Universitat Autònoma de Barcelona

Genetic dissection of aroma and other fruit quality traits in cultivated strawberry

Pol Rey Serra

ADVERTIMENT. L'accés als continguts d'aquesta tesi queda condicionat a l'acceptació de les condicions d'ús establertes per la següent llicència Creative Commons: http://cat.creativecommons.org/?page_id=184

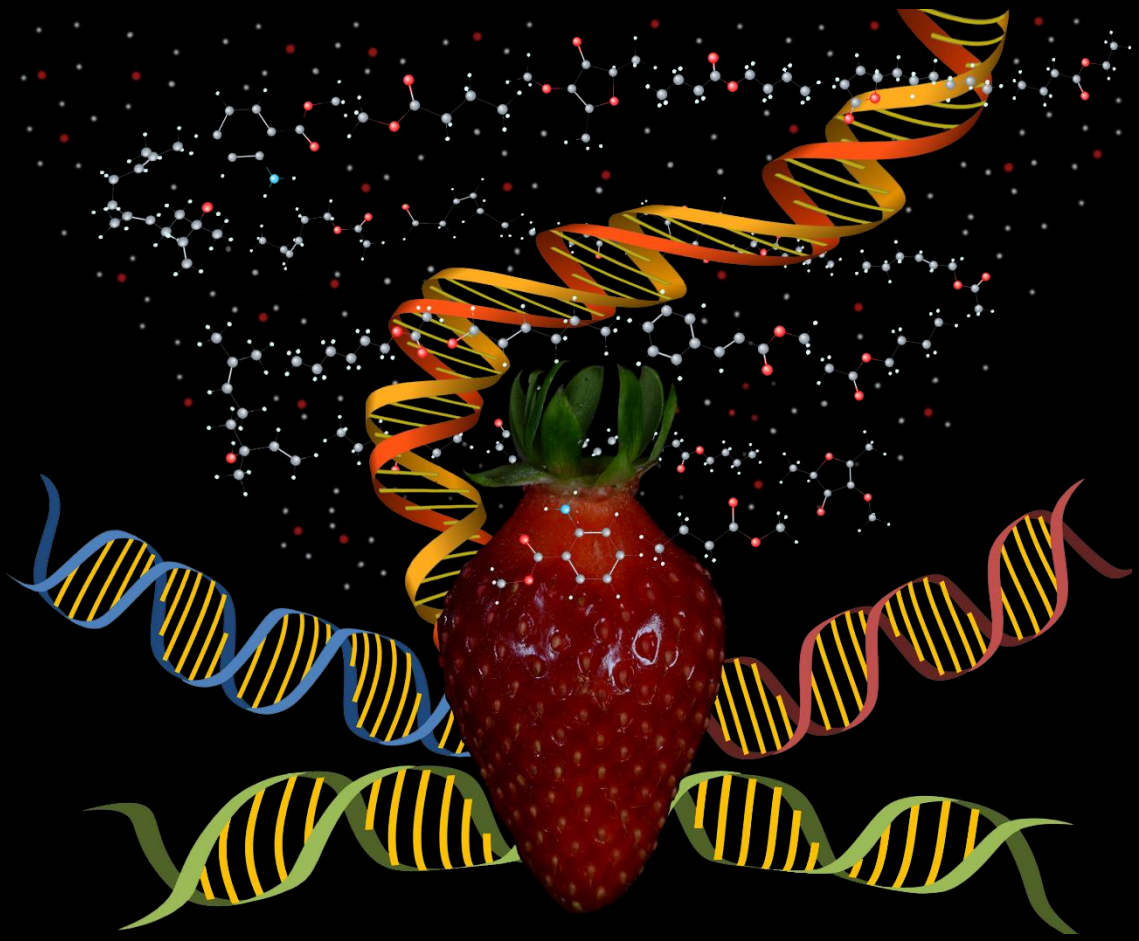
ADVERTENCIA. El acceso a los contenidos de esta tesis queda condicionado a la aceptación de las condiciones de uso establecidas por la siguiente licencia Creative Commons: http://es.creativecommons.org/blog/licencias/

WARNING. The access to the contents of this doctoral thesis it is limited to the acceptance of the use conditions set by the following Creative Commons license: https://creativecommons.org/licenses/?lang=en

Tesis doctoral

Genetic dissection of aroma and other fruit quality traits in cultivated strawberry

Pol Rey Serra



Barcelona, 2020

UNIVERSITAT AUTONOMA DE BARCELONA

Facultat de Biociències
Dept. Biologia Animal, Biologia Vegetal i Ecologia

PhD thesis

Genetic dissection of aroma and other fruit quality traits in cultivated strawberry

Dissertation presented by Pol Rey Serra for the degree of Doctor in Plant Biology and Biotechnology by Universitat Autònoma de Barcelona (UAB)

This work was performed in Centre de Recerca en Agrigenòmica (CRAG), Bellaterra.

Thesis director	Tutor	PhD candidate
Dr. Amparo Monfort Vives	Dr. Amparo Monfort Vives	Pol Rey Serra

Barcelona, September 2020

This thesis has been carried out in the department of Plant and Animal Genomics of the Centre de Recerca en Agrigenòmica (CRAG) CSIC-IRTA-UAB-UB, Barcelona, and in a short stage in Gulf Coast Research and Education Center (GREC), from the University of Florida (UF) at the Strawberry genetics and genomics lab, Wimauma (FL, US). It has been funded by the Spanish Ministerio de Economía y Competitividad with the FPI-INIA grant CPD2015-0185 and a short research stay grant associated to RTA2013-00010-C02-00 project.

Per Assumpta i Josep

Vaig començar l'aventura de la tesis doctoral amb el mateix entusiasme amb la que he acabat, gràcies a la implicació d'un gran nombre de persones del CRAG-IRTA que han ajudat tan en la part científica com en establir un bon ambient de treball. Voldria agrair a cada una de les persones que heu fet que aquesta etapa fos tan especial.

A la directora de tesis, Dra. Amparo Monfort, per concedir-me la possibilitat d'endinsar-me en aquest projecte i per la amabilitat en que m'has atès cada cop que picava a la porta del teu despatx per comentar i avançar els resultats obtinguts.

A tots els Investigadors Principals, Dra. Amparo Monfort, Dra. Txosse Aranzana, Dr. Pere Arús, Dr. Ibo Eduardo, Dra. Montse Martín, Dr. Jordi Garcia i Dr. Werner Howard per l'aportació de noves idees per tirar endavant la tesis. A tots els tècnics del laboratori, Àngel, Fuensi, Esteve, Vanesa, Elena, Sara, Jose, Nuria, Anna que en algun moment o altre he necessitat de la vostra ajuda i experiència i l'heu donada de bon grat. Al Dr. Mourad i la Dra. Silvia pels consells de redacció i la revisió critica del manuscrit d'aquesta tesis doctoral. Us estic molt agraït i mai us ho podré recompensar suficientment. A la Dra. Elli i l'ensenyança de tècniques moleculars durant l'estada d'un any al grup de maduixa.

A tot el personal de Torre Marimon, en especial a en Josep Maria i a la Manoli per cuidar tan bé les maduixes i al personal de PLANASA, especialment a en Romain, per cuidar també la població d'estudi i pels viatges a Carcassona per intercanviar caixes/maduixes, sense vosaltres les maduixeres no haguessin sobreviscut aquests quatre anys.

A l'Eloy d'Odourmet, per entrenar-me amb el GC-MS, a la Dra. Lara per ajudar-me a perdre la por a l'aparell i transferir-me alguns trucs per la anàlisis i a en Carlos per compartir les ensenyances i preocupacions del GC-MS, ara ja en saps més que jo. A en Chema, per ensenyar-me els primers conceptes de la anàlisis genètic. Al Dr. Kostas, pels esforços en el complex anàlisis bioinformàtic que suposa una espècie poliploide.

Al Dr. Lee i Dr. Whitaker, per poder formar part del vostre laboratori del GCREC de la Universitat de Florida (UF) i aprendre del vostre programa de millora. A tots els companys de la residència, en especial a José, i Krishna amb les múltiples sol·licituds de taxista, als partits de futbol fins que el sol es ponia i tantes altres aventures. Per descomptat, no em puc oblidar del company d'habitació i de despatx, Prashant, i la facilitat per conviure-hi.

A tots els companys de despatx, a l'Elena, Yu, Rong, Carlos, Natalia i Lorena i també pel bàndol animal en Manu, Lourdes, Yron, Magí i Jesús pels moments distesos durant les llargues hores d'oficina. A la resta dels doctorands del departament a Octavio, Lara, Chema, Christian, Nuria, Miguel, Neus, Arnau, Naveen, Felipe i Laura pels riures durant l'hora d'esmorzar i de dinar, per

les cerveses al bar de Biociencies i les quedades a fora del centre. Amb companys així, es gaudeix millor de la tesis. A la Neus, companya de beca compartint cada una de les etapes de la tesis i la seva burocràcia, un plaer haver compartit tots aquests anys.

A tots els companys de futbolí, que amb aquests quatre anys han passat d'haver overbooking a haver de buscar jugadors a sota de les pedres. En especial a l'Arnau i la nostra cita obligada de les 5 de la tarda.

A tots els companys del CRAG que en algun moment o altre ens hem creuat a les happy hours, els partits de basquet, etc. A la Marina, pel suport donat els primers anys, els riures i les escapades, gràcies.

A tota la família, en especial els meus pares, Assumpta i Josep i a la meva germana Blanca, pel suport donat en tots aquests anys, pels intens d'entendre la recerca que feia. També agrair a la Txell, que malgrat estar lluny i haver de fer escapades, sempre et tenia al meu costat.

Gràcies a totes les persones els que he compartit algun moment durant aquests quatre anys, aquesta tesis no hagués sigut al mateix sense vosaltres!

Indexes

Index of contents

Index of Figures

Index of Tables

List of Digital Supplementary tables

Main abbreviations

AAT	Alcohol acyl transferase	HG	Homeologous group
ABA	Abscisic acid	KW	Kruskal-Wallis
BA	Butyl acetate	L	Linalool
BB	Butyl butanoate	LG	Linkage group
bp	Base pair	LOD	Marker assisted selection
Cir	Circular	LOX	Lipoxygenases
cM	centiMorgan	M	Mesifurane
CNA	Cluster Network Analysis	MA	Methyl anthranilate
E2HA	E-2-hexenyl acetate	MB	Methyl butanoate
E2Hal	E-2-Hexenal	Mbp	Megabase pair
EB	Ethyl butanoate	MH	Methyl hexanoate
EH	Ethyl hexanoate	N	Nerolidol
Ell	Ellipsoid	NES	Nerolidol synthase
F	Furaneol	Ovo	Ovoid
FA	Fruit area	PCA	Principal Component Analysis
FIR	Firmness	ppm	Parts per million
FL	Fruit length	QTL	Quantitative Trait Loci
FP	Fruit perimeter	Rec	Rectangular
FS	Fruit shape ratio	SAR	Sweetness-acidity ratio
Fv	*Fragaria vesca*	SNP	Single Nucleotide Polymorphism
FW	Fruit weight	SSC	Soluble Solid Content
Fxa	*Fragaria x ananassa*	SSR	Single Sequence Repeat
GA	Gibberellic acid	TA	Titratable acidity
GC-MS	Gas chromatography – mass spectrometry	VA	Vertical asymmetry
		WWP	Width-widest position
GD	γ-decalactone	Z3HA	Z-3-hexenyl acetate
HA	Hexyl acetate or Horizontal asymmetry		

Summaries

SUMMARY

Cultivated strawberry, *Fragaria x ananassa,* is an important crop cultivated worldwide. Its fruits are appreciated for their balance between sweetness and acidity and for their characteristic aroma. Although, traditionally, strawberry breeding programs were essentially focused on yield improvement and disease resistance, currently, there is also an interest in improving some fruit quality traits.

Cultivated strawberry is allo-octoploide (2n = 8x = 56) and its genome is highly diploidised. Its polyploid nature increases the difficulty to analyse the genetic heritability of quantitative traits. The knowledge of the model species *F. vesca* can be easily transferred to cultivated strawberry, since its genome constitute one of its subgenomes.

In order to elucidate the heritability of strawberry aroma and other fruit quality traits, we have studied in depth two breeding populations, an F1 and an F2. The F1 population ('FC50xFD54') was produced through a cross between the breeding lines 'FC50', selected for its wild strawberry aroma, and 'FD54', selected for its fruity aroma. We have constructed a saturated 'FC50xFD54' genetic map with 14595 polymorphic markers, from IStraw35k array, grouped in 28 LGs and spanning 3451cM. The maps comparison shows high collinearity between marker position in 'FC50xFD54' genetic map and the *F. x ananassa* consensus map and with the diploid and octoploid *Fragaria* genomes. Additionally, we improved the genetic map of the F2 population ('21AF'), derived from a cross between two elite lines 'Camarosa', selected for its fruit quality, and 'Dover', selected for its robustness. The resulting map is composed of 7977 polymorphic markers grouped in 28 LGs and covers a total length of 2056cM.

The phenotypic variability of ripe fruits in these two populations for volatile compounds accumulation (VOCs), taste (acidity and SSC) and appearance traits such as shape, colour and firmness has been thoroughly studied in different harvests. The analysis of segregation and the study of correlation between traits allowed the detection of possible co-regulated traits and the QTL analysis located hundreds of regions that explain the variance of these traits.

The main organic compounds contributing to consumer preferences are sugars, acids and VOCs. A total of 58 different VOCs was identified in 'FC50xFD54' and 179 stable QTLs were localised. Of these QTLs, 22 presenting high stability were related to key volatile compounds (KVCs) which are target of some aroma breeding programs. Additionally, some of these KVC QTLs were validated in the '21AF' population. We highlight the detection of a QTL for terpene compounds in LG3B, different QTLs for ester compounds in LG1A, LG4B, LG6A and LG7D, a QTL for methyl anthranilate in LG7A and another for γ-decalactone in LG3D.

The analysis of taste and appearance traits reported a total of 77 stable QTLs in the 'FC50xFD54' population and 31 QTLs in the '21AF' population. We highlight an internal colour QTL found in LG1A and two fruit shape QTLs in LG4C and LG6D. 'FC50xFD54' population segregated for an undesirable neck phenotype whose QTLs were mapped in LG3A and LG4B.

The aim of this thesis was to increase the genetic knowledge of fruit quality, specifically aroma, taste and appearance traits. Furthermore, the selected SNPs linked to a desired trait could be applied in marker assisted selection (MAS) to satisfy the consumer demand.

RESUMEN

La fresa cultivada, *Fragaria x ananassa*, es un cultivo importante mundialmente. Sus frutas son apreciadas por su sabor, balance entre dulzura y acidez y por su característico aroma. Tradicionalmente, los programas de mejora de fresa han estado enfocados en la mejora del rendimiento y de la resistencia a enfermedades, pero actualmente, el interés en la calidad de la fruta se ha incrementado.

La fresa cultivada es una especie allo-octoploide (2n = 8x = 56) y su genoma está altamente diploidizado. Su naturaleza poliploide dificulta el análisis de la herencia genética de los caracteres cuantitativos. El conocimiento en la especie modelo *F. vesca* se puede transferir fácilmente a la fresa cultivada, ya que su genoma es también uno de los subgenomas en las fresas octoploide.

Con el objetivo de conocer la herencia del aroma y de otros caracteres de la calidad de la fruta, hemos estudiado en profundidad dos poblaciones de mejora, una F1 y una F2. La población F1 ('FC50xFD54') fue obtenida del cruzamiento entre las líneas de mejora 'FC50', seleccionada por su aroma a fresa del bosque, y 'FD54', seleccionada por su aroma afrutado. Hemos construido un mapa genético saturado para esta población con 1495 marcadores polimórficos del chip IStraw35k agrupados en 28 LGs y repartidos en 3451cM. La comparación de los mapas muestra una alta colinealidad entre las posiciones de los marcadores del mapa genético 'FC50xFD54' y del mapa consenso de *F. x ananassa* y también con los genomas diploide y octoploide de *Fragaria*. Además, hemos mejorado el mapa genético de la población F2 ('21AF'), obtenida del cruzamiento entre las líneas 'Camarosa', seleccionada por la calidad de su fruta, y 'Dover', por ser robusta. El mapa resultante está compuesto por 7977 marcadores polimórficos, agrupados en 28 LGs y cubriendo 2056cM.

La variación fenotípica de la acumulación de los compuestos volátiles (VOCs), gusto (acidez y SSC) y caracteres de apariencia como la forma, el color o la firmeza en las frutas maduras de las dos poblaciones ha sido estudiada detalladamente en diferentes cosechas. Los análisis de la segregación y correlación entre estos caracteres han permitido la detección de posibles caracteres co-regulados y los análisis de QTL han localizado cientos de regiones explicando la variación observada.

Los compuestos orgánicos que contribuyen en la preferencia de los consumidores son los azucares, los ácidos y los VOCs. Un total de 58 VOCs fueron identificados en la población 'FC50xFD54' localizando 179 QTLs estables. De entre estos QTLs, 22 con alta estabilidad fueron relacionados con los compuestos volátiles clave (KVCs), objetivo de algunos programas de

mejora. Además, algunos de estos QTLs para los KVCs fueron validados en la población '21AF'. Resaltamos la detección de un QTL para los compuestos terpenoides localizado en LG3B, diferentes QTLs para compuestos de tipo éster en LG1A, LG4B, LG6A y LG7D, un QTL para metil antranilato en LG7A y otro para γ-decalactona en LG3D.

Los análisis del sabor y la apariencia han reportado un total de 77 QTLs estables en la población 'FC50xFD54' y 31 QTLs en la población '21AF'. Resaltamos un QTL para el color interno encontrado en LG1A y dos QTLs para la forma de la fruta localizados en LG4C y LG6D. La población 'FC50xFD54' segregaba por un carácter no deseado, el alargamiento del cuello, y el análisis de QTLs ha mapeado QTLs en LG3A y LG4B.

Esta tesis ha contribuido a mejorar el conocimiento genético sobre la calidad de la fruta, específicamente el aroma, el gusto y la apariencia. Además, los SNPs seleccionados ligados a un carácter de interés podrían ser usados en la selección asistida por marcadores (MAS) para satisfacer la demanda de los consumidores.

RESUM

La maduixa cultivada, *Fragaria x ananassa*, és un cultiu mundialment important. Les seves fruites s'aprecien pel balanç entre la dolçor i la acidesa, i pel seu aroma característic. Malgrat que tradicionalment els programes de millora de maduixa estaven enfocats en el rendiment i la resistència a les malalties, actualment, l'interès per la qualitat de la fruita ha augmentat.

La maduixa cultivada és una espècie allo-octoploid (2n = 8x = 56) i el seu genoma és altament diploiditzat. La naturalesa poliploide augmenta la dificultat per analitzar l'herència genètica dels caràcters quantitatius. El coneixement que es té sobre l'espècie model *F. vesca* es pot transferir fàcilment a la maduixa cultivada, ja que el seu genoma és un dels subgenomes de les espècies octoploides.

Amb l'objectiu de determinar l'herència de l'aroma de la maduixa i d'altres caràcters de la qualitat de la fruita, hem estudiat amb profunditat dues poblacions de millora, una F1 i una F2. La població F1 ('FC50xFD54') s'ha obtingut amb l'encreuament de les línies de millora 'FC50', seleccionada per l'aroma a maduixa del bosc, i 'FD54', seleccionada per l'aroma afruitat. Hem construït un mapa genètic saturat per a la població 'FC50xFD54' amb 1495 marcadors polimòrfics, del xip IStraw35k, agrupats en 28 LGs i repartits en 3451cM. La comparació del mapa mostra una alta col·linealitat entre les posicions dels marcadors del mapa genètic del 'FC50xFD54' amb el mapa consens de *F. x ananassa* i amb els genomes diploide i octoploide de *Fragaria*. A més, hem millorat el mapa genètic de la població F2 ('21AF'), obtinguda de l'encreuament entre les línies d'elit 'Camarosa', seleccionada per la qualitat de la fruita i 'Dover', seleccionada per ser robusta. El mapa resultant està compost per 7977 marcadors polimòrfics, agrupats en 28 LGs i cobreix una longitud de 2056cM.

La variació fenotípica de l'acumulació dels compostos volàtils (VOCs), gust (acidesa i SSC) i caràcters d'aparença com la forma, el color o la fermesa s'ha estudiat detalladament en les fruites madures d'aquestes dues poblacions en diferents collites. L'anàlisi de la segregació i de la correlació entre aquests caràcters han permès la detecció de possibles caràcters co-regulats i l'anàlisi de QTL han permès localitzar centenars de regions que expliquen la variació observada.

Els compostos orgànics majoritaris que contribueixen en la preferència dels consumidors són els sucres, els àcids i els VOCs. S'han identificat un total de 58 VOCs en 'FC50xFD54' i s'han localitzat 19 QTLs estables. D'entre aquests, 22 QTLs amb alta estabilitat estan relacionats amb els compostos volàtils claus (KVCs) que són objectiu d'alguns programes de millora de l'aroma. A més, alguns d'aquests QTLs pels KVCs han estat validats en la població '21AF'. Ressaltem la

detecció d'un QTL pels terpens localitzat al LG3B, diferents QTLs per èsters als LG1A, LG4B, LG6A i LG7D, un QTL pel metil antranilat al LG7A i un altre per la γ-decalactona al LG3D.

Les anàlisis del sabor i de l'aparença han permès localitzar un total de 77 QTLs estables a la població 'FC50xFD54' i de 31 QTLs a la població '21AF'. Ressaltem un QTL pel color intern trobat al LG1A i dos QTLs per la forma de la fruita localitzats al LG4C i LG6D. La població 'FC50xFD54' segrega per un allargament del coll no desitjat, mapant dos QTLs als LG3A i LG4B.

Aquesta tesi ha contribuït en la millora del coneixement genètic sobre la qualitat de la fruita, específicament l'aroma, el gust i l'aparença. A més, els SNPs seleccionats lligats a un caràcter d'interès podrien utilitzar-se per la selecció assistida per marcadors (MAS) per satisfer la demanda dels consumidors.

General introduction

1. General introduction

Cultivated strawberry (*Fragaria x ananassa* Duch) is an important crop worldwide appreciated for its fruits that are a good source of essential nutrients such as sugars, fibre, minerals and antioxidants. Strawberries are consumed principally as fresh fruit but they are also used in processed food like shakes, ice-creams or jams. Consumers prefer strawberries with a good balance between sweetness and acidity, and a nice flavour provided by several volatile compounds.

Strawberry production experienced a huge increment during the last 50 years. For example, China and USA triplicated and duplicated their production respectively during the last 25 years. However, in the last years, USA has suffered a reduction of the production, while at the same time Mexico showed an increment. In Europe, Spain started to increase the strawberry production in 1978, reaching around 300,000 tonnes in the 1990s, being the third producer country in the world. Since then, its production has been stable and other countries, such as Mexico, Turkey or Egypt, had climbed this third position in the world ranking. Other countries like Japan and South Korea showed little variation in their production during the last 50 years (Fig. I-1). In 2018, worldwide production of strawberry was around 8,337,098 tonnes, being China the main producer by far with near 3 million tonnes. Spain was the sixth producer in the world and the first in European Union region with 344,679 tonnes (FAOSTAT, http://www.fao.org/faostat/en/#data/QC) (Fig. I-1).

In Spain, cultivated strawberries were distributed in seven thousand hectares and generated 492,271 thousand euros in 2019. The main producer region is Andalucía, specifically Huelva province, with 333,795 tonnes (97% of Spanish production), being Catalonia the second producer with 1,925 tonnes (0.56%) ("Avances de Hortícolas", 2019). The main cultivated varieties in Huelva were 'Florida Fortuna' (University of Florida), 'Rociera' (FNM SA) and 'Victory' (Plant Science/Berry Genetics) accounting for 75.9% of cultivated strawberries in Huelva (IFAPA). Nine out of ten strawberries harvested in Huelva were exported principally to European countries such as Germany, France, United Kingdom, Italy and Portugal (FEPEX).

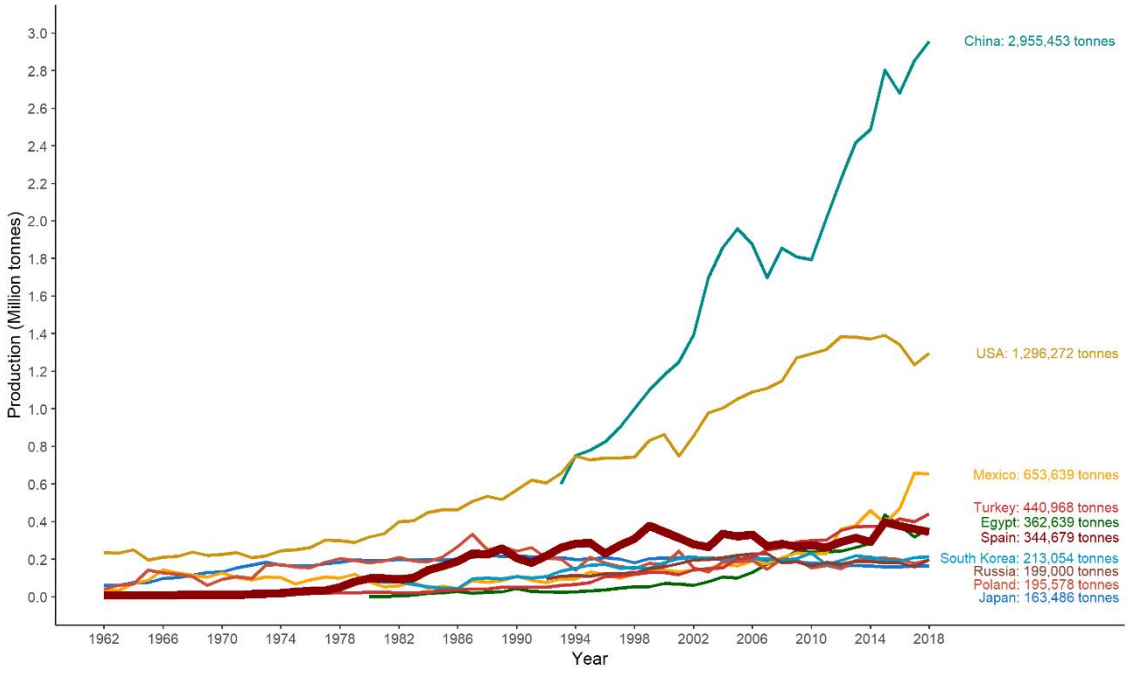


Figure I-1. World strawberry production showing the top ten country producers in 2018 (FAOSTATS).

2. *Fragaria* genus

2.1 Morphology and organoleptic characteristics

Strawberry is a small perennial plant. Leaves, inflorescences, runners and roots emerge from a crown or short thickened stem. Petioles, which can be 20 cm long, grow from the crown to form a trifoliate leaf. Flowers are bisexual with five mostly-white petals and a big yellow receptacle. This receptacle consists of hundreds of pistils surrounded by several anthers. Once the flowers have been pollinized, an increased red and fleshy receptacle will develop within four weeks with multiple surface-fixed dry achenes. These achenes are the real fruits and each one contains one seed. Calix remains attached to the receptacle. White internal fibrovascular strands supply nutrients to achenes. Strawberry can also be multiplied by vegetative cloning through monopodial runners. Of each runner, one or more plantlets can grow as vegetative reproduction (Fig. I-2).

Little differences could be found between *Fragaria* species such as plant habit, foliage, inflorescence structure, flowers and fruit (Hummer and Hancock, 2009). These characteristics help to distinguish between them. Moreover, sex dimorphism is found in wild octoploid species (Spigler *et al.*, 2011).

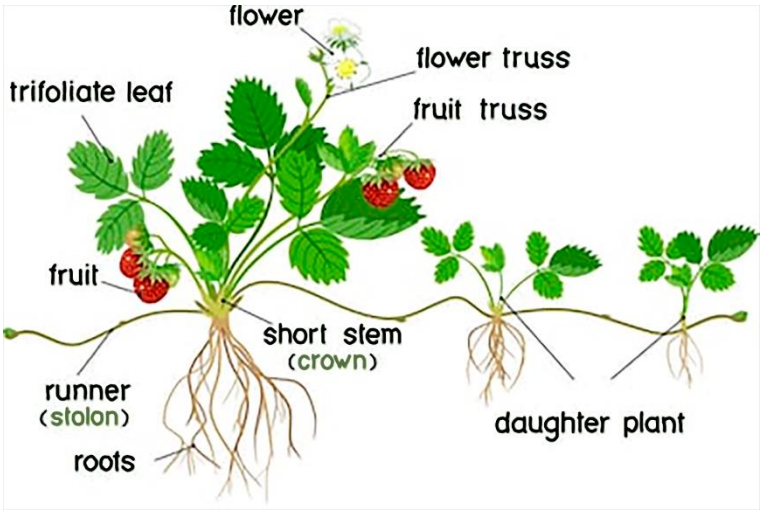


Figure I-2. Strawberry plant diagram showing its different parts (Applied from https://www.dreamstime.com/).

2.1.1. Characteristics of cultivated strawberry fruits

Once the flowers have been fertilised, fruit development followed by fruit ripening will end with a sweet, red, soft and healthy fruit. Fruit development is characterised by the increase of fruit size through cell division and enlargement by the simultaneous seed maturation (Knee *et al.*, 1977). When the seed has developed, the maturation of the receptacle triggers some physiological and metabolic changes finishing with an appetizing fruit for consumers, humans or animals, which will disperse its seeds.

Strawberry flavour is a key trait for consumer preferences (Schwieterman *et al.*, 2014 and Ulrich and Olbricht, 2016). Flavour can be understood as the combination between taste and aroma and it is modulated by different factors. A critical one is the balance between sweetness and acidity, being predominantly sweet with some acid notes. Moreover, volatile compounds, which are responsible for aroma, can modify taste perception, principally sweetness (Schwieterman *et al.*, 2014 and Ulrich and Olbricht, 2016) and adding a special flavour, such as fruity, herbal and flowery notes. Volatile compounds provide information about healthy nutrient source (Goff and Klee, 2006). Over 350 compounds had been identified in strawberry, but only 19 have been described as the main contributors of strawberry aroma (Schieberle and Hofmann, 1997). In the last years, flavour importance of these volatile compounds has been further studied (Schieberle and Hofmann, 1997, Ulrich *et al.*, 1997, Jetti *et al.*, 2007, Ulrich *et al.*, 2007, Nuzzi *et al.*, 2008, Olbricht *et al.*, 2008 and Schwieterman *et al.*, 2014).

Cultivated strawberry has a characteristic shape. For strawberries description, small differences can be found between lines or cultivars, and for this reason the UPOV has developed some

guidelines to differentiate them, finding nine categories of fruit shape, from conical to circular, cordate or oblate (Fig. I-3A).

Strawberry colour can vary, going from whitish to orangish or reddish (Fig. I-3B). Consumer colour preferences are not clear and vary according to the word region, but the main one is reddish without being too dark. The metabolites responsible for fruit colour providing red, blue or black pigments are the anthocyanins.

During fruit ripening, cell-wall degradation, triggering a soft fruit, is primordial to get a pleasant bite. However, an excessively soft fruit will shorten its shelf-life period, enhance pathogen infection and therefore consumers will discard it. Despite firmness is an important trait for both consumers and producers, little is known about its regulation.

Strawberry has an incredible antioxidant capacity, anti-inflammatory properties and anti-atherosclerotic effects thanks to the high polyphenols content and (L)-acid ascorbic (also called vitamin C). The strawberry consumption is beneficial for human health preventing several diseases such as cancer, cardiovascular diseases, neurodegenerative diseases or some chronic pathologies as demonstrate in several studies (Reviewed in Giampieri *et al.*, 2014 and Forbes-Hernandez *et al.*, 2016).

Strawberry plants are also sensible to short and long days period, being able to produce flowers during longer period and consequently increasing the harvest season. This trait is called everbearing (production in spring and autumn) and it is similar to neutral-day flowery and opposite to short day flowering.

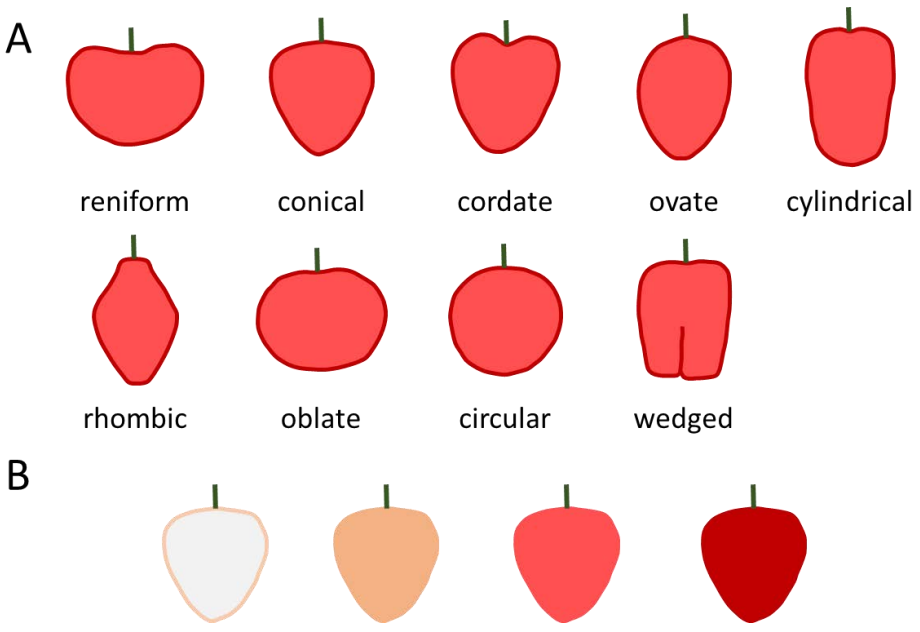


Figure I-3. Strawberry fruit shape (A) and external colour (B) according to UPOV Ad 30 and 32.

2.2 Taxonomy, distribution and phylogeny

F. x ananassa is an allo-octoploid specie originated from a human-led hybridization between two octoploid species (2n = 8x = 56): a female *Fragaria chiloensis* and a male *Fragaria virginiana*. This hybridization took place in a garden in Versailles (France) between 1714 and 1759 (Staudt, 1962 and Darrow, 1966). *F. virginiana* was introduced to Europe one century earlier from Virginia, North America and *F. chiloensis* in 1714 from Chile, South America. This new specie was rapidly commercialised for its fruit size and aroma (Darrow, 1966).

F. x ananassa belongs to the *Rosaceae* family which consists of 90 genera and around 3000 species distributed throughout almost all parts of the planet. Recently, several phylogenetic studies in *Rosaceae* family have been performed using nuclear and chloroplast genes (Potter *et al.*, 2007), next generation sequencing of plastid genomes (Zhang *et al.*, 2017b) and low-copy nuclear genes (Xiang *et al.*, 2017). All these studies reported that some important crops such as apples, pears, cherries or peaches cluster in the *Spiraeoideae* subfamily, whereas strawberries, roses and blackberries belong to the *Rosoideae* subfamily. Cultivated strawberry was classified in the *Rosodae* supertribe with roses, *Potentilleae* tribe, *Fragariinae* subtribe and *Fragaria* genus. The origin of *Roseaceae* family is estimated to be around 95-101Ma ago and *Rosoideae* tribes gradually diverged between 82Ma and 62Ma ago (Xiang *et al.*, 2017 and Zhang *et al.*, 2017b).

Fragaria genus comprises around 20 species with the basic number of chromosome n=7 and different degree of ploidy, ranging from diploid (2n=2x=14) to decaploid (2n=10x=70). This genus had been phylogenetically classified using nuclear and chloroplast genes (Potter *et al.*, 2000, Rousseau-Gueutin *et al.*, 2009 and DiMeglio *et al.*, 2014), chloroplast genomes (Njuguna *et al.*, 2013), low-copy nuclear genes (Yang and Davis, 2017) and low coverage whole-genome sequencing (Tennessen *et al.*, 2014). Almost all phylogenetic studies separated *Fragaria* species into three clades. Clade A includes European, America and west-Asia species with different level of ploidy. Clade B was only represented by a single diploid specie (*F. iinumae*) and clade C contains diploid and tetraploid species from south-east of China and Japan (Table I-1). The origin of *Fragaria* genus was dated around 24Ma and *Fragaria* divergence around 2.12Ma. The appearance of ploidy species was dated around 1.02Ma (Njuguna *et al.*, 2013).

The octoploid species, *F. virginiana* and *F. chiloensis*, might have been originated from a common ancestor and they could have diverged by adapting to wet or dry environments (Harrison *et al.*, 1997). Some phylogenetic analyses were performed to unravel the subgenomes of the octoploid strawberry. The first conclusion was that *F. vesca* was the mother donor and *F.*

iinumae constituted another subgenome (Potter *et al.*, 2007, Rousseau-Gueutin *et al.*, 2009, Tennessen *et al.*, 2014 and Yang and Davis, 2017). There is not a unanimous hypothesis for the two subgenomes left. One of the hypotheses claimed that the other two subgenomes are close to *F. iinumae,* (A B1 B2 Bi) (Tennessen *et al.*, 2014 and Liston *et al.*, 2020). The other hypothesis appeared with the sequenced octoploid genome and concluded that the other two subgenomes are from *F. nipponica* and *F. viridis* (Edger *et al.*, 2019 and Edger *et al.*, 2020). Differences in *F x ananassa* cultivars were caused by breeding programs.

