from reactions catalyzed by the same enzyme, as seems to be the case for LeCCD1 enzymes with 9,10 cleavage specificity acting on both cyclic and non-cyclic carotenoid substrates (Simkin et al., 2004).

Two of the volatile compounds located in the fatty acid derivatives cluster are 2-ethylfuran and 2-pentylfuran. A metabolic pathway has not been elucidated for these compounds, but cluster analysis on independent materials of tomato (Tikunov et al., 2005; own unpublished data) and strawberry (Zorrilla-Fontanesi et al., 2012) have repeatedly suggested that both could be derived from fatty acids. This hypothesis is strengthened by the fact that C<sub>9</sub> volatile 2-pentylfuran clustered with C<sub>8</sub>-C<sub>10</sub> fatty acid derivatives in our RIL population, whilst 2-ethylfuran (a C<sub>6</sub> compound) clustered with double-bonded C<sub>6</sub> fatty acid derivatives. The latter volatile showed a remarkably high correlation with (*Z*)-3-hexenal (0.95), therefore suggesting that (*Z*)-3-hexenal might be postulated as a plausible precursor of 2-ethylfuran.

### **Identification of volatile QTLs**

The 8K SolCap Illumina Infinium SNP tomato array (Sim et al., 2012) containing 7720 SNPs was used to genotype the *S. lycopersicum* x *S. pimpinellifolium* RIL population. Out of the 7720 markers, 2941 (38%) were monomorphic in the RILs and 538 (7%) could not be reliably genotyped. Therefore, the remaining 4241 markers (circa 55% of the SNPs included in the array) were used for the generation of a genetic linkage map. The markers in the resulting map were assigned to 1704 unique map locations at an average genetic distance of 1.0 cM and a maximum distance of 8 cM. For QTL analysis, a skeleton map was developed consisting of 297 markers evenly distributed along the tomato genome.

A total number of 102 metabolite QTLs distributed along all the 12 tomato chromosomes were identified for 39 of the 52 volatile compounds identified in the samples obtained from the 169 RILs (Figure 4, Table S4).

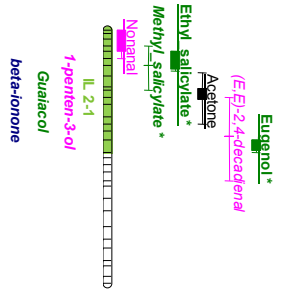
Chromosome 1 contained a particularly high number of QTLs – a total of 31 for compounds synthesized via different metabolic pathways. In total, the alleles inherited from *S. pimpinellifolium* (SP) produced an increase in the levels of volatiles in 42 loci, whilst it produced reduced levels in the remaining 60 loci. Regarding the metabolic pathways, the production of phenolic volatile compounds tended to be increased by *S. lycopersicum* (SL) alleles, while levels of apocarotenoid volatiles were more likely to be increased by the alleles coming from *S. pimpinellifolium*. In the remaining metabolic pathways, no clear tendency was observed. These results point to *S. pimpinellifolium* as a species with a potential in terms of breeding for cultivated tomato flavour, as apocarotenoid volatiles have been highlighted as important contributors to tomato liking (Vogel et al., 2010, Tieman et al., 2012).

### *Fatty acid derivatives*

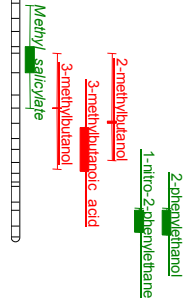
QTLs modifying the levels of volatiles derived from fatty acids were identified in five different chromosomes. It may be highlighted that a region located at the top of chromosome 1 severely affected the levels of 10 different volatile compounds. The SP allele decreased levels of six C<sub>6</sub> volatiles plus C<sub>7</sub> (*E*)-2-heptenal, while the levels of the non-C<sub>6</sub> aliphatic aldehydes pentanal, octanal and nonanal were increased, indicating an opposed effect of this region on the production of these two groups of compounds, which actually clustered independently and in most cases (with the exception of pentanal) showed significant negative correlations (Table S3). The metabolic pathway of many of these compounds is known and some key genes involved in their biosynthesis in tomato fruit have been described. Free fatty acids are first oxidized by means of lipoxygenase enzymes (LOX), and the resulting molecule is subsequently broken