Table I-1. Species included in *Fragaria* genus, their ploidy, geographic distribution (Folta and Davis, 2006) **and phylogenetic clustering** (Njuguna *et al.*, 2013).

Species	Ploidy	Geographic distribution	Phylogenetic clustering
F. vesca	2x	Northern hemisphere	
ssp. vesca	2x	Europe to Siberia	A
ssp. bracteata	2x	North America	A
ssp. americana	2x	North America	A
ssp. californica	2x	California	
F. x bifera	2x, 3x	Europe	A
F. bucharica	2x	West Himalayan region	A
F. mandshurica	2x	Northeast Asia	A
F. chiloensis	8x	Western N. and South America, Hawaii	A
F. virginiana	8x	North America	A
F. iturupensis	8x	Iturup Island	A
F. x ananassa	8x	Worldwide cultivated	A
F. orientalis	4x	Northeast China	A
F. moschata	6x	Europe	A
F. viridis	2x	Europe to Siberia	A
F. nilgerrensis	2x	Central Asia into China	A
F. iinumae	2x	Southern and central Sahalin, Russia, Japan	B
F. nipponica	2x	Japan	C
F. yezoensis	2x	Japan	
F. daltoniana	2x	East Himalayan region	C
F. nubicola	2x	East Himalayan region	
F. corymbosa	4x	Northern China	C
F. gracilis	2x, 4x	Northwest China	C
F. pentaphylla	2x	Southwest China, Himalayan region	C
F. tibetica	4x	Eastern Himalayan Region	C
F. moupinensis	4x	Southern China	C
F. x bringthursii	5x, 6x, 9x	California	

F. x ananassa subgenomes are in bold and underline.

2.3 Genome

Cultivated strawberry is an allo-octoploid specie with a highly diploidised behaviour. It means that for each linkage group, there are a total of four homeologous linkage groups that derivate from different putative diploid species.

The first *Fragaria* reference genome was obtained from the wild diploid specie *F. vesca* cv. 'Hawaii-4' (Shulaev *et al.*, 2011). It was sequenced in short reads using different platforms 454 (Roche), Solexa (Illumina) and SOLiD (Life Technologies) with a final coverage of x39. Genome size was predicted to be 240Mb and it had 25,050 annotated genes (Shulaev *et al.*, 2011). This genome had been improved several times until the actual version 4.0, which was sequenced using single-molecule real-time sequencing (SMRT) from Pacific Bioscience providing an 80.8x coverage and included 1,496 new genes (Edger *et al.*, 2018). A new annotation for v4.0 was improved by Li *et al.* (2019). Other genome sequencing projects in *Fragaria* have been performed on the diploid species *F. iinumae*, *F. orientalis*, *F. nipponica* and *F. nubicola* (actually classified as *F. bucharica*) (Hirakawa *et al.*, 2014). Moreover, the last genome sequenced was *F. nilgerrensis* obtained using SMRT sequencing and chromosome conformation capture (Hi-C) genome scaffolding (Zhang *et al.*, 2020).

The first draft of an octoploid *F. x ananassa* cv. 'Reikou' was sequenced using Roche 454 GS FLX+ and Illumina GAIIx/Hiseq 1000 platform (Hirakawa *et al.*, 2014). These short-reads sequences and the high homology between the homeologous groups, made the assembly difficult and provided low quality. Very recently, the octoploid reference genome was released in *F. x ananassa* cv. 'Camarosa' using SMRT sequencing (Pacbio) which provided 82.4x coverage and 10X Gemcode library sequenced in HiSeqX system (Illumina) providing 117x coverage (Edger *et al.*, 2019). This reference genome could identify the four subgenomes of *F. x ananassa* and how they hybridise. Moreover, cultivated strawberry is highly heterozygous which adds more difficulties to distinguish between homeologous groups. With this problem in mind, other octoploid strawberry projects are focused on sequencing and distinguishing haplotypes. One of them is the new version of *F. x ananassa* cv. 'Reikou' genome (unpublished, Isobe *et al.*, 2020 available in http://strawberry-garden.kazusa.or.jp/). Another one is the sequencing of the 'Redgauntlet' cultivar, with the participation of our group (not publish, Harrison *et al.*, 2019). In the framework of this collaboration, we could also sequence in low coverage 'FC50' and 'FD54', the parental lines of our main working population, and get an *F. x ananassa* consensus map.

3. Resources

3.1 Breeding history

Although *F. x ananassa* was first hybridised in France, the first breeding programs were carried out in England, around 1810s. The first cultivated strawberry from North America was dated in 1836. Breeders worked with many different wild *F. virginiana* cultivars. Only a few *F. chiloensis* varieties were used in England, because it had bad yield, lighter fruit colour, soft texture, plane

flavour and it was not adapted to inland. However, it produced larger fruits (Darrow, 1966 and Hummer and Hancock, 2009). It was not until the appearance of a breeding program in California, dated around 1836, that *F. chiloensis* adapted to coastal condition was used for breeding. California varieties diverged from their North American and European founders around 1950s (Hardigan *et al.*, 2018). In the second half of 20th century, new breeding programs in Japan, Scotland, Germany, France and Holland emerged (Hummer and Hancock, 2009).

Due to its highly heterozygous genotypes, all these breeding programs were based on pedigree breeding and elite parental lines were selected for each generation. Elite cultivars were propagated asexually by runners. However, there are few examples of backcross strategies incorporating specific traits, such as aphid resistance from *F. chiloensis* or day-neutrality from *F. virginiana*, to cultivated strawberry (Hummer and Hancock, 2009).

3.2 Markers

Molecular markers are tags in the DNA which explain some variability between individuals. These markers could be phenotypic, biochemical or DNA-based. Phenotypic markers are the ones which could be visually characterised. Biochemical markers are based on the detection of different enzyme isoforms. As these two kinds of markers need to be expressed, they can be dependent on environmental conditions. DNA markers, hereafter molecular markers, are insertions, deletions or duplications of one or several DNA base pairs that modify the original sequence. These modifications are spread around the whole nuclear and plastid genome. There are several types of markers according to the genome reduction or amplification strategy.

In the 2000s, the most common type of markers used in *Fragaria* genus was the Simple Sequence Repeat (SSR), also called microsatellites. SSRs are repetitions of a 2-6 bp DNA motif randomly distributed throughout the genome. Differences in the length of SSRs emerged naturally by adding or reducing the repeated motif. The SSR discovery and its first used as molecular markers dated to the 1980s (Hamada *et al.*, 1982 and Weber and May, 1989). SSRs markers can be amplified with specific primers designed in the conserved flanking regions and length sequence differences are detected by high resolution gels or in capillary electrophoresis sequencing machines. The use of SSRs is a reproducible, reliable, transferable and constitutes a cheap genotyping method. SSRs are codominant markers, meaning that all marker alleles could be distinguished and genotyped. This is a clear advantage compared to dominant markers based on presence or absence of an allele, and therefore unable to distinguish between heterozygous or homozygous genotypes. Several SSRs were developed from *F. virginiana* (Ashley *et al.*, 2003) and *F. x ananassa* (Folta *et al.*, 2005 and Bassil *et al.*, 2006). Thanks to the good SSR

transferability between *Fragaria* species (Monfort *et al.*, 2006), others sets of SSRs developed from diploids could be used in octoploid strawberry (Govan *et al.*, 2008).

In the 2010s, Single Nucleotide Polymorphic marker (SNP) became the most used marker in strawberries. SNPs are single base changes between individuals that provide variability. This kind of polymorphism is the most abundant type of variation in DNA and they are discovered by sequence alignment. Hence, SNPs enable to construct higher density genetic maps. These SNPs can be highly correlated to a desired trait. Since there is huge number of SNPs, there is an interest in genotyping them simultaneously. The first SNPs array was built with 1477 SNPs in human (Wang *et al.*, 1998). Since then, several SNP arrays have been developed for rosaceous crops such as the 9k peach array (Verde *et al.*, 2012), 6k cherry array (Peace *et al.*, 2012), 68k array in rose (Koning-Boucoiran *et al.*, 2015), 90k SNP array in strawberry (Bassil *et al.*, 2015), 480k apple array (Bianco *et al.*, 2016) and 70k SNP in pear (Montanari *et al.*, 2019). Rose and strawberry arrays were developed in polyploid species.

Different methods are available to genotype SNPs. KASPar™ genotyping is based on competitive allele-specific PCR amplification of target sequences and reading the genotype at the endpoint fluorescent using a FRET capable plate. This method can be used in a high-throughput platform called Fluidigm's dynamic array. Moreover, TaqMan is also a competitive PCR based assay but with real-time fluorescent reader and it can be multiplexed on a high-throughput platform called OpenArray.

Another PCR-based method to detect SNPs or indels is the High-Resolution Melting (HRM). This molecular technique is based on the comparison of denaturalization curves of DNA strands monitored in real time. Differences in DNA strands induce differences in the annealing temperature.

Genotyping octoploid strawberry is really complicated and challenging since its genome presents four homeologous linkage groups (HGs) (Fig. I-4). To properly genotype any SNP, this should be segregating only in one linkage group. Otherwise, it will be extremely difficult to differentiate among the numerous genotypic clusters resulting from homeologous alleles. An idealistic marker should be the one that only segregates in one HG and do not hybridised with the other three HGs (Fig. 1-4, I). However, the most common situation is that sequences from other HGs can hybridize with one allele probe, without showing segregation. For example, one HG has the marker segregating, one HG hybridize with allele A and the other two do not hybridised, thus it increases the dosage of A allele (Fig. I-4, II), or even worse if two or three HGs have allele A (Fig. I-4, III and V). Another possibility may be that one HG has the marker

segregating, other HG show A allele and another HG B allele, increasing A and B allele dosages and clustering closer to heterozygous AB (Fig. I-4, IV and VI). As genotyping is based on fluorescent differences between both alleles, an increment of allele dosage will dilute the differences caused by the marker and genotyping clusters will be closer to each other.

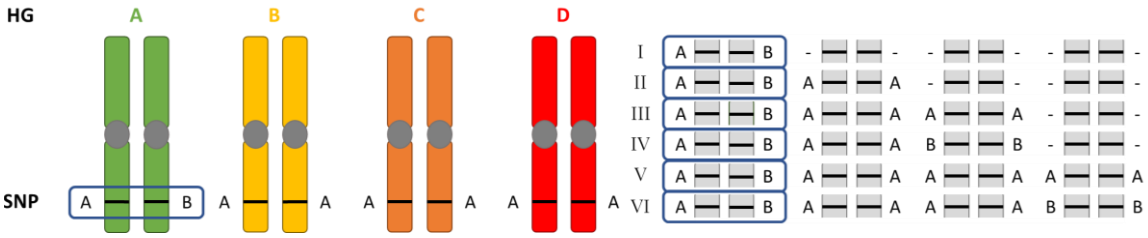


Figure I-4. Diagram of a segregating marker in HG and example of allele dosage.

The strawberry 90k SNP (IStraw90k) array was developed by sequencing 19 octoploid and six diploid accessions and aligning these sequences to *Fragaria vesca* diploid genome v1.1 (Shulaev *et al.*, 2011) leading to discover 95.062 polymorphic sites. IStraw90k SNP array was built using Axiom MyDesign™ custom genotyping platform based on KASPar™ technology. SNP genotypes were analysed by clustering as a diploid and they can be classify as: 1) MonoHighResolution (MHR) showing only one cluster indicating that this marker is monomorphic (Fig. I-5: A), 2) Heterozygous no-segregating, showing all siblings clustering in heterozygous cluster and each parental line in one homozygous cluster (Fig. I-5: B), 3) PolyHighResolution (PHR) showing three clusters being two homozygous and one heterozygous (Fig. I-5: C), 4) NoMinorHomozygote (NMH) showing two clusters with one homozygous and one heterozygous (Fig. I-5: D), 5) Off-Target Variant (OTV) presenting an additional cluster resulting from mismatches between probes and sequences (Fig. I-5: E), 6) Call Rate Below Threshold (CRBT) showing good clusters but the call rate is below 97% and 7) Other not being classified in previous categories. Moreover, in polyploid, diversity on genotyping clusters can be found, such as PHR showing two heterozygous clusters (Fig. I-5: F), heterozygous cluster being closer to one homozygous cluster (Fig. I-5: G) or NMH showing both genotyping clusters really close to each other (Fig. I-5: H).

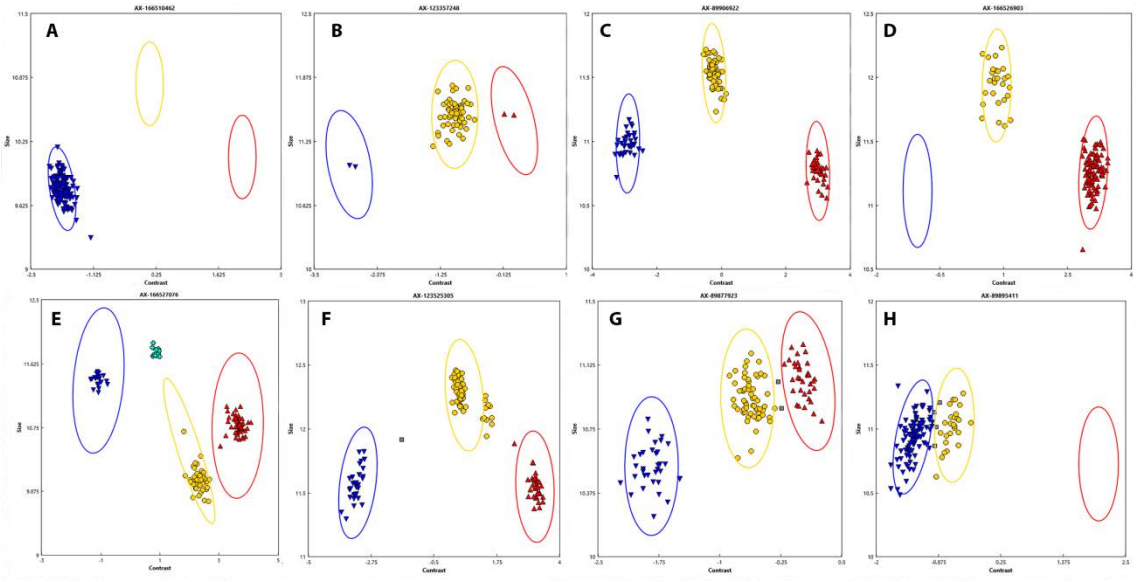


Figure I-5. Marker allele cluster plot from two segregating population A: Homozygous marker, B: Heterozygous no segregating marker, C: PolyHighResolution marker, as a diploid marker, D: NoMinorHomozigous marker, as a diploid marker, E: Off-Target Variant (OTV), F: PolyHighResolution marker with two heterozygous clusters, G and H: markers with homeologous dosage marker for PolyHighResolution and NoMinorHomozigous, respectively. Genotypes: AA (dark blue), AB (yellow), BB (red), OTV (light blue).

The IStraw90k SNP array had a problem of high cost and low number of informative SNPs. For this reason, some SNPs were selected for mapping in at least two out of eight mapping populations ('FL_08-10' x '12.115-19', 'Holiday' x 'Korona', 'Treasure' x 'Winter Dawn', 'Camarosa' x 'Dover', genotyped by our group, and four more populations) and segregating in other accessions to build a more efficient IStraw35k 384 HT array (IStraw35k) with 38506 SNPs using the same chemistry and analysis method (Verma *et al.*, 2017a).

Since IStraw90k and IStraw35k SNP array are based on *Fragaria vesca* genome v1.1 (Shulaev *et al.*, 2011), the dominant *F. vesca*-like subgenome is overrepresented. In addition, markers are not subgenome specific and it is not always clear to which chromosome of *F. x ananassa* genome belong.

Very recently, new SNP arrays with subgenome specific markers have been developed. Whole-genome sequencing in octoploid species for cultivated *F. x ananassa* (47) and wild ancestors *F. chiloensis* (24) and *F. virginiana* (22) were used for DNA variant calling against *F. x ananassa* genome (Edger *et al.*, 2019). The 850k screening SNP array was built with selected polymorphic and subgenomic specific markers in *F. x ananassa* and inheriting 16k markers from IStraw35k array. It was tested in 384 octoploid strawberry accessions. A 52.54% were clustered as disomic, from them they were classify as PHR (78.3%), NMH (18.8%) and OTV (2.9%). A derivate 50k SNP

production array was tested with 1421 octoploid samples including two F1 populations (Hardigan *et al.*, 2020).

3.3 Genetic maps

A genetic map is the ordered position of markers within one linkage group. Distance between markers is measured as the frequency of crossing over events between markers. The larger the physical distance between markers, the higher the probability that a crossing over event (or recombination) occurs during meiosis. The first genetic map was constructed by Sturtervant (1913), who ordered linearly six traits in Drosophila. Since then, a huge number of new markers and genetic maps has been developed.

The first *F. vesca* linkage map was constructed from an F2 population of 80 siblings. It consisted of 80 Randomly Amplify Polymorphic DNA (RAPD) markers and its size was of 445cM (Davis and Yu, 1997). An F2 genetic map of a cross between *F. vesca* and *F. bucharica* has also been constructed (Sargent *et al.*, 2004, Sargent *et al.*, 2006 and Sargent *et al.*, 2011). This map was used to anchor *F. vesca* genome (Shulaev *et al.*, 2011).

The first *F x ananassa* parental linkage map was constructed using a cross population of 'Capitola' and 'CF1116' with Amplified Fragment Length Polymorphism (AFLP) dominant markers. Female map had 235 markers with 30 linkage groups and a map size of 1604cM, while the male map had 280 markers in 28 linkage groups (LGs) and 1496cM (Lerceteau-Kohler *et al.*, 2003). Several genetic maps have been developed using SSRs. A linkage map of 'Redgauntlet' x 'Hapil' population had 549 SSRs in 28 linkage groups and 2140cM length (Sargent *et al.*, 2012). An integrated linkage map was designed with three populations: '02-19' x 'Sachinoka', 'Kaorino' x 'Akihime' and '0212921' self-pollination population. This integrated map had 1856 loci spanning 28 linkage groups covering a total genetic length of 2364cM (Isobe *et al.*, 2013). Another SSR genetic map from a 'Sonata' x 'Babette' cross was built, resulting in 907 markers distributed in 31 linkage groups and 1581cM genetic distances (Davik *et al.*, 2015). More SSRs genetic maps have been reported (Castro *et al.*, 2015 and Castro and Lewers, 2016).

Genetic maps were not only constructed for octoploid *F. x ananassa* but also for the wild octoploid species, *F. virginiana*. This map was build using 212 SSRs clustered in 42 linkage groups covering 2373cM length (Spigler *et al.*, 2008).

Once the IStraw90k SNP array appeared, denser SNP-based genetic maps have been released. In order to validate this array, an F1 population from a cross between 'Holliday' and 'Korona' was genotyped and a genetic map of 6593 SNPs was built in 35 linkage groups and ranging

2050cM in size (Bassil *et al.*, 2015). The second genetic map was constructed using an F1 population from the cross 'Monterey' x 'Darselect' with 8407 SNPs clustering in 32 LGs and spanning 1820cM length (Sargent *et al.*, 2016). An S1 population derived from a self-pollination of the Japanese cultivar 'Reikou' was used to construct a genetic map using SNPs and SSRs. A total of 11574 markers were mapped in 31 LGs covering 2816.5cM (Nagano *et al.*, 2017). More genetic maps from different parental lines have been built using both IStraw90k and IStraw35k SNP arrays (Anciro *et al.*, 2018, Cockerton *et al.*, 2018 and Sargent *et al.*, 2019).

IStraw90k SNP array was also used in diploid species. A preliminary SSR genetic diversity study between Japanese species (Njuguna *et al.*, 2011) was useful to select two distant *F. iinumae* accessions to cross them and genotype their F2 population. This genetic map had a size of 451cM in 28 LGs and consisted of 3280 SNPs from IStraw90k SNP array and 893 SNPs derived from a genotyping by sequencing strategy (Mahoney *et al.*, 2016).

Recently, new arrays with 850k and 50k SNPs have been developed. Two F1 population were hybridised with 50k SNP array: *F. x ananassa* cultivar 'Camarosa' x *F. chiloensis* ecotype 'Del Norte' with 182 siblings and *F. virginiana* subsp. *virginiana* accessions, 'PI552277' and 'PI612493' with 96 siblings, constructing each parental genetic map. 'Camarosa' genetic map was built with 9062 markers in 30 LGs spanning 1712cM and 'Del Norte' genetic map had 2325 markers in 28 LGs spanning 1959cM. 'PI552277' genetic map was constructed with 1767 markers in 28 LGs spanning 3020cM and 'PI612493' genetic map has 1844 markers in 28 LGs spanning 2995cM. Since this array was based on *F. x ananassa* genome, the 'Camarosa' genetic map displays higher number of segregating markers (Hardigan *et al.*, 2020).

3.4 Genetic studies of traits

Once a desired trait has been phenotyped and its segregation has been studied, a Quantitative Trait Locus (QTL) analysis will reveal the genetic regions causing this variation. A QTL is a genetic region (locus) that explains the variation of a quantitative trait (phenotype). Each locus has one or multiple genes involved with the variation of a trait. To achieve a reliable QTL analysis, a sufficiently large mapping population is required and a saturated genetic map has to be constructed using informative molecular markers, such as SNPs and SSRs. In addition to the genetic information, consistent phenotypic data, which could be environment-dependent, must be minutely collected over time. Since harvesting in open field presents different environmental characteristics, QTL analysis in different years will be more reliable.

QTL analyses can be performed using different approaches. The most used is the Interval Mapping (IM) analysis which estimates the QTL position between two markers by calculating the logarithm of the odds ratio (LOD score).

There are other statistical analyses such as non-parametric Kruskal-Wallis (KW) or Genome-Wide Association Study (GWAS) which inspect the association between each marker independently and the phenotype. These approaches are useful once the marker position is known, such as using the consensus genetic map positions or the physical position from a genome. Moreover, it is very powerful when QTL discovering is performed in several populations.

QTLs should be compared between different populations. However, in some cases, this comparison is complicated in polyploid strawberry. Since SSRs and other markers are not subgenome specific and their alleles could diversify across populations, it is difficult to know in which specific HG belongs. QTL comparisons between SNP genetic maps are easier, because SNPs are segregating in one LG. Hardigan *et al.* (2020) found the correspondence between LG nomenclature from the most used genetic maps.

3.4.1. Fruit quality traits and flowering behaviour

During fruit development, strawberry displays a firm texture in green, white and turning stages, then it begins to soften at the turning and red phases and it becomes extremely soften in overripe fruits. During this process, several metabolic changes occur to achieve a pleasant and tasty fruit.

The first QTL analysis in *F. x ananassa* was carried out using an F1 population from a cross between 'Capitola' and 'CF1116' (Lerceteau-Köhler *et al.*, 2004 and Lerceteau-Köhler *et al.*, 2012). Nineteen traits related to fruit development, texture, colour, anthocyanin, sugar and organic acid contents were mapped in 115 QTLs. Some of these QTLs co-localised in LGIIIa and LGVIa. Another progeny resulting from a cross between '232' and '1392' has been analysed in detail. A total of 33 QTLs for agronomic and fruit quality traits were mapped but only 12 of them were stable (Zorrilla-Fontanesi *et al.*, 2011). Other 70 QTLs responsible for the genetic variation of 48 different volatile compounds (Zorrilla-Fontanesi *et al.*, 2012) were mapped and, recently, 133 QTLs for 44 primary metabolites have been detected, but only 20 were stable (Vallarino *et al.*, 2019).

Sugars are the main contributors to sweetness, being glucose, sucrose and fructose the most abundant ones. A good approximation and cheap method to quantify the sugar content is measuring the soluble solids content (SSC) in a juicy sample. Soluble solids content QTLs were

detected in LGII-1, LGV-4 and LGVI-3 (Zorrilla-Fontanesi *et al.*, 2011 and Vallarino *et al.*, 2019), LG3a, LG5a and LG6a (Lerceteau-Köhler *et al.*, 2012) and LG2-5, LG5-1 and LG6-3 (Castro and Lewers, 2016). In addition, sucrose and raffinose were mapped in LGV-4, being glucose-6-P 1 epimerase a candidate gene (Vallarino *et al.*, 2019).

The main metabolites providing acidity to strawberry fruit are citric and malic acid which are intermediate metabolites of the tricarboxylic acid cycle. QTLs for titratable acidity or pH were mapped in LGIV-2 and LGV-2 (Zorrilla-Fontanesi *et al.*, 2011 and Vallarino *et al.*, 2019), in LG3a, LG5a and LG6a (Lerceteau-Köhler *et al.*, 2012) and LG4-4 and LG6-4 (Castro and Lewers, 2016). Citric acid QTLs were located in all HGs except for HG4 and malic acid QTLs were located in LG5d, LG6a, LG6b and LG7d (Lerceteau-Köhler *et al.*, 2012). Additionally, other acids, such as glyceric, succinic, fumaric, citric and threonic acid, were mapped in LGV-2 (Vallarino *et al.*, 2019).

Fruit firmness is related to cell-wall degradation, achieved by modifying cellulose and hemicellulose and releasing pectin (Knee *et al.*, 1977 and Perkins-Veazie, 1995). Little is known about the mechanism and the regulation of this degradation. Recently, some enzymes have been found to be involved in cell-wall modification, such as pectate lyases, polygalacturonases, pectin methylesterases, rhamnogalacturonan lyase (Draye and Van Cutsem, 2008, Molina-Hidalgo *et al.*, 2013, Pose *et al.*, 2015, Paniagua *et al.*, 2017 and Mendez-Yanez *et al.*, 2020).

Wild octoploid strawberries have also been investigated for fruit quality. In *F. chiloensis*, fruit traits such as firmness (Figueroa *et al.*, 2010), aroma (Prat *et al.*, 2014) and colour (Salvatierra *et al.*, 2013) have been studied. Two loci determining sex (male, female or hermaphrodite) were linked in *F. virginiana* (Spigler *et al.*, 2008 and Spigler *et al.*, 2011) and they were mapped in HGVI in both *F. virginiana* and *F. chiloensis* (Tennessen *et al.*, 2014).

Polyphenols are a good source of healthy metabolites that help to prevent several diseases. Many cultivars have been analysed to find the most representative compounds in strawberry. Anthocyanins are the most abundant being pelargonidin-3-O-glucoside (Pg-3-O-Glu) the most synthesised following by perlargonidin-3-rutinoside, cyaniding-3-glucoside and cyaniding-3-malonylglucoside. Quercetin glycosides and kaempferol glycosides are the most important flavonols. Flavan-3-ols group, such as procyanidins and (+)-catechin, have been also identified. The main ellagitannin is agrimoniin, and ellagic acid and their glycosides are also present. Finally, cinnamic acid conjugates, coumaroyl hexoses and cinnamoyl glucoses are others polyphenols compounds in strawberries (Buendia *et al.*, 2010 and Aaby *et al.*, 2012). With the aim of increasing polyphenol contents, some inter-specific crosses with wild accessions have been performed (Diamanti *et al.*, 2012). Seventy-six stable QTLs were mapped in a NILs collection with

F. vesca, as the recurrent parental, and *F. bucharica* as the donor (Urrutia *et al.*, 2016). Moreover, QTL mapping in an F2 population between 'Camarosa' and 'Dover' have been performed (unpublished).

The (L)-ascorbic acid (L-AA, vitamin C) is also an important nutritional compound which has high antioxidant activity. L-AA QTLs have been mapped in LGIV-2 and LGV-1 (Zorrilla-Fontanesi *et al.*, 2011), being mannose-6P isomerase a candidate gene in LGV-1 (Vallarino *et al.*, 2019). Transgenic Arabidopsis plants expressing the *FaGalUR* gene, encoding a D-galacturonate reductase, increased vitamin C content via D-galacturonic acid (Agius *et al.*, 2003). Moreover, it has been shown that the *FaMYOX* gene, encoding a monodehydroascorbate reductase, is also correlated with L-AA content (Cruz-Rus *et al.*, 2011).

The everbearing trait is a flowering behaviour highly interesting for strawberry production. Everbearing strawberries display more than one production peak during a year. This trait has been located in HG4 (Castro *et al.*, 2015 and Honjo *et al.*, 2015) and *Flowering locus T* (*FT*) has been suggested as a candidate gene (Perrotte *et al.*, 2016). There are other genes described in *F. vesca* which contribute to flower induction in *F. x ananassa*, namely *FaFT1* and *FaSOC1* (Koskela *et al.*, 2016).

3.4.2. Aroma studies

Aroma can be detected via orthonasal route, directly from nose to brain, or via retronasal route, once volatiles are released by chewing (Fig. I-6). Volatiles could also inform about fruit nutrient content (Goff and Klee, 2006). Aroma perception depends on the concentration of its components which has to exceed the odour threshold. This odour threshold depends on the interaction of each compound with a specific olfactory receptor protein. These proteins show a high variability between humans (Review by Hasin-Brumshtein *et al.*, 2009). In addition to this complex sensory processing system, the diversity of large volatile compounds in strawberry makes it necessary to perform panel assays within any aroma breeding program. Different studies looking for the preferred volatile compounds have been done in tomato (Tieman *et al.*, 2012) and strawberry (Schwieterman *et al.*, 2014 and Ulrich and Olbricht, 2016).

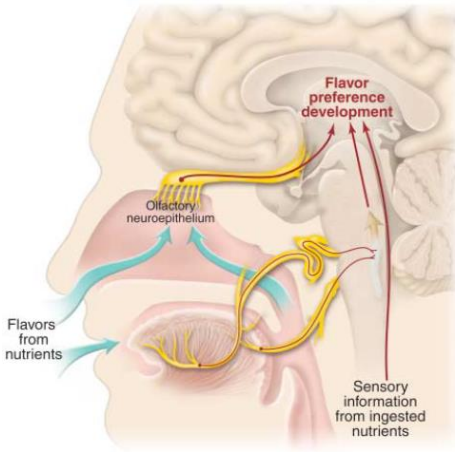


Figure I-6. Taste and olfactory sensory through orthonasal and/or retonasal routes, stimulates brain and discern flavour preferences (Figure applied from Goff and Klee, 2006).

Volatile organic compounds (VOCs) in fruits can be dependent on genetic variability, environmental effect (Samykanno *et al.*, 2013) and fruit ripening stage (Vandendriessche *et al.*, 2013). Sugars and esters increase during ripening, whereas aldehydes, alcohols, acidity and firmness decrease (Vandendriessche *et al.*, 2013).

During the last 30 years, several VOCs analyses have been performed in strawberry and they are categorised as aldehydes, esters, furans, lactones and terpenes (Reviewed by Ulrich *et al.*, 2018). From over 350 VOCs detected in strawberry, nineteen key volatile compounds (KVCs) were found to mainly contribute to the strawberry aroma (Schieberle and Hofmann, 1997). Ester volatile compounds provide fruity and sweet notes like methyl butanoate, methyl hexanoate, ethyl butanoate, ethyl hexanoate, butyl acetate, butyl butanoate and hexyl acetate. Moreover, myrtenyl acetate confers herbaceous notes and methyl cinnamate and methyl anthranilate are described as the unique compounds producing a strawberry-like flavour. These last three compounds add an interesting flavour but they are present only in few cultivars. γ-decalactone (a lactone compound) provides a peach-like aroma. Aldehydes, such as E-2-hexenal and Z-3-hexenal and their derivative esters E-2-hexenyl acetate and Z-3-hexenyl acetate, confer herbaceous notes. Linalool and nerolidol (terpene compounds) produce flowery notes. Finally, furans, principally furaneol and mesifurane, contribute with a caramel note (Schieberle and Hofmann, 1997, Ulrich *et al.*, 1997, Jetti *et al.*, 2007, Ulrich *et al.*, 2007, Nuzzi *et al.*, 2008, Olbricht *et al.*, 2008 and Schwieterman *et al.*, 2014) (Fig. I-7).

Fatty acids pathway is the main route producing strawberry volatile compounds. This pathway generates aliphatic compounds belonging to different chemical families, such as acids, alcohols, aldehydes, esters, ketones and lactones (Fig. I-7).

Esters are the products of the esterification between alcohol and acyl-CoA catalysed by alcohol acyl transferases (AAT). Alcohols are, previously, dehydrogenated by alcohol dehydrogenase (ADH). FaSAAT, an alcohol acyl transferase identified in *F. x ananassa*, has a wide range of substrates (Aharoni *et al.*, 2000, Beekwilder *et al.*, 2004 and Navarro-Retamal *et al.*, 2016). Acyl-CoA is synthesised through β-oxidation or the lipoxygenases (LOX) pathway (Schwab *et al.*, 2008). The LOX pathway produces aldehydes and alcohols, such as hexanal and 2-hexenal. Additionally, 3Z-2E-hexenal isomerases are responsible for converting 2-E-hexenal to 3-Z-hexenal in cucumber (Spyropoulou *et al.*, 2017). Lactones are also derived from fatty acids (Fig. I-7). In cultivated strawberry, *FaFAD1*, encoding an omega-6 fatty acid desaturase, has an important role in γ-decalactone synthesis (Chambers *et al.*, 2014 and Sanchez-Sevilla *et al.*, 2014).

Other volatiles are synthetized via the degradation of aromatic amino acids, such as phenylalanine and tyrosine, or branched-chain amino acids, such as leucine, isoleucine, valine and alanine. Methyl anthranilate and methyl cinnamate are synthesised through phenylalanine pathway (Fig. I-7). A methyltransferase (*F. x ananassa anthranilic acid methyl transferase*, *FaAAMT*) is responsible for the last step of methyl anthranilate synthesis (Pillet *et al.*, 2017).

Terpenes are synthetized through isopentenyl pyrophosphate (IPP) and dimethylallyl pyrophosphate (DMAPP) pathways producing monoterpenes, such as linalool, and sesquiterpenes, such as nerolidol, respectively. In strawberry, *FaNES1*, encoding a nerolidol synthase, catalyse the synthesis of linalool and nerolidol (Aharoni *et al.*, 2004). Some important compounds for wild strawberry aroma such as myrtenol, precursor of myrtenyl acetate, are not present in cultivated strawberry (Ulrich *et al.*, 2007), due to the loss of some functional enzymes (Aharoni *et al.*, 2004).

Finally, furans can also produce volatile compounds such as mesifurane and furaneol. A quinone oxidoreductase identified in octoploid cultivated strawberries (FaQR) is required for the synthesis of furaneol (Raab *et al.*, 2006), whereas O-methyltransferase (FaOMT) methylates furaneol to produce mesifurane (Wein *et al.*, 2002 and Zorrilla-Fontanesi *et al.*, 2012).

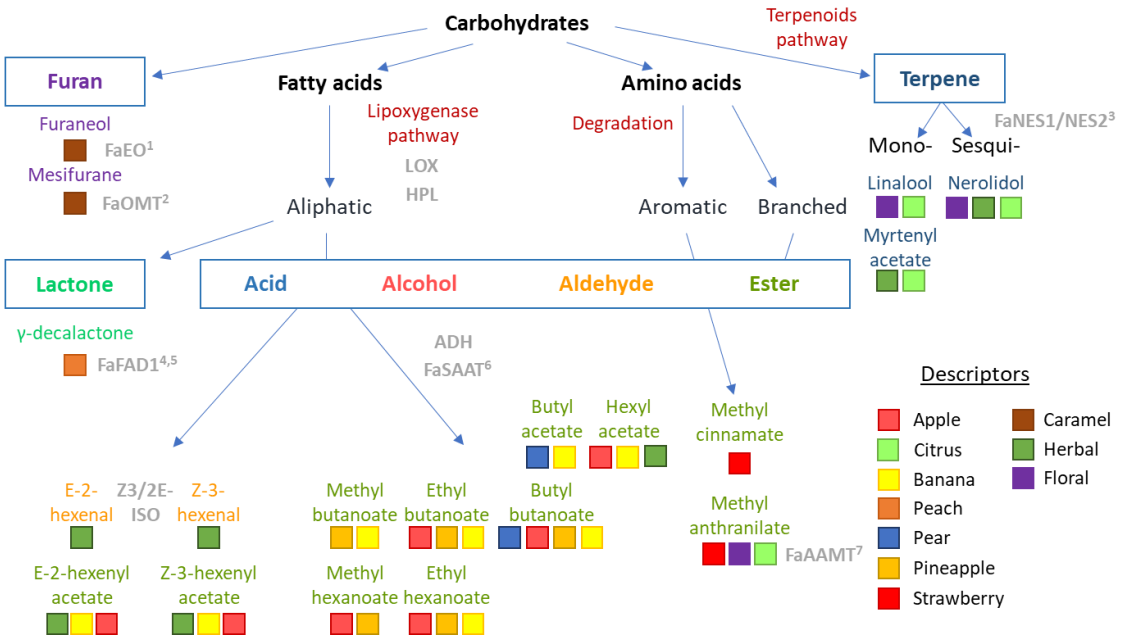


Figure I-7. Simplified diagram of volatile compound biosynthetic pathways showing KVCs, genes involved in their pathways and their aroma descriptors. [1] Raab *et al.*, 2006, [2] Wein *et al.*, 2002, [3] Aharoni *et al.*, 2004, [4] Chambers *et al.*, 2014, [5] Sanchez-Sevilla *et al.*, 2014, [6] Aharoni *et al.*, 2000, [7] Pillet *et al.*, 2017.

3.4.3. Fruit shape studies

Little is known about the regulation of strawberry shape. Final shape can vary depending on many factors, such as flower bud position, duration of cell division, degree of cell enlargement, number of cells, size of intercellular air space, number of pollinated achenes, temperature during fruit set and plant nutrition (Perkins-Veazie, 1995).

Some hormones play an important role in fruit development. In strawberry, gibberellins appear in the early stage of fruit development, whereas indole-3-acetic acid (IAA) and gibberellic acid (GA) have their peak before colour accumulation and abscisic acid (ABA) in last stages (Symons *et al.*, 2012). A cross-talk between hormones take place during fruit development. In *F. vesca*, auxins promote diameter enlargement of the fruit while GAs control the length enlargement, at the same time that they inhibit the synthesis of ABA during early stages. On the other hand, at later stages, ABA content increases, inhibiting fruit enlargement and triggering the ripening process (Liao *et al.*, 2018) (Fig. I-8).

QTL analysis from a diploid NILs collection revealed two QTLs for round-fruit shape in LG2 and LG4, and an elongated-fruit shape QTL in LG5 (Urrutia *et al.*, 2015).

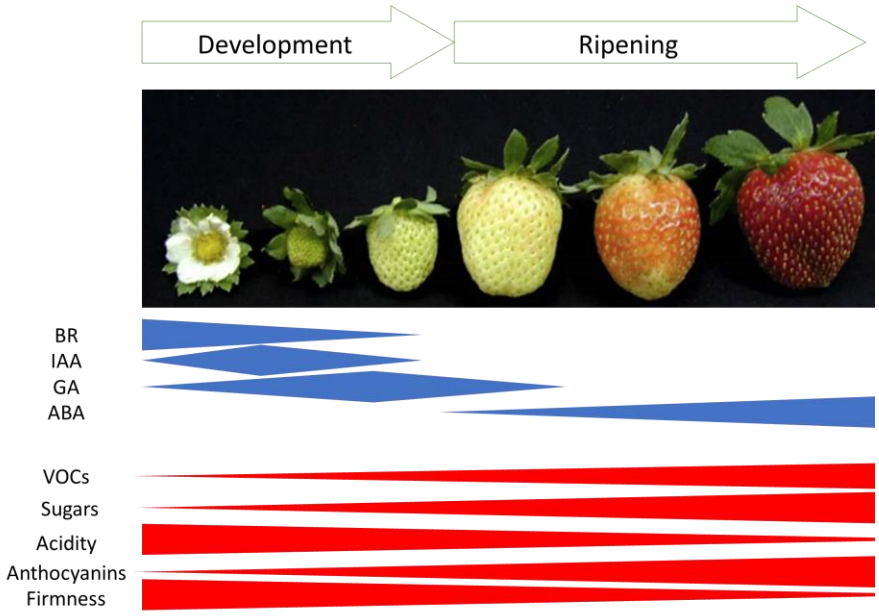


Figure I-8. Scheme of some development and ripening changes on hormones (blue) and metabolites (red).

3.4.4. Fruit colour studies

Strawberry has a characteristic red colour. However, there are accessions of *F. chiloensis* spp. *chiloensis* and some cultivated varieties which are whitish. This red-colour trait is related to the accumulation of flavonoid controlled by ABA during ripening process.

The anthocyanins content has been analysed in different *Fragaria* species (Muñoz *et al.*, 2011 and Urrutia *et al.*, 2015). Strawberry fruit development goes from green fruits to white and then to a turning stage before reaching the red fruit colour stage (Fig. I-8). During fruit development, an early peak of proanthocyanidins and ellagic acid is followed by a late peak of anthocyanins and flavonols accumulation (Halbwirth *et al.*, 2006, Fait *et al.*, 2008 and Baldi *et al.*, 2018). Pelargonidin-3-glucoside (Pg-3-Glu) is the most abundant anthocyanin compound that provide red colour. Flavonoids pathway is well known and mostly all genes involved have been studied in *F. x ananassa* (Almeida *et al.*, 2007, Carbone *et al.*, 2009, Buendia *et al.*, 2010, Thill *et al.*, 2013 and Baldi *et al.*, 2018). Anthocyanins are unstable molecules that need to be stored in the controlled acid environment of a vacuole.

Anthocyanins are transcriptional regulated by the MYB-bHLH-WDR complex (Xu *et al.*, 2015). R2R3 MYB family genes have been reported to regulate strawberry anthocyanin pathway, such as FaMYB1 (Aharoni *et al.*, 2001), FaMYB10 (Medina-Puche *et al.*, 2015), FaMYB9 and FaMYB11 (Schaart *et al.*, 2013). Key amino acid residues from MYB10 sequences were identified in *Rosaceae* family crops (Lin-Wang *et al.*, 2010). In cultivated strawberry, FaMYB10 was shown to upregulate the majority of genes involved in flavonoid / phenylpropanoid pathways and it is

repressed by auxin and activated by ABA (Medina-Puche *et al.*, 2015). A truncated FaMYB10 gene by the insertion of a short sequence "ACTTATAC" and located in LG1 produces a white fruit phenotype (Wang *et al.*, 2019), whereas in wild strawberry fruit, an amino acid W12S mutation (G/C) in the FvMYB10 protein is responsible for the red and yellow accessions (Hawkins *et al.*, 2016). Other MYB proteins are located in LG5 and LG6 (Wang *et al.*, 2019). *FaMYB9* and *FaMYB11,* from LG6, are involved in the accumulation of proanthocyanidins (Schaart *et al.*, 2013).

Few QTLs for colour traits have been reported. Anthocyanin QTLs were detected in HG1, HG2, HG3, HG5 and HG6 in different studies (Lerceteau-Köhler *et al.*, 2012, Zorrilla-Fontanesi *et al.*, 2011 and Castro and Lewers, 2016). Additionally, external fruit colour assay using CIELab parameters which compromise of 'a' going from red (+a) to green (-a), 'b' going from yellow (+b) to blue (-b) and 'L' going from brightness (+L) to darkness (-L) (Fig. I-9). QTLs were mapped for L value in LGIIa, LGIIIa, LGIVd and LGVIa, for 'a' value in all HG except of HGIII and HGVII and for 'b' value in LGIa, LGIIIa, LGVIa and LGVIb (Lerceteau-Köhler *et al.*, 2012). The only QTLs found in Zorrilla-Fontanesi *et al.* (2011) were for 'a' and 'b' value located in LGV-2.

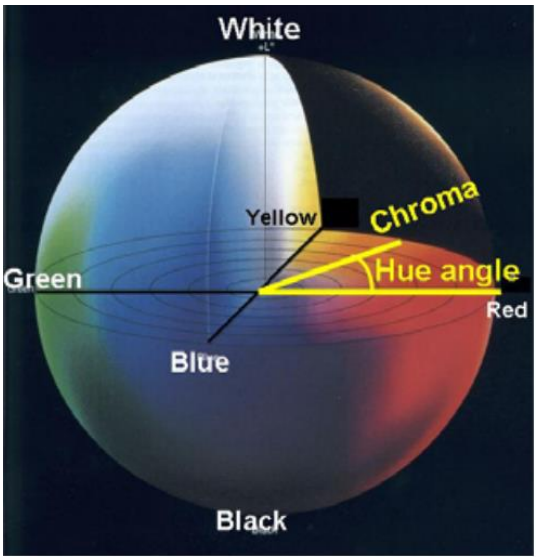


Figure I-9. Representation of CIElab values, Hue and Chroma (Strecker *et al.*, 2010).

As explained above, fruit quality traits are really important for consumer acceptance, but little is known about the genetic regions responsible for these traits in strawberry. To explore these genetic regions in a polyploid species, we must generate populations with seedlings that segregate for traits such as aroma, taste, colour and shape. To identify these regions through QTL analysis, it is essential to have a complete genotyping of the populations that will generate the saturated genetic maps. The analysis of VOC segregation in the population will allow us to identify the QTL that should be validated in other populations. The comparison of taste, colour

and shape in different populations together with the QTL analysis, should allow us to identify the genetic regions that explain their phenotypic variation and, in the end, determine the associated genetic markers that could be used in MAS.

Objectives

The general objective of this PhD thesis is to advance knowledge and genetic characterization of fruit quality traits in cultivated strawberry, focusing on three main characteristics: aroma, taste and appearance.

To achieve this main goal, three specific objectives have been stablished:

1. Generation of an F1 mapping population, originated from a cross between two breeding lines 'FC50' and 'FD54', and construction of a saturated genetic map by genotyping with SNPs. Validation of the genetic map coverage by comparison with strawberry genomes.
2. Identification and characterization of volatile strawberry compounds by GC-MS in the F1 mapping population and detection of different aroma QTLs. Validation of these QTLs in a second available strawberry population, F2, obtained from a cross between 'Camarosa' and 'Dover'.
3. Deep characterization of strawberry taste and appearance traits in the F1 and F2 populations, QTL mapping for segregating traits and comparison between both populations.

Chapter 1:

‘FC50xFD54’ genetic map construction and comparison with an octoploid consensus map and *Fragaria* genomes

Introduction

Cultivated strawberry (*Fragaria x ananassa*) is an allo-octoploid species (2n = 8x = 56) with an estimated genome size of 813Mb (Edger *et al.*, 2019). *F. x ananassa* is a hybrid between two wild octoploid species, *Fragaria chiloensis* and *Fragaria virginiana*. Despite being a polyploid specie formed by four subgenomes, its genome is highly diploidised. There are two hypotheses finding out the origins of each subgenome. The first one identified *Fragaria vesca* (AA) and *Fragaria iinumae* (BiBi) as two subgenomes, and the two others as a mixture of previous species but closer to *F. iinumae* (B1B1 B2B2) (Tennessen *et al.*, 2014 and Liston *et al.*, 2020). The other hypothesis appeared with the release of the cultivated strawberry genome and proposed that two of the subgenomes were from *F. vesca* and *F. iinumae* and the other two from *Fragaria nipponica* and *Fragaria viridis* (Edger *et al.*, 2019 and Edger *et al.*, 2020).

In order to properly detect a QTL, it is essential to construct a saturated genetic map. The first genetic maps in *F. x ananassa* were mainly built with SSRs markers designed in *F. vesca*, *F. iinumae*, *F. virginiana* and *F. x ananassa* (Sargent *et al.*, 2012, Isobe *et al.*, 2013 and Davik *et al.*, 2015) These maps were possible thanks to the high transferability of these markers to the *Fragaria* genus (Monfort *et al.*, 2006).

When a large number of SNPs started to be produced in plants, an SNP platform with 95062 strawberry polymorphic markers (IStraw90k) was released (Bassil *et al.*, 2015). As a consequence, high-density genetic maps began to be built in cultivated strawberry, as explained in the General introduction. The majority of them were built in an F1 population, such as 'Holiday' x 'Korona' (Bassil *et al.*, 2015 and van de Weg, *et al.* unpublished) and 'Monterey' x 'Darselect' (Sargent *et al.*, 2016). Moreover, an S1 genetic map from self-pollination of the Japanese cultivar 'Reikou' (Nagano *et al.*, 2017) and an F2 genetic map from 'Camarosa' and 'Dover' have been constructed, the latter by our group (Hidalgo *et al.*, 2014).

In order to reduce cost and get a higher proportion of informative SNPs, a subset of IStraw90k markers was selected to construct a new IStraw35k SNP array, with 38503 markers (Verma *et al.*, 2017a) and additional F1 genetic maps were built such as 'Emily' x 'Fenella' (Cockerton *et al.*, 2018), 'Redgauntlet' x 'Hapil' (Cockerton *et al.*, 2018) and 'Flamenco' x 'Chandler' (Cockerton *et al.*, 2019). Another genetic map of an F1 population from a cross between 'Sonata' and 'Babette' was firstly constructed with doble digest restriction-associated DNA (ddRAD) (Davik *et al.*, 2015) and recently fed with SNPs (Sargent *et al.*, 2019).

An alternative to the genetic map construction is looking for the marker position from available genetic map, such as 'Holiday' x 'Korona' (van de Weg *et al.* unpublished) or 'FL_08-10' x '12.115-

10' genetic maps (Verma *et al.* unpublished) and perform QTL analysis based on pedigree or GWAS analysis (Roach *et al.*, 2016, Mangandi *et al.*, 2017, Anciro *et al.*, 2018 and Salinas *et al.*, 2019).

Since genetic maps are limited by the marker segregation within a specific population, merging markers' position from different genetic maps will increase the genetic knowledge. With this aim, marker's positions from four F1 populations 'Redgauntlet' x 'Hapil' (Cockerton *et al.*, 2018), 'Emily x Fenella' (Cockerton *et al.*, 2018), 'Flamenco x Chandler' (Cockerton *et al.*, 2019), 'Capitola' x 'CF1116' (Lerceteau-Köhler *et al.* 2003) and an F2 'Camarosa' x 'Dover' (Hidalgo *et al.*, 2014) were used to generated an *F. x ananassa* consensus map (Cockerton *et al.*, 2018).

As already explained in the General introduction, new SNP arrays have been recently developed, such as a 850k SNP array and a derivate from the previous one called 50k SNP array (Hardigan *et al.*, 2020). Both of them have inherited markers from IStraw35k array. The first genetic maps using 50k SNP array were constructed with two F1 populations from a cross between *F. x ananassa* cultivar 'Camarosa' and *F. chiloensis* ecotype 'Del Norte' and from a cross between two *F. virginiana* subsp. *virginiana* accessions (Hardigan *et al.*, 2020). In addition to previous genetic maps, the F1 'Camarosa' x 'Del Norte' population was sequenced by low-coverage sequencing. These sequences were aligned against *F. x ananassa* genome (Edger *et al.*, 2019) discovering 3.7M subgenomic SNPs (Hardigan *et al.*, 2020).

In this study, we present a new genetic map using the IStraw35k SNP array for an F1 population derived from a cross between 'FC50' and 'FD54' breeding lines. This genetic map is compared with a *F. x ananassa* consensus map and *F. vesca* and *F. x ananassa* genomes.

Materials and Methods

Plant material

Plant material was generated through a cross between two breeding commercial lines. 'FC50' was chosen for presenting the characteristic aroma of wild strawberry and 'FD54' for its fruity aroma. The cross, named 'FC50xFD54', produced 70 individuals, which segregate, mainly, for volatile compound content.

Young leaves were used to extract DNA following a modified method of Doyle and Doyle, 1990 by adding 2% PVP at CTAB solution. DNA concentration (30ng/µl) and its quality were checked by NanoDrop spectrophotometer and PicoGreen.

Genetic map

High quality DNA extractions from 56 selected progenies and both parental lines were hybridised with Axiom® IStraw35 384HT array (hereafter, IStraw35k array) (Verma *et al.*, 2017a) in Thermo Fisher Scientific company. Markers were clustered using the default parameters from Axiom Suite software (Thermo Fisher Scientific, MA, USA). We excluded non-segregating markers and the ones presenting more than 10 missing data (>17% missing data). The segregating markers were classified into three categories: 1) codominant markers when genotypes of both parental lines were heterozygous, 2) 'FC50' segregation markers when the genotype of 'FC50' was heterozygous while 'FD54' was homozygous, 3) 'FD54' segregation markers when the genotype of 'FD54' was heterozygous while 'FC50' was homozygous. Markers from each category were filtered based on a chi-square test for its segregation distortion (p-value < 0.05). To create our genetic map, we used the JoinMap®5 software (van Ooijen, 2018). Since identical markers does not provide relevant information for genetic map construction, they were discarded to reduce computational power and added afterwards to have all segregating marker in our genetic map. The segregating markers were grouped using the Maximum Likelihood (ML) and LOD < 2 was used to split different linkage groups (LGs). To identify and orientate the different LGs, we adopted those from the *F. x ananassa* consensus map, which followed the subgenome nomenclature from van Dijk *et al.* (2014). Genetic distances were estimated using the mapping function Kosambi method (Kosambi, 1944). To have an accurate genetic map, we discarded possible wrong genotyped markers using the *Nearest neighbour stress* and probabilities of wrong genotype $-\log_{10}(P)$ guides provided by the software. To reduce the oversized genetic distances, all unlikely double recombination were checked in the genotype matrix and transformed into missing data. This dataset was used for genetic map construction. The 'FC50'

and 'FD54' genetic maps were constructed previously to the combined 'FC50xFD54' genetic map (hereafter, 'FC50xFD54' genetic map).

As genetic distances remained oversized, a reduced 'FC50xFD54' genetic map was created excluding, manually, the SNPs which had some missing data and unlikely genotypes.

'FC50xFD54' genetic map and reduced 'FC50xFD54' genetic map were drawn using Genetic-Mapper software with -bar and -pos parameters (Bekaert, 2016).

Synteny between parental genetic maps, 'FC50xFD54' genetic map and reduced 'FC50xFD54' genetic map was visualised using geom_jitter for marker distribution and geom_line for shared markers. Both functions are from ggplot2 (version 3.3.0) package from R (version 3.6.2) (RCoreTeam, 2019) and R Studio software (version 1.2.5033) (RStudio, 2019).

Dendrogram for 'FC50xFD54' population

The dendrogram of our mapping population 'FC50xFD54' was generated with the genotypes of the mapped IStraw35k markers. A total of 2882 SNPs were selected for having neither redundancies nor missing data. These markers were used to calculate distance matrix by Euclidean method from 'dist' function and hierarchical clustering by UPGMA algorithm from 'hclust' function (Supplementary table 1-4). Dendrogram plot was made up by 'dendeextend' (version 1.13.4) package. All these functions and package were from R software.

Map comparison to Fragaria resource

Blast+ application was installed in a local computer. IStraw35k markers were blast using default parameters to *F. vesca* cv. 'Hawaii-4' v4 genome (Edger *et al.*, 2018) and *F. x ananassa* cv. 'Camarosa' v1 genome (Edger *et al.*, 2019) to identify the SNP position in diploid and octoploid genomes.

Different steps have been followed to get the maximum number of marker position in *F. x ananassa* genome. We identified 4060 markers as unique positions in the genome. In many cases, markers located on different homeologous chromosomes of *F. x ananassa* genome were mapped into the same LG in our map. In this case, we accepted as corresponding, the most abundant chromosome in each LG. Afterwards, markers with two hits were added if they were mapped in the corresponding LG and chromosome.

We have obtained an *F. x ananassa* consensus map from a collaboration with EMR (UK), INRA (France), WUR (Netherland) and IRTA-CRAG (Spain) and several private companies such as PLANASA (Harrinson *et al.* unpublished).

‘FC50xFD54’ genetic map was compared with both *Fragaria* genomes obtained from GDR (Jung *et al.*, 2019) and *F. x ananassa* consensus map through dot plot using ggplot2 (version 3.3.0) and gridExtra (version 2.3) packages from R software.

High-density predicted map

Dried young leaves from ‘FC50’ and ‘FD54’ lines were sent to Thermo Fisher Scientific company. High quality DNAs were extracted using a column-based kit adding a Proteinase K in the lysis step following Affymetrix recommendation. These DNAs were hybridised with the 850k SNP array (Hardigan *et al.*, 2020) via RosBREED consortium. The clustering and genotyping were done by Knapp and collaborators using Axiom Suite software (Thermo Fisher Scientific, MA, USA).

The same segregation classification of the IStraw35k array were used to establish three categories for segregating markers such as codominant, ‘FC50’ segregation and ‘FD54’ segregation. The remaining markers that didn’t segregate were classified within two groups: Homozygous, if the genotypes of both parental lines were grouped in the same homozygous cluster, and Heterozygous no-segregating, if the genotype of each parental line belong to different homozygous clusters. Finally, genotypes with missing data for one or both parental lines were classified as Others. Predicted maps was visualised using geom_jitter from ggplot2 package from R software.

Results

IStraw35k filtering

'FC50xFD54' population was developed to segregate for aroma and fruit quality traits and we have genotyped this population to explore the segregating regions which explain the variance of some trait. A total of 18513 markers of the IStraw35k array (48.08% of total array markers) did not segregate. Focusing on the no-segregating markers, 17487 markers (45.41%) were Homozygous and 1026 markers (2.66%) were Heterozygous no-segregating. To sum up, a total of 19993 (51.93%) markers segregated. We discarded 2477 markers for having more than ten missing data or having error genotype in the parental lines. Moreover, we excluded 1309 more markers for distortional segregation. A total of 16207 markers (81.06% segregating markers) were introduced into JoinMap®5. We finally discarded 1612 markers for introducing stress in 'FC50xFD54' genetic map. The Others category consisted of markers discarded for containing >10 missing data, segregation distortion and introducing stress to genetic map (Fig. 1-1).

Focusing on the mapped markers, we found 3122 codominant markers (21.39% mapped markers) whereas NoMinorHomozygous was the most abundant marker category with 11473 markers (78.61%), as it was expected. NoMinorHomozygous can be classified in two groups depending on parental line segregation. We counted 5982 markers (41.17% mapped markers) for 'FC50' segregation and 5491 markers (37.79%) for 'FD54' segregation. Segregating markers from 'FC50' were slightly more abundant compared to the ones segregating in 'FD54' (Fig. 1-1).

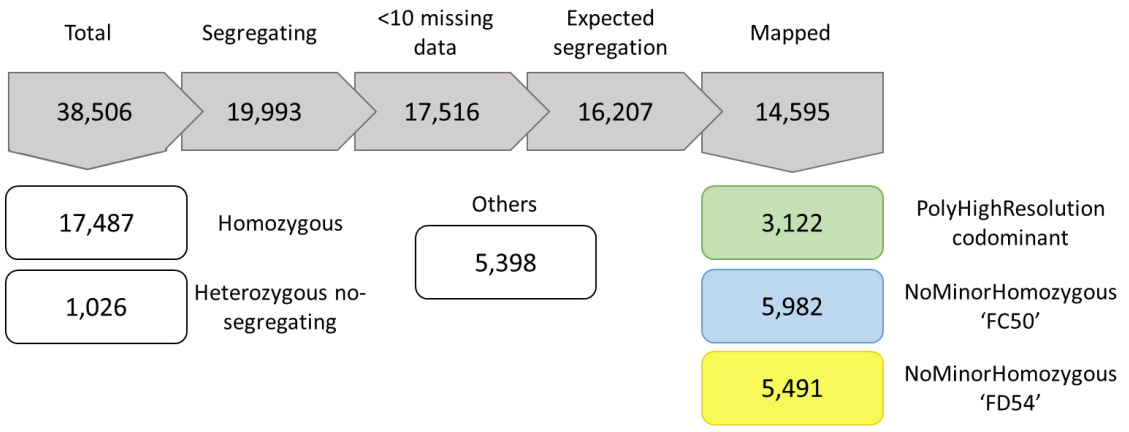


Figure 1-1. Summary of marker filtering flow in IStraw35k array and number of segregating markers in 'FC50xFD54' genetic map for each marker class.

‘FC50xFD54’ genetic maps construction

In order to construct a combined ‘FC50xFD54’ genetic map, we firstly build the parental genetic maps. A total of 6196 markers segregated on the ‘FC50’ (mother) genetic map, grouped into 31 LGs and giving a total size of 2234.66cM. Three of the expected 28 LGs, LG1D, LG4A and LG6B, presented two fragments each. On the other hand, the ‘FD54’ (father) map was constructed with 5490 markers, divided into 32 LGs and spanning 2000.63cM. This map presented four split LGs, LG2A, LG2C, LG3C and LG5B, and markers from LG4C and LG5B1 fell into a single map position (Table 1-1, Supplementary table 1-1 and 1-2).

In addition to the parental genetic maps, we constructed a combined ‘FC50xFD54’ genetic map. This combined genetic map merge ‘FC50’ and ‘FD54’ genetic maps using codominant markers as anchors. The resulting ‘FC50xFD54’ genetic map consisted of 14595 markers in 28 LGs and spanning 3451.38cM. A total of eight LGs (LG1B, LG1D, LG3B, LG3C, LG4A, LG5C, LG6B and LG7B) presented a gap greater than 25cM (Table 1-1). LG1C showed an unexpected short size. The largest LGs in our map were LG6A and LG6B with a total size of 193cM (Fig. 1-2). Despite a high number of segregating markers, a total of 2090 unique positions were present in the ‘FC50xFD54’ genetic map (Supplementary table 1-3).

Table 1-1. Summary of ‘FC50xFD54’ and both parental genetic maps.

	‘FC50xFD54’			‘FC50’			‘FD54’		
LG	Markers	Size (cM)	Loci	Markers	Size (cM)	Loci	Markers	Size (cM)	Loci
1A	827	160.90	137	159	101.57	30	282	93.36	29
1B	282	129.59	65	189	98.02	33	31	31.23	8
1C	151	57.16	34	109	55.39	27	16	5.53	3
1D	345	146.08	41	120 | 69	36.75 | 15.01	20 | 6	133	64.82	21
2A	854	140.19	108	351	87.02	42	340 | 78	52.21 | 20.18	21 | 12
2B	406	97.93	52	250	90.28	28	121	59.10	19
2C	326	105.66	72	100	100.35	13	41 | 81	5.49 | 18.36	4 | 10
2D	342	91.99	52	94	57.52	23	209	90.23	27
3A	885	170.51	122	215	57.16	18	400	122.70	42
3B	643	106.44	74	228	101.86	29	303	93.32	32
3C	567	172.65	66	404	99.83	35	26 | 61	1.818 | 60.84	2 | 9
3D	782	144.89	110	260	113.05	32	439	122.27	54
4A	175	79.70	26	67 | 25	11.66 | 7.33	5 | 5	58	76.24	8
4B	467	118.48	78	301	123.24	53	129	80.65	23
4C	242	93.77	49	153	78.83	25	6	0.00	1
4D	388	103.84	63	168	102.85	28	150	66.67	15
5A	821	154.34	109	275	80.81	22	170	87.09	18
5B	747	153.53	110	412	99.12	48	6 | 171	0.00 | 38.92	1 | 24
5C	430	96.04	57	212	82.76	21	123	68.58	12
5D	753	125.23	98	279	61.95	24	315	91.81	34
6A	843	193.55	114	238	119.89	32	320	94.45	35
6B	612	193.34	84	42 | 74	42.29 | 9.16	7 | 5	409	131.80	49
6C	680	132.69	86	439	57.65	25	320	95.05	41
6D	501	133.66	70	335	110.69	34	29	50.83	6
7A	401	75.47	51	113	48.74	21	254	68.02	21
7B	418	96.69	45	173	74.67	15	179	88.29	24
7C	225	60.84	26	45	11.08	7	170	54.15	21
7D	482	116.23	91	297	98.13	46	120	66.64	17
all	14595	3451.38	2090	6196	2234.66	759	5490	2000.63	643

Vertical bars (|): separation between the two fragments from the same LG.

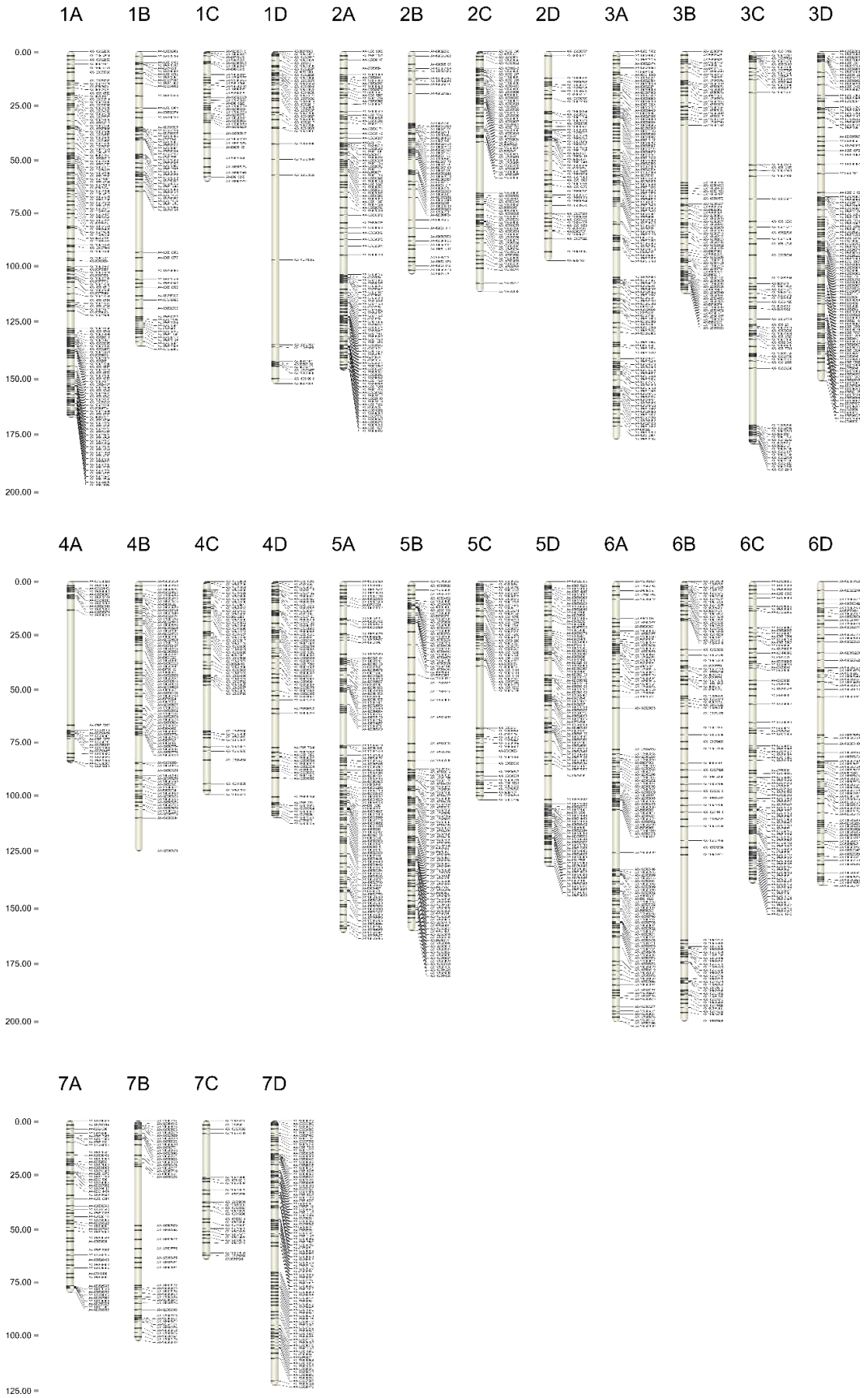


Figure 1-2. Representation of the 'FC50xFD54' genetic map with unique positions distributed in seven LGs (1-7) and four HGs (A-D).

Although the construction of the ‘FC50xFD54’ genetic map was guided by the previously constructed parental genetic maps, little differences could be observed. Markers showed a perfect relative position in all three genetic maps. Regarding to the presence of gaps in the genetic maps, all three genetic maps showed the same four big gaps (LG1B, LG3B, LG4A and LG7B). Furthermore, a big gap in ‘FC50xFD54’ genetic map was observed in LG1D and LG6B, whereas in the ‘FC50’ genetic map, both LGs were fragmented in two pieces. In the ‘FD54’ genetic map, they were smaller in size and covered along ‘FC50xFD54’ genetic map. Opposite situation than previous LGs, LG3C from ‘FD54’ genetic map was divided in two fragments. Whereas, markers from LG3C in ‘FC50’ genetic map were also spread over ‘FC50xFD54’ genetic map but being smaller in size. Concerning to LG2C, ‘FC50’ genetic map presented three gaps and ‘FD54’ genetic map were split in two fragments, whereas ‘FC50xFD54’ genetic map presented only one gap. ‘FD54’ markers contributed poorly in LG4C appearing in one single locus (Fig. 1-3).

Moreover, some large genetic distances in the ‘FC50xFD54’ genetic map were caused by missing data or double recombination errors. For this reason, it was necessary to apply a stricter marker filter. These selected markers are all around ‘FC50xFD54’ genetic map (Fig. 1-4) and they have been highlighted (Fig. 1-3).

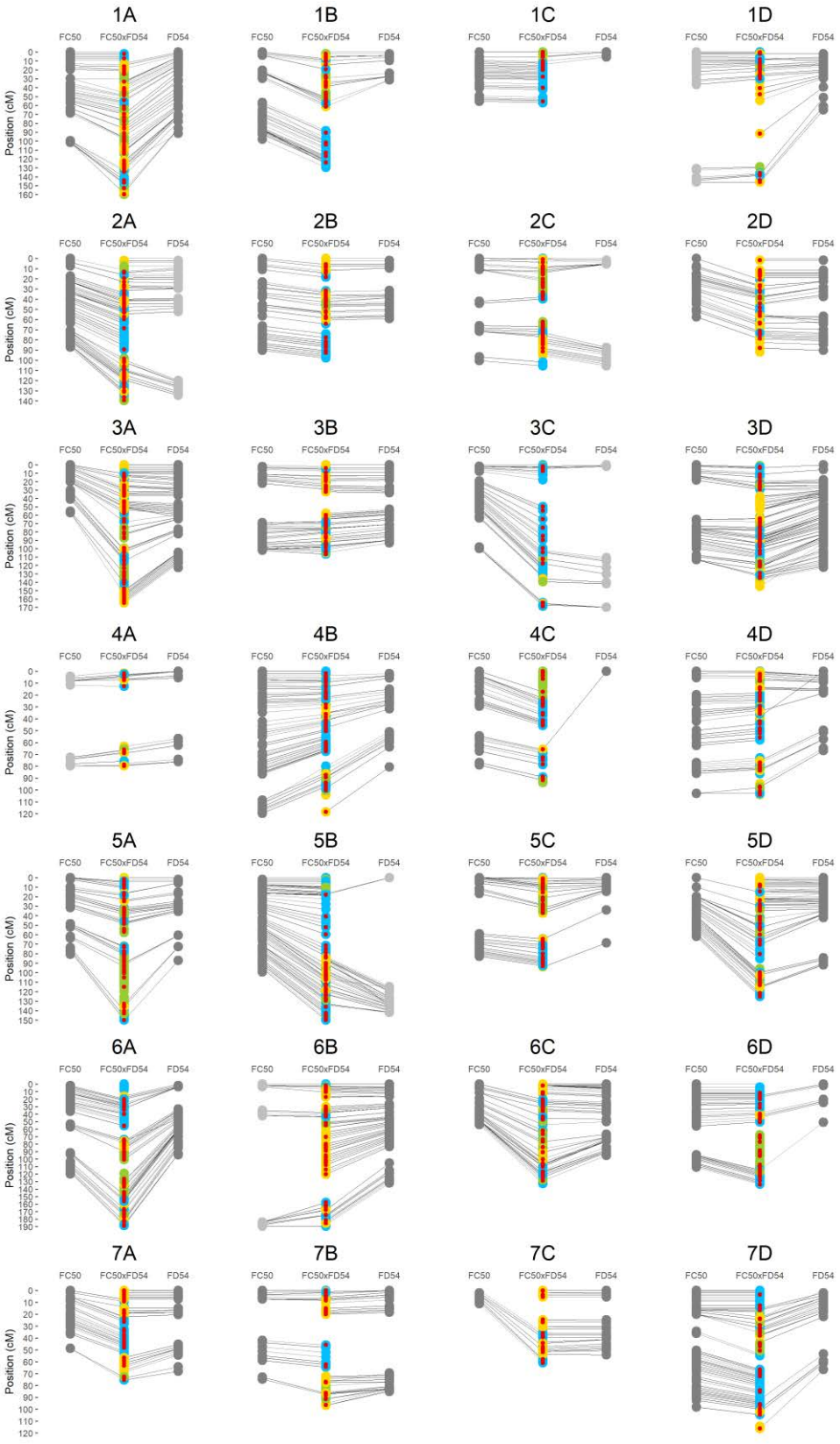


Figure 1-3. Marker position comparison between ‘FC50xFD54’ and parental genetic maps. Colours indicate: codominant markers (green), ‘FC50’ segregating markers (blue), ‘FD54’ segregating markers (yellow), markers position in each parental genetic map in entire LGs (dark grey) and divided LGs (light grey), markers selected in reduced ‘FC50xFD54’ genetic map (red).

After excluding SNPs with missing data, unlikely genotypes and redundant markers, we built a new combined map, called reduced ‘FC50xFD54’ genetic map. This map was constructed with 1461 markers, grouped into 28 LGs and a total size of 2332.50cM (Fig- 1-4). The total size was reduced by one third in comparison with the original genetic map. In fact, LGs over 120cM were decreased from 14 to only three LGs (LG3A, LG6A and LG6B). Most of the gaps size were also diminished. However, six LGs (LG1B, LG1D, LG3C, LG4A, LG6B and LG7B) out of the eight of the first map had still gaps larger than 25cM (Fig. 1-4).