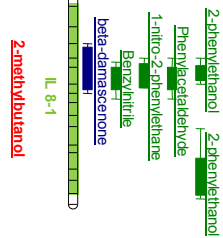
Chr2



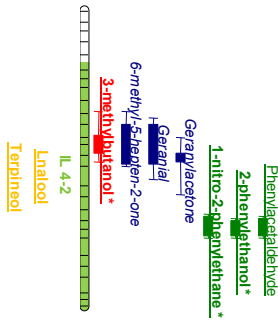
Chr5



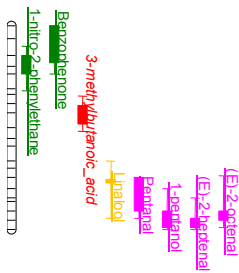
Chr8



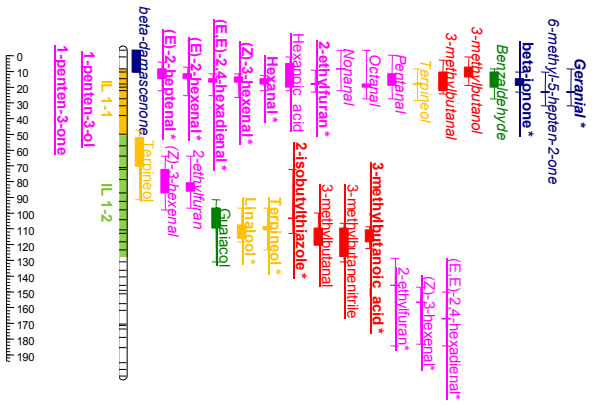
Chr4



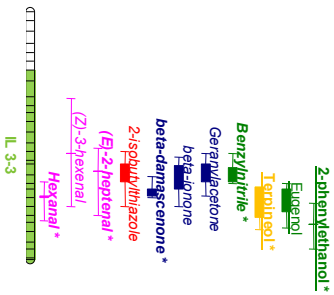
Chr7



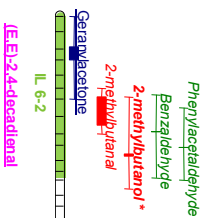
Chr1

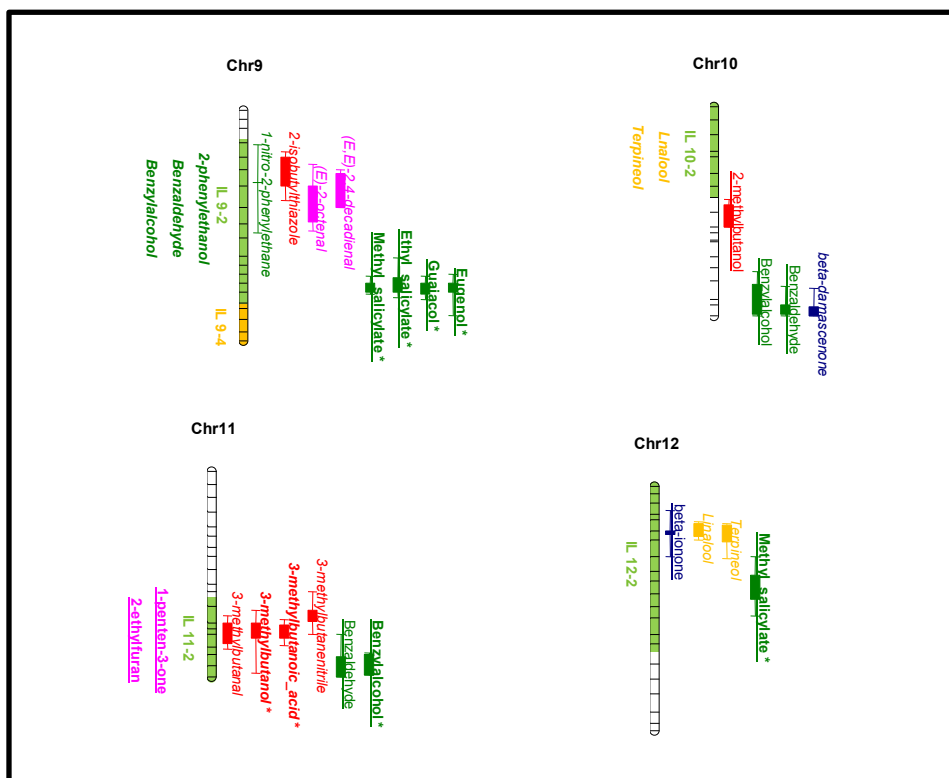


Chr3



Chr6





**Figure 4.** Volatile QTL map obtained from the RIL and IL populations. Volatiles are identified with colours according to their metabolic pathway: pink, fatty acid derivatives; green, phenolic compounds; red, branched-chain amino acid related compounds; orange, terpenoids; blue, apocarotenoids. Loci in which SP alleles induced higher levels of volatiles are marked in italics; those in which SP alleles induced lower levels are marked with plain underlined text. QTLs identified in the RIL population are shown on the right side of each chromosome. The interval corresponds to extreme values of LOD-2 and LOD+2 in the different seasons; the central solid area corresponds to the interval overlapping in the different seasons. QTLs confirmed in the ILs are highlighted in bold and marked with an asterisk. New QTLs identified in the ILs are shown in the left side of each chromosome.

down by means of hydroperoxide lyases (HPLs) to produce one non-polar volatile metabolite and one polar non-volatile compound (Rambla et al., 2014). Lipoxygenase gene *TomloxC* (Chen et al., 2004) has been mapped in this region of chromosome 1, being a reasonable candidate gene for this major QTL, which in some seasons accounts for over 50% of the total variability of some compounds. Hydroperoxide lyase gene *LeHPL* (Matsui et al., 2000) mapped in the bottom of the same chromosome in a region where three C<sub>6</sub> volatile QTLs were identified. Therefore, it is possible that this gene also has an effect on the levels of fatty acid-derived volatiles, although at a much lower extent than *TomloxC*, as revealed by the R<sup>2</sup> value (Table S4).

### *Phenolic compounds*

Over thirty volatile QTLs were identified for phenolic compounds spread across all the tomato chromosomes. Major QTLs accounting for over one-third of total variability (Table S4) were identified for guaiacol, eugenol, methyl salicylate and, in a lesser extent, ethyl salicylate in the bottom of chromosome 9, in the same region where the *NSGT1* glycosyltransferase is mapped (Tikunov et al., 2013). Many minor QTLs for these compounds were identified in other locations, including one for guaiacol in the same region of chromosome 1 where *SIUGT5*, a gene involved in the glycosilation of many phenylpropanoids, is located (Louveau et al., 2011). Additionally, QTLs for phenylalanine-derived volatiles benzylnitrile, 1-nitro-2-phenylethane, phenylacetaldehyde and 2-phenylethanol mapped in the same region of chromosome 8 as *LeAADC* genes, which had been reported to be involved in their biosynthesis (Tieman et al., 2006a).

### *Branched-chain amino acid related compounds*

QTLs for several branched-chain volatiles with a chemical structure resembling those of amino acids leucine or isoleucine have been identified in chromosomes 1, 3, 4, 5, 6, 7, 9, 10 and 11. The biosynthetic pathway of these volatiles has not been established in plants yet, but it had been considered that these branched-chain amino acids would be their direct precursors (Mathieu et al., 2009). Nevertheless, recent evidence indicates that it is more likely that these volatiles are derived from their respective keto acids rather than from the amino acids (Kochevenko et al., 2012). Although QTLs controlling branched-chain volatiles were identified in most of the chromosomes, those for structurally leucine-related volatile compounds (3-methylbutanal, 3-methylbutanol, 3-methylbutanenitrile, 3-methylbutanoic acid and 2-isobutylthiazole) were in the same position than those isoleucine-related (2-methylbutanal, 2-methylbutanol, (*E*)-2-methyl-2-butenal) only in chromosome 5. This is indicative that most of the regulation of each group of compounds is independent from one another, which is confirmed by the fact that significant but low positive correlation between both groups was observed.