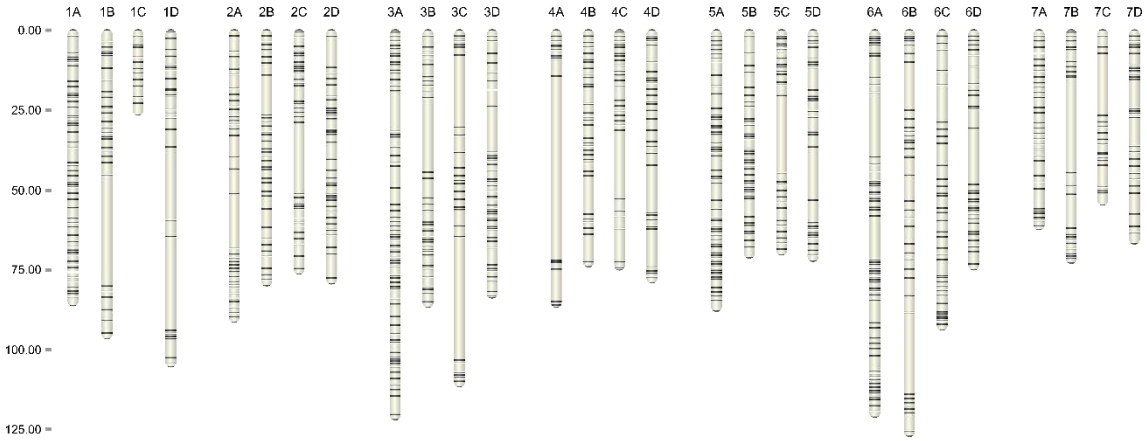


Figure 1-4. Reduced ‘FC50xFD54’ genetic map distributed in seven LG (1-7) and four HGs (A-D).

Genetic distances between ‘FC50xFD54’ progenies

Knowing the genetic distances between progenies and parental lines are really important for planning future crosses or understanding phenotypic distances. Since it is an F1 population, large genetic distances were observed between progenies being far related (Fig. 1-5). We observed two asymmetrically differentiated clusters in the ‘FC50xFD54’ dendrogram. Cluster A was compromised with 45 progenies, representing the 82% of them, and being closely related to ‘FD54’ parental line. Whereas, cluster B had 11 progenies being grouped together with the ‘FC50’ parental line. As ‘FD54’ had a higher ratio of homozygous marker than ‘FC50’ (Fig. 1.1), we expected a greater number of siblings related to ‘FD54’ (Fig. 1-5).

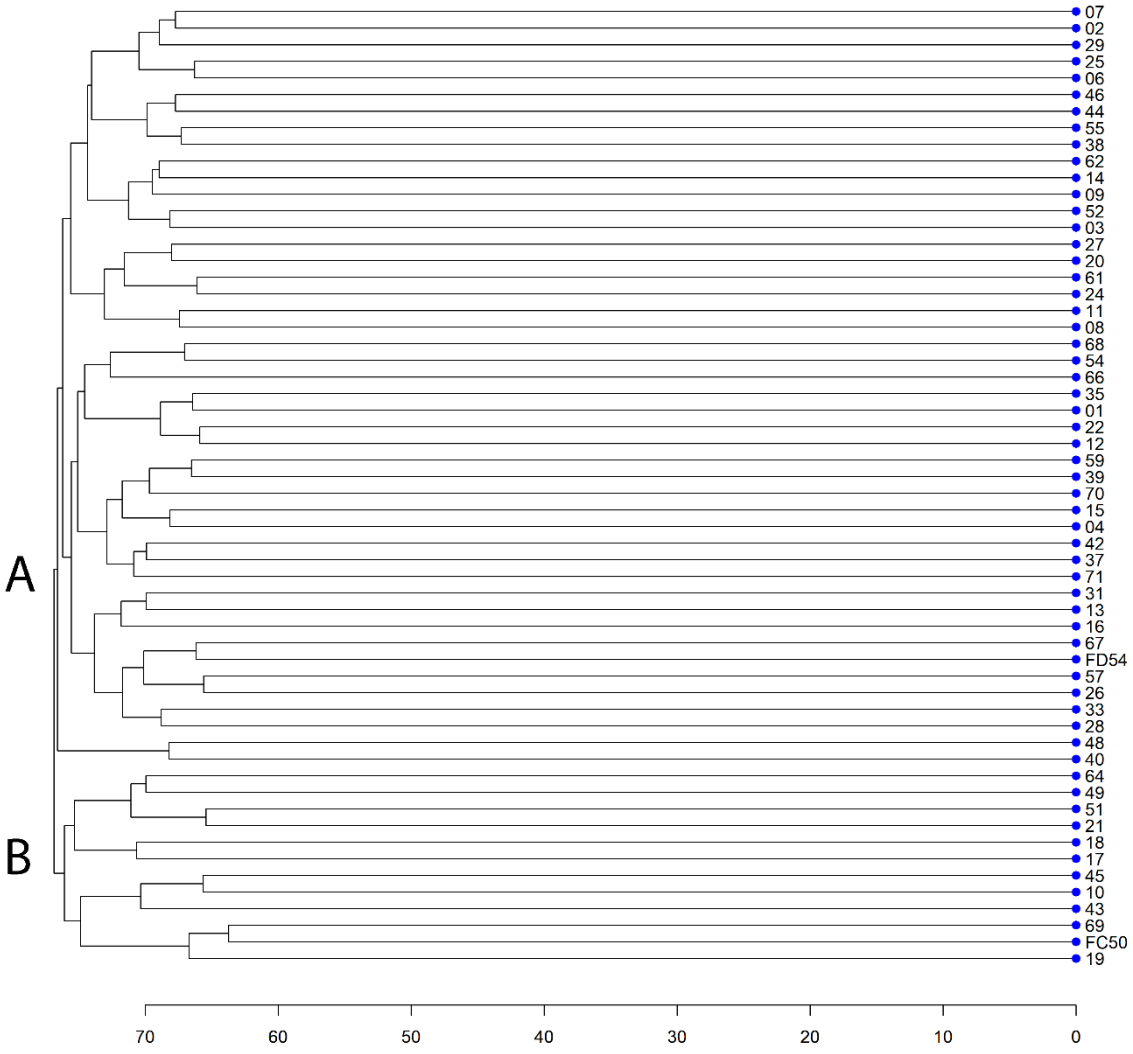


Figure 1-5. 'FC50xFD54' dendrogram. Distance matrix was calculated with Euclidean method and hierarchical clustering with UPGMA algorithm. Letters represent the two cluster groups.

Map comparison to *Fragaria vesca* genome

F. vesca is the dominant subgenome from the cultivated strawberry. Furthermore, SNPs from IStraw90k were developed by aligning sequences against its first genome version. Hence, the comparison with this diploid genome should be informative and confirm the linkage groups (LGs) without differentiating the homeologous groups (HGs). We have localised the majority of IStraw35k markers in the *F. vesca* v4 and almost all of them presented a unique hit providing an unmistakable physical genome position. Even though some markers should not come from *F. vesca*-like subgenome, they could have a physical position without showing segregation (see General introduction Fig. I-4).

A dot plot using markers from the 'FC50xFD54' genetic map visualizes the collinearity between our genetic map and the physical map of *F. vesca*. Additionally, markers are differentiated by

their segregation category. Since few markers were blasted in an unexpected chromosome, we detected a good collinearity between maps. The only exception was LG2D which presented two lineal tendencies. We found that large LGs, such as HG3 and HG6, corresponded to large *F. vesca* chromosomes and small LGs, such as LG1C and LG7C, corresponded to a lack of fragment at the bottom of LG1C and at the top of LG7C. We also noticed that usually big gaps in different LGs correspond to big fragments of chromosomes without markers in *F. vesca*, such as LG1B, LG3B, LG3C, LG4A, LG5C and LG6B. However, the big gaps of the LG1D and LG7B were much smaller in *F. vesca* genome (Fig. 1-6).

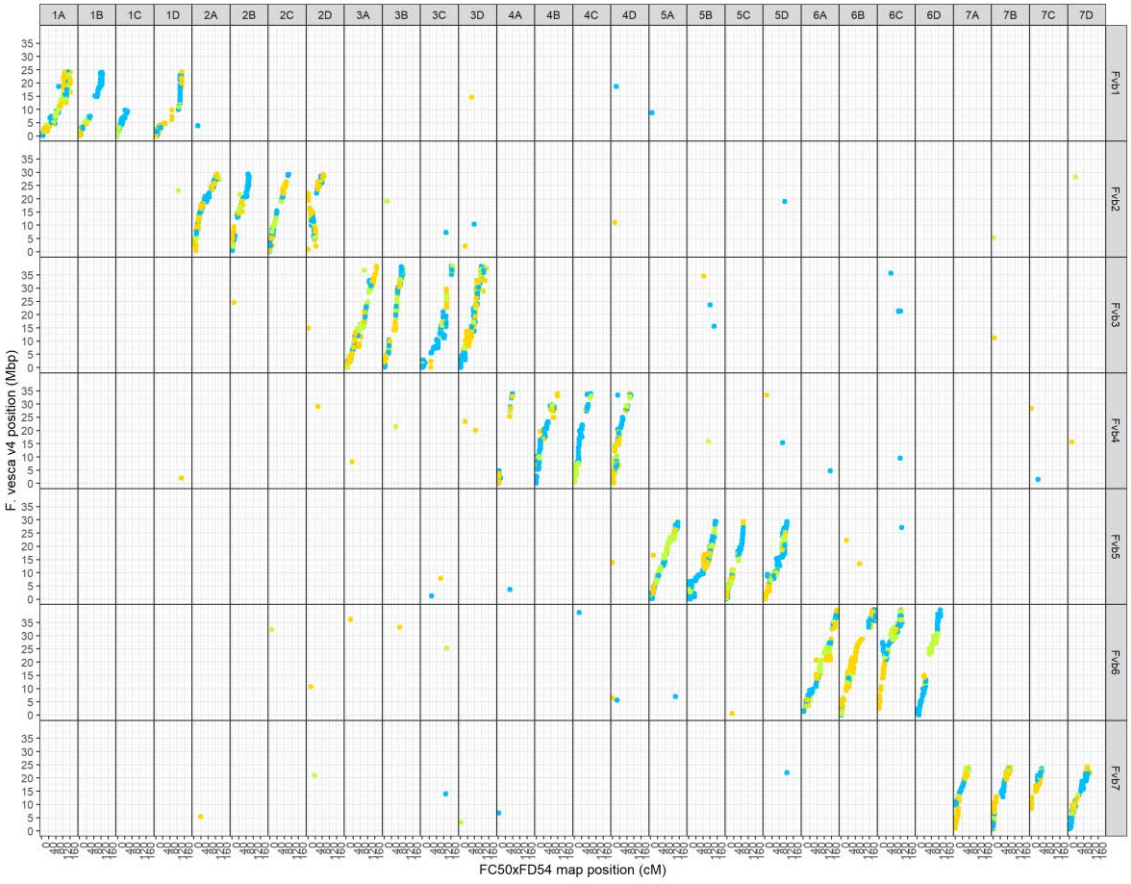


Figure 1-6. Dot plot comparison between 'FC50xFD54' genetic map and *F. vesca* genome common markers grouped in four homeologous group for each *F. vesca* chromosome. Colours indicate marker segregation: green (codominant), blue ('FC50') and yellow ('FD54').

Map comparison to *F. x ananassa* consensus map

The consensus map is an artificial map generated by selecting common markers from different *F. x ananassa* genetic maps and calculating their relative position. Since the SNP distribution from this consensus map is validated in four genetic maps, looking for the collinearity with this consensus map will reinforce the SNPs order of our genetic map.

We found a high collinearity between 'FC50xFD54' genetic map and the consensus map (Fig. 1-7). However, few markers from LG6C, showing 'FC50' segregation, broke the LG trend. Concerning the big gaps in the LGs, most of them were shared in both maps (LG1B, LG1D, LG2C, LG4A, LG4C, LG5C, LG6B and LG7C). However, those big gaps in LG3B, LG3C and LG7B were smaller in the consensus map. As in the comparison with *F. vesca* genome, we also observed the missing fragment at the bottom of our LG1C (Fig. 1-7).

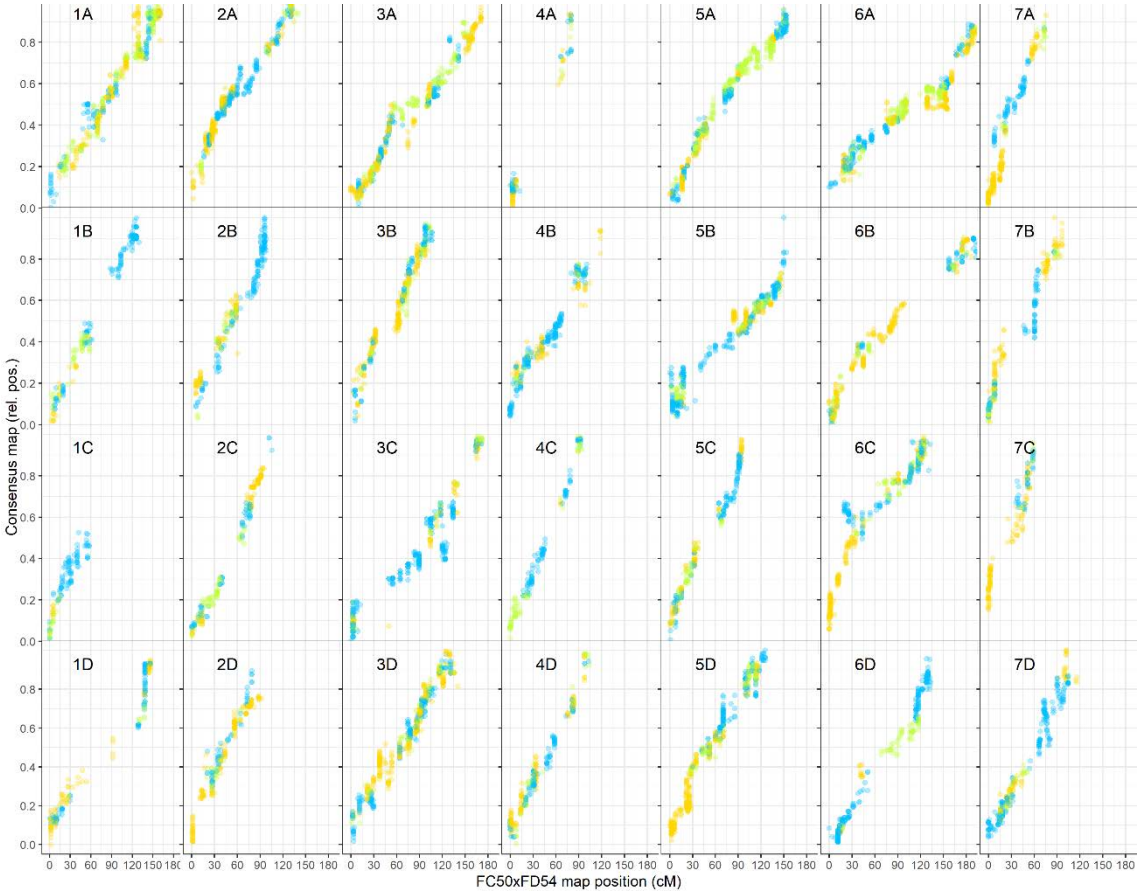


Figure 1-7. Dot plot comparison between 'FC50xFD54' genetic map and the consensus map showing common marker from the same LGs. Colours indicate markers segregation: green (codominant), blue ('FC50'), yellow ('FD54').

Map comparison to *F. x ananassa* genome

With the recent apparition of the octoploid genome, we can compare our genetic map with its physical position, distinguish between HGs and look for the correspondence between them. However, most of the markers from 'FC50xFD54' genetic map showed multiple hits in homeologous chromosomes from *F. x ananassa* genome. Luckily, 4060 markers were located in a unique position. Although LGs had markers from the four homeologous chromosomes, a majority of the markers corresponded to a single one. Correspondence between LGs and

chromosomes were relatively easy, especially in the case of the LGs that corresponded to *F. vesca*-like subgenome because they had much more markers than the rest of the homeologous groups. Each correspondence between LGs and chromosomes are shown in Fig 1-8.

Once we compared each LGs with their correspondent chromosome, we observed a good correlation in most of the cases. However, a total of 15 inversions were observed (LG1D, LG2A, LG2B, LG2D, LG3A, LG3C, LG4A, LG4B, LG4D, LG5B, LG6A, LG6B, LG6D, LG7B and LG7D). Moreover, some LGs such as LG1C, LG2D and LG6C had small inversions. Some markers at the end of LG6C and LG6D were mapped in the opposite side of each chromosome. Furthermore, LG4B showed a clear tendency but some markers escaped of it (Fig. 1-7).

We observed some regions showing multiple position in the genome but no differences were observed in the genetic map such as LG1B (end), LG1D (end) and LG2B (end). Moreover, some regions showed greater distances in the genetic map than in the genome, such as LG1D (beginning) and LG3C (beginning) (Fig. 1-7).

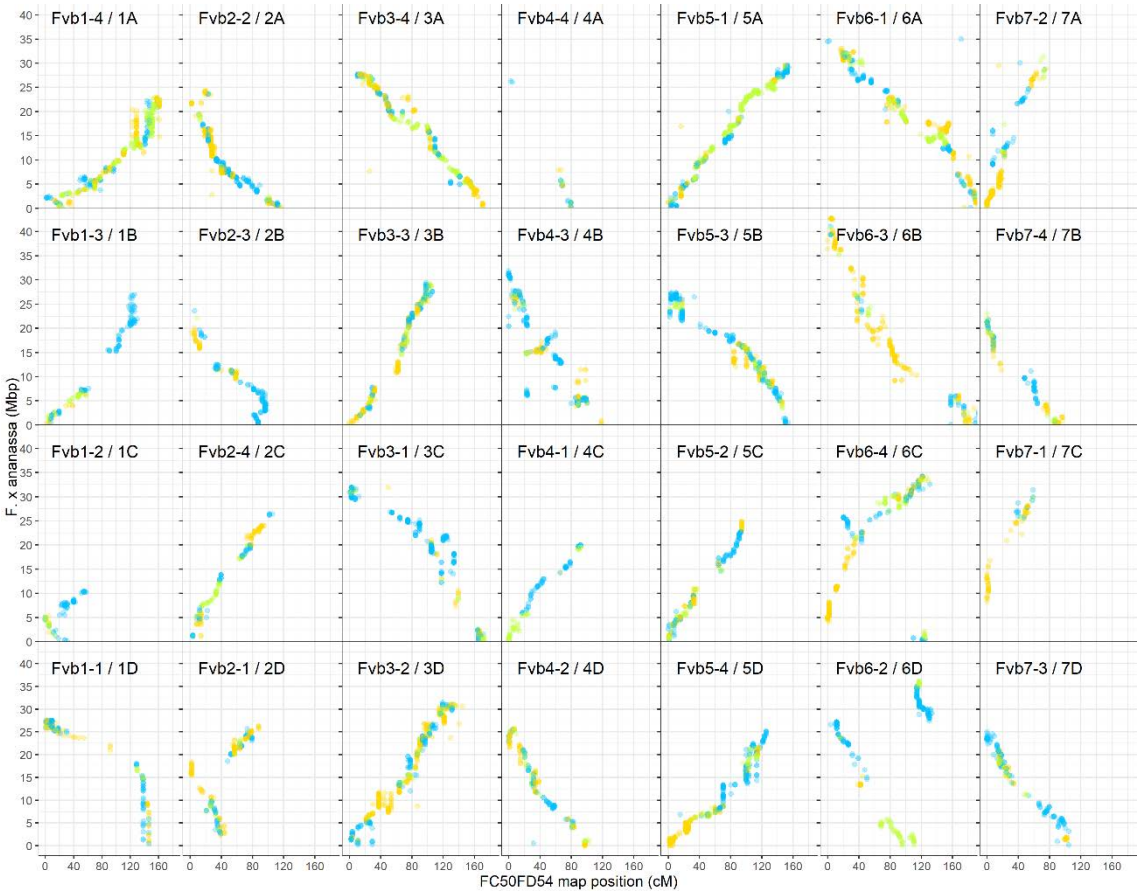


Figure 1-8. Dot plot comparison between 'FC50xFD54' genetic map and *F. x ananassa* genome showing common markers in the correspondence between LGs and chromosomes. Colours indicate markers segregation: green (codominant), blue ('FC50') and yellow ('FD54').

Synteny between genetic maps and *F. x ananassa* genome

To further compare our genetic map with *F. x ananassa* consensus map and *F. x ananassa* genome, we found high synteny between them. The main exceptions were observed with LG2D and LG6D which showed high synteny with consensus map but they had partial inversion in *F. x ananassa* genome. Moreover, few markers from LG1C, LG2A, LG4B and LG6C showed a wrong *F. x ananassa* position (Fig. 1-9).

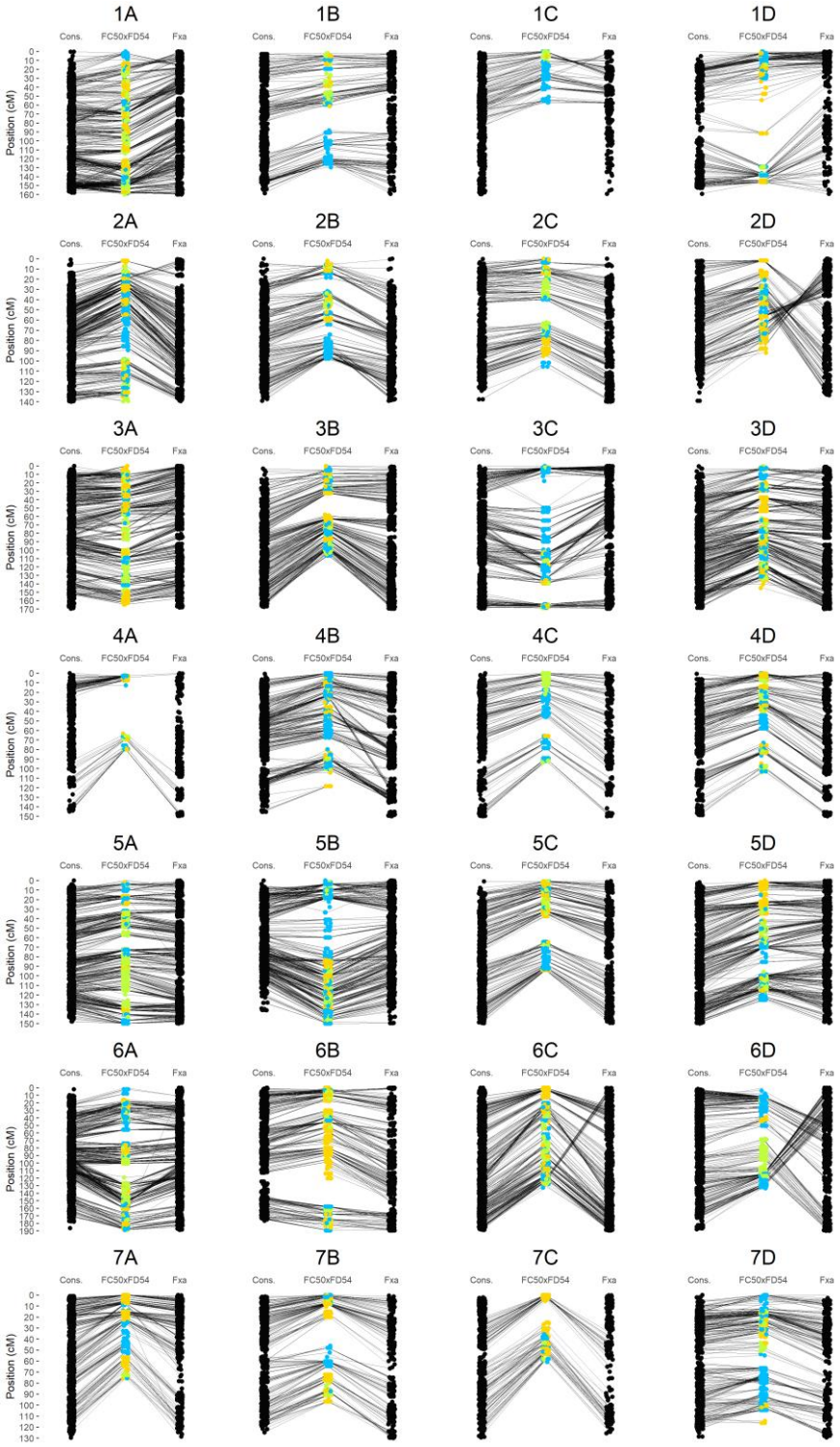


Figure 1-9. Synteny between genetic maps and the *F. x ananassa* physical genome position. Markers position from 'FC50xFD54' genetic map (FC50xFD54), *F. x ananasa* consensus map (Cons.) and *F. x ananasa* genome (Fxa) in each linkage group. Position axis is related to 'FC50xFD54' genetic map and consensus map and *F. x ananassa* position are normalised to maximum HG. Colours indicate markers segregation: green (codominant), blue ('FC50'), yellow ('FD54') and black (no segregation information).

Putative map comparison between IStraw35k and 850k SNP arrays

The main difference between IStraw35k and 850 SNP arrays is that the former was designed using diploid *F. vesca* genome and the latter using the octoploid *F. x ananassa* genome providing subgenome specific markers. As the parental lines, 'FC50' and 'FD54', had been genotyped in both arrays, we can predict the expected marker segregation for a F1 population and compare the ratio of each category to both SNP arrays. However, a total of 487082 markers (49.88%) from 850k SNP array (976478) were well genotyped.

By comparing the ratio of marker segregation classes, we are not able to detect huge differences between genotyping with IStraw35k or 850k SNP arrays. However, looking at the different ratios in detail, we observe that there is almost half number of markers of Heterozygous no-segregating class (2.66% for the IStraw35k and 4.94% for the 850k SNP arrays) and almost three times markers for the Others class (2.89% and 1.16%). Nevertheless, comparing the total of non-segregating marker ratios, we find similar values in both arrays (Table 1-2).

Markers which segregate from 'FC50' parental line were the most abundant, as we observed with the mapped markers from IStraw35k array, reaffirming that the 'FC50' parental line is more heterozygous than the 'FD54' parental line (Fig. 1-1 and Table 1-1).

Table 1-2. Classification of 'FC50xFD54' expected segregation in IStraw35k and 850k SNP arrays inferred from 'FC50' and 'FD54' genotypes.

Marker segregation class	IStraw35k	850k
Codominant	5,011 (13.01%)	66,591 (13.67%)
'FC50' segregation	7,235 (18.79%)	93,835 (19.26%)
'FD54' segregation	6,636 (17.23%)	83,541 (17.15%)
Heterozygous no segregating	1,026 (2.66%)	24,083 (4.94%)
Homozygous	17,487 (45.41%)	213,366 (43.80%)
Others	1.114, (2.89%)	5,666 (1.16%)
Total	38,506 (100%)	487,082 (100%)

To deeply compare both SNP arrays, we have visualised a predicted genetic map using physical marker positions from the 850k and the IStraw35k SNP arrays. Additionally, we also used the relative consensus map positions for the IStraw35k array. In this last map, fifteen LGs had to be inverted and all LGs had to reordered as shown in Fig. 1-8.

As already noticed, higher number of markers from IStraw35k array are mapped in the LGs related to *F. vesca*-like subgenome compared to the other HGs. Additionally, some regions with few markers using the IStraw35k array are filled with segregating markers in the 850k SNP array, such as the middle region of Fvb1-1 (LG1D), the beginning of Fvb2-1 (LG2D) and the middle of Fvb7-1 (LG7C). Big regions whose markers are not segregating, such as the end of Fvb1-2 (LG1C) and the middle of Fvb4-4 (LG4A), are observed in both arrays. We noticed the same with the gap in Fvb7-1 (LG7C).

Focusing on the segregation classes, high 'FC50' marker ratio is found in both arrays, such as in some regions from Fvb1-3 (LG1B), Fvb2-3 (LG2B), Fv3-1 (LG3C), Fvb3-2 (LG3D), Fv4-1 (LG4C), Fvb4-3 (LG4B), Fvb5-3 (LG5B), Fvb6-2 (LG6D), Fvb7-3 (LG7D) and Fvb7-4 (LG7B). Therefore, high number of 'FD54' markers in both arrays are observed in Fv3-2 (LG3D), Fvb5-4 (LG5D), Fvb6-3 (LG6B), Fvb6-4 (LG6C) and Fvb7-2 (LG7A). Other chromosomes with high quantities of codominant markers in both arrays are Fvb1-4 (LG1A), Fvb3-3 (LG3B), Fvb3-4 (LG3A), Fvb4-1 (LG4C), Fvb5-1 (LG5A), Fvb6-1 (LG6A) and Fvb6-2 (LG6D). To sum up, similar predicted segregation markers are mapped in both arrays (Fig. 1-10).

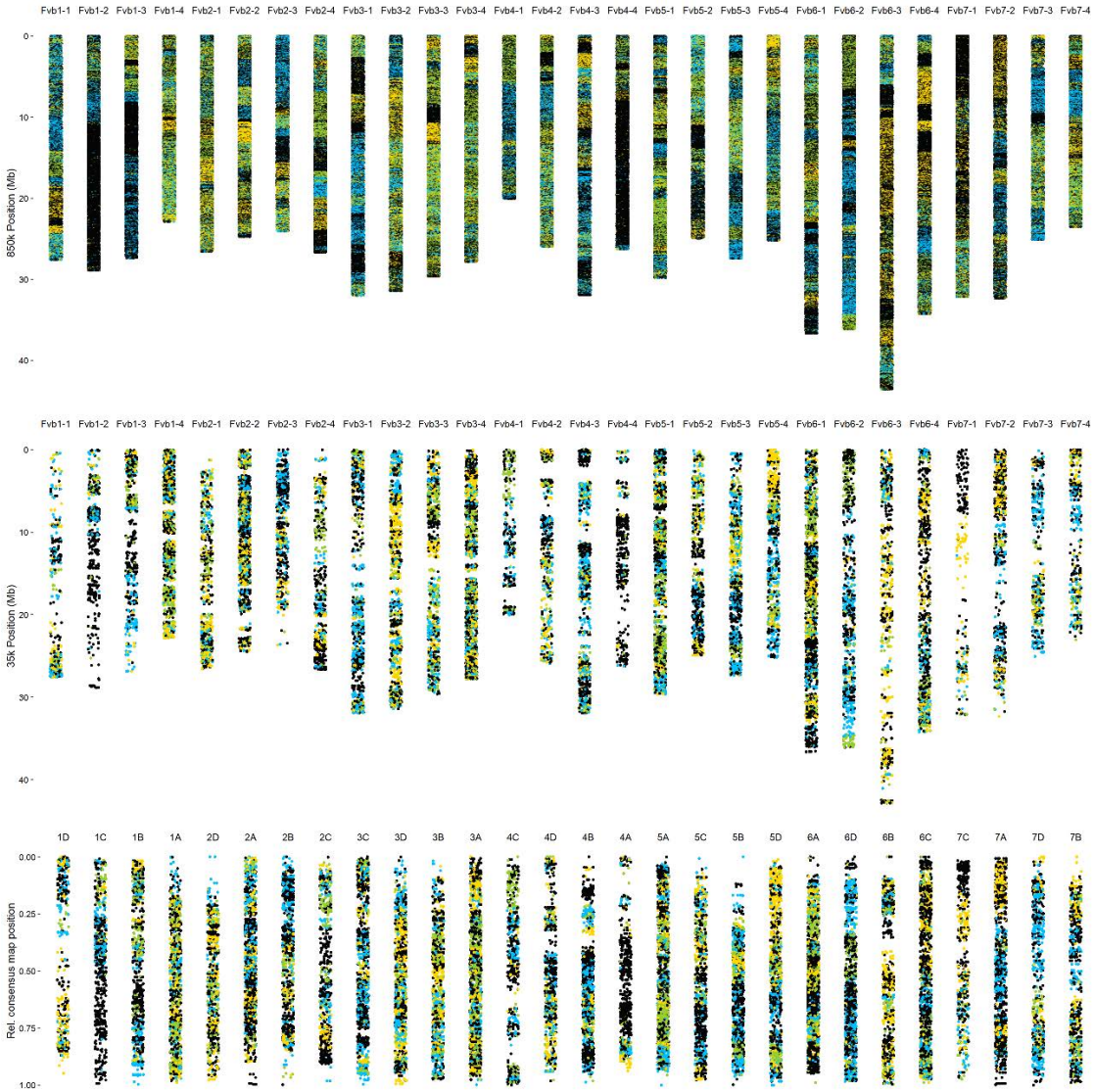


Figure 1-10. Predicted ‘FC50xFD54’ segregation in 850k and IStraw35k SNP arrays. Upper: 850k SNP array. Middle: IStraw35k in physical position. Lower: IStraw35k array in consensus map position. Colours indicate markers segregation: green (codominant), blue (‘FC50’ segregating markers), yellow (‘FD54’ segregating markers) and black (no segregating markers).

Discussion

Fragaria x ananassa is an octoploid specie with a high interest in breeding programs for its largely consumed and appreciate fruits. Its polyploid nature adds difficulty to analyse the genetic heritability of quantitative traits. The construction of saturated genetic maps from breeding lines is a previous effort to the QTL analyses on the traits of interest. The breeding lines ‘FC50’ and ‘FD54’ have been selected for their genetic distance and fruit quality differences, to generate a new F1 population.

A population of 56 individuals and both parental lines were hybridised with the IStraw35k array. From the mapped markers, we have classified them as codominant (3122), ‘FC50’ segregation (5982) and ‘FD54’ segregation (5491) (Fig. 1-1). Additionally, expected ‘FC50xFD54’ using parental lines genotypes from 850k SNP array showed 13.67% of codominant markers, 19.26% of ‘FC50’ segregation’ and 17.15% of ‘FD54’ segregation similar to the obtain from IStraw35k array (Table 1-2). These results suggest that the parental line ‘FC50’, mother line, is higher heterozygous than the parental line ‘FD54’. Moreover, ‘FC50xFD54’ dendrogram showed one little cluster (having ‘FC50’ line) and one big cluster (having ‘FD54’ line) (Fig. 1-5) which could be explained by the parental heterozygosity ratio.

The ‘FC50xFD54’ genetic map has 14595 markers in 28 LGs and spanning 3451cM. However, eight LGs show a gap bigger than 25cM and other two LGs lack a fragment in the extreme of each LGs. ‘FC50xFD54’ siblings present the highest quantity of segregating markers compared to previous genetic maps. Moreover, ‘FC50xFD54’ together with ‘Redgauntlet’ x ‘Hapil’, ‘FL_08-10’ x ‘12.115-10’ and ‘Camarosa’ x ‘Dover’ are the only genetic maps with only 28 LGs (Table 3).

Finally, SNP arrays constitute a high throughput genotyping platform which could provide a high number of missing data or wrong genotypes. These genotyping errors could enlarge genetic maps giving unrealistic genetic distances. For this reason, ‘FC50xFD54’ genetic map shows an oversized genetic distance (Table 1-3). To solve this problem, we selected markers around ‘FC50xFD54’ genetic map and constructed a reduced ‘FC50xFD54’ genetic map (Fig.1-4). This reduced ‘FC50xFD54’ genetic map decreases by two third the previous genetic map distance. Therefore, this reduced ‘FC50xFD54’ genetic map is suitable for QTL analysis.

Table 1-3. Summary of genetic maps

Cross	Type	Siblings	Markers	LGs	Size	Reference
'Holiday' x 'Korona'	F1	75	6593	35	2050	Bassil *et al.*, 2015
			12966	37		van de Weg, *et al.* unpublished
'Darselect' x 'Monterey'	F1	86	8407	32	1820	Sargent *et al.*, 2016
'Reikou'	S1	161	11574	31	2816	Nagano *et al.*, 2017
'Redgauntlet' x 'Hapil'	F1	168	11581	28	2545	Cockerton *et al.*, 2018
'Emily' x 'Fenella'	F1	181	11552	29	3207	Cockerton *et al.*, 2018
'FL_08-10' x '12.115-10'	F1	165	14332	28		Anciro *et al.*, 2018
'Flamenco' x 'Chandler'	F1	140	9937	32	3069	Cockerton *et al.*, 2019
'Camarosa' x 'Dover'	F2	117	13270	36		Hidalgo *et al.*, 2014
			7977	28	2056	This thesis (Annex 1)
'FC50' x 'FD54'*	F1	56	14595	28	3451	This thesis
			1461	28	2332	

* Data from both 'FC50xFD54' and reduced 'FC50xFD54' genetic maps

Regarding to the comparison with *F. x ananassa* consensus map, more than one quarter of their makers (9982 of 35441) are in the 'FC50xFD54' genetic map. Using these common markers, we found high collinearity between both maps (Fig. 1-7). This collinearity verifies the order of markers in both maps. Furthermore, 4613 markers, which are not in the consensus map, could be added.

Since IStraw90k array, and consequently IStraw35k array, were built using the diploid *F. vesca* genome v1 (Shuev, *et al.* 2011), high collinearity between 'FC50xFD54' genetic map and *F. vesca* genome v4 (Edger *et al.* 2018) is observed. However, this collinearity does not verify the HGs. To clarify them, we also analysed the collinearity with the recent release octoploid genome (Edger *et al.*, 2019). Remarkedly, a total of 15 LG inversions, such as LG1D, LG2A, LG2B, LG2D, LG3A, LG3C, LG4A, LG4B, LG4D, LG5B, LG6A, LG6B, LG6D, LG7B and LG7D, were observed in comparison between 'FC50xFD54' genetic map and *F. x ananassa* genome, in agreement with chromosome scale collinearity (Hardigan *et al.*, 2020). Moreover, the small inversions at the beginning of LG1C, middle of LG6C and end of LG6D are also in agreement with Hardigan *et al.* (2020). Additionally, the LG6D inversion was also observed in the *F. x ananassa* genome (Edger *et al.*, 2019).