The only sulfur compound identified in our samples was the branched-chain volatile 2-isobutylthiazole. QTLs for this compound were identified in chromosomes 1, 3 and 9. It co-localized in chromosome 1 with QTLs for 3-methylbutanol, 3-methylbutanoic acid and 3-methylbutanenitrile, the SP alleles decreasing the levels of them all. In the other two QTLs identified for 2-isobutylthiazole there was no co-localization with loci controlling other volatiles in the same pathway, which points to those regions as the ones harbouring the genes responsible of the addition of the sulfur moiety.

## *Terpenoids*

Very little is known about terpenoid biosynthesis in the tomato fruit, and only two of the compounds in this pathway were identified in the samples analyzed. Three and five QTLs have been identified regulating the levels of terpenic alcohols linalool and terpineol respectively, in chromosomes 1, 3, 7 and 12. Both compounds provide floral notes, although it is considered that they do not have an effect on tomato flavour (Buttery, 1993). SP alleles are associated with reduced levels in all the QTLs identified except for those in chromosome 12. Therefore, this would be the target of choice for breeding enhanced levels of these potentially desirable volatiles. The chromosome regions where a QTL for this compound map may contain terpene synthases or other genes involved in their biosynthesis and can be used for the search of candidate genes in these regions of the tomato genome.

## *Apocarotenoids*

Fourteen quantitative trait loci for a set of volatiles derived from carotenoids were identified in chromosomes 1, 3, 4, 6, 8, 10 and 12. A cluster of QTL controlling the levels of linear apocarotenoids geranylacetone, geranial and 6-methyl-5-hepten-2-one was identified in chromosome 4. Another cluster of QTL was identified in chromosome 3 with an effect of both linear and cyclic C<sub>13</sub> apocarotenoid volatiles. In both cases the SP alleles induced increased levels. Interestingly, chromosome 1 harbours a region containing QTLs with an opposed effect on linear compounds 6-methyl-5-hepten-2-one (C<sub>8</sub>) and geranial (C<sub>10</sub>) (alleles from SP increasing their levels) and cyclic C<sub>13</sub> compound β-ionone (decreased by SP alleles). It has to be noted that carotenoid cleavage dioxygenase 1 genes (*LeCCD1A* and *LeCCD1B*), which have been described to

participate in the biosynthesis of C<sub>13</sub> apocarotenoids (Simkin et al., 2004), also map in this region of chromosome 1.

Up to 9 QTLs for the carotenoids lycopene and  $\beta$ -carotene had been identified previously in the same RIL population (Capel et al., 2015). Nevertheless, apocarotenoid volatiles QTLs in general tended not to co-localize with carotenoid QTLs, with the exception of the QTLs for 6-methyl-5-hepten-2-one, geranial and geranylacetone in chromosome 4 which co-localized with QTLs for both lycopene (the plausible precursor of them all) and  $\beta$ -carotene. This indicates the main determinants of the production of apocarotenoid volatiles are factors other than substrate accumulation. Therefore, all the other QTLs identified could be due to still unknown genes located in these regions implied in either volatile biosynthesis final steps directly, or in the accumulation of possible precursors other than the already described lycopene and  $\beta$ -carotene, or else in other processes facilitating the interaction between substrates and biosynthetic enzymes.

### **Introgression and validation of the volatile QTLs in a fresh market tomato genetic background**

To validate the effects of these selected QTLs, twelve ILs previously generated from the same cross (Barrantes et al., 2014) with SP introgressions covering QTL regions described above (Figure 4) were grown at three different locations, and the fruit volatile composition determined.

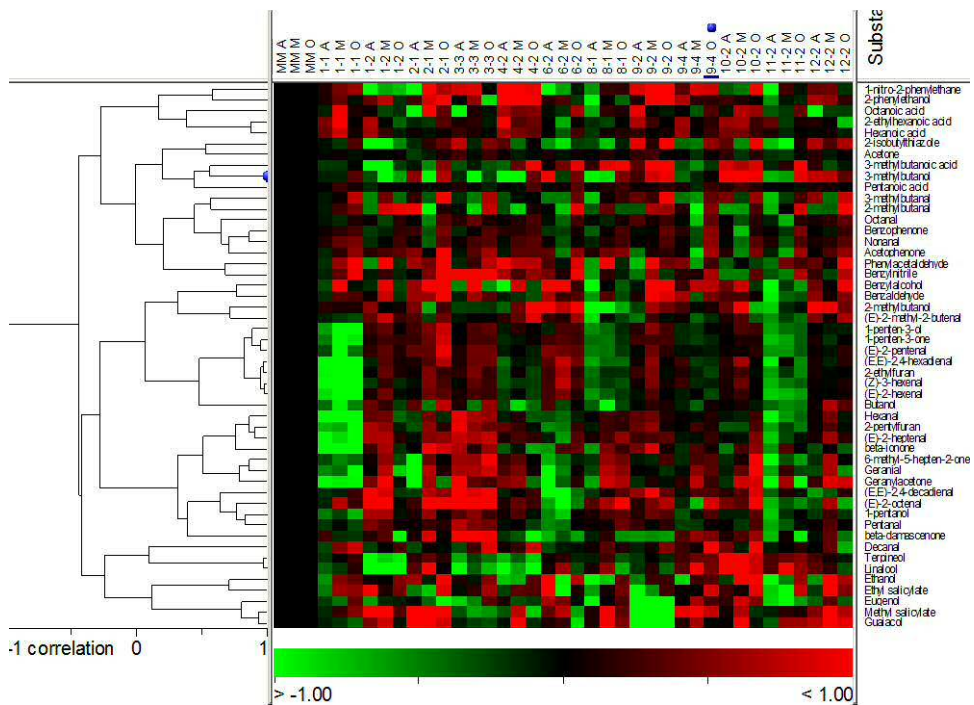
Unlike in the RIL population, most of the variability observed for the volatile levels in the fruit of these introgression lines was due to the environment and the genotype, and only for a few compounds presented a significant genotype x environment interaction (Table S5). This could be

attributed to the high isogenicity among ILs, with an average genetic difference among ILs over 10 %, whereas between RILs it is about 50 %. This reduction of genotypic variability among RILs may also decrease the power of GxE interaction detection.