Focusing on the gaps in LG1C, LG4A and LG7C, they are observed in all three studied resources. However, a small inversion in LG2D was noticed comparing both genomes, but it was not in the *F. x ananassa* consensus map (Fig. 1-7). Since genetic maps are based on the studied cultivars and genomes are from other cultivars such as *F. vesca* cv. 'Hawaii-4' and *F. x ananassa* cv. 'Camarosa', some genetic differences might be found between them.

Nowadays, there are four different SNP arrays for strawberry being two of them, IStraw35k and 50k SNP array, derivatives of IStraw90k and 850k SNP array, respectively. Despite we have not hybridised with the 50k SNP array, we could predict its results by selecting SNPs from 850k which

are also in the 50k SNP array. Thus, the predicted 'FC50xFD54' map using 50k SNP array would increase two to three percent in codominant, 'FC50' segregation and 'FD54' segregation categories. The first genetic maps, using 50k SNP array, did not increased the genetic map density (Hardigan *et al.*, 2020), because these populations were not derivate from two *F. x ananassa* parental lines. Furthermore, we are considering to increase the resolution in our map enlarging the progenies and genotyping them with 50k SNP array.

Despite much more polymorphisms have been found in cultivated strawberry, IStraw35k array has enough markers spread across the whole genome and higher density of markers are mapped in the *F. vesca*-like subgenome, which is the dominant subgenome (Edger *et al.*, 2019 and Hardigan *et al.*, 2020). The 850k array will be highly suitable to find a really linked or causative SNP involved in a desired trait. Moreover, markers from 850k and 50k SNP arrays have been designed to be subgenome specific (Hardigan *et al.*, 2020) and knowing their physical positions will help to identify the candidate genes.

The 'FC50xFD54' genetic map is a necessary tool for accurate QTL analyses of several segregating traits that this population present, such as aroma and other fruit quality traits. With these analyses, it will be possible to select some SNPs linked to the desired trait and use them in marker assisted selection (MAS).

Annex chapter 1:

‘21AF’ genetic map reconstruction

Introduction

Previously to the 'FC50xFD54' genetic map, our group generated the '21AF' population. This second population is an F2 from a cross between two elite cultivars 'Camarosa' and 'Dover'. A total of 93 siblings were genotyped by 192 loci, principally SSRs, building a genetic map and discovering a firmness QTL in LG1B (Molina-Hidalgo *et al.*, 2013).

Moreover, 117 siblings were also genotyped using the IStraw90k array (Bassil *et al.*, 2015). The first analysis showed that 14295 SNPs were segregating within the mapping population. From them, 9356 were classified as PolyHighResolution and 4939 as NoMinorHomozygous (Hidalgo *et al.*, 2014). Polyphenol content from 83 selected progenies were analysed over two years by LC-MS, mapping 24 QTLs (Hidalgo, *et al.* unpublished).

In order to validate some QTLs found in the 'FC50xFD54' population and discover new QTLs, we constructed an improved '21AF' genetic map and compared with the 'FC50xFD54' genetic map and the *F. x ananassa* consensus map.

Materials and Methods

Previously to this thesis, DNA from 117 progenies F2, 'Camarosa' and 'Dover' parental lines and the hybrid ('H-21') were hybridised with the IStraw90k array and genotype analysis was performed using Genotyping console™ and SNPpolisher © (Affymetrix, CA, US). Genotype data from external 'Dover' line ('Dover_E') was also used.

To reconstruct the F2 '21AF' genetic map, we used genotypes from different versions of this genetic map and introduced to JoinMap®5 software (van Ooijen, 2018) following the same criteria used for the 'FC50xFD54' genetic map.

We visualised '21AF' genetic map using Genetic-Mapper software (Bekaert, 2016). Synteny analysis between '21AF', 'FC50xFD54' genetic map and *F. x ananassa* consensus map, and '21AF' dendrogram were done following the same pipeline and R packages used for the 'FC50xFD54' population analysis (Chapter 1, Materials and methods). A total of 971 SNPs was used for the '21AF' dendrogram.

Results

'21AF' genetic map

From the total of 13270 segregating markers in '21AF' population, only 6730 markers were used for building a genetic map spanning 3287cM in 33LGs. This map was oversized and some LGs were divided.

We collected genotypes from previous versions of this genetic map selecting 7977 markers to reconstruct this '21AF' genetic map. Our new version of '21AF' genetic map was distributed in 28 LGs and spanned 2056cM (Fig. A1-1 and Supplementary table A1-1). Map size was reduced more than one third compared to the original genetic map. Six LGs showed a gap bigger than 25cM, such as LG1A, LG3A, LG5D, LG6A, LG6B, LG6C and LG7A. The total size of LG1B, LG2C and LG7C was especially small (<35cM), whereas LG3A and the homeologous LG6A, LG6B, LG6C and LG6D were bigger than 100cM in size.

Although '21AF' genetic map had a correct genetic distance, the reduced '21AF' genetic map decreased the total size in 248cM showing a reduction greater than 10cM in nine LGs, such as LG3B, LG3C, LG4A, LG5A, LG5B, LG5C, LG5D and LG6D. All big gaps in the '21AF' genetic map were still present in this genetic map, with the exception of LG6C which was split in two pieces (Table A1-1).

Table A1-1. Summary of '21AF' genetic maps.

	Original '21AF' genetic map			'21AF' genetic map			Reduced '21AF' genetic map	
LG	Markers	Size (cM)	Loci	Markers	Size (cM)	Loci	Markers	Size (cM)
1A	246	80.41	98	332	76.03	64	53	72.659
1B	25	16.71	14	26	9.682	10	8	8.823
1C	192	100.97	103	229	45.732	64	40	39.755
1D	155	126.26	101	183	55.599	59	43	54.985
2A	231	129.44	92	282	89.207	67	55	79.892
2B	189	133.52	105	219	71.925	63	49	68.033
2C	-	-	-	78	32.908	25	20	33.615
2D	200	106.28	110	232	74.810	66	52	70.005
3A	117 | 105	35.18 | 64.04	41 | 38	266	109.79	53	41	101.502
3B	133 | 162	65.44 | 76.26	61 | 74	347	92.984	79	58	74.040
3C	228	122.85	118	268	53.989	71	44	43.819
3D	421	204.98	221	482	85.007	121	93	79.290
4A	221	128.56	96	257	59.978	70	40	48.460
4B	267	114.8	113	313	70.787	78	63	65.734
4C	207	95.44	98	237	55.386	69	51	50.254
4D	132	98.25	75	156	57.861	41	21	51.610
5A	450	150.31	159	539	77.887	100	80	66.679
5B	325	139.41	161	389	67.508	94	74	57.292
5C	293	166.06	150	342	95.695	106	72	76.429

	Original ‘21AF’ genetic map			‘21AF’ genetic map			Reduced ‘21AF’ genetic map	
LG	Markers	Size (cM)	Loci	Markers	Size (cM)	Loci	Markers	Size (cM)
5D	136 | 69	89.36 | 22.20	67 | 31	220	95.384	56	31	83.991
6A	359 | 81	121.35 | 23.54	123 | 32	548	137.552	114	97	131.438
6B	107 | 88	80.37 | 42.24	51 | 46	220	117.868	59	44	116.498
6C	95 | 227	47.54 | 124.88	51 | 129	362	104.090	101	14 | 41	8.689 | 36.147
6D	378	199.24	181	439	111.402	117	87	100.482
7A	305	105.57	93	385	69.782	71	54	64.638
7B	260	133.79	123	260	63.037	75	52	55.091
7C	105	44.42	51	120	22.584	33	27	19.122
7D	221	97.97	102	246	51.578	64	53	49.081
All	6730	3287.64	3108	7977	2056.042	1990	1457	1808.053

Vertical bars (|): separation between the two fragments of the divided LG.

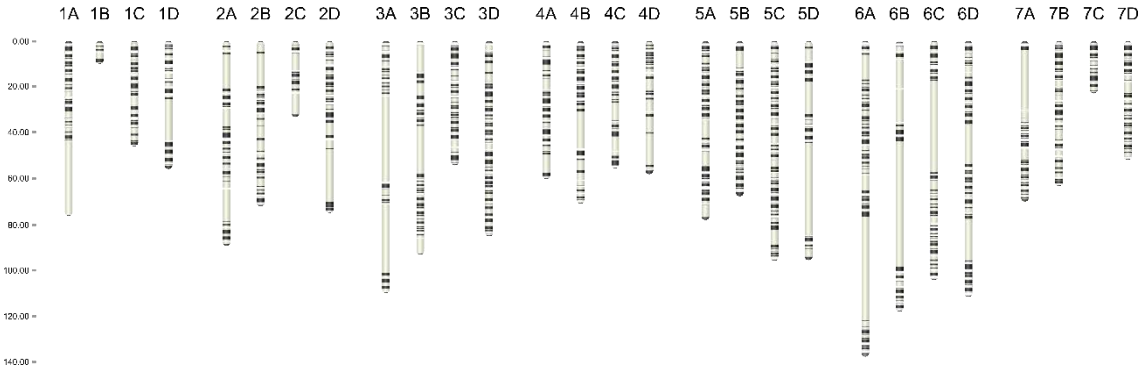


Figure A1-1. Representation of ‘21AF’ genetic map distributed in seven LG (1-7) and four subgenome groups each (A-D).

Synergy between ‘21AF’, ‘FC50xFD54’ genetic maps and consensus map

High synergy could be observed among all the three maps. The beginning or the end of some LGs from ‘21AF’ genetic map did not have segregating markers such, as LG1B, LG2C, LG3C and LG7C. Big gaps in LG1A, LG1D, LG3A, LG5D, LG6A, LG6B and LG6C corresponded to real genetic distances (Fig. A1-2). Moreover, markers selected for the reduced ‘21AF’ genetic map were spread all over the ‘21AF’ genetic map.

Genetic distances between ‘21AF’ progenies

Dendrograms provide genetic relations between siblings and the parental lines. This analysis showed that two pairs of progenies were close to each other (058 - 059 and 175 - 176) (Fig. A1-3). Moreover, the parental lines, ‘Camarosa’ and ‘Dover’, clustered together and far from progenies (Cluster A). However, genotype from the external ‘Dover_E’ sample and the hybrid (‘H-21’) were found in the middle of the progeny cluster (cluster B).

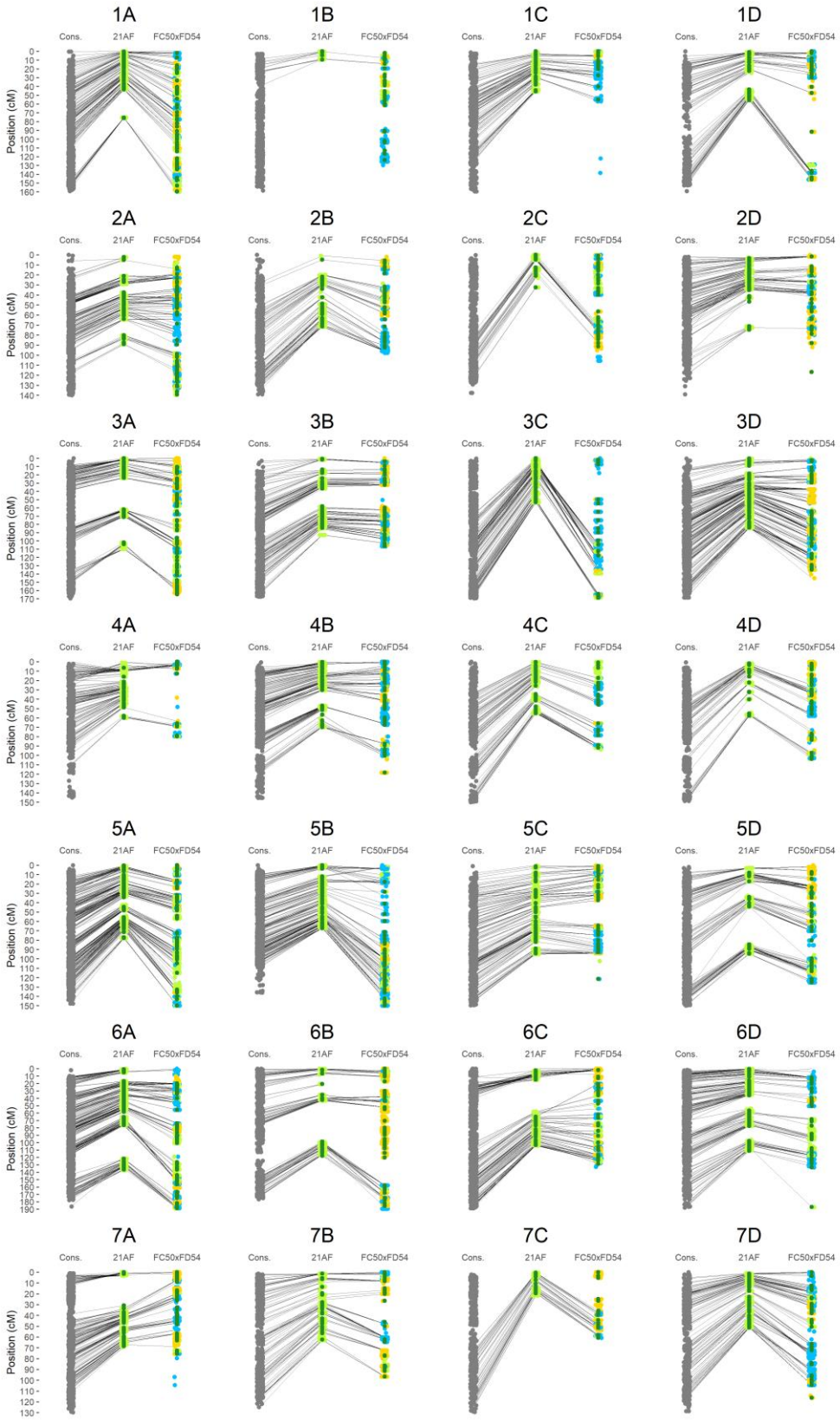


Figure A1-2. Synergy between '21AF', 'FC50xFD54' genetics maps and consensus map. Colour dots indicate the segregation type: codominant (light green), 'FC50' segregation (blue), 'FD54' segregation (yellow), consensus map (grey). Little dark green dots indicate markers selected for reduced genetic maps in '21AF' and 'FC50xFD54' populations. Grey lines indicate synergy between common markers.

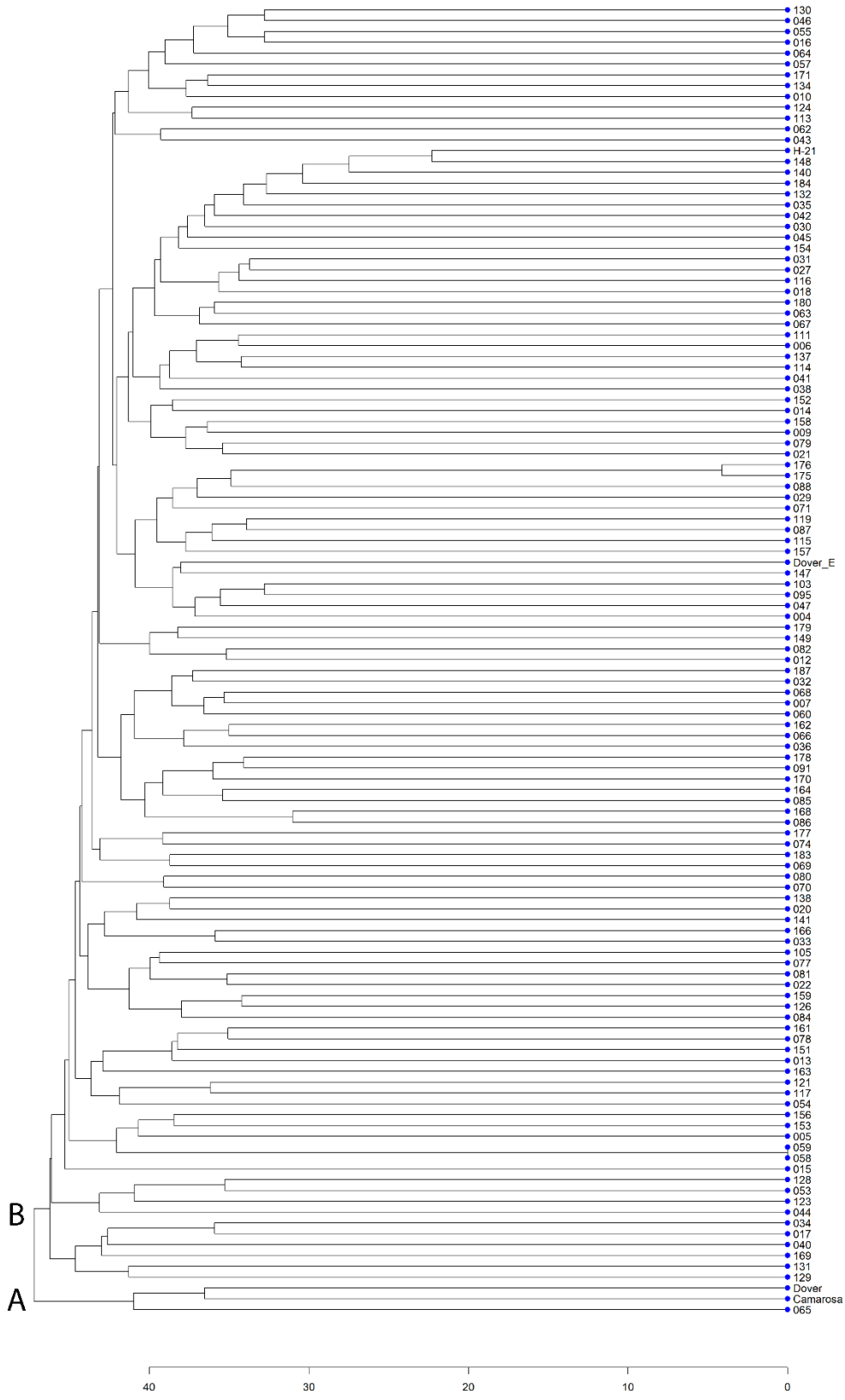


Figure A1-3. ‘21AF’ dendrogram. Parental lines ‘Camarosa’, ‘Dover’, ‘H-21’ (hybrid) and the external ‘Dover’ (‘Dover_E’) were clustered with ‘21AF’ progenies. Distance matrix were calculated with Euclidean method and hierarchical clustering with UPGMA algorithm. Letters represent the two major cluster groups.

Discussion

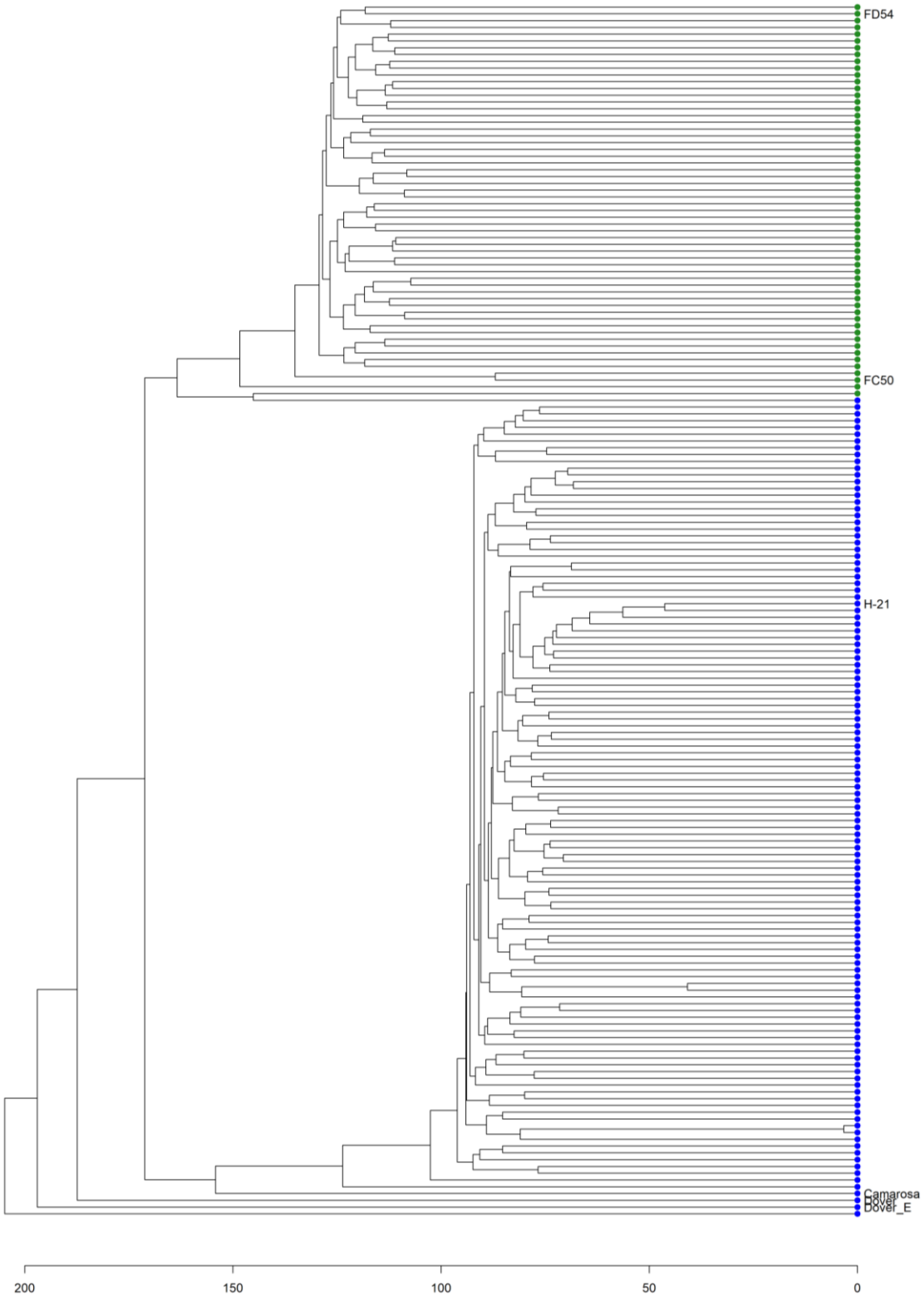
We improved considerably the ‘21AF’ genetic map by reducing one third the total size and grouped in the expected 28 LGs. Both the ‘21AF’ and the reduced ‘21AF’ genetic maps have a length size appropriated for QTL analysis. Moreover, the length reduction in the reduced ‘21AF’ genetic map was smaller than the observed in the reduced ‘FC50xFD54’ genetic map (Fig. 1-4).

‘Camarosa’ and ‘Dover’ clustered near each other (Fig. A1-3), indicating that the majority of markers were identical and some of them were homozygous. In this case, fewer markers would have segregated in the progenies. On the other hand, ‘Dover_E’ clustered in the middle of the progenies. Therefore, we suggested that the parental line called ‘Dover’ might not be the real ‘21AF’ parental, while ‘Dover_E’ might be a closer one. Moreover, two pairs of progenies were unexpectedly close and numerically contiguous, indicating that both DNA may be collected from the same individual.

The ‘21AF’ population show lower genetic differences between progenies than the ‘FC50xFD54’ population (SF. A1-1). This is explained because an F1 cross between two heterozygous lines diverges more than a F2 population from a heterozygous hybrid.

This improved ‘21AF’ genetic map will be a useful tool to validate QTLs detected in the ‘FC50xFD54’ analysis and found new ones.

Supplementary figure



Suppl. figure A1-1. Dendrogram of ‘FC50xFD54’ (green dots) and ‘21AF’ (blue dots) populations showing the progenitors’ names. 9711 markers segregating in both populations were used to calculate distance matrix with Euclidean method and hierarchical clustering with UPGMA algorithm.

Chapter 2:

Genetic analysis of volatile organic compounds in cultivated strawberry fruits

Introduction

Strawberries are one of the most worldwide consumed berries. During the last decades, strawberry breeding had been focused principally on fruit quality and plant resistance. However, consumers are demanding better tasting fruits (Folta and Klee, 2016). During the last years, quality traits such as volatiles and polyphenolics content gained increasing importance in breeding programs. It is well known that volatiles and sugars increase during fruit ripening. Volatile compounds are perceived as a healthy nutrients label (Goff and Klee, 2006). Therefore, animals, including humans, will prefer eating it and, consequently, disperse their seeds. The large variability in strawberry volatile compounds can be explained by genetics, maturity stage and postharvest factors (Schwieterman *et al.*, 2014).

Strawberries produce more than 350 volatile organic compounds (VOCs), but only few of them contribute to the strawberry aroma (Schieberle and Hofmann, 1997, Ulrich *et al.*, 1997, Nuzzi *et al.*, 2008 and Urrutia *et al.*, 2017). In particular, wild strawberry has higher intense flavour than cultivated strawberry.

As explained in the General introduction, fatty acids are the major precursor of aromatic compounds. Fatty acids are precursors of straight-chain acids, alcohols, aldehydes, esters, ketones and lactones. Esters, the most abundant chemical family in strawberry fruits, are produced by the esterification of alcohols, dehydrogenated by alcohol dehydrogenase (ADH), and acyl-CoA catalysed by the alcohol acyl transferases (AAT). In this last step, FaSAAT show a wide range of substrate affinity (Aharoni *et al.*, 2000 and Beekwilder *et al.*, 2004). Aldehydes and alcohols are decreasing during ripening at the same time that transcription level of AAT is increased in apricot (Gonzalez-Aguero *et al.*, 2009) and during postharvest in peach (Zhang *et al.*, 2010). Acyl-coA can be synthesised by β-oxidation or the lipoxygenases (LOX) pathway being aldehydes and alcohols intermediaries. Additionally, 3-Z/2-E-hexenal isomerases are responsible to convert 2-E-hexenal to 3-Z-hexenal in cucumber (Spyropoulou *et al.*, 2017). Different studies reported that an omega-6 fatty acid desaturase (FaFAD1) is necessary for the lactone γ-decalactone biosynthesis (Chambers *et al.*, 2014 and Sanchez-Sevilla *et al.*, 2014).

Branched and aromatic ester compounds are synthesised through amino acid metabolism. In strawberry, two eugenol synthases (*FaEGS1* and *FaEGS2*) are involved in the biosynthesis of eugenol and both are controlled by a R2R3 MYB transcription factor (*FaEOBII*) (Aragüez *et al.*, 2013 and Medina-Puche *et al.*, 2015). An anthranilate acid methyl transferase (FaAAMT) controls the last step of methyl anthranilate synthesis (Pillet *et al.*, 2017). However, this

compound is only detected in few cultivars such as 'Charlotte' and 'Mara des Bois', and in wild strawberry (Ulrich *et al.*, 1997 and Schwieterman *et al.*, 2014).

Terpenes are derived from isopentenyl pyrophosphate (IPP) and dimethylallyl pyrophosphate (DMAPP) which produce monoterpenes and sesquiterpenes, respectively. In strawberry, nerolidol synthase (FaNES1) can synthesise the monoterpene linalool and the sesquiterpene nerolidol (Aharoni *et al.*, 2004). This gene is present only in cultivated but not in wild strawberries (Aharoni *et al.*, 2004). Moreover, this gene has been widely screened in cultivated, wild octoploid and diploid accessions, reporting that only octoploid strawberries have the functional allele (Chambers *et al.*, 2012). On the other hand, some compounds that are important for the wild strawberry aroma such as myrtenol are not present in cultivated strawberry (Aharoni *et al.*, 2004 and Ulrich *et al.*, 2007).

Finally, furanones can also produce volatile compounds such as mesifurane and furaneol. A quinone oxidoreductase (FaQR) is required for furaneol biosynthesis (Raab *et al.*, 2006) and an O-methyltransferases (FaOMT) methylate furaneol to produce mesifurane (Wein *et al.*, 2002).

Nineteen volatile compounds had been described as the most relevant ones in fruits and are considered as key volatile compounds (KVCs). Its importance lies on overcoming the threshold of odour perception in human, and the combination of these 19 compounds produces the characteristic strawberry aroma (Schieberle and Hofmann, 1997).

These KVCs are categorised as aldehydes, esters, furans, lactones and terpenes. Ester volatile compounds, such as butyl acetate, hexyl acetate, methyl butanoate, ethyl butanoate, butyl butanoate, methyl hexanoate or ethyl hexanoate, deliver fruity and sweet notes. Other esters are present only in few cultivars such as methyl cinnamate which provides spice notes, myrtenyl acetate conferring herbaceous notes and methyl anthranilate that provides strawberry-like flavour. The γ-decalactone is a lactone which contribute with a peach-like aroma. Aldehydes, such as E-2-hexenal and Z-3-hexenal and their ester derivative, E-2-hexenyl acetate and Z-3-hexenyl acetate, deliver an herbaceous note. Terpene compounds, such as linalool and nerolidol, provide a floral aroma. Finally, furans, mainly furaneol and mesifurane, contribute with a caramel notes (Schieberle and Hofmann, 1997, Ulrich *et al.*, 1997, Jetti *et al.*, 2007, Ulrich *et al.*, 2007, Nuzzi *et al.*, 2008, Olbricht *et al.*, 2008 and Schwieterman *et al.*, 2014).

Strawberry has an allo-octoploid genome composed by four subgenomes, one related to *Fragaria vesca*, another to *Fragaria innumae* and the last two subgenomes are under discussion. The recent release of the *F. x ananassa* genome (Edger *et al.*, 2019) will increase the genomic knowledge.

As explained in Chapter 1, several saturated genetic maps were constructed using the IStraw90k (Bassil *et al.*, 2015) and IStraw35k SNP arrays (Verma *et al.*, 2017a). Recently, new 850k and 50k SNP arrays have been developed (Hardigan *et al.*, 2020).

Different mapping populations, such as F1 and NIL collection, were used to detect volatile QTLs. Zorrilla-Fontanesi *et al.* (2011) screened an F1 population by GC-MS and mapped 70 QTLs for 48 different volatile compounds, but only 35 of them were found to be stable over time. The NIL collection using *F. vesca*, as the recurrent parent, and *F. bucharica*, as the donor parent, revealed 14 QTLs for the different key volatile compounds (Urrutia *et al.*, 2017). Moreover, two genes involved in the accumulation of two compound were detected using RNAseq (Chambers *et al.*, 2014, Sanchez-Sevilla *et al.*, 2014 and Pillet *et al.*, 2017).

This chapter will increase the knowledge on genetic regions responsible for the complex aroma of cultivated strawberry fruits by analysing the VOCs diversity in an F1 cultivated strawberry population. This study was performed over six harvests in three consecutive years by GC-MS analysis. From this screening, stable and major QTLs have been mapped for esters and terpenes and some of them have been validated in another F2 population.

Materials and Methods

Plant material

An F1 cross between two breeding lines, 'FC50' and 'FD54', was made to produce 70 progeny lines (hereafter, 'FC50xFD54' population). 'FC50' has a wild strawberry aroma, whereas 'FD54' has a characteristic fruity aroma. At least six clonal runner plants per progeny or parental lines were grown using standard cultivation practices in the south-west of France (Le Barp, latitude: 44°67'N, longitude 0°73'W) during three successive years (2016-18). Grandparental lines ('FD019', 'FD016', 'FC030' and FD197') were also analysed. Ripe fruits were collected in six harvests during three years. Harvests from the same year were collected in one to two weeks interval in May and June. Each year, the plants have been replicated clonally by runner.

An F2 cross between two commercial cultivars, 'Dover' and 'Camarosa' produced a hybrid 'H-21' which was self-pollinated, obtaining 117 progeny lines (hereafter, '21AF' population). 'Dover' was selected for its plant agronomic quality and resistance to diseases and 'Camarosa' for its fruit quality such as high polyphenol content. At least two clones per progeny or parental lines were grown using standard cultivation practices in the north-east of Spain (Caldes de Montbui, latitude: 41°36'N, longitude 2°10'E) during three years (2014, 2015 and 2018). Ripe fruits were picked up at different times over three years, between May and July. Each year, clonal runner plants were replicated. All the harvested fruits from the same year were considered as a unique harvest.

Volatile compound analysis by GC-MS

A pool of three to five strawberry fruits from each genotype and harvest were frozen in $N_{2\ (liq.)}$, powdered with a coffee grinder and stored immediately at - 80°C. Each harvest sample was analysed as an independent sample. To perform the GC-MS assay, we weighted 1 g of frozen powdered strawberry fruits in a 10 ml screw cap headspace vial and added 1 ml of saturated NaCl solution with 10 ppm of internal standard (3-hexanone; Sigma-Aldrich, MO, USA). A GC Sampler 80 (Agilent Technology, CA, USA) was used for incubation, extraction and desorption. Vials were first heated at 50°C during 10 min with agitation at 500 rpm, then SPME fibre (50/30µm DVB/CAR/PDMS; Supelco, PA, USA) was exposed into the headspace vial for 30 min in the same conditions to extract the volatile compounds. The extracted volatiles were desorbed in the GC injection port at 250°C for 5 min in splitless mode. Volatile compounds were analysed on a 7890A gas chromatograph coupled with a 5975C mass spectrometer (Agilent Technology, CA, USA). An HP-5MS UI GC column (30 m, 0.250 mm, 0.25 µm) (J&W, CA, USA) and 1.2mL/min

constant helium flow was used for chromatographic separation. Oven conditions started at 40°C for 2 mins, then increasing by 5°C/min ramp until reaching the temperature of 250°C and ending at that temperature for 5 min.

To identify different volatile compounds, we used two methods. One of them was the identification of the 19 key volatile compounds (KVCs) by comparing the retention time and ion fragmentation with each respective commercial standard (Sigma-Aldrich, MO, USA). The other method was used for the rest of the compounds by comparing Kovats retention index (KI). The KI (Kovats, 1958) is calculated using C7-C30 saturated standards (Sigma-Aldrich, MO, USA). Our results were compared with KI from NIST database (https://webbook.nist.gov/chemistry/gc-ri/). Only peaks that could be clearly identified were used for further analysis. Peak areas were normalised to the internal standard peak. All profiles were analysed using Enhanced ChemStation software (Agilent Technologies, CA, USA) and compared with mass spectra libraries, NIST08 and NIST11.

Data analysis

The statistical analysis of GC-MS data was performed using R v3.6.2 (RCoreTeam, 2019) with the Rstudio v1.2.5033 interface (RStudio, 2019). A violin plot obtained by 'ggplot2' (version 3.3.0) visualised relative VOCs content distribution. Pearson correlation between harvests and VOCs were calculated using 'cor' function and visualised by 'corrplot' R package (version 0.84). Pearson correlation between VOCs were visualised as cluster network analyses (CNA) using 'qgraph' function from 'qgraph' R package (version 1.6.5). Heatmap representation of hierarchical cluster analysis was performed using 'heatmap.2' function from 'gplots' R package (version 3.0.3). Principal component analysis (PCA) was done with 'fviz_pca_biplot' function from 'factorextra' R package (version 1.0.7). Omega squared values (ω^2) were derived from ANOVA residuals calculated with 'aov' function and using the formula: $(SSi - df_i * MS_{err}) * (MS_t + MS_{err})^{-1}$.

Volatile QTL analysis

To analyse the genetic regions responsible of VOCs variance, we constructed a reduced 'FC50xFD54' genetic map, as explained in the Chapter 1, consisted of 1464 markers grouped in 28 LGs and extending a total size of 2273cM (Fig. 1-4). Likewise, we constructed a reduced genetic map from the F2 'Camarosa' x 'Dover' population ('21AF') with 1457 markers grouped in 29 LGs and spanning 1808.05cM (Table A1-1 and Fig. A1-1).

VOC data was transformed with log_2 and analysed separately for each harvest in MapQTL®6 software (van Ooijen, 2009) using Interval Mapping method (IM) and non-parametric Kruskal-Wallis test (KW). A QTL with a minimum LOD score of 2.5 in at least two harvests were considered significant and stable, when analysed by IM. A QTL was classified as major QTL if its LOD score was higher than four. QTL stability was considered high if it was detected in three or four harvests and totally stable if it was identified in five or six harvests. Since '21AF' QTL analysis was performed with three different harvests (one harvest each year), a QTL was considered as highly stable QTL when it was identified in two harvests and totally stable when it was detected in all three harvests. When analysing a QTL with KW test, markers with the highest values were the most linked to each trait. The MapChart2.2 software (Voorrips, 2002) was used to visualize each QTL in our genetic map with fill and colour parameters that distinguish the stability and chemical family of each VOC, respectively. Box size was related to 1-LOD confidence interval.

High Resolution Melting (HRM) marker development

To design HRM markers, we selected subgenomic specific markers from 50k SNP array (Hardigan *et al.*, 2020) which fall within the 1-LOD confidence interval in *Terpene_3B* and *Hexanoate_4B* QTLs. We examined each flanking region in the *F. x ananassa* genome (Edger *et al.*, 2019) using CLC genomic workbench 5, in limited configuration, to extract a bigger flanking region. Primers using this increased flanking region were designed in the PrimerQuest software, www.idtdna.com/Primerquest/Home/Index, (Integrated DNA Technologies, Inc. IA, USA).