The 12 selected ILs harboured introgression on chromosomic regions covering 85 of the volatile QTLs identified above. Thirty-four of these 85 QTLs were confirmed in the ILs population, maintaining the direction of the effect on the volatile levels after a single introgression of the *S. pimpinellifolium* in the fresh market tomato variety 'Moneymaker' (Figure 4). This can be considered as indicative of the potential of the QTLs identified for their use in breeding tomato flavour. Nevertheless, it has been described that the introgression of a genetic region has a different effect on the volatile profile depending on the genetic background where it is introgressed. Therefore, it seems plausible that the introgression of the same regions of the chromosomes in a different tomato variety would alter different sets of volatiles, as previously observed when introgressing selected regions of cherry tomato in different elite beef tomato varieties (Zanor et al., 2009).

Additionally, 16 new QTL for volatile compounds corresponding to different metabolic pathways were identified in chromosomes 1, 2, 4, 6, 8, 9, 10 and 11, represented by eight ILs (Figure 4). Some of them corresponded to volatiles synthesized by the same metabolic pathway than a QTL previously detected in the RIL population, such as in the case of fatty acid derivatives 1-penten-3-ol and 1-penten-3-one in chromosome IL 1-1, which also contained QTLs for other six metabolically related volatiles, suggesting that the QTL located in that region is enhancing in general the accumulation of volatiles of that metabolic pathway. Nevertheless, some new QTLs were detected in the ILs unrelated to those mapped in the RIL population. This is the case of terpenoids

linalool and terpineol (in IL 4-2 and IL 10-2), apocarotenoid  $\beta$ -ionone (IL 2-1), branched-chain volatile 2-methylbutanol (IL 8-1) and fatty acid derivatives (*E,E*-2,4-decadienal (IL 6-2) and 1-penten-3-one and 2-ethylfuran (IL 11-2). Additionally, we found significant IL effects for some individual compounds in only one location, but other volatiles belonging to the same metabolic pathway changed significantly in the same ILs in other locations. Although such loci were not included in the QTL map presented here, we consider that they would correspond to minor genes with an effect on early stages of the metabolic pathway, therefore affecting the levels of a number of compounds downstream.



**Figure 5.** Heat map showing the volatile levels in the ILs. Values correspond to the  $\log_2$  of the ratio of the levels of each volatile in the ILs in relation to those in *S. lycopersicum* cv. ‘Moneymaker’ cultivated in each location. Red colour corresponds to higher levels; green corresponds to lower levels; black corresponds to identical levels. The letter after the IL descriptor indicates the location where it was cultivated: ‘A’, Alginet; ‘M’, Malaga; ‘O’, Orihuela.



The complete data of the volatile levels in the ILs are shown in Figure 5 and Table S6.

### **Volatile QTLs identified in tomato show a low degree of overlapping**

QTLs for volatile compounds in tomato fruit have been previously reported from a RIL population developed from a cross between *S. lycopersicum* var. *cerasiforme* inbred line ‘Cervil’ and tomato cultivar ‘Levovil’ (Saliba-Colombani et al., 2001) or selected ILs generated from the same cross (Zanor et al., 2009), on a particular IL (Tadmor et al., 2002) or an IL population derived from *S. pennellii* in the M82 processing tomato background (Tieman et al., 2006b), and also from an IL population generated from *S. habrochaites* in the *S. lycopersicum* cv. E6203 background (Mathieu et al., 2009). Surprisingly, there is a very low degree of overlapping between the QTLs described. Only 26 out of the 102 QTLs described in this manuscript had been reported in any of the previous research, whilst the other 76 are described here for the first time (Table 1). Moreover, only 2 of those 26, the QTLs for guaiacol in chromosome 1 (probably corresponding to the *SIUGT1* gene) and phenylacetaldehyde in chromosome 8 (probably due to *LeAADc* genes) had been reported in more than one previous experiment. This lack of consistent results could be explained in part by the different analytical methods employed for the determination of the volatile compounds in each of the experiments. The precise method used for volatile determination, including both sample preparation and the technique of extraction, has a major effect on the compounds that can be detected, as recently stated (Rambla et al., 2015). Therefore, a perfect match in the QTLs identified by different research groups should not be expected even if obtained from the same biological material. Nevertheless, even in the cases where some