The parameters for primers design were customised by primer temperature ranging from 55-65°C, being the optimal at 62°C, and by amplicon size ranging from of 50 to 150 bp, being preferred the shortest one. Primer sequences were checked in CLC genomic workbench 5 to ensure that they are unique in the *F. x ananassa* cv. 'Camarosa' genome. As well, we added primers with more than one position if the undesired one was only found in the forward or reverse primers. We also check them in *F. x ananassa* cv.'Reikou' genome, which has haplotypes, to ensure primer success (Isobe *et al.*, 2020, http://strawberry-garden.kazusa.or.jp/). The selected primers are listed in Supplementary table 2-1.

The HRM assay was performed using 68 progenies and the two parental lines of 'FC50xFD54' population. A total volume of 5µl PCR-HRM reaction was reached by adding 2.5 µl 2x AccuStart II PCR ToughMix master mix (Quantabio, MA, USA), 1,5 µl at 5 µM primer mix (forward and reverse), 0.25 µl 10x LCGreen Plus Melting Dye (BioFire Defense, UT, USA), 0.5 µl DNA (50ng/µl) and 0.75 µl deionised H_2O. The reactions were carried out in a 384-well PCR plate in a LightCycler 480 II instrument (Roche, Switzerland) for both PCR and HRM. PCR conditions started with a

denaturation step at 95ºC for 30s, following by 55 cycles consisting of denaturation step at 95ºC for 15s, annealing at each primer Tm for 15s and extension step at 72ºC for 15s. After amplification, the PCR product was first denatured at 95ºC for 60s, then cooled to 40ºC for 60s and lastly, a ramp step was performed beginning at 55 ºC and increasing by 1ºC per second until reaching 95ºC. The melting curves were analysed and classified by the differences in their shapes and Tm values, allowing the detection of different genotypes.

Analysis of candidate genes

To identify the physical position of each marker, SNP probes were blasted to the *F. x ananassa* cv. 'Camarosa' genome (Edger *et al.*, 2019) as explained in Chapter 1. Once the genome distances of the QTLs had been delimited, genes inside each QTL were analysed by gene homology using NCBI, SwissProt, Tair10 and TrEMBL database obtained from https://www.rosaceae.org/species/fragaria_x_ananassa/genome_v1.0.a1 (Jung *et al.*, 2019).

Results

Volatile compounds composition

To identify the genomic regions which contribute to the strawberry aroma, we analysed the volatile organic compounds (VOCs) content by GC-MS of full ripe fruits from 68 progenies of the 'FC50xFD54' population collected in six different harvests during three consecutive years (2016-2018) (Supplementary table 2-2). From hundreds of peaks present in strawberry aroma profile, 58 VOCs could be clearly identified (Table 2-1) including 33 (57%) esters, nine (16%) terpenes, seven (12%) aldehydes, three (5%) lactones, two (3%) furans and four different compounds (7%) from other families. Nineteen compounds are considered as key volatile compounds (KVCs) of which we could identify sixteen. The three KVCs left could not be measured because myrtenyl acetate and methyl cinnamate peaks were not present and the 3-Z-hexenal peak was almost undetectable and overlapping with the major peak of hexanal.

Despite the differences among the VOC content in different harvests (Figure 2-1), the 'FD54' parental line displayed a significantly higher concentration of most of the VOCs compared to the 'FC50' parental line, such as some esters (methyl 3-methyl butanoate, ethyl octanoate, methyl decanoate and methyl benzeneacetate), aldehydes (hexanal, heptanal and Z-2-heptanal), lactones (γ-octalactone and γ-decalactone) and terpenes (β-pinene, D-limonene, α-terpineol and linalool), whereas methyl anthranilate and furaneol were significantly lower than the other parental line (Table 2-1).

The VOC variance of 'FC50xFD54' progeny confirms the segregating behaviour (Fig. 2-1). The most abundant compounds were methyl butanoate, ethyl butanoate, butyl butanoate, methyl hexanoate, ethyl hexanoate, E-2-hexanal, mesifurane, γ-decalactone and linalool which are KVCs (Table 2-1 and Fig. 2-1).

In order to check if VOCs content showed a normal distribution, we calculated the Shapiro-Wilk test for each VOCs and harvest. Few VOCs (methyl butanoate, methyl hexanoate, hexanal, E-2-hexenal, Z-2-heptanal and γ-decalactone) revealed a normal distribution according to this test. When VOCs content was transformed to $\log_2$, the majority of them (42) presented a normal distribution in at least three harvests (Supplementary table 2-3).

Table 2-1. Summary of VOCs in 'FC50xFD54' population. Volatile compound marking the most relevant ones (KVCs), their abbreviation and chemical family are presented. Means and standard deviations (SD) of six harvests of each compound in the 'FC50' and 'FD54'. For each VOCs, mean, standard deviation and range for F1 population and harvest correlation are annotated.

VOC	KVC	Abb.	Type	FC50	FD54	FC50xFD54		Correlation	
				Mean ± SD	Mean ± SD	Mean ± SD	Range	Mean ± SD	Range
Butyl acetate	**X**	BA	Ester	0.092 ± 0.043	0.118 ± 0.058	0.104 ± 0.075	0.007 - 0.528	0.364 ± 0.106	0.198 - 0.586
Isopentyl acetate		IPA	Ester	0.042 ± 0.011	0.074 ± 0.061	0.065 ± 0.060	0.005 - 0.537	0.461 ± 0.097	0.238 - 0.666
Hexyl acetate	**X**	HA	Ester	0.152 ± 0.062	0.461 ± 0.385	0.332 ± 0.534	0.013 - 6.350	0.167 ± 0.155	-0.041 - 0.528
Octyl acetate		OA	Ester	0.026 ± 0.042	0.129 ± 0.103	0.045 ± 0.099	0.002 - 0.953	0.663 ± 0.108	0.415 - 0.807
Methyl butanoate	**X**	MB	Ester	2.938 ± 1.132	2.104 ± 0.605	3.507 ± 1.580	0.140 - 9.569	0.403 ± 0.117	0.237 - 0.549
Ethyl butanoate	**X**	EB	Ester	2.724 ± 1.837	1.630 ± 0.886	3.598 ± 3.182	0.064 - 21.773	0.234 ± 0.151	-0.049 - 0.447
Isopropyl butanoate		IPB	Ester	0.411 ± 0.122	0.417 ± 0.253	0.443 ± 0.292	0.012 - 1.445	0.376 ± 0.153	0.086 - 0.641
Methyl 3-methyl butanoate		M3MB	Ester	0.041 ± 0.013	**0.126 ± 0.064**	0.083 ± 0.060	0.011 - 0.438	0.555 ± 0.099	0.376 - 0.689
Propyl butanoate		PB	Ester	0.056 ± 0.029	0.056 ± 0.033	0.065 ± 0.048	0.003 - 0.262	0.358 ± 0.120	0.122 - 0.595
Butyl butanoate	**X**	BB	Ester	0.419 ± 0.205	0.511 ± 0.478	0.630 ± 0.705	0.002 - 4.076	0.570 ± 0.124	0.323 - 0.736
Hexyl butanoate		HB	Ester	0.073 ± 0.040	0.418 ± 0.509	0.239 ± 0.399	0.003 - 4.972	0.370 ± 0.201	0.032 - 0.726
Octyl butanoate		OB	Ester	0.031 ± 0.042	0.680 ± 0.718	0.380 ± 0.843	0.002 - 7.411	0.336 ± 0.291	-0.011 - 0.849
1-Methyloctyl butanoate		MOB	Ester	0.350 ± 0.690	0.040 ± 0.046	0.060 ± 0.226	0.001 - 3.183	0.132 ± 0.163	-0.091 - 0.405
Octyl 3-methylbutanoate		O3MB	Ester	0.004 ± 0.002	0.274 ± 0.275	0.046 ± 0.100	0.002 - 1.113	0.733 ± 0.106	0.528 - 0.937
Methyl 4-methylpentanoate		M4MP	Ester	0.017 ± 0.012	0.017 ± 0.017	0.024 ± 0.024	0.002 - 0.167	0.290 ± 0.131	0.060 - 0.532
Methyl hexanoate	**X**	MH	Ester	2.876 ± 1.078	2.961 ± 1.802	3.840 ± 2.258	0.028 - 12.808	0.503 ± 0.100	0.339 - 0.643
Ethyl hexanoate	**X**	EH	Ester	1.592 ± 1.552	1.777 ± 1.478	3.231 ± 3.626	0.009 - 20.876	0.326 ± 0.215	0.004 - 0.619
Isopropyl hexanoate		IPH	Ester	0.084 ± 0.030	0.116 ± 0.075	0.143 ± 0.145	0.002 - 1.186	0.627 ± 0.080	0.516 - 0.799
Isoamyl hexanoate		IAH	Ester	0.014 ± 0.007	0.021 ± 0.016	0.045 ± 0.063	0.001 - 0.419	0.632 ± 0.099	0.465 - 0.772
Hexyl hexanoate		HH	Ester	0.007 ± 0.003	0.101 ± 0.140	0.032 ± 0.048	0.003 - 0.372	0.368 ± 0.223	-0.038 - 0.754
Octyl hexanoate		OH	Ester	0.005 ± 0.002	0.202 ± 0.228	0.091 ± 0.248	0.002 - 2.833	0.589 ± 0.191	0.217 - 0.897
Methyl E-2-hexenoate		M2H	Ester	0.100 ± 0.039	0.054 ± 0.035	0.069 ± 0.063	0.003 - 0.388	0.539 ± 0.110	0.359 - 0.707
Ethyl 2-hexenoate		E2H	Ester	0.006 ± 0.004	0.010 ± 0.008	0.020 ± 0.031	0.002 - 0.381	0.174 ± 0.150	-0.072 - 0.467
Methyl octanoate		MO	Ester	0.062 ± 0.026	0.363 ± 0.310	0.206 ± 0.234	0.001 - 1.768	0.571 ± 0.091	0.390 - 0.754
Ethyl octanoate		EO	Ester	0.007 ± 0.005	**0.036 ± 0.024**	0.067 ± 0.142	0.001 - 1.273	0.296 ± 0.190	0.070 - 0.638
Methyl decanoate		MD	Ester	0.001 ± 0.000	**0.023 ± 0.011**	0.011 ± 0.013	0.001 - 0.097	0.479 ± 0.118	0.304 - 0.662
Methyl benzene acetate		MBA	Ester	0.037 ± 0.011	**0.081 ± 0.033**	0.054 ± 0.031	0.007 - 0.266	0.449 ± 0.155	0.141 - 0.745
Methyl nicotinate		MN	Ester	0.038 ± 0.027	0.002 ± 0.002	0.016 ± 0.028	0.001 - 0.236	0.438 ± 0.130	0.161 - 0.649

VOC	KVC	Abb.	Type	FC50	FD54	FC50xFD54		Correlation	
				Mean ± SD	Mean ± SD	Mean ± SD	Range	Mean ± SD	Range
Methyl salicylate		MS	Ester	0.090 ± 0.054	0.040 ± 0.032	0.090 ± 0.085	0.002 - 0.506	0.570 ± 0.091	0.376 - 0.706
Methyl anthranilate	**X**	MA	Ester	**0.252 ± 0.252**	0.009 ± 0.008	0.065 ± 0.180	0.001 - 1.752	0.423 ± 0.219	0.018 - 0.695
Z-3-hexenyl acetate	**X**	Z3HA	Ester	0.028 ± 0.024	0.039 ± 0.036	0.030 ± 0.019	0.004 - 0.134	0.156 ± 0.169	-0.186 - 0.349
E-2-hexenyl acetate	**X**	E2HA	Ester	0.069 ± 0.065	0.156 ± 0.192	0.166 ± 0.199	0.009 - 2.509	0.162 ± 0.161	-0.048 - 0.507
E-2-hexenyl butanoate		E2HB	Ester	0.032 ± 0.021	0.086 ± 0.159	0.105 ± 0.182	0.002 - 2.070	0.132 ± 0.159	-0.056 - 0.442
Hexanal		Hex	Aldehyde	0.144 ± 0.050	**0.526 ± 0.266**	0.353 ± 0.220	0.064 - 1.272	0.227 ± 0.179	-0.103 - 0.494
Heptanal		Hep	Aldehyde	0.026 ± 0.001	**0.042 ± 0.008**	0.043 ± 0.018	0.017 - 0.139	0.233 ± 0.119	0.061 - 0.408
Nonanal		Non	Aldehyde	0.230 ± 0.051	0.166 ± 0.103	0.209 ± 0.153	0.007 - 1.070	0.283 ± 0.145	0.075 - 0.498
Decanal		Dec	Aldehyde	0.011 ± 0.007	0.043 ± 0.062	0.021 ± 0.023	0.003 - 0.315	0.065 ± 0.158	-0.111 - 0.491
E-2-Hexenal	**X**	E2Hal	Aldehyde	1.990 ± 0.521	2.290 ± 1.430	1.865 ± 1.434	0.064 - 9.109	0.208 ± 0.128	-0.051 - 0.430
Z-2-Heptenal		Z2H	Aldehyde	0.024 ± 0.002	**0.048 ± 0.019**	0.048 ± 0.018	0.014 - 0.121	0.206 ± 0.133	-0.088 - 0.403
Benzaldehyde		B	Aldehyde	0.154 ± 0.122	0.187 ± 0.150	0.200 ± 0.160	0.007 - 1.260	0.288 ± 0.141	0.052 - 0.511
Dodecane		Dode	Alkane	0.039 ± 0.042	0.011 ± 0.010	0.029 ± 0.056	0.001 - 1.017	0.190 ± 0.239	-0.054 - 0.538
Octanoic acid		Oac	Acid	0.074 ± 0.072	0.203 ± 0.209	0.176 ± 0.312	0.003 - 2.936	0.600 ± 0.151	0.297 - 0.828
Octanol		Ool	Alcohol	0.082 ± 0.063	0.071 ± 0.049	0.068 ± 0.053	0.009 - 0.386	0.314 ± 0.184	0.047 - 0.572
2-Heptanone		Hep2	Ketone	0.156 ± 0.083	0.291 ± 0.268	0.169 ± 0.146	0.010 - 1.293	0.560 ± 0.139	0.333 - 0.792
γ-octalactone		GO	Lactone	0.004 ± 0.004	**0.043 ± 0.030**	0.023 ± 0.021	0.001 - 0.180	0.553 ± 0.086	0.340 - 0.691
γ-decalactone	**X**	GD	Lactone	0.046 ± 0.063	**4.046 ± 2.068**	2.720 ± 1.901	0.001 - 14.162	0.409 ± 0.093	0.230 - 0.565
γ-dodecalactone		GDo	Lactone	0.124 ± 0.189	0.292 ± 0.398	0.195 ± 0.288	0.001 - 1.591	0.446 ± 0.117	0.249 - 0.628
Mesifurane	**X**	M	Furan	3.375 ± 1.271	1.828 ± 1.410	2.419 ± 2.242	0.023 - 13.937	0.361 ± 0.222	0.040 - 0.710
Furaneol	**X**	F	Furan	**0.280 ± 0.087**	0.111 ± 0.079	0.118 ± 0.144	0.006 - 1.018	0.369 ± 0.137	0.082 - 0.574
β-Pinene		BP	Terpene	0.096 ± 0.079	**0.278 ± 0.127**	0.199 ± 0.162	0.001 - 1.002	0.746 ± 0.078	0.579 - 0.867
D-Limonene		DL	Terpene	0.035 ± 0.030	**0.115 ± 0.054**	0.080 ± 0.063	0.002 - 0.415	0.653 ± 0.100	0.480 - 0.851
α-Terpineol		AT	Terpene	0.076 ± 0.057	**0.375 ± 0.186**	0.171 ± 0.162	0.002 - 1.242	0.559 ± 0.155	0.223 - 0.770
Linalool	**X**	L	Terpene	0.854 ± 0.395	**2.421 ± 0.691**	1.488 ± 0.919	0.010 - 5.554	0.732 ± 0.093	0.531 - 0.867
Geraniol		G	Terpene	0.002 ± 0.001	0.008 ± 0.006	0.004 ± 0.005	0.000 - 0.034	0.594 ± 0.178	0.207 - 0.837
β-Farnesene		BF	Terpene	0.161 ± 0.182	0.414 ± 0.338	0.296 ± 0.278	0.003 - 1.460	0.615 ± 0.093	0.420 - 0.750
α-Curcumene		AC	Terpene	0.032 ± 0.044	0.063 ± 0.073	0.052 ± 0.064	0.001 - 0.403	0.608 ± 0.095	0.377 - 0.739
α-Farnesene		AF	Terpene	0.096 ± 0.097	0.292 ± 0.217	0.203 ± 0.191	0.003 - 1.268	0.595 ± 0.090	0.426 - 0.757
Nerolidol	**X**	N	Terpene	0.382 ± 0.300	0.962 ± 0.670	0.603 ± 0.529	0.003 - 2.706	0.602 ± 0.107	0.397 - 0.772

Bold in the VOC names and X represent the KVCs detected with commercial standards. Significant differences between 'FC50' and 'FD54' are in bold in the parental with highest concentration.

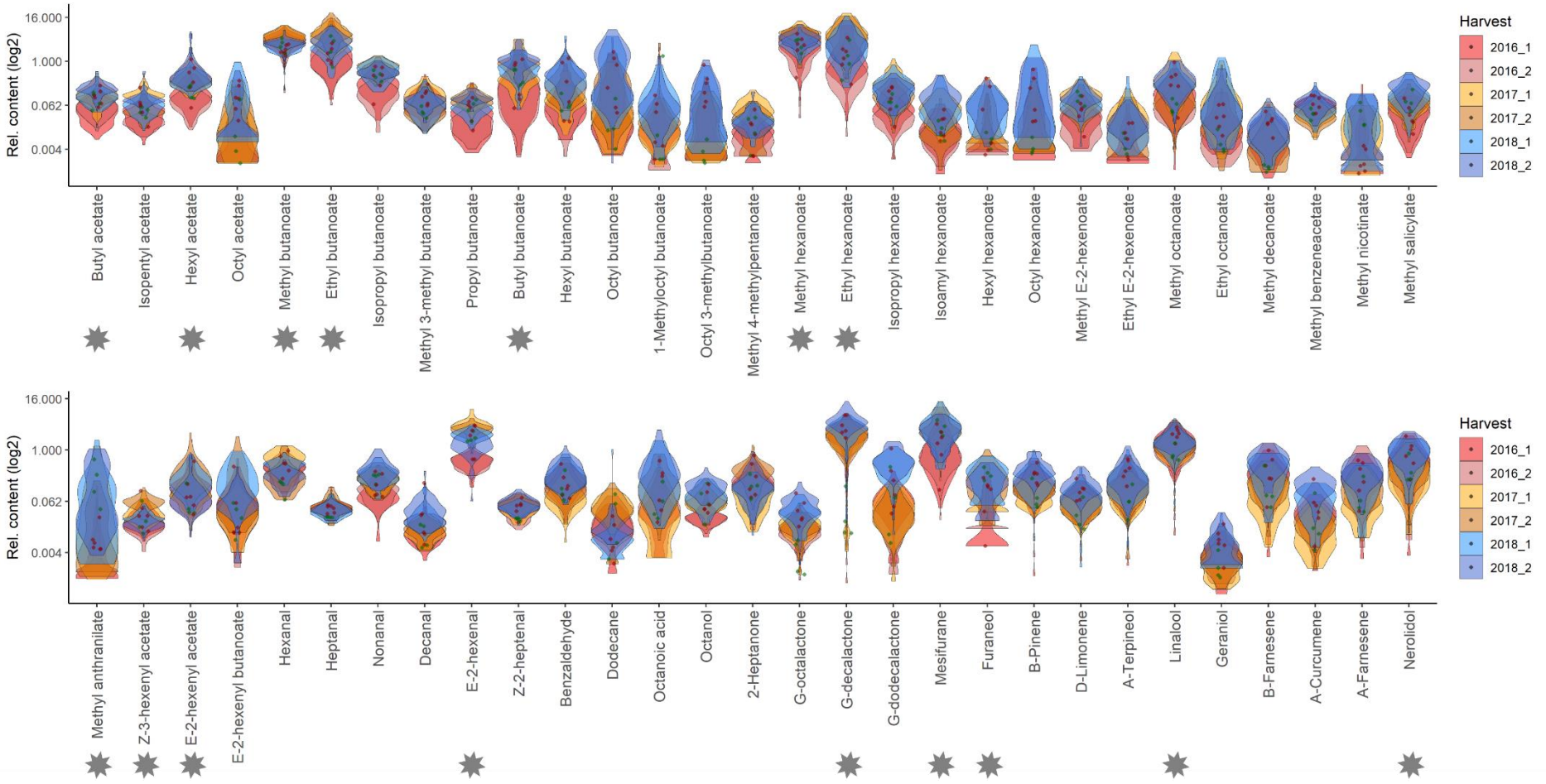


Figure 2-1. Violin plot representing VOC distributions. Harvest colours: 2016 (reddish), 2017 (orangish) and 2018 (bluish). Dots: 'FC50' (green), 'FD54' (red). Stars indicate KVCs.

Volatile correlation analysis

To further understand the relation between compounds in different harvests, each VOC independently of the harvest was analysed with Pearson correlation (Supplementary table 2-4). The results showed that ester compounds sharing the same acyl-CoA group were highly correlated between them, such as propyl butanoate with ethyl butanoate (0.64 – 0.90 in range), isopropyl butanoate with butyl butanoate (0.43 – 0.67), methyl hexanoate with propyl hexanoate (0.64 – 0.87) and with isoamyl hexanoate (0.43 – 0.78) and ethyl hexanoate with isoamyl hexanoate (0.20 – 0.77), or having the same alcohol group, such as methyl hexanoate with methyl octanoate (0.61 – 0.75) and with methyl decanoate (0.53 – 0.73), ethyl hexanoate with ethyl 2-hexenoate (0.30 – 0.92), with ethyl octanoate (0.68 – 0.89) and with ethyl butanoate (0.65 – 0.93), methyl octanoate with methyl decanoate (0.86 – 0.97), hexyl butanoate with hexyl acetate (0.52-0.93) and octyl acetate with octyl 3-methyl butanoate (0.41 – 0.83) and with octyl hexanoate (0.59 – 0.87) (Fig. 2-2).

Lactones had a high correlation between them being the highest one with γ-nonalactone and γ-decalactone (0.52 – 0.83). Also, furans showed higher correlation with each other than with other compounds (Fig. 2-2).

The highest stable correlations in different harvest were detected between terpene compounds. The minimum correlation between monoterpenes was 0.74, while in the case of sesquiterpene compounds it was of 0.77. However, minimum correlation between monoterpene and sesquiterpene compounds was lower (0.32). Linalool showed the highest correlation between all terpene compounds ranging from 0.60 to 0.99 (Fig. 2-2).

We also detected a high correlation between compounds from different families in most of the harvests. For instance, some butanoate related esters and furans presented high correlation between them, such as butyl butanoate with mesifurane (0.38 – 0.61) and furaneol (0.11 – 0.60). Moreover, some lactones and ester compounds were also correlated, such as γ-octalactone with methyl hexanoate (0.43 – 0.67), isopropyl hexanoate (0.50 – 0.76) and hexyl hexanoate (0.29 – 0.71). Terpenes also exhibited high correlation with other families, such as linalool with methyl hexanoate (0.33 – 0.67), methyl decanoate (0.44 – 0.61), methyl salicylate (0.42 – 0.60) and γ-decalactone (0.37 – 0.57) and nerolidol with methyl salicylate (0.32 – 0.63) and γ-octalactone (0.39 – 0.62) (Fig. 2-2).

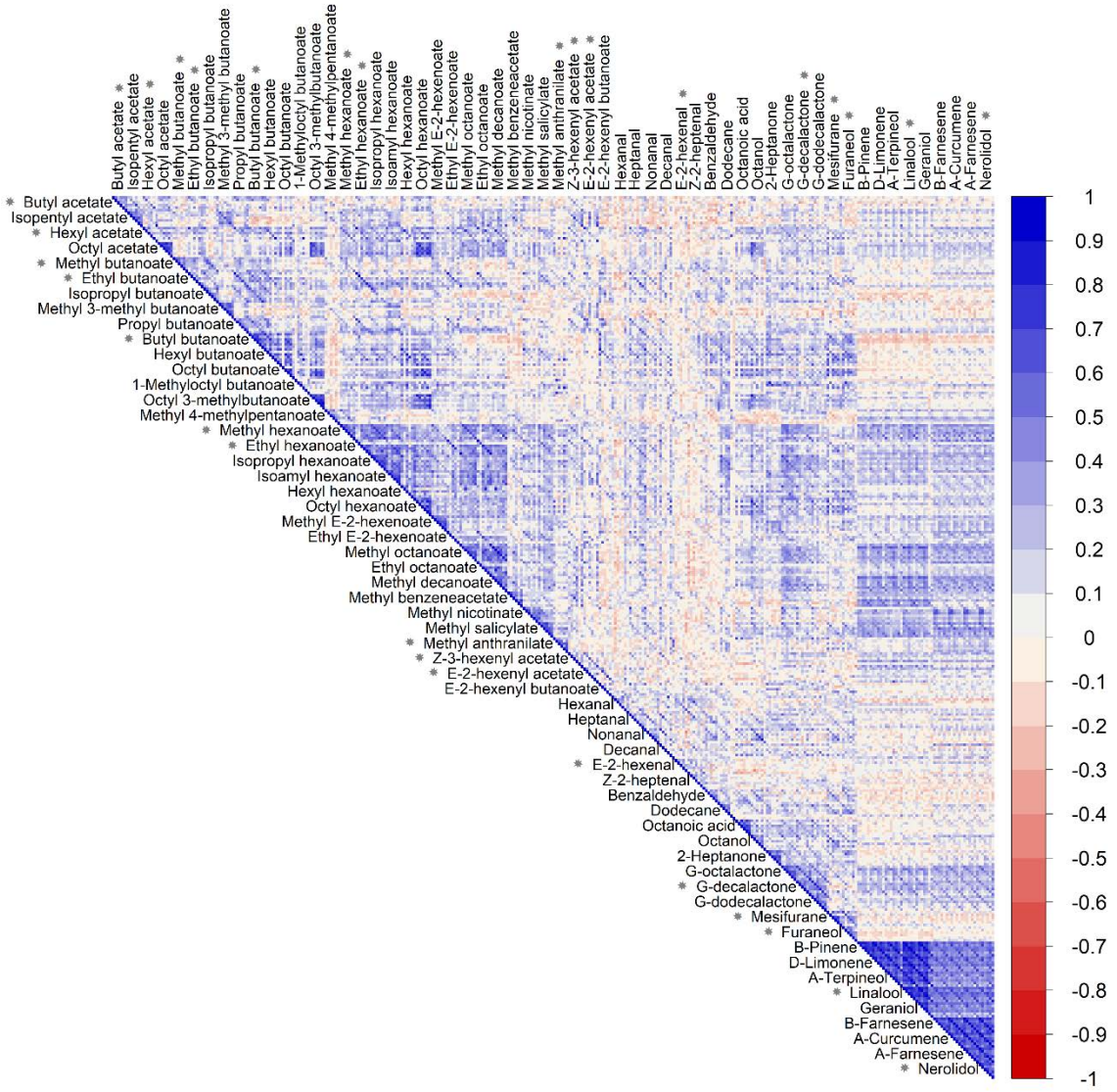


Figure 2-2. Heatmap visualization of Pearson correlation between VOCs in six harvests.

The Cluster network analysis (CNA) indicates the established relation between different VOCs (Fig. 2-3 and Supplementary table 2-5). This analysis suggested the existence of a strong correlation between the different terpene compounds which were clustered together and somehow separated to other VOCs. Esters, which are the most abundant compounds, were highly correlated with each other, while branched esters displayed less correlation with other compounds. Lactones showed a high correlation between them and also with some esters. Regarding to the furan compounds, they revealed a high correlation between them and also with hexyl butanoate. Only, the aldehyde family presented a low correlation between their components. Nevertheless, few weak negative correlations were detected.

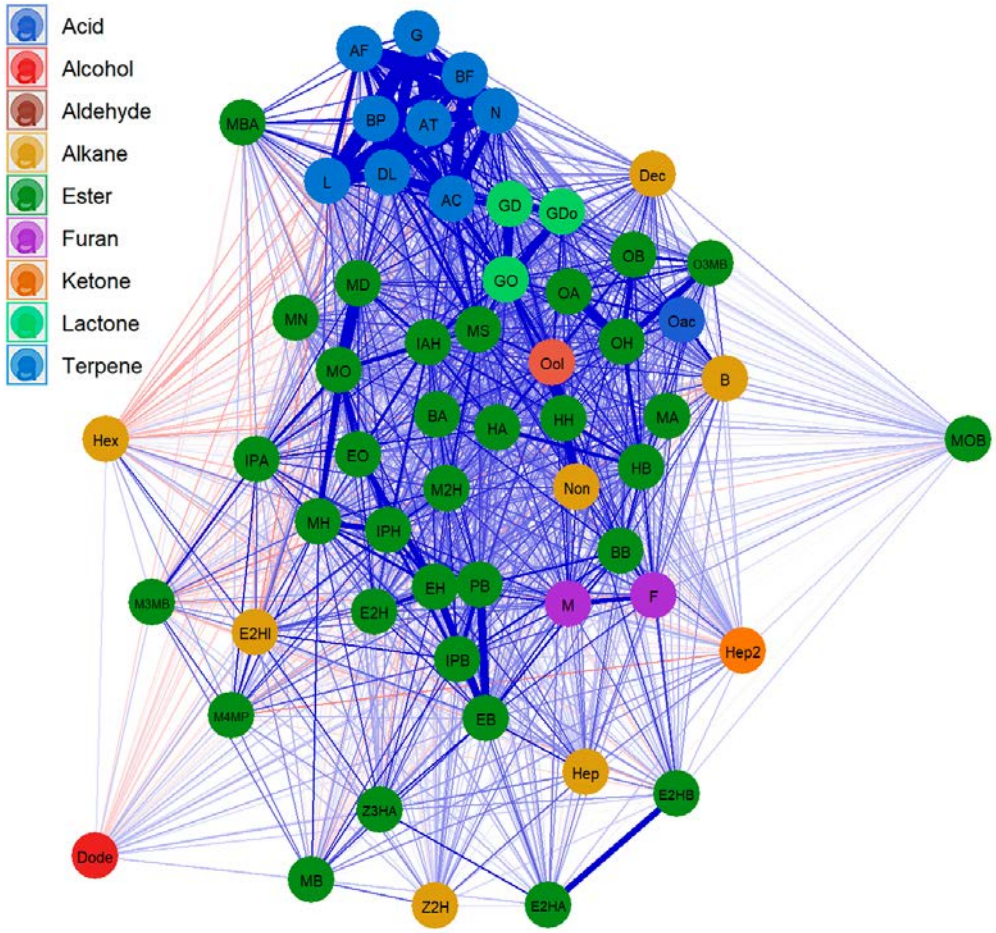


Figure 2-3. Cluster network analysis (CNA) in VOCs. Compounds are represented as nodes and coloured according to the chemical families. Positive (blue) and negative (red) correlation are represented as link between compounds, the widest links being the strongest correlation. For the correspondences between the full name and their abbreviation names consult Table 2-1.

To better understand the relationship between VOCs and progenies, parental lines ('FC50' and 'FD54') and grandparental lines ('FD019', 'FD016', 'FC030' and 'FD197'), a hierarchical cluster analysis (HCA) using the average from six harvests had been performed. A heatmap representation of the HCA shows three compound and four individual clusters (Fig. 2-4). In relation to the compound clusters, cluster 1 which represented 43% of the VOCs could be divided in two subclusters, 1A and 1B. Cluster 1A, containing 21% of the VOCs, grouped ester compounds derived from octyl-CoA and lactones. Cluster 1B, comprising 22% of the VOCs, clustered aldehydes and hexyl-CoA esters. Cluster 2, the smallest one, grouped 9% of the VOCs, being all butanoate derived esters. Finally, in cluster 3, we found 48% of the VOCs, grouping terpenes, hexanoate derived esters and branched ester, where the latter is the further from the others.

Regarding the relationship between lines, the hierarchical cluster analysis defines four groups, which could be interesting for breeding programs. The majority of the lines, 64% of the total analysed individuals, were grouped together in a big cluster A, which we separated in two subclusters, A1 and A2. Cluster A1, including 18% of individuals such as 'FC50' parental line and 'FD016' grandparental line, was characterised by its low relative VOC content. Cluster A2, the largest one comprising 46% of individuals such as 'FC030' grandparent line, was distinguished by its high content of several VOCs, but also by its low terpene content. Cluster B, represented by a unique sample, the 'FD197' grandparental line, was characterised by its high content of octanol, methyl nicotinate, butyl acetate and ethyl E-2-hexanoate. Cluster C, represented by the 28% of individuals and including the 'FC54' parental line, was characterised of having high terpene and several ester content. The last group, cluster D with 7% of individuals including the 'FD019' grandparental line was characterised by its high content of the majority of VOCs except for terpenes (Fig. 2-4).

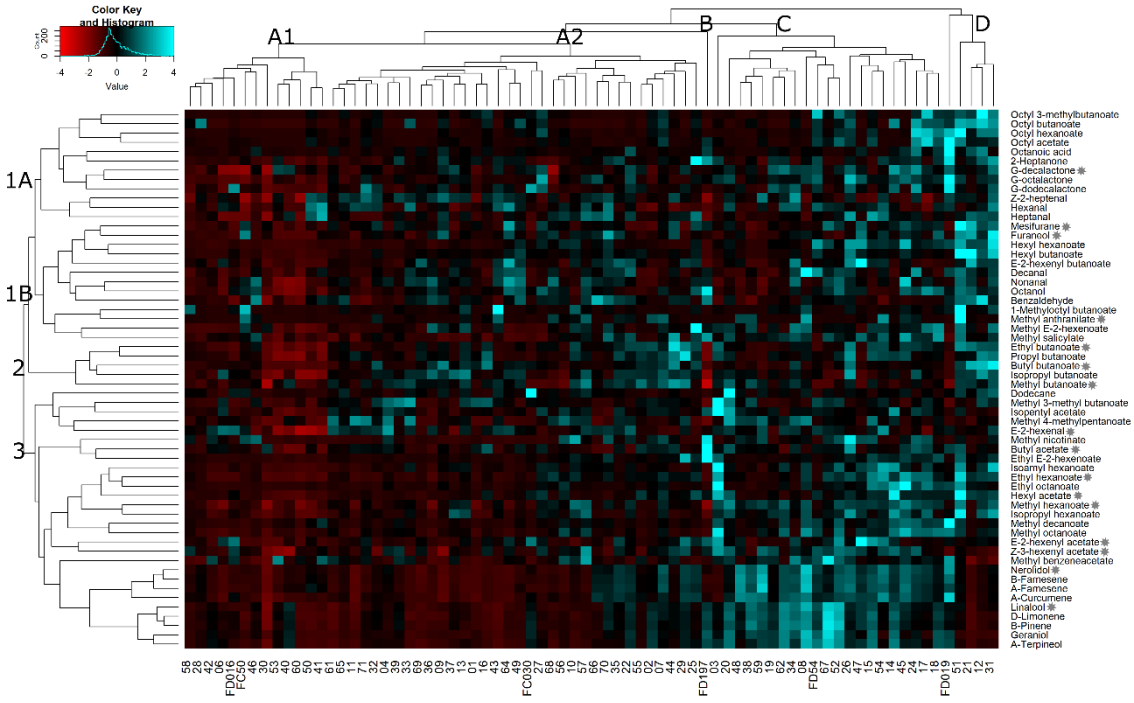


Figure 2-4. Heatmap representation of hierarchical clustering analyses (HCA) in VOCs. Clusters are named by capital letters for individuals and numbers for VOCs. Stars indicate the KVCs.

Analysis of environmental effect in VOCs

Plants need to respond to different environmental conditions which could alter in different ways the biosynthesis of some VOCs. We assumed that each harvest, cultivated in open field, behaves according to the environmental conditions which could modify the VOCs accumulation. To better understand how different environmental conditions could affect the VOC content, we

performed a Principal component analysis (PCA). The first three dimensions explained 26.5% (Dim1), 11.9% (Dim2) and 8.2% (Dim3) of the VOCs variance. In general, each harvest partially overlaps with the others in both PCA plots, but specifically, the two harvests of the same year are grouped closely, while slight differences arise between years, being 2018 the most divergent one (Fig. 2-5). This analysis reveals that the environmental effect, specially from different years, modified the accumulation of some volatile compounds.

Regarding the contribution of the VOCs to each PCA dimension, we did not find a clear pattern in Dim 1 and Dim 3. This indicates a complex relationship between VOCs. Focusing on Dim 2, we found terpene, lactone and aldehyde compounds clustering together, whereas esters such as the octyl group and hexanoate related esters were forming a small and dense group. Other ester compounds were scattered.