**Table 1.** Volatile QTLs with possible overlap to others previously identified.

	Saliba-Colombani <i>et al.</i> , 2001 Cherry tomato	Zanor <i>et al.</i> , 2009 Cherry tomato	Tadmor <i>et al.</i> , 2002 <i>S. pennellii</i>	Tieman <i>et al.</i> , 2006 <i>S. pennellii</i>	Mathieu <i>et al.</i> , 2009 <i>S. habrochaites</i>
<b>Chr 1</b>	pentanal hexanal	guaiacol		guaiacol	guaiacol (Z)-3-hexenal 3-methylbutanol
		terpineol			
<b>Chr 2</b>		eugenol			
<b>Chr 4</b>	phenylacetaldehyde	phenylacetaldehyde 2-phenylethanol			3-methylbutanol
<b>Chr 6</b>				2-methylbutanol benzaldehyde	
<b>Chr 8</b>	phenylacetaldehyde		phenylacetaldehyde 2-phenylethanol	phenylacetaldehyde 2-phenylethanol	
<b>Chr 9</b>	guaiacol eugenol	guaiacol eugenol methyl salicylate			
<b>Chr 10</b>				2-methylbutanol	benzaldehyde benzyl alcohol
<b>Chr 11</b>				3-methylbutanal 3-methylbutanol 3-methylbutanenitrile	benzaldehyde benzyl alcohol

discrepancy exists between the results available, co-localization of different QTLs from metabolically related compounds can be used as a clue for their probable association to the same gene. For example, we identified QTLs for four phenylalanine-derived volatiles: 1-nitro-2-phenylethane, phenylacetaldehyde, 2-phenylethanol and benzylnitrile, in the same region of chromosome 8 where Saliba-Colombani *et al.* (2001) found a QTL for only phenylacetaldehyde, and Tadmor *et al.* (2002) and Tieman *et al.* (2006) described QTLs for phenylacetaldehyde and 2-phenylethanol. In our opinion, the most suitable

explanation for this is that there is a single gene in that region influencing the levels of at least all these four compounds, and the ability to identify a lower or a higher number of volatile QTLs largely relies on the particular analytical method employed by each research group.

Nevertheless, even after considering that part of the lack of redundancy is due to methodological factors, a remarkable part of it still can be claimed to respond to specific genetic variation, as no QTLs for volatiles in the same metabolic pathway had been reported earlier for most of the genetic regions where volatile QTLs have been described here. This is in consonance with the results derived from the comparison between the QTLs reported in ILs derived from *S. pennellii* (Tieman et al., 2006b) and *S. habrochaites* (Mathieu et al., 2009), where only a minor part of them (10 out of the total 60) were overlapping, despite the same analytical method was employed in both studies. Taken all together, this suggests that there is ample genetic variability in wild species of the tomato clade readily available for breeding increased / new fruit flavor and aroma by modulating the levels of volatile compounds.

## CONCLUSION

One hundred and two volatile compound QTLs were identified in a RIL population generated after the cross of red-fruited wild tomato relative *S. pimpinellifolium* with fresh market tomato variety 'Moneymaker'. Seventy-six of these QTLs had not been described before and represent the largest set of loci for volatile compounds described to date. An additional advantage for the use of these QTLs in breeding programs is the fact that *S. pimpinellifolium* is the closest wild relative of cultivated tomato, which reduces the probability that linked loci producing undesirable phenotypic effects are incorporated during introgression. Results revealed that slightly less than half of the QTLs retained their effect on the production of volatiles after introgression, while some new others appeared, indicating the potential of the loci identified for breeding the flavor of fresh market tomato. SP alleles induced higher levels of apocarotenoid volatiles in ten loci, and the effect of the one identified in chromosome 3 for  $\beta$ -damascenone was confirmed in the ILs in all the locations of this study. Considering that apocarotenoids have been described to have a favourable effect on tomato liking, this one could be proposed as a good candidate for breeding tomato flavour.

## Acknowledgements

We thank Erika Moro for valuable help in volatile analysis of the ILs.

## REFERENCES

- ALBA, J.M., MONTSERRAT, M., FERNANDEZ-MUÑOZ, R. 2009. Resistance to the two-spotted spider mite (*Tetranychus urticae*) by acylsucroses of wild tomato (*Solanum pimpinellifolium*) trichomes studied in a recombinant inbred line population. *Experimental and Applied Acarology*, 47, 35-47.
- ABEGAZ, E. G., TANDON, K. S., SCOTT, J. W., BALDWIN, E. A. & SHEWFELT, R. L. 2004. Partitioning taste from aromatic flavor notes of fresh tomato (*Lycopersicon esculentum*, Mill) to develop predictive models as a function of volatile and nonvolatile components. *Postharvest Biology and Technology*, 34, 227-235.
- BALDWIN, E. A., GOODNER, K. & PLOTTO, A. 2008. Interaction of volatiles, sugars, and acids on perception of tomato aroma and flavor descriptors. *Journal of Food Science*, 73, S294-S307.
- BARRANTES, W., FERNANDEZ-DEL-CARMEN, A., LOPEZ-CASADO, G., ANGELES GONZALEZ-SANCHEZ, M., FERNANDEZ-MUNOZ, R., GRANELL, A. & MONFORTE, A. J. 2014. Highly efficient genomics-assisted development of a library of introgression lines of *Solanum pimpinellifolium*. *Molecular Breeding*, 34, 1817-1831.
- CHEN, G. P., HACKETT, R., WALKER, D., TAYLOR, A., LIN, Z. F. & GRIERSON, D. 2004. Identification of a specific isoform of tomato lipoxygenase (TomloxC) involved in the generation of fatty acid-derived flavor compounds. *Plant Physiology*, 136, 2641-2651.
- BUTTERY, R.G. 1993. Quantitative and sensory aspects of flavor of tomato and other vegetables and fruits. In: Acree T.E., Teranishi, R. (Eds), *Flavor Science: Sensible Principles and Techniques*. American Chemical Society: Washington, DC, pp. 259-286.
- CAPEL, C., FERNÁNDEZ-DEL-CARMEN, A., ALBA, J.M., LIMA-SILVA, V., HERNÁNDEZ-GRAS, F., SALINAS, M., BORONAT, A., ANGOSTO, T., BOTELLA, M.A., FERNÁNDEZ-MUÑOZ, R., GRANELL, A., CAPEL, J., LOZANO, R. 2015. Wide-genome QTL mapping of fruit quality traits in a tomato RIL population derived from the wild-relative species *Solanum pimpinellifolium* L. *Theoretical and Applied Genetics*. doi: 10.1007/s00122-015-2563-4. [Epub ahead of print]
- EISEN, M. B., SPELLMAN, P. T., BROWN, P. O. & BOTSTEIN, D. 1998. Cluster analysis and display of genome-wide expression patterns. *Proceedings of the National Academy of Sciences of the United States of America*, 95, 14863-14868.

- FOWLKES, E. B. & MALLOWS, C. L. 1983. A METHOD FOR COMPARING 2 HIERARCHICAL CLUSTERINGS. *Journal of the American Statistical Association*, 78, 553-569.
- GOULET, C., KAMIYOSHIHARA, Y., LAM, N. B., RICHARD, T., TAYLOR, M. G., TIEMAN, D. M. & KLEE, H. J. 2015. Divergence in the Enzymatic Activities of a Tomato and *Solanum pennellii* Alcohol Acyltransferase Impacts Fruit Volatile Ester Composition. *Molecular Plant*, 8, 153-162.
- GOULET, C., MAGEROY, M. H., LAM, N. B., FLOYSTAD, A., TIEMAN, D. M. & KLEE, H. J. 2012. Role of an esterase in flavor volatile variation within the tomato clade. *Proceedings of the National Academy of Sciences of the United States of America*, 109, 19009-19014.
- KLEE, H. J. & TIEMAN, D. M. 2013. Genetic challenges of flavor improvement in tomato. *Trends in Genetics*, 29, 257-262.
- KOCHEVENKO, A., ARAUJO, W. L., MALONEY, G. S., TIEMAN, D. M., PHUC THI, D., TAYLOR, M. G., KLEE, H. J. & FERNIE, A. R. 2012. Catabolism of Branched Chain Amino Acids Supports Respiration but Not Volatile Synthesis in Tomato Fruits. *Molecular Plant*, 5, 366-375.
- LOUVEAU, T., LEITAO, C., GREEN, S., HAMIAUX, C., VAN DER REST, B., DECHY-CABARET, O., ATKINSON, R. G. & CHERVIN, C. 2011. Predicting the substrate specificity of a glycosyltransferase implicated in the production of phenolic volatiles in tomato fruit. *Febs Journal*, 278, 390-400.
- MAGEROY, M. H., TIEMAN, D. M., FLOYSTAD, A., TAYLOR, M. G. & KLEE, H. J. 2012. A *Solanum lycopersicum* catechol-O-methyltransferase involved in synthesis of the flavor molecule guaiacol. *Plant Journal*, 69, 1043-1051.
- MANUEL ALBA, J., MONTSERRAT, M. & FERNANDEZ-MUNOZ, R. 2009. Resistance to the two-spotted spider mite (*Tetranychus urticae*) by acylsucroses of wild tomato (*Solanum pimpinellifolium*) trichomes studied in a recombinant inbred line population. *Experimental and Applied Acarology*, 47, 35-47.
- MATHIEU, S., CIN, V. D., FEI, Z. J., LI, H., BLISS, P., TAYLOR, M. G., KLEE, H. J. & TIEMAN, D. M. 2009. Flavour compounds in tomato fruits: identification of loci and potential pathways affecting volatile composition. *Journal of Experimental Botany*, 60, 325-337.

- MATSUI, K., ISHII, M., SASAKI, M., RABINOWITCH, H. D. & BEN-OLIEL, G. 2007. Identification of an allele attributable to formation of cucumber-like flavor in wild tomato species (*Solanum pennellii*) that was inactivated during domestication. *Journal of Agricultural and Food Chemistry*, 55, 4080-4086.
- MATSUI, K., MIYAHARA, C., WILKINSON, J., HIATT, B., KNAUF, V. & KAJIWARA, T. 2000. Fatty acid hydroperoxide lyase in tomato fruits: Cloning and properties of a recombinant enzyme expressed in *Escherichia coli*. *Bioscience Biotechnology and Biochemistry*, 64, 1189-1196.
- RAMBLA, J. L., TIKUNOV, Y. M., MONFORTE, A. J., BOVY, A. G. & GRANELL, A. 2014. The expanded tomato fruit volatile landscape. *Journal of Experimental Botany*, 65, 4613-4623.
- RAMBLA, J.L., ALFARO, C., MEDINA, A., ZARZO, M., PRIMO, J., GRANELL, A. 2015. Tomato fruit volatile profiles are highly dependent on sample processing and capturing methods. *METABOLOMICS*, doi: 10.1007/s11306-015-0824-5.
- SALIBA-COLOMBANI, V., CAUSSE, M., LANGLOIS, D., PHILOUZE, J. & BURET, M. 2001. Genetic analysis of organoleptic quality in fresh market tomato. 1. Mapping QTLs for physical and chemical traits. *Theoretical and Applied Genetics*, 102, 259-272.
- SALINAS, M., CAPEL, C., MANUEL ALBA, J., MORA, B., CUARTERO, J., FERNANDEZ-MUNOZ, R., LOZANO, R. & CAPEL, J. 2013. Genetic mapping of two QTL from the wild tomato *Solanum pimpinellifolium* L. controlling resistance against two-spotted spider mite (*Tetranychus urticae* Koch). *Theoretical and Applied Genetics*, 126, 83-92.
- SHEN, J. Y., TIEMAN, D., JONES, J. B., TAYLOR, M. G., SCHMELZ, E., HUFFAKER, A., BIES, D., CHEN, K. S. & KLEE, H. J. 2014. A 13-lipoxygenase, TomloxC, is essential for synthesis of C5 flavour volatiles in tomato. *Journal of Experimental Botany*, 65, 419-428.
- SIM, S. C., DURSTEWITZ, G., PLIESKE, J., WIESEKE, R., GANAL, M. W., VAN DEYNZE, A., HAMILTON, J. P., BUELL, C. R., CAUSSE, M., WIJERATNE, S. & FRANCIS, D. M. 2012. Development of a large SNP genotyping array and generation of high-density genetic maps in tomato. *PLoS One*, 7, e40563.
- SIMKIN, A. J., SCHWARTZ, S. H., AULDRIDGE, M., TAYLOR, M. G. & KLEE, H. J. 2004. The tomato carotenoid cleavage dioxygenase 1 genes contribute