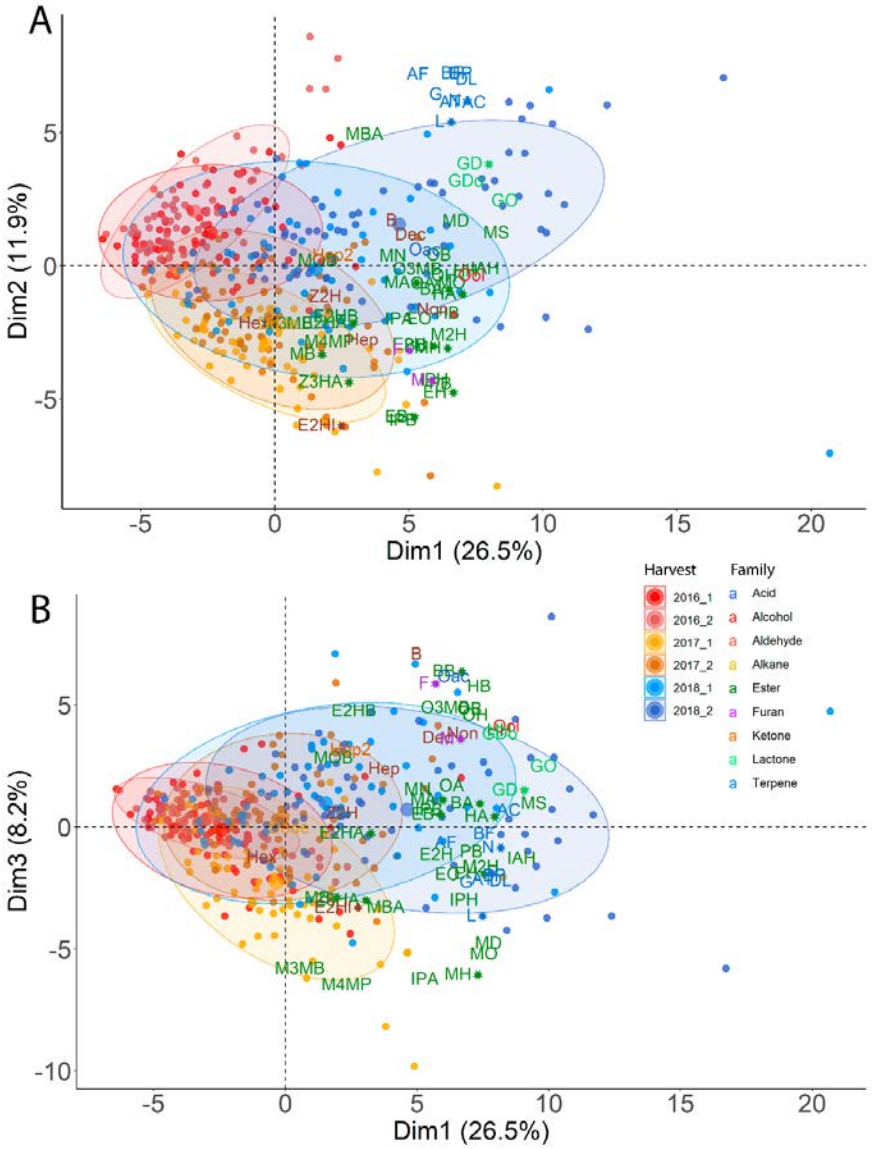


Figure 2-5. Principal component analysis (PCA) plot in VOCs: A) Dim1 vs Dim2 and B) Dim1 vs Dim3. Harvest: 2016 (reddish), 2017 (orangish) and 2018 (bluish).

To break down the variance caused by the genotype (G), the environment (E) or the interactions between them (GxE) in VOCs accumulation, we evaluated their significance and quantified their effect for each trait (Supplementary table 2-6). The analysis of variance (ANOVA), taking into account the genotype, year and interaction between them (G + E + GxE), revealed 50 VOCs showing significant genetic differences (p-value<0.05), 48 VOCs depending significantly on the environmental effect and only 18 VOCs exhibiting a significant effect of GxE interaction. Additionally, 15 VOCs showed to be significantly influenced by these three factors, and 28 of them were only influenced by genotype and environmental factors including most of KVCs. This demonstrate that most of the VOCs, 43 out of 56, were affected by environmental changes.

In general, the genotypic factor was the main contributor to the observed variance over environmental and GxE factors. However, a high proportion of the VOCs variability cannot be attributed to any of these factors which we named 'error'. At least 25% of the observed variance was explained by the G factor in 26 VOCs. This genetic variance was particularly high in linalool, nerolidol and octyl 3-methyl butanoate, explaining about 40% of their variance. In contrast, we found a high influence of the E factor explaining more than 30% of their total observed variance in some VOCs, such as methyl salicylate, nonanal, octanol and γ-nonalactone. Nevertheless, we did not observe a GxE effect higher than 20% of the total observed variance in any VOCs (Fig. 2-6).

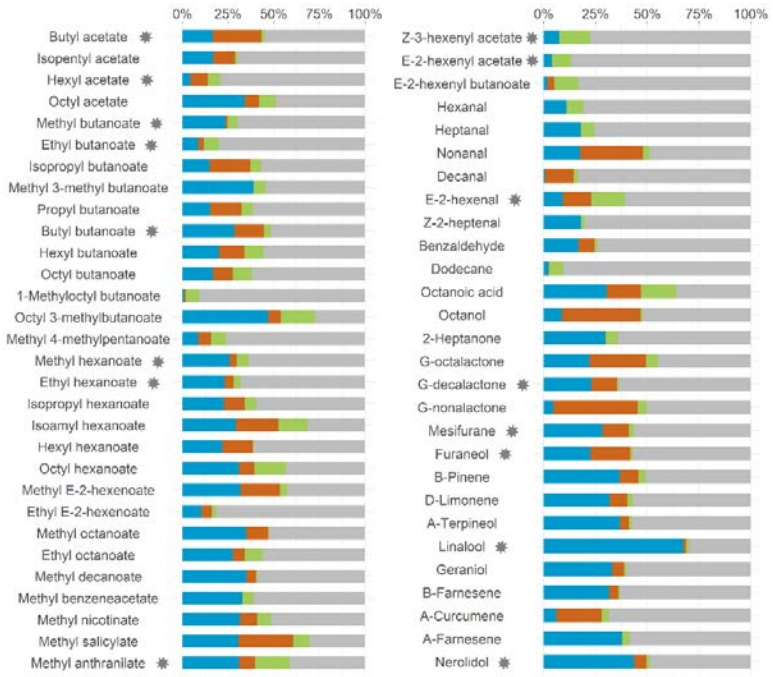


Figure 2-6. ω^2 values for VOCs. Percentages of variance for G factor (blue), E factor (orange), GxE factor (green) and error (grey) for each VOC obtained by ANOVA. Stars indicate KVCs.

Volatile QTLs detection in 'FC50xFD54' genetic map

Despite more than 350 volatile compounds have been detected in strawberry, only 58 VOCs has been well identified by GC-MS in the 'FC50xFD54' population, of which 16 were considered KVCs. Nonetheless, a total of 178 QTLs were found with LOD > 2.5 in at least two different harvests for 55 different VOCs (Supplementary table 2-7 and Supplementary fig. 2-1). From them, 94 QTLs were highly stable (three to four harvests) including 22 of them being totally stable (five to six harvests) (Table 2-2). Only in the case of ethyl butanoate and E-2-hexenyl acetate any QTL was detected. Concerning to KVCs, a total of 44 QTLs were mapped from 14 KVCs.

In relation to the KVCs, we detected 22 highly stable QTLs for eleven compounds. Focusing on the ester family, two butyl acetate QTLs were mapped being *BA_5A*, which explained a maximum of 27% of the observed variance, and *BA_6A*, a major and totally stable QTL explaining a range between 13.8-35.2% of the phenotypic variance. Significant QTLs for hexyl acetate were detected such as *HA_1A* and *HA_5A* which explained a maximum observed variance of 24.5%. For methyl butanoate, we found different QTLs such as *MB_2C* a major QTL that explained a maximum of 28.5% of the variance and a significant *MB_3A* QTL that explains a maximum of 20% of the variance. For butyl butanoate, we mapped a single significant QTL in LG6A explaining a maximum of 27.4% of the phenotypic variance. Two major and highly stable QTLs for methyl hexanoate were *MH_4B* which explained a range between 13.6-44.5% of the variance and *MH_7C* explaining a range between 13.6-36.3% of variance. Although ethyl hexanoate did not pass our QTL criteria, a stable QTL but less significant was also mapped in LG4B. Six different QTLs were detected for methyl anthranilate. From them, two were totally stable QTLs like the *MA_1A,* explaining a maximum of 34.6% of phenotypic variance and *MA_5C* showing a maximum of 29% of variance and a totally stable and major QTL in LG7A explaining a range of 22.7-46.4% of the phenotypic variance.

With respect to lactone family compounds, γ-decalactone showed a stable and significant QTL, mapped in LG6C and explained a maximum of 25.5% of the observed variance. For mesifurane, the stable *M_6A* QTL explained a maximum of 25.1% of its variance. For furaneol, a stable and major QTL was mapped in LG1B, explaining a maximum of 33.2% of the variance.

Focusing on the terpene family, we mapped a major and totally stable QTLs for linalool and nerolidol in LG3B explaining a range of 16.2-44.0% of the linalool variance. In addition to *N_3B* QTL, we found a major *N_4B* explaining a maximum of 33.4% of the variance and a significant *N_1C* QTL (Table 2-2).

Table 2-2. vocQTLs detected in 'FC50xFD54' population. List of highly and totally stable and significant vocQTLs ordered by compounds. Maximum values of LOD score, % of explanation, Kruskal-Wallis test, QTL size determined by 1-LOD confidence interval, QTL position in *F x ananassa* genome and number of genes in each region.

QTL ID	KVC	Harv.	LOD	% Expl.	KW test	Genetic map			*F. x ananassa* genome				
						LG	QTL interval (cM)	QTL size (cM)	CHR	QTL interval (pb)		QTL size (pb)	nº genes
BA_5A	X	3	3.76	27.0	13.187	5A	26.351	38.274	11.923	Fvb5-1	6049805	10631744	4581939	850
BA_6A	X	**5**	**5.19**	**35.2**	**20.049**	6A	68.031	77.919	9.888	Fvb6-1	12278924	17819932	5541008	753
IPA_3B		3	3.17	23.3	11.945	3B	80.003	86.536	6.533	Fvb3-3	25521482	29430243	3908761	557
HA_1A	X	4	3.36	24.5	12.546	1A	56.671	61.246	4.575	Fvb1-4	11053422	12328746	1275324	172
HA_5A	X	3	3.36	24.6	12.322	5A	48.617	65.788	17.171	Fvb5-1	13455988	20026637	6570649	831
OA_5C		4	3.13	23.0	6.258	5C	55.636	62.159	6.523	Fvb5-2	17732180	19025224	1293044	147
OA_6A		**6**	**14.41**	**69.4**	**31.480**	6A	73.390	92.943	19.553	Fvb6-1	6732393	17819932	11087539	1652
OA_6D		3	3.33	24.7	11.420	6D	61.154	74.899	13.745	Fvb6-2	27424479	36095144	8670665	1453
OA_7D		3	3.56	26.2	13.015	7D	3.637	7.274	3.637	Fvb7-3	15343169	21304648	5961479	705
MB_2C	X	4	**4.07**	**28.5**	**13.388**	2C	58.732	76.172	17.440	Fvb2-4	17897006	23918678	6021672	1021
MB_3A	X	3	2.68	20.0	9.743	3A	10.070	27.157	17.087	Fvb3-4	23563664	26780849	3217185	546
IPB_4A		3	**5.53**	**37.6**	**10.012**	4A	1.855	12.144	10.289	Fvb4-4	26113700	26266631	152931	13
M3MB_3D		4	**6.52**	**41.5**	**20.607**	3D	64.920	69.993	5.073	Fvb3-2	20985006	26768589	5783583	653
PB_1A		3	3.70	27.0	10.320	1A	60.337	77.252	16.915	Fvb1-4	11397822	20339707	8941885	1047
PB_5A		4	**4.13**	**29.2**	**14.942**	5A	53.099	84.112	31.013	Fvb5-1	13455988	29624690	16168702	2154
PB_5C		3	**4.91**	**33.2**	**18.621**	5C	5.910	15.937	10.027	Fvb5-2	1136082	8091321	6955239	1108
BB_6A	X	3	3.76	27.4	11.222	6A	8.226	36.499	28.273	Fvb6-1	25985516	32354801	6369285	1096
HB_1A		4	**5.06**	**35.1**	**16.116**	1A	28.151	56.671	28.520	Fvb1-4	3737049	9864602	6127553	1037
HB_6A		4	**4.64**	**32.7**	**18.232**	6A	3.646	31.499	27.853	Fvb6-1	25985516	32354801	6369285	1096
OB_1A		3	**5.72**	**38.6**	**11.102**	1A	43.144	50.984	7.840	Fvb1-4	7646768	9864602	2217834	339
OB_6A		**6**	**11.26**	**61.7**	**17.699**	6A	77.919	90.129	12.210	Fvb6-1	11796362	17819932	6023570	830
O3MB_2B		4	**4.51**	**31.9**	**11.819**	2B	3.705	13.372	9.667	Fvb2-3	15761148	22097247	6336099	653
O3MB_3A		4	3.45	24.7	13.372	3A	0.000	2.818	2.818	Fvb3-4	27377241	27736960	359719	62
O3MB_3D		3	2.96	21.9	4.226	3D	67.926	77.191	9.265	Fvb3-2	23891004	30491301	6600297	898
O3MB_6A		**6**	**14.40**	**70.7**	**20.006**	6A	73.390	85.129	11.739	Fvb6-1	11796362	17819932	6023570	830
O3MB_6B		3	3.25	24.2	10.171	6B	68.768	107.162	38.394	Fvb6-3	10324770	11930571	1605801	202
O3MB_7D		4	3.05	22.9	10.752	7D	5.455	13.744	8.289	Fvb7-3	13522328	20762643	7240315	866
M4MP_6A		4	3.84	27.5	14.022	6A	98.893	107.236	8.343	Fvb6-1	77396	10670615	10593219	1945
MH_4B	X	4	**6.91**	**44.5**	**25.431**	4B	57.484	62.900	5.416	Fvb4-3	4032763	11434813	7402050	1233
MH_7D	X	3	**5.49**	**36.3**	**19.673**	7D	27.134	49.803	22.669	Fvb7-3	927546	11272705	10345159	1754

QTL ID	KVC	Harv.	LOD	% Expl.	KW test	Genetic map			*F. x ananassa* genome				
						LG	QTL interval (cM)	QTL size (cM)	CHR	QTL interval (pb)		QTL size (pb)	nº genes
IPH_4B		4	**4.51**	**31.9**	**16.692**	4B	57.484	61.900	4.416	Fvb4-3	4032763	11434813	7402050	1233
IPH_7D		**6**	**9.08**	**53.9**	**29.927**	7D	27.134	44.570	17.436	Fvb7-3	6319658	11272705	4953047	765
IAH_3C		4	3.70	27.1	10.260	3C	42.075	54.849	12.774	Fvb3-1	18530045	25219765	6689720	849
IAH_4B		4	**5.25**	**36.1**	**18.859**	4B	54.488	60.082	5.594	Fvb4-3	4032763	11434813	7402050	1233
HH_3C		3	3.42	25.3	11.485	3C	96.677	109.491	12.814	Fvb3-1	98571	2627253	2528682	350
HH_4B		3	**4.48**	**31.8**	**17.026**	4B	54.488	62.900	8.412	Fvb4-3	4032763	11434813	7402050	1233
OH_6A		**6**	**12.59**	**65.2**	**21.408**	6A	73.390	90.129	16.739	Fvb6-1	11796362	17819932	6023570	830
OH_6B		3	3.90	27.5	11.412	6B	83.764	106.162	22.398	Fvb6-3	11,470,022	11,494,187	24165	5
OH_7D		**5**	3.34	24.4	12.579	7D	5.455	9.274	3.819	Fvb7-3	15343169	20762643	5419474	647
M2H_2A		4	3.55	26.1	14.501	2A	82.164	91.181	9.017	Fvb2-1	25,214,351	26,213,900	999549	204
M2H_4B		3	**4.32**	**30.8**	**16.600**	4B	56.488	61.900	5.412	Fvb4-3	4032763	11434813	7402050	1233
MO_2B		3	3.51	25.5	11.745	2B	1.887	7.362	5.475	Fvb2-3	15761148	23613740	7852592	823
MO_3C		4	3.22	24.0	10.694	3C	64.677	103.033	38.356	Fvb3-1	231771	21949202	21717431	2370
MO_4B		4	**7.22**	**46.0**	**25.730**	4B	58.484	61.900	3.416	Fvb4-3	4032763	11434813	7402050	1233
MO_7B		**5**	**4.03**	**28.6**	**16.451**	7B	6.364	20.696	14.332	Fvb7-4	11334719	18394187	7059468	849
MO_7D		4	3.23	24.1	12.256	7D	25.316	44.454	19.138	Fvb7-3	8401319	11272705	2871386	395
MD_2A		4	2.90	21.9	11.555	2A	1.957	19.391	17.434	Fvb2-2	2790981	24369919	21578938	3040
MD_3B		3	3.86	27.2	6.562	3B	30.350	52.249	21.899	Fvb3-3	10894646	18844018	7949372	812
MD_4B		**6**	**9.98**	**57.3**	**31.520**	4B	57.484	62.900	5.416	Fvb4-3	4032763	11434813	7402050	1233
MBA_1A		3	**4.37**	**31.1**	**14.419**	1A	1.818	8.849	7.031	Fvb1-4	258081	2140039	1881958	382
MBA_3B		**6**	**4.45**	**31.1**	**12.102**	3B	53.249	86.536	33.287	Fvb3-3	14538899	29430243	14891344	1682
MN_3B		**6**	**14.93**	**72.0**	**35.521**	3B	18.550	35.350	16.800	Fvb3-3	5280472	7799558	2519086	358
MS_3B		**6**	**13.63**	**68.0**	**27.947**	3B	18.550	21.350	2.800	Fvb3-3	5280472	7799558	2519086	358
MA_1A	X	3	**5.07**	**34.6**	**16.360**	1A	31.865	38.923	7.058	Fvb1-4	3737049	6576705	2839656	526
MA_4D	X	3	3.33	24.7	13.875	4D	49.275	67.145	17.870	Fvb4-2	3979108	5942673	1963565	322
MA_5A	X	3	3.93	28.5	13.069	5A	17.580	28.170	10.590	Fvb5-1	5157884	8627726	3469842	656
MA_5B	X	4	3.74	26.9	9.399	5B	17.009	31.613	14.604	Fvb5-3	12122130	18591433	6469303	767
MA_5C	X	3	**4.16**	**29.0**	**7.819**	5C	4.849	10.092	5.243	Fvb5-2	1136082	6216063	5079981	809
MA_7A	X	**6**	**7.30**	**46.4**	**23.438**	7A	24.262	46.524	22.262	Fvb7-2	13707587	30080597	16373010	2767
B_2B		3	3.76	27.4	13.356	2B	10.372	31.782	21.410	Fvb2-3	11376509	19079082	7702573	835
B_4C		3	3.75	27.4	10.467	4C	54.514	62.010	7.496	Fvb4-1	15125510	16588955	1463445	194
Dode_6B		3	**4.15**	**28.9**	**10.334**	6B	92.162	126.816	34.654	Fvb6-3	96647	6088328	5991681	1078
Dode_7D		4	**6.85**	**44.2**	**25.838**	7D	19.562	40.221	20.659	Fvb7-3	8401319	11272705	2871386	395
Oac_1B		3	**4.85**	**33.9**	**15.942**	1B	0.000	21.933	21.933	Fvb1-3	384740	3939728	3554988	667

QTL ID	KVC	Harv.	LOD	% Expl.	KW test	Genetic map			*F. x ananassa* genome				
						LG	QTL interval (cM)	QTL size (cM)	CHR	QTL interval (pb)		QTL size (pb)	nº genes
Oac_5D		3	3.68	26.1	15.350	5D	55.000	72.244	17.244	Fvb5-4	13639694	25171401	11531707	1235
Oac_7B		**5**	**6.25**	**41.3**	**22.426**	7B	0.000	14.053	14.053	Fvb7-4	11334719	23072006	11737287	1243
Ool_6A		3	3.85	28.0	14.764	6A	90.129	120.837	30.708	Fvb6-1	77396	34979000	34901604	5660
Ool_6C		3	**4.54**	**31.6**	**8.111**	6C	50.903	61.337	10.434	Fvb6-4	18263601	22737430	4473829	502
Ool_6D		3	3.45	25.1	3.978	6D	6.320	12.598	6.278	Fvb6-2	22040645	23358583	1317938	187
GO_2A		3	3.27	24.3	11.225	2A	14.045	24.767	10.722	Fvb2-2	2790981	12838545	10047564	1641
GO_4B		4	3.72	27.2	14.415	4B	56.488	73.959	17.471	Fvb4-3	103787	11434813	11331026	1980
GO_6B		3	2.74	20.8	2.411	6B	56.072	63.370	7.298	Fvb6-3	9180853	16191056	7010203	808
GD_6C	X	3	3.51	25.5	12.229	6C	53.812	80.743	26.931	Fvb6-4	15043440	31604845	16561405	1951
GDo_6C		4	**4.06**	**28.8**	**12.229**	6C	53.812	66.629	12.817	Fvb6-4	16212039	25317092	9105053	1010
M_6A	X	3	3.39	25.1	11.691	6A	112.537	120.837	8.300	Fvb6-1	137388	3658007	3520619	703
F_1B	X	3	**4.74**	**33.2**	**8.970**	1B	0.000	5.524	5.524	Fvb1-3	384740	1890336	1505596	283
BP_3B		**6**	**7.15**	**45.6**	**26.307**	3B	1.818	7.433	5.615	Fvb3-3	1003939	1958161	954222	171
DL_3B		**5**	**7.32**	**45.2**	**27.413**	3B	1.818	7.433	5.615	Fvb3-3	1003939	1958161	954222	171
AT_3B		**5**	**6.20**	**39.9**	**12.430**	3B	1.818	5.524	3.706	Fvb3-3	1003939	1958161	954222	171
L_3B	X	**5**	**7.06**	**44.0**	**25.171**	3B	1.818	7.433	5.615	Fvb3-3	1003939	1958161	954222	171
G_3B		**6**	**7.01**	**43.8**	**25.835**	3B	1.818	7.433	5.615	Fvb3-3	1003939	1958161	954222	171
BF_1C		3	3.75	27.4	12.710	1C	11.104	26.607	15.503	Fvb1-2	142477	10369708	10227231	1847
BF_3B		4	**7.23**	**44.8**	**23.846**	3B	1.818	35.350	33.532	Fvb3-3	1003939	7799558	6795619	1077
BF_4B		3	**5.00**	**34.7**	**15.490**	4B	57.484	73.959	16.475	Fvb4-3	103787	11434813	11331026	1980
AC_1C		3	3.64	26.7	12.344	1C	10.104	26.607	16.503	Fvb1-2	142477	10369708	10227231	1847
AC_3B		**5**	**7.25**	**44.9**	**23.610**	3B	1.818	34.350	32.532	Fvb3-3	1003939	7799558	6795619	1077
AC_4B		4	**4.94**	**34.4**	**15.764**	4B	57.484	73.959	16.475	Fvb4-3	103787	11434813	11331026	1980
AC_4D		3	3.03	22.8	8.314	4D	23.161	27.786	4.625	Fvb4-2	11234456	16566656	5332200	513
AF_1C		3	3.09	23.2	9.914	1C	10.104	26.607	16.503	Fvb1-2	142477	10369708	10227231	1847
AF_3B		**5**	**7.31**	**45.2**	**23.846**	3B	1.818	35.350	33.532	Fvb3-3	1003939	7799558	6795619	1077
AF_4D		3	2.85	21.5	10.258	4D	11.206	28.786	17.580	Fvb4-2	557831	21805796	21247965	2595
N_1C	X	3	3.51	25.9	11.385	1C	8.585	26.607	18.022	Fvb1-2	142477	10369708	10227231	1847
N_3B	X	**5**	**7.17**	**44.6**	**23.687**	3B	1.818	35.350	33.532	Fvb3-3	1003939	7799558	6795619	1077
N_4B	X	3	**4.76**	**33.4**	**15.901**	4B	57.484	73.959	16.475	Fvb4-3	103787	11434813	11331026	1980

Bold and underline numbers in harvest indicate totally stable QTLs and in LOD, % Expl. and KW test major QTLs.

The genetic regions responsible for the accumulation of the VOCs are usually shared with compounds that have the same biosynthesis pathway. For instance, major and highly stable QTLs for octyl ester compounds, such as *OA_6A*, *OB_6A*, *OMBA_6A* and *OH_6A*, were mapped at the same region, but this was not the case for octanoic acid and octanol (Fig. 2-7). Other QTLs for octyl esters were located in LG6B and LG7D (Fig. 2-7). Furthermore, QTLs for the acetate ester compounds such as *BA_6A* and *HA_6A* were also mapped in the same region, although, they had a smaller LOD score (Fig. 2-7). Additionally, in the same region of the LG1A, we found different stable and significant QTLs for octanoic acid, octanol, octyl butanoate, octyl 3-methylbutanoate and octyl hexanoate (Fig. 2-7).

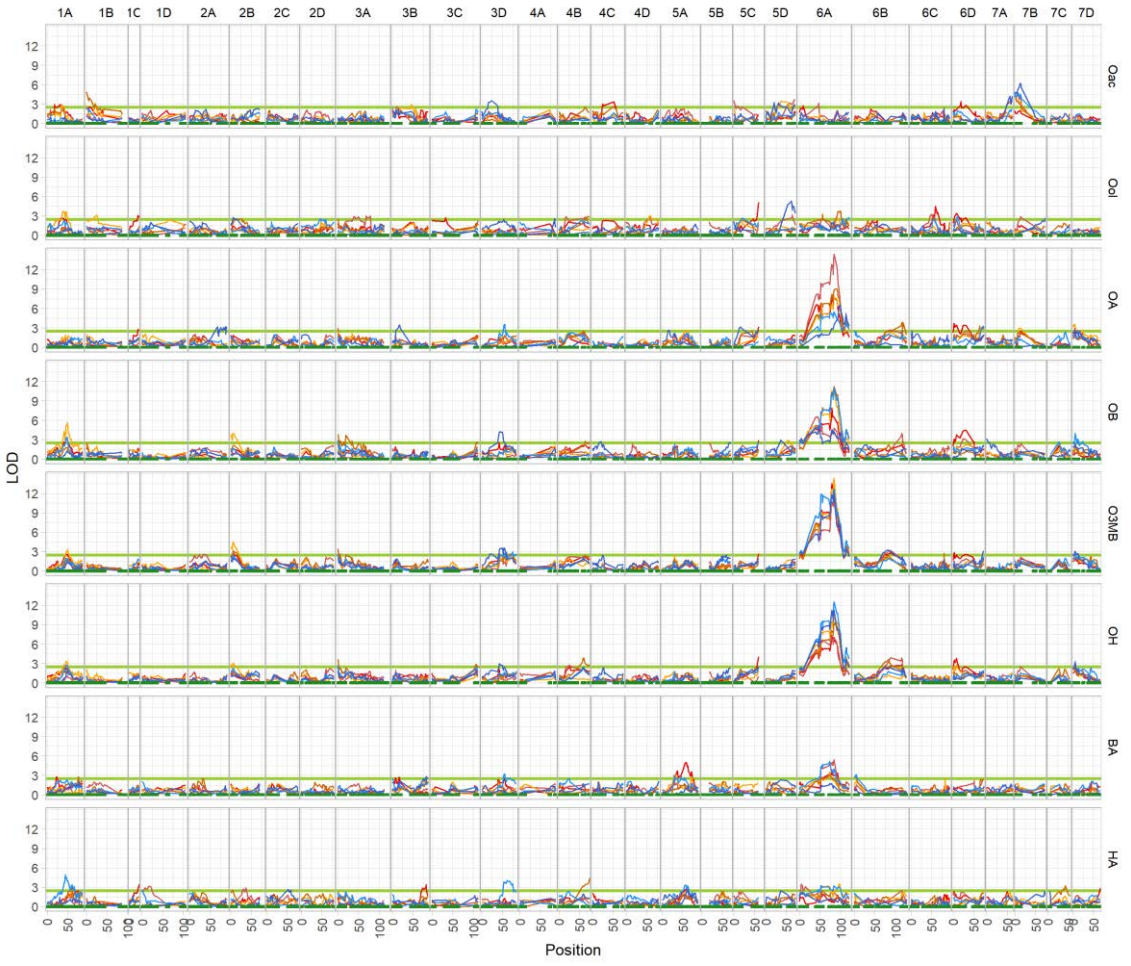


Figure 2-7. Graphical representation of LOD score for octyl related compounds in six harvests and all LGs. Harvest colour: 2016 (reddish), 2017 (orangish) and 2018 (bluish). Horizontal green line indicates 2.5 LOD score threshold.

The QTLs for ester compounds are present in almost all linkage groups. As it is shown in Fig. 2-7, major QTLs of octyl ester compounds are mapped in LG6A, LG6B and LG7D, being the QTL in LG6A the most significant and stable one. In LG1A, few ester QTLs are mapped closely to the *OB_1A* QTL, such as *HA_1A, HB_1A, PB_1A* and *MA_1A* (Fig. 2-8). Several QTLs mainly related to

hexanoate but also to octanoate and decanoate (*MH_4B*, *IAH_4B*, *IPH_4B*, *HH_4B*, *M2H_4B*, *MO_4B* and *MD_4B*) are mapped at the end of LG4B. Moreover, QTLs of *MH_7D*, *IPH_7D* and *MO_7D* are detected in the same LG7D than some octyl ester QTLs but in different regions.

The QTLs of the lactone compounds are located in few LGs, being a major *GDo_6C* QTL located at the same region than *GD_6C*. Regarding the QTLs for the terpenes, we found a major and highly stable QTL shared with all the identified terpenes (*L_3B*, *BP_3B*, *DL_3B*, *AT_3B*, *G_3B*, *N_3B*, *AF_3B*, *BF_3B* and *AC_3B*) at the beginning of LG3B. In addition to these major QTLs, we detected stable QTLs for monoterpene compounds in LG3A and for sesquiterpenes in LG1C, LG4B and LG4D (Supplementary table 2-6).

As we already explained in the case of monoterpenes (LG3A and LG3B) and sesquiterpenes (LG4B and LG4D), QTLs may be located in the same HG but in different LGs showing one of them higher significance and stability than the other. We also noticed the same QTL characteristic for the QTLs of octyl hexanoate and octyl 3-methylbutanoate mapped in LG6A and LG6B. Another example is methyl anthranilate QTLs mapped in three different LGs of the same HG5, but in this case the most stable QTL is located in LG7A.

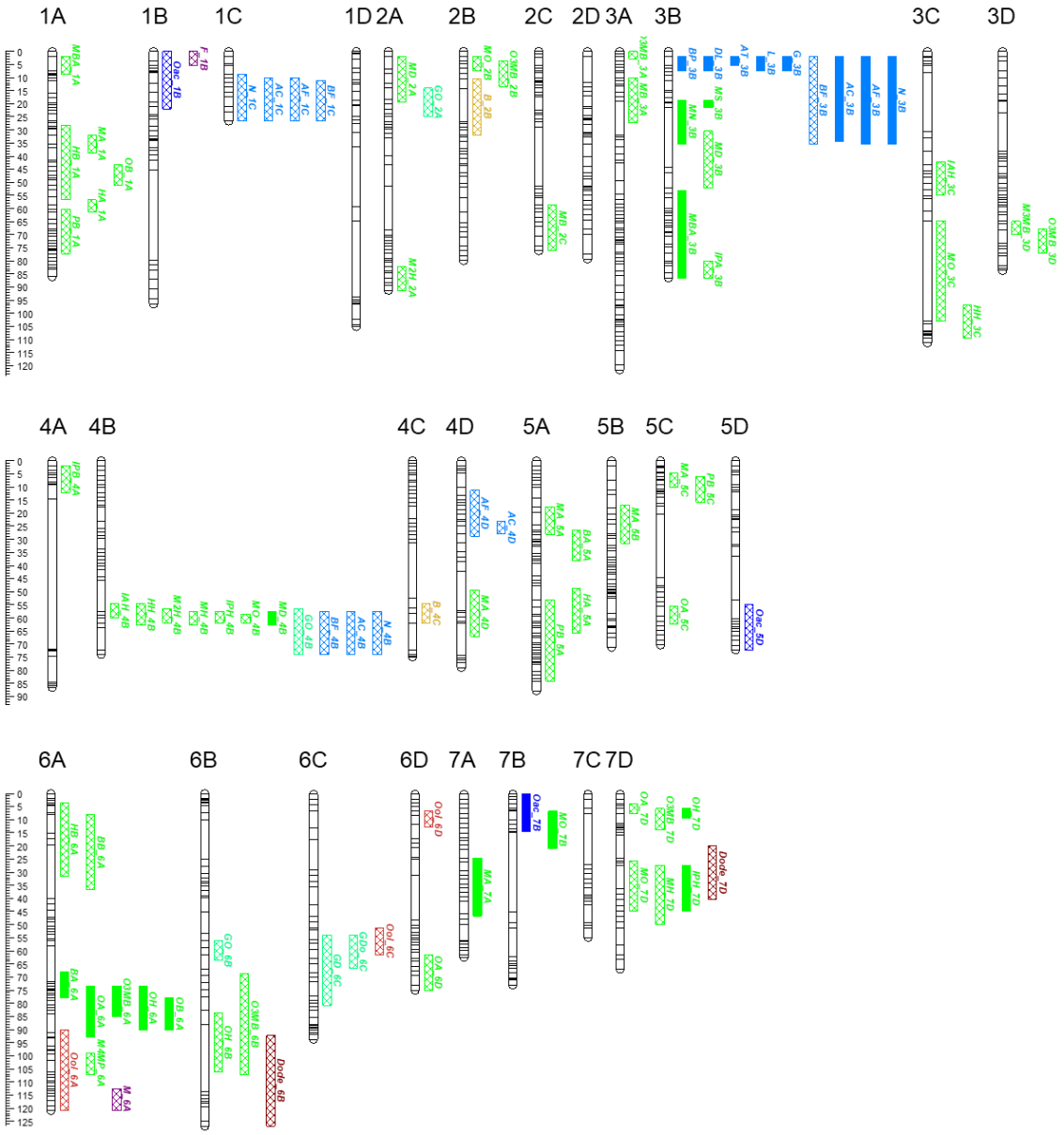


Figure 2-8. vocQTLs located in the 'FC50xFD54' genetic map. Box size correspond to the 1-LOD confidence interval. Stability is represented by box filling degree: totally stable (filled) and highly stable (semi-filled). Chemical family compounds are differentiated by colours: acids (light blue), alcohols (red), aldehydes (yellow), alkanes (dark red), esters (green), furans (purple), lactones (turquoise) and terpenes (dark blue).

Terpenes, which we identified five monoterpenes (linalool, β-pinene, D-limonene, α-terpineol and geraniol) and four sesquiterpenes (nerolidol, α-farnesene, β-farnesene and α-curcumene), provide a floral aroma. We have further studied the *Terpene_3B*, the most significant and stable QTL located at the beginning of LG3B detected in this population (Fig. 2-8 and Fig. 2-9A).

Focusing on the parental genetic maps, we found that this QTL was segregating in 'FC50' genetic map, just before a gap that split this LG. Moreover, any 'FD54' marker was significant in this region (Fig. 2-9B).

Since markers from IStraw35k array are not subgenome-specific, blast analysis showed that markers of linalool QTL could correspond either to the Fvb3-3 (*F. nipponica-like*) or Fvb3-4 (*F. vesca-like*) chromosomes. As Fvb3-4 is inverted comparing to our LG3B, these markers were located at the end of this chromosome. With the recent appearance of the 50k SNP array whose markers are supposed to be subgenomic-specific, we designed HRM markers using their SNP flanking sequences located inside the region of interest in Fvb3-3 and Fvb3-4 chromosomes and tested them in our 68 progenies and parental lines. The HRM markers designed from Fvb3-3 showed the same segregation than the markers mapped in LG3B. So, this ensure that our QTL is really located in the Fvb3-3 chromosome. Moreover, the TP_3.3_2.623_50k marker was the most significant HRM marker, located at the 2.623 Mbp position in Fvb3-3 chromosome. In addition, progenies with XX genotype contains significantly (p-value <0.001) higher linalool levels than those of the XY genotype in all six harvests (Fig. 2-9C), meaning that the homozygote genotype of the 'FD54' allele increase linalool accumulation.