- to the formation of the flavor volatiles beta-ionone, pseudoionone, and geranylacetone. *Plant Journal*, 40, 882-892.
- SKOUROUMOUNIS, G. K., MASSYWESTROPP, R. A., SEFTON, M. A. & WILLIAMS, P. J. 1993. beta-DAMASCENONE FORMATION IN JUICES AND WINES. *Progress in Flavour Precursor Studies: Analysis - Generation - Biotechnology*, 275-278.
- SPEIRS, J., LEE, E., HOLT, K., YONG-DUK, K., SCOTT, N. S., LOVEYS, B. & SCHUCH, W. 1998. Genetic manipulation of alcohol dehydrogenase levels in ripening tomato fruit affects the balance of some flavor aldehydes and alcohols. *Plant Physiology*, 117, 1047-1058.
- TADMOR, Y., FRIDMAN, E., GUR, A., LARKOV, O., LASTOCHKIN, E., RAVID, U., ZAMIR, D. & LEWINSOHN, E. 2002. Identification of malodorous, a wild species allele affecting tomato aroma that was selected against during domestication. *Journal of Agricultural and Food Chemistry*, 50, 2005-2009.
- TIEMAN, D., BLISS, P., MCINTYRE, L. M., BLANDON-UBEDA, A., BIES, D., ODABASI, A. Z., RODRIGUEZ, G. R., VAN DER KNAAP, E., TAYLOR, M. G., GOULET, C., MAGEROY, M. H., SNYDER, D. J., COLQUHOUN, T., MOSKOWITZ, H., CLARK, D. G., SIMS, C., BARTOSHUK, L. & KLEE, H. J. 2012. The Chemical Interactions Underlying Tomato Flavor Preferences. *Current Biology*, 22, 1035-1039.
- TIEMAN, D., TAYLOR, M., SCHAUER, N., FERNIE, A. R., HANSON, A. D. & KLEE, H. J. 2006a. Tomato aromatic amino acid decarboxylases participate in synthesis of the flavor volatiles 2-phenylethanol and 2-phenylacetaldehyde. *Proceedings of the National Academy of Sciences of the United States of America*, 103, 8287-8292.
- TIEMAN, D., ZEIGLER, M., SCHMELZ, E., TAYLOR, M. G., RUSHING, S., JONES, J. B. & KLEE, H. J. 2010. Functional analysis of a tomato salicylic acid methyl transferase and its role in synthesis of the flavor volatile methyl salicylate. *Plant Journal*, 62, 113-123.
- TIEMAN, D. M., LOUCAS, H. M., KIM, J. Y., CLARK, D. G. & KLEE, H. J. 2007. Tomato phenylacetaldehyde reductases catalyze the last step in the synthesis of the aroma volatile 2-phenylethanol. *Phytochemistry*, 68, 2660-2669.



- TIEMAN, D. M., ZEIGLER, M., SCHMELZ, E. A., TAYLOR, M. G., BLISS, P., KIRST, M. & KLEE, H. J. 2006b. Identification of loci affecting flavour volatile emissions in tomato fruits. *Journal of Experimental Botany*, 57, 887-896.
- TIKUNOV, Y., LOMMEN, A., DE VOS, C. H. R., VERHOEVEN, H. A., BINO, R. J., HALL, R. D. & BOVY, A. G. 2005. A novel approach for nontargeted data analysis for metabolomics. Large-scale profiling of tomato fruit volatiles. *Plant Physiology*, 139, 1125-1137.
- TIKUNOV, Y. M., MOLTHOFF, J., DE VOS, R. C. H., BEEKWILDER, J., VAN HOUWELINGEN, A., VAN DER HOOFT, J. J. J., NIJENHUIS-DE VRIES, M., LABRIE, C. W., VERKERKE, W., VAN DE GEEST, H., ZAMORA, M. V., PRESA, S., LUIS RAMBLA, J., GRANELL, A., HALL, R. D. & BOVY, A. G. 2013. NON-SMOKY GLYCOSYLTRANSFERASE1 Prevents the Release of Smoky Aroma from Tomato Fruit. *Plant Cell*, 25, 3067-3078.
- VAN OOIJEN, J. W. 2006. JoinMap® 4. Software for the calculation of genetic linkage maps in experimental populations. *Kyazma B.V., Wageningen, Netherlands*.
- VOGEL, J. T., TIEMAN, D. M., SIMS, C. A., ODABASI, A. Z., CLARK, D. G. & KLEE, H. J. 2010. Carotenoid content impacts flavor acceptability in tomato (*Solanum lycopersicum*). *Journal of the Science of Food and Agriculture*, 90, 2233-2240.
- VOORRIPS, R. E. 2002. MapChart: Software for the graphical presentation of linkage maps and QTLs. *Journal of Heredity*, 93, 77-78.
- ZANOR, M. I., RAMBLA, J. L., CHAIB, J., STEPPA, A., MEDINA, A., GRANELL, A., FERNIE, A. R. & CAUSSE, M. 2009. Metabolic characterization of loci affecting sensory attributes in tomato allows an assessment of the influence of the levels of primary metabolites and volatile organic contents. *Journal of Experimental Botany*, 60, 2139-2154.
- ZORRILLA-FONTANESI, Y., RAMBLA, J.-L., CABEZA, A., MEDINA, J. J., SANCHEZ-SEVILLA, J. F., VALPUESTA, V., BOTELLA, M. A., GRANELL, A. & AMAYA, I. 2012. Genetic Analysis of Strawberry Fruit Aroma and Identification of O-Methyltransferase FaOMT as the Locus Controlling Natural Variation in Mesifurane Content. *Plant Physiology*, 159, 851-870.



## GENERAL DISCUSSION

There exist many different methods for the analysis of volatile compounds in the tomato fruit, each research group using a particular one and producing their own sets of results. These analytical methods show important differences between them regarding both the sampling procedure and the technique used for the acquisition of the volatile compounds. The most commonly used methods were compared in a systematic way and, surprisingly, results revealed that both the sampling procedure and the technique used for volatile acquisition have a dramatic effect on the volatile profile detected. In our opinion, this fact has important consequences, as it means that comparison between the results obtained from research groups using different techniques is not affordable, and also that the direct use of 'quantitative' analytical results for the prediction of fruit flavour is questionable. We consider that the analytical approach is still a valuable tool for the determination of fruit flavour and aroma, but the simplistic approach of using 'odour units' to predict the impact of each volatile on aroma may easily lead to wrong conclusions. The support of taste panels in combination with instrumental analysis would instead be a more reliable approach to understand and predict tomato flavour.

One of the objectives of this research was to develop an analytical tool which would allow the comparison of several hundreds of independent samples coming from different – and often distant – locations, which was capable of identifying as many compounds as possible. After comparing different techniques, we decided to flash-freeze the samples in origin and analyze them by headspace solid phase microextraction coupled to gas chromatography/mass spectrometry. This allowed us to process a high number of perfectly comparable samples as they reached the right ripening stage independently of the capacity

of instrumental analysis. The high sensitivity in volatile acquisition provided by solid phase microextraction allowed us to unequivocally identify over fifty compounds in the tomato fruit chromatograms. Therefore, we decided to use these procedures for the analysis of volatiles in the ripe tomato fruit.

Once established the analytical platform, it was used for the determination of volatiles in a collection of NILs generated from a cross between the cherry tomato line 'Cervil' and the large size tomato lines 'Levovil' and subsequent backcrossing to large size tomato lines 'Levovil' and 'Vil B' in two independent seasons. Results revealed that the QTLs detected in the NILs from 'Levovil' and those from 'VilB' showed a poor match one to another. Therefore, it was concluded that expression of the QTLs is highly dependent on the genetic background. When comparing these results with those obtained by a collaborator research group for primary metabolites, a similar poor match between NILs developed from different tomato lines was observed.

The availability of these large sets of data from both volatile compounds and primary metabolites allowed us to study the relationship between these two sets of metabolites. Correlation analysis revealed that the levels of the precursor metabolites cannot be used to predict the levels of the volatiles derived from them. Therefore, it was concluded that the rate of volatile production is generally not governed by precursor levels but by some other downstream processes (precursors or intermediates transport and availability; variability on specific steps involved in the precursor conversion into volatiles; or other unknown regulatory mechanisms).

The analytical platform developed for the analysis of volatile compounds was also used for the determination of the levels of these specialized metabolites in a RIL population developed from *S. pimpinellifolium*

accession TO-937 and *S. lycopersicum* cv. 'Moneymaker', a fresh market variety. *S. pimpinellifolium* is a red-fruited wild tomato species, the one most closely related to cultivated tomato. Therefore, the QTLs identified in this material would be more easily introgressed in the cultivated tomato with a reduced risk of introgressing linked genes in the flanking genomic region producing undesirable phenotypes. This RIL population was evaluated for three independent seasons, and 102 volatile QTLs were detected, consisting in the largest set described up to date. About sixty per cent of the *S. pimpinellifolium* alleles induced reduced levels of volatiles, while volatile production was enhanced by the remaining forty per cent. Taking into account the current knowledge about volatile compounds and their impact in tomato flavour and aroma and consumer acceptance, several of the QTLs identified seem to be interesting candidates for breeding tomato flavour.

The stability of the QTLs identified was evaluated by determining the levels of volatiles in twelve selected lines of an IL population generated from the same parentals, cultivated during the same season at three different locations. Results revealed that almost half of them retained their effect after introgression in the cultivated tomato variety 'Moneymaker'. Additionally, sixteen new QTLs were detected in the ILs which had not previously been detected after analysis of the RIL population.

Finally, the comparison of the localization of the loci identified with an effect on the tomato volatile profile in both the three populations analyzed and those available in the literature revealed a low degree of overlapping among them. This suggests that there exists in the tomato clade ample genetic variability readily available for breeding fruit flavour and aroma by modulating the levels of volatile compounds.



## CONCLUSIONS

1. The precise method used for volatile analysis, including both sample processing and the capturing technique, has a dramatic effect on the volatile profile obtained. This analytical bias has to be considered when comparing results obtained by means of different methods, and when using these results to predict their effect on consumer preference.
2. A sample processing method starting with flash-freezing the fruit coupled to a capturing method based on solid phase microextraction would be the most adequate procedure for the comparison of vast sets of samples from a metabolomics approach. This would additionally allow a multi-omics approach from exactly the same pool of vegetative material.
3. The expression of QTLs for both volatile compounds and primary metabolites revealed to be highly dependent on the genetic background.
4. Considering that the levels of volatile compounds did not correlate with those of their precursors, it can be concluded that the rate of volatile production is generally not governed by precursor levels but by some other downstream processes (precursors or intermediates transport and availability; variability on specific steps involved in the precursor conversion into volatiles; or other unknown regulatory mechanisms).

5. One hundred and two volatile QTLs for thirty-nine different compounds were identified in a RIL population generated after the cross of *S. pimpinellifolium* with fresh market tomato cv. 'Moneymaker', several of which seem to be promising for breeding tomato flavour.
  
6. Almost half of the QTLs identified retained their effect on the production of volatiles after introgression in *S. lycopersicum* cv. 'Moneymaker' while sixteen new QTLs appeared, indicating the potential of the loci identified for breeding the flavor of fresh market tomato.
  
7. There is a low degree of co-localization in the volatile QTLs identified up to date from different species in the tomato clade. This suggests that there is ample genetic variability readily available for breeding fruit flavour and aroma by modulating the levels of volatile compounds.







UNIVERSITAT  
POLITÈCNICA  
DE VALÈNCIA