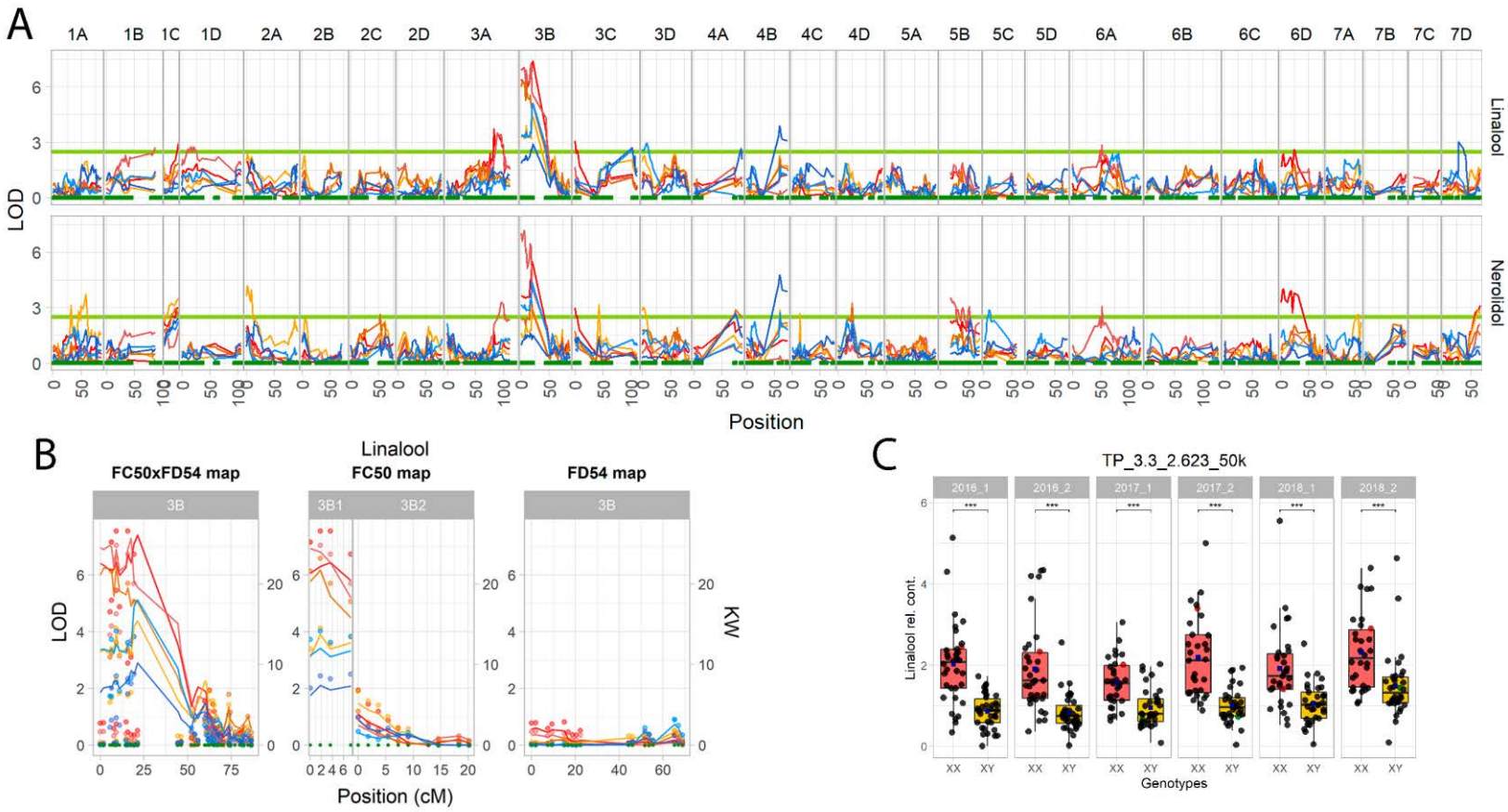


Figure 2-9. *Terpene_3B* QTL. A) LOD score for linalool and nerolidol for all LGs of the 'FC50xFD54' genetic map. Horizontal green line corresponds to the threshold of 2.5 LOD score. B) Linalool LOD score (lines) and KW statistic test (dots) in LG3B for 'FC50xFD54', 'FD54' and 'FC50' genetic maps. Harvest: 2016 (reddish), 2017 (orangish) and 2018 (bluish). C) Boxplot of the HRM marker genotypes with the linalool relative content. Dots: 'FC50xFD54' progenies (black), 'FC50' (green) and 'FD54' (red). Blue square is the average of each genotype. Significant level <0.001 (***).

Concerning to the major ester QTLs involving nine different compounds related to hexanoate group, they were located at the end of the LG4B and we named these group of QTLs as *Hexanoate_4B.* However, different stability degrees were observed in these compounds. Four of these compounds are directly related to hexanoate group (methyl hexanoate, ethyl hexanoate, isoamyl hexanoate and hexyl hexanoate), two are related to hexenoate group (methyl 2-hexanoate and ethyl 2-hexanoate), one to octanoate group (methyl octanoate) and one to decanoate group (methyl decanoate) (Fig. 2-8). However, octyl hexanoate and ethyl octanoate compounds, which may share the same pathway, did not show a QTL in this region (Fig. 2-8).

Focusing on the KVC methyl hexanoate QTL, we demonstrate that markers mapping in the 'FC50' genetic map were responsible for this QTL located at the end of the LG just after a gap (Fig. 2-10B). In this LG of the 'FD54' genetic map, we did not find any significant QTL for the methyl hexanoate.

The LG4B was clearly identified to correspond to the Fvb4-3, *F. vesca-like* subgenome. As Fvb4-3 was inverted, this QTL was located at the beginning of this chromosome. We designed HRM markers using the SNPs flanking regions from the corresponding genome regions of this QTL taken from the new 50k SNP array. The HRM showed a segregation similar to the mapped markers. One of these markers located in the 4.204Mbp position showed the highest association to the phenotype. Moreover, the progenies with the XX genotype contains significant higher levels of methyl hexanoate (p-value <0.001) than those of the XY genotype in all six harvest, indicating that the homozygote genotype of the 'FD54' allele enhance methyl hexanoate accumulation (Fig. 2-10C).

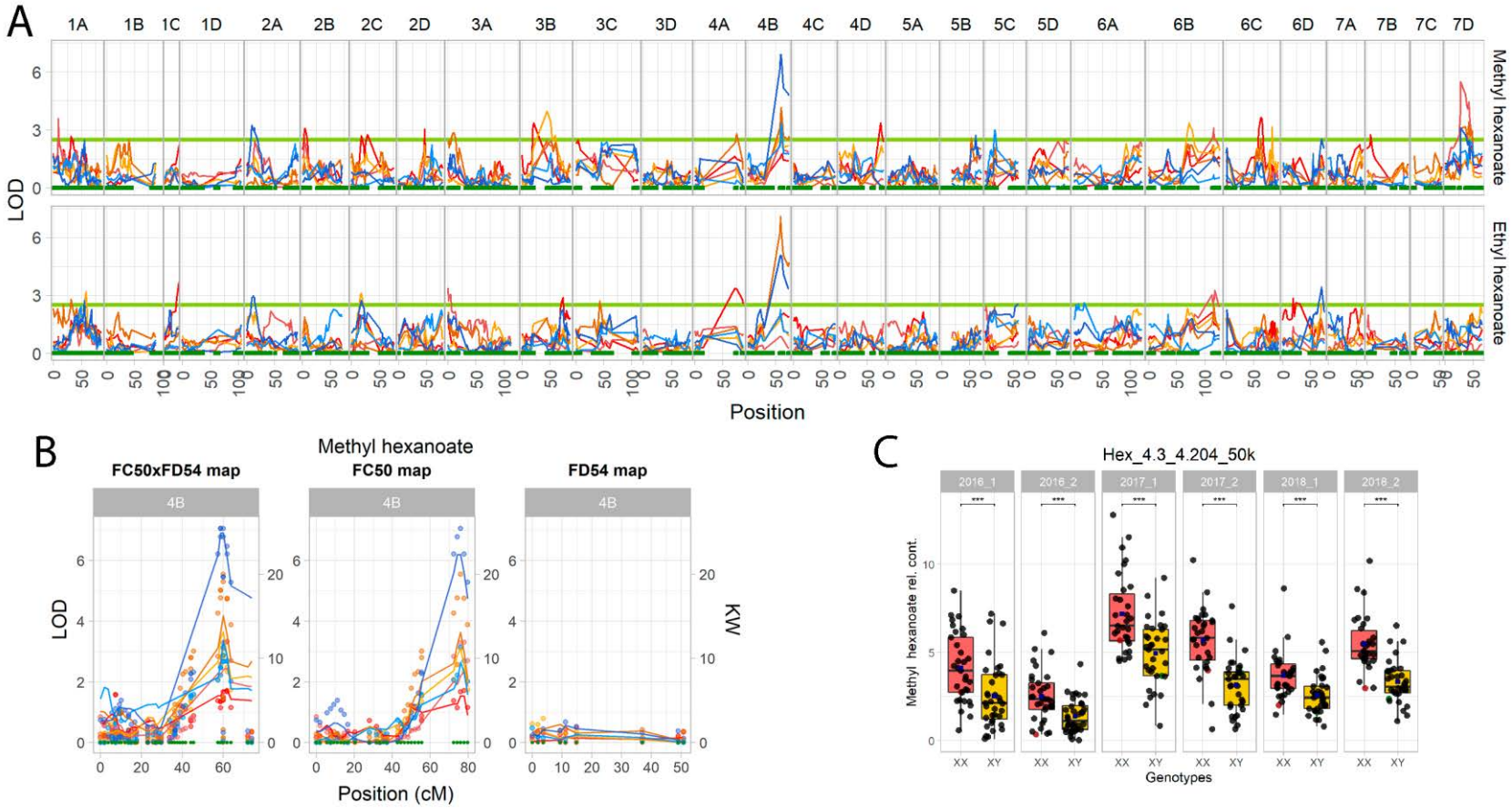


Figure 2-10. *Hexanoate_4B* QTL. A) LOD score for methyl hexanoate and ethyl hexanoate for all LGs from 'FC50xFD54' genetic map. Horizontal green line corresponds to the threshold of 2.5 LOD score. B) Methyl hexanoate LOD score (lines) and KW statistic test (dots) for FC50xFD54, FD54 and FC50 genetic map. Harvest: 2016 (reddish), 2017 (orangish) and 2018 (bluish). C) Boxplot of the HRM marker genotypes with the methyl hexanoate relative content. Dots: 'FC50xFD54' progenies (black), 'FC50' (green) and 'FD54' (red). Significant level <0.001.

Validation of QTLs for KVCs in '21AF' population

To validate the QTLs for KVCs, we analysed full ripe fruits by GC-MS in three harvest (2014, 2015 and 2018) of the '21AF' population (Supplementary table 2-9). A total of 51 QTLs for 14 KVCs were detected, but only nine of them were major and totally stable QTLs. Six of these QTLs are for ester compounds, one for lactones and two for terpenes (Table 2-3 and Supplementary table 2-10).

Concerning the ester QTLs, methyl butanoate and methyl hexanoate QTLs were mapped in the same region in LG1A (*MB_1A* and *MH_1A*) and LG7D (*MB_7D* and *MH_7D*), whereas two other ester compounds, butyl acetate and hexyl acetate, were located in LG6A (*BA_6A* and *HA_6A*). For lactones, the γ-decalactone QTL was mapped at the end of the LG3D, and for terpenes, linalool and nerolidol QTLs were located at the beginning of LG3B (Table 2-3).

Table 2-3. QTLs for KVCs detected in '21AF' population. List of major QTLs showing maximum LOD score, % of explanation, Kruskal-Wallis test, 1-LOD confidence interval, QTL size, QTL position in *F x ananassa* genome and number of genes in each region.

QTL ID	Harv.	LOD	Expl.	KW	Genetic map			*F. x ananassa* genome			
					LG	QTL int. (cM)	Size (cM)	CHR	QTL int. (bp)		Genes
MB_1A	3	4.6	17.3	18.3	1A	24.32	31.79	7.46	Fvb1-4	5480538	8482745	481
MH_1A	3	4.4	17.4	13.8	1A	21.25	37.88	16.63	Fvb1-4	5480538	9319444	635
L_3B	3	7.6	27.3	47.7	3B	0.00	14.02	14.02	Fvb3-3*	48389	79762	9
N_3B	3	9.6	34.4	49.8	3B	0.00	0.86	0.86	Fvb3-3*	48389	79762	9
GD_3D	3	7.6	30.0	58.7	3D	72.76	79.29	6.53	Fvb3-2	28716822	31158030	368
BA_6A	3	5.1	21.5	42.1	6A	59.39	66.61	7.22	Fvb6-1	13552808	17175832	481
HA_6A	3	5.7	22.1	37.4	6A	60.99	70.18	9.19	Fvb6-1	14866830	16606205	237
MB_7D	3	9.1	32.1	66.1	7D	21.18	26.40	5.22	Fvb7-3	8386863	8981806	93
MH_7D	3	10.6	37.1	56.4	7D	21.18	26.40	5.22	Fvb7-3	8386863	8981806	93

*SNPs also mapped in Fvb3-1 (30339108-31797408pb and including 260 genes) and Fvb3-2 (1828610-3346348pb and 307 genes).

In order to validate the detected QTLs in the two populations under study, we carried out a synteny analysis and checked if the behaviour of the markers is consistent over different harvests and populations. The terpene QTL in LG3B and two ester QTLs *BA_6A* and *MH_7D*, had been mapped in the same LGs in 'FC50xFD54' and '21AF' populations. To get higher synteny, we have used the saturated genetic maps with maximum segregating SNPs (See Chapter 1 and Annex 1).

Nine different terpenes were studied in the 'FC50xFD54' population and only the KVCs, linalool and nerolidol, were studied in the '21AF'. Therefore, we focused on the *L_3B* mapped at the beginning of LG3B (Fig. 2-11). We observed similar linalool relative contents in both populations. This QTL in the '21AF' genetic map showed a gap providing few markers for the synteny analysis.

Despite of that, good synteny is observed between both genetic maps (Fig. 2-11 and Fig. A1-2). Looking at a common marker inside this QTL, such as the Affx-88832495 marker, we found that individuals owning AA genotype showed significant differences in the accumulation of linalool compared to individuals having AB genotype in the 'FC50xFD54' or BB genotype in '21AF'. In fact, the A allele may increase linalool content in both populations, while the B allele in homozygosity appears to correlate with a very low linalool accumulation (Fig. 2-11).

As we have already mentioned, the *Terpene_3B* QTL correspond to Fvb3-3 (*F. nipponica-like*) from *F. x ananassa* genome. In the 'FC50xFD54' genetic map, the shortest QTL interval size was observed in monoterpene compounds with an interval of 0.9Mbp (1003939-1958161bp) and includes 171 annotated genes. Eighteen of them did not show homology to any database and sixteen encoded an uncharacterised protein. Looking at the gene description of the remaining 137 genes, any of them seemed to be related to the terpene pathway.

Regarding to this QTL in the '21AF' genetic map, few markers were located in the expected Fvb3-3 chromosome. The corresponding fragment is only 43kbp in physical size (45389-79762) and includes nine genes. However, other markers were located in the Fvb3-1 and Fvb3-2 chromosomes. Focusing on the Fvb3-1 chromosome (*F. viridis-like*), the QTL physical size was of 1.4Mbp (30339108-31797408pb) with 260 annotated genes. Of these genes, we highlight the maker-Fvb3-1-augustus-gene-304.55-mRNA-1 gene which is described as *nerolidol synthase 1* (*NES1*). Looking at the Fvb3-2 chromosome (*F. iinumae-like*) which covered 1.5Mbp (1828610-3346348pb), we counted 307 genes in this interval in which we underlined two genes, namely maker-Fvb3-2-augustus-gene-19.39-mRNA-1 and maker-Fvb3-2-augustus-gene-19.40-mRNA-1 described as *nerolidol synthase* (*NES1-like* and *NES2*, respectively). All the annotated genes are listed in the Supplementary table 2-11.

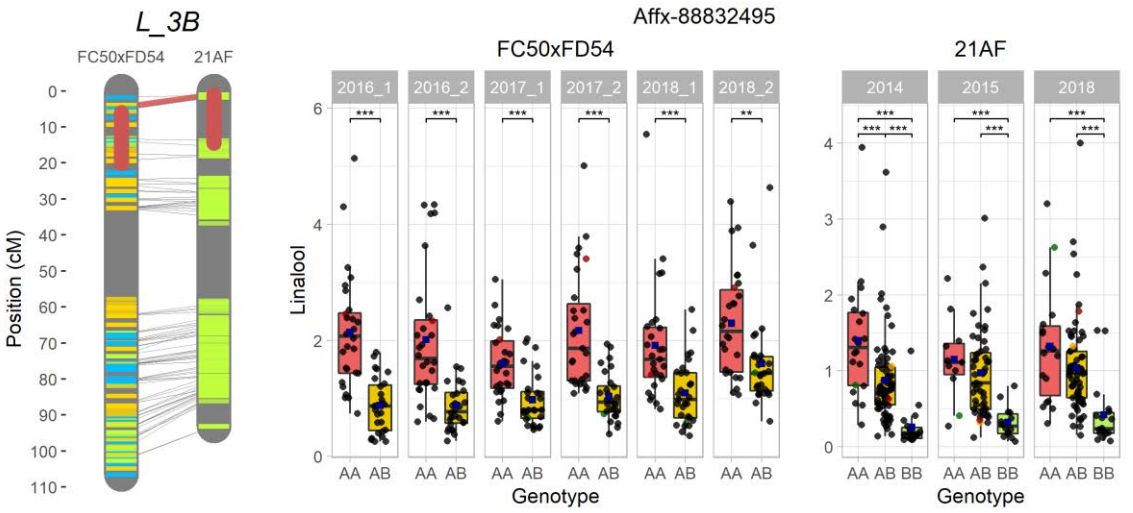


Figure 2-11. *L_3B* QTL synteny and boxplot of the Affx-88832495 marker in the 'FC50xFD54' and '21AF' populations. Left: red boxes cover 1-LOD confidence intervals and red line is the Affx-88832495 position. Marker class represented as codominant (green), 'FC50' segregating (blue) and 'FD54' segregating (yellow). Right: Boxplot with each sample represented as dots and the blue square the average of each group. Significant levels <0.001 (***) and <0.01 (**).

The ester QTLs were spread over different LGs. Focusing on the cluster QTLs in LG6A, *BA_6A* and *HA_6A* detected in the '21AF' genetic map were located in the same region as *BA_6A*, *O3MB_6A, OA_6A, OB_6A* and *OH_6A* from the 'FC50xFD54' genetic map (Fig. 2-12). These QTLs are related to acetate and octyl groups (*Acetate_6A* and *Octyl_6A*). Since hexyl acetate QTLs in 'FC50xFD54' were mapped in LG1A and LG5A, *BA_6A* was the only QTL shared in both populations. Relative butyl acetate content is around ten times higher in '21AF' than in 'FC50xFD54' (Fig. 2-12).

In the case of the Affx-8884155 marker, we observed different segregation behaviour in both populations. Actually, the AA genotype of the '21AF' genetic map corresponds to a higher butyl acetate accumulation (p value<0.001) than the AB and BB genotypes. However, in 'FC50xFD54' genetic map, we observed the opposite behaviour (Fig. 2-12).

The *Acetate_6A* is localised in Fvb6-1 (*F. vesca-like*). The smallest QTL interval was detected for hexyl acetate in the '21AF' genetic map which corresponds to a physical distance of 1.7Mbp (14866830-16606205bp) and comprises 237 annotated genes (Supplementary table 2-11). Thirty-six of these annotated genes are not found in any database and another 25 encode an uncharacterised protein. None of the remaining 176 genes seems to be related to ester accumulation.

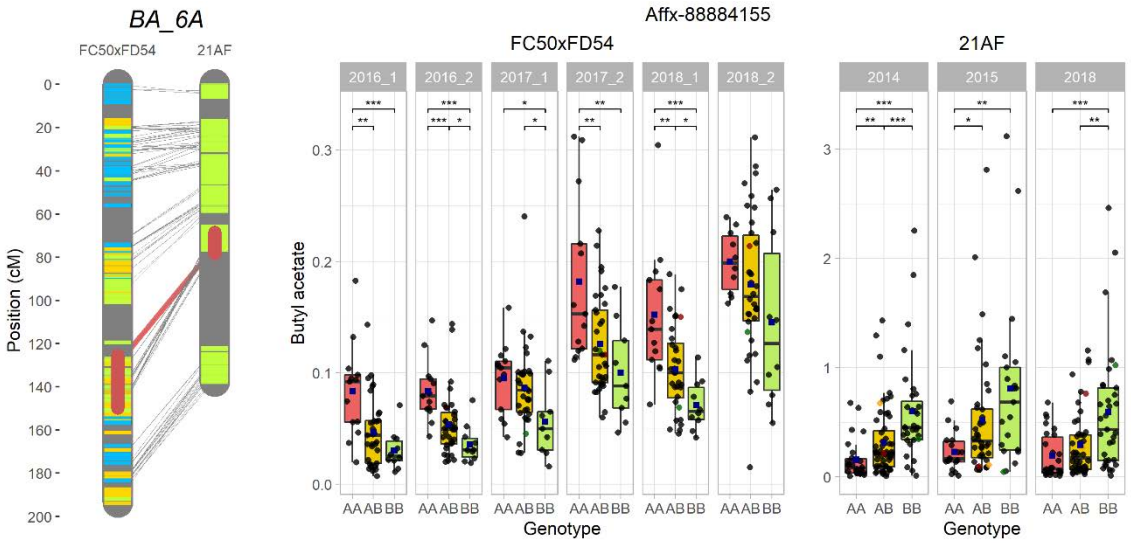


Figure 2-12. *BA_6A* QTL synteny and boxplot of the Affx-8884155 marker in the 'FC50xFD54' and '21AF' populations. Left: red boxes cover 1-LOD confidence intervals and red line is the Affx-8884155 position. Marker class represented as codominant (green), 'FC50' segregating (blue) and 'FD54' segregating (yellow). Right: Boxplot with each sample represented as dots and the blue square the average of each group. Significant levels <0.001 (***), <0.01 (**) and <0.05 (*).

Different ester QTLs were mapped in LG7D such as *MH_7D*, *IPH_7D* and *MO_7D* in the 'FC50xFD54' genetic map and *MH_7D* and *MB_7D* in the '21AF' genetic map. Despite these differences between both maps, we highlight that a methyl group is a common part of all these compounds (*Methyl_7D* QTL). However, the only QTL related to KVC detected in both populations was for methyl hexanoate. Relative methyl hexanoate content was twice to three times higher in the 'FC50xFD54' population than in the '21AF' population. When observing the Affx-88897295 marker, we noticed that the individuals carrying BB genotype showed higher methyl hexanoate content than those having AB genotypes in both populations. In addition, individuals with AA genotype, from '21AF' population, revealed nearly undetectable methyl hexanoate content (Fig. 2-13). Therefore, the B allele of this marker seems to be responsible for the accumulation of methyl hexanoate.

The LG7D in our genetic maps corresponds to the Fvb7-3 chromosome which is *F. iinumae-like* subgenome. The smallest QTL size for the *Methyl_7D* was 0.6Mbp (8386863-8981806bp) and comprises 93 genes (Supplementary table 2-11). Looking at the homology of these genes, we underline the maker-Fvb7-3-snap-gene-89.68-mRNA-1 gene, which is described as an *alcohol acyltransferase* (*AAT*) and might be a candidate gene.

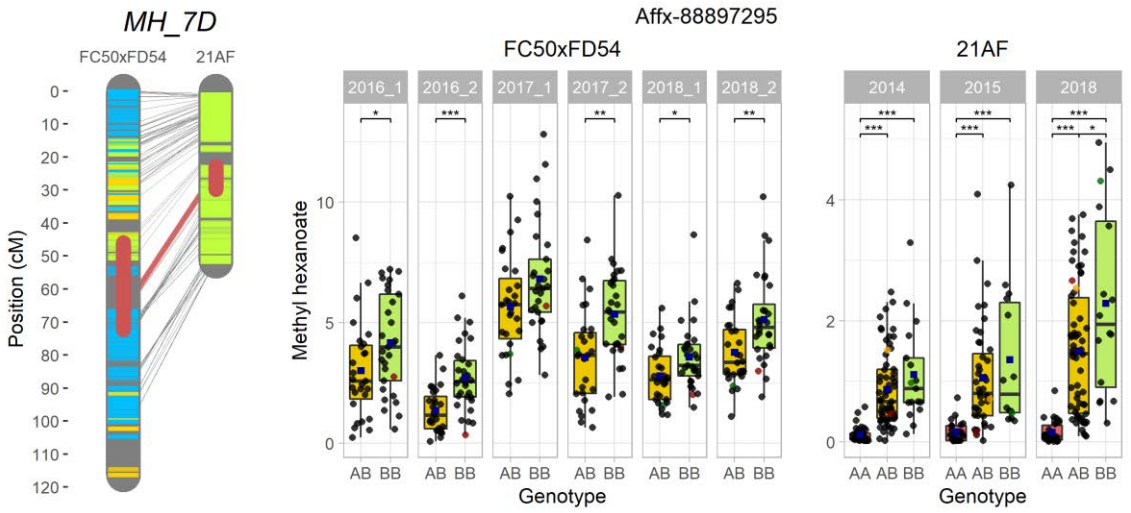


Figure 2-13. *MH_7D* QTL synteny and boxplot of the Affx-88897295 marker in the 'FC50xFD54' and '21AF' populations. Left: red boxes cover 1-LOD confidence intervals and red line is the Affx-88897295 position. Marker class represented as codominant (green), 'FC50' segregating (blue) and 'FD54' segregating (yellow). Right: Boxplot with each sample represented as dots and the blue square the average of each group. Significant levels <0.001 (***), <0.01 (**) and <0.05 (*).

In addition to the KVC QTLs mentioned before, we identified other major QTLs which may be interesting for breeding. However, these QTLs were significant in only one of the studied populations, such as *GD_3D* and *MH_1A* detected in the '21AF' genetic map, and *Hexanoate_4B*, (Fig. 2-10) and *MA_7A* discovered in the 'FC50xFD54' genetic map.

The γ-decalactone is known to provide strawberries with a peach-like aroma. The major and highly stable QTL in '21AF' genetic map was located in LG3D. Analysing the progeny segregation in the Affx-88845940 marker and comparing with the γ-decalactone content, it revealed that the BB genotype correlates with a very low content of γ-decalactone contrasting with the little differences observed between genotypes AA and AB which are related to high content of γ-decalactone. For that reason, we were not able to detect this QTL in the 'FC50xFD54' population. This means that A allele is responsible for the γ-decalactone accumulation (Fig. 2-14).

This QTL is located at the end of the Fvb3-2 chromosome (*F. iinumae-like*) spanning a physical size of 2.4Mbp (28716822-31158030). A total of 368 genes are annotated in this region (Supplementary table 2-11) and the homology analysis of these genes reveals 49 of them without any homology in database and another 23 genes encoding an uncharacterised protein. A priori, any gene seems to be involved to γ-decalactone pathway.

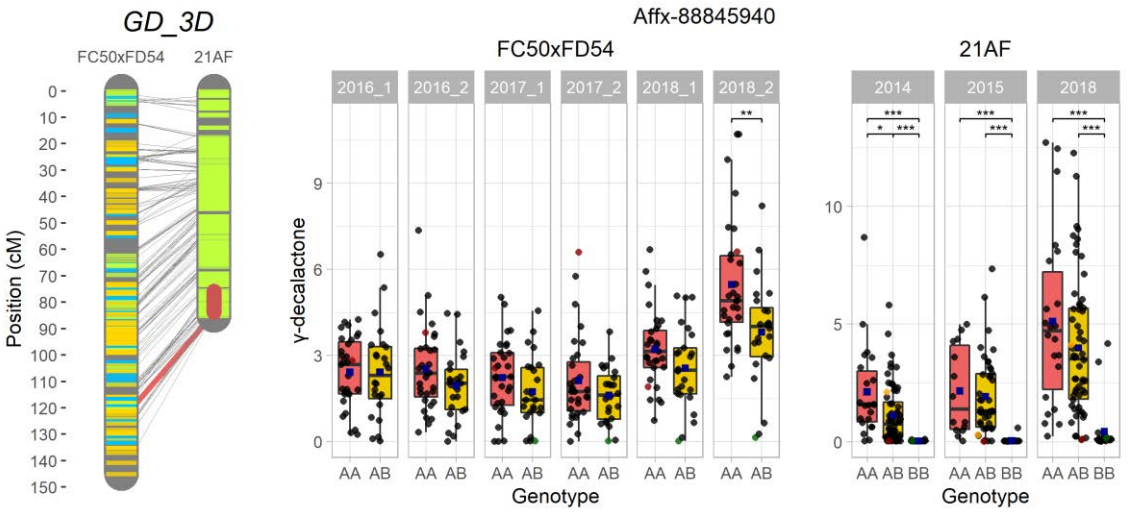


Figure 2-14. *GD_3D* QTL synteny and boxplot of the Affx-88845940 marker for 'FC50xFD54' and '21AF' population. Left: red boxes cover 1-LOD confidence intervals and red line is the Affx-88845940 position. Marker class represented as codominant (green), 'FC50' segregating (blue) and 'FD54' segregating (yellow). Right: Boxplot with each sample represented as dots and the blue square the average of each group. Significant levels <0.001 (***), <0.01 (**) and <0.05 (*).

The other major QTLs detected only in the '21AF' population were *MH_1A* and *MB_1A,* both sharing a methyl group, so we named it *Methyl_1A*. However, we detected different QTLs related to acetate group in the 'FC50xFD54' population not being as major and stable as the previous ones. Focusing on the Affx-88812680 located in the *Methyl_1A* QTL, we noticed that the progenies with the homozygous A allele showed significant higher methyl hexanoate content than the ones with B allele. Since the BB genotype is not present in the 'FC50xFD54' population, no significant differences were observed between 'FC50xFD54' genotypes (Fig. 2-15).

The *Methyl_1A* QTL physical size was of 3Mbp (5480538 – 8482745 bp) and contains 481 annotated genes (Supplementary table 2-11). However, any of these showed homologies with any known gene involved in the ester biosynthesis.

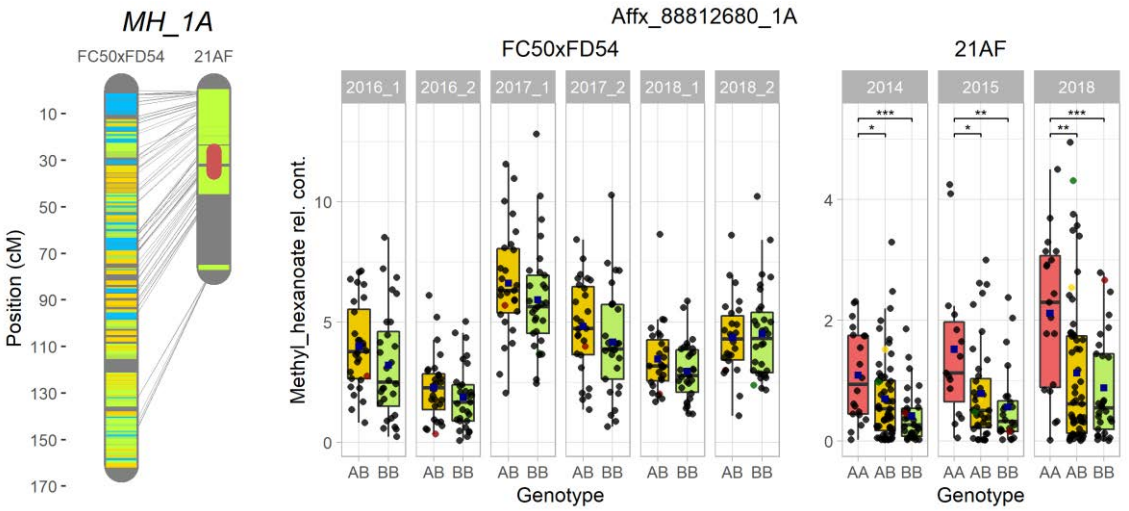


Figure 2-15. *MH_1A* QTL synteny and boxplot of the Affx-88812680 marker for 'FC50xFD54' and '21AF' population. Left: red boxes cover 1-LOD confidence intervals and red line is the Affx-88812680 position. Marker class represented as codominant (green), 'FC50' segregating (blue) and 'FD54' segregating (yellow). Right: Boxplot with each sample represented as dots and the blue square the average of each group. Significant levels <0.001 (***), <0.01 (**) and <0.05 (*).

Methyl anthranilate is an aromatic ester compound providing wild strawberry aroma. Since this compound was not detected in the '21AF' population, we could not validate any methyl anthranilate QTL. Even so, we were able to elucidate which allele is related to high methyl anthranilate content in the 'FC50xFD54' population. The AB genotype of the Affx-8897207 marker located in LG7A shows a significant higher methyl anthranilate content than the AA genotype (p-value <0.001) (Fig. 2-16).

This *MA_7A* QTL is located in the middle of the Fvb7-2 chromosome (*F. vesca-like*) and have a physical size of 16.4Mbp (13707587-30080597). This QTL interval is relatively large, so we found 2767 annotated genes (Supplementary table 2-11). The homology analysis of these genes allowed to identify 326 genes without any match in any database and 200 genes that encode an uncharacterised protein. A priori, any gene is related to amino acid degradation or ester biosynthesis.

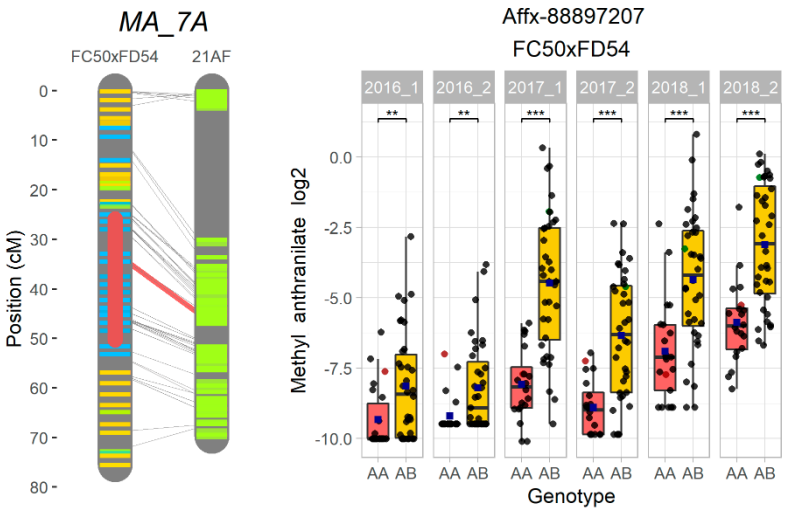


Figure 2-16. *MA_7A* QTL synteny for 'FC50xFD54' and '21AF' population and boxplot of the Affx-88897207 marker. Left: red boxes cover 1-LOD confidence intervals and red line is the Affx-88897207 position. Marker class represented as codominant (green), 'FC50' segregating (blue) and 'FD54' segregating (yellow). Right: Boxplot with each sample represented as dots and the blue square the average of each group. Significant levels <0.001 (***), <0.01 (**) and <0.05 (*).

The *Hexanoates_4B* QTL, discovered in the 'FC50xFD54' genetic map, was located at the end of LG4B which correspond to the beginning of the Fvb4-3 chromosome (*F. vesca-like*). The QTL interval size is 7.4Mbp (4032723-11434813bp) containing a total of 1233 annotated genes (Supplementary table 2-11). Of these genes, 167 are not found in any database and 108 genes does not have a known function. None of the remaining genes seems to be related to ester biosynthesis.

Discussion

Volatile organic compounds (VOCs) are key factors for consumer acceptance. For this reason, breeding programs are demanding new tools for marker assisted selection (MAS) in order to satisfy the market demand. In this study, we analysed the aromatic profile of two populations and revealed the genome regions responsible of VOCs accumulation. Additionally, we suggest some SNP markers for MAS.

The volatilome analysis for 'FC50xFD54' population during six harvests identified 58 VOCs, sixteen of them being KVCs. The total number of VOCs in the 'FC50xFD54' population appears to be lower than previous studies such as the 87 VOCs found in an F1 population (Zorrilla-Fontanesi *et al.*, 2012), the 81 identified in a parental collection (Schwieterman *et al.*, 2014) or the 100 discovered in an *F. vesca* NILs collection (Urrutia *et al.*, 2017). These differences might be caused by metabolite identification and by the differences existing between wild and cultivated strawberry. Actually, in these previous studies, commercial standards were used to detect all the VOCs, thus providing high confidence identification, while we used this method only to identify the KVCs. Similar to our results, most of volatile assays using few accessions lines led to the identification of 50 to 75 compounds (Ulrich *et al.*, 2018). In relation to the proportion of VOCs for each chemical family, we found that more than half of them are esters (57%), followed by terpenes (16%), aldehydes (12%) and lactones (5%), in agreement with previous studies (Zorrilla-Fontanesi *et al.*, 2012, Schwieterman *et al.*, 2014 and Urrutia *et al.*, 2017).

Generally, we observed a high correlation among the compounds of the same chemical family suggesting that these compounds are regulated in the same way. Moreover, some high correlations were observed between compounds of different chemical families. Although these correlated compounds should not share the same biosynthetic pathway, they might have some common transcription factors regulating their activity or some linked genes involved in different volatile pathways. Hence, when we select for one compound, we may take with other related ones. Although the genetic factor was the main contributor to the phenotypic variance, several compounds were also dependent on the environment.

In our analysis, we discovered a high number of significant and stable QTLs (179) for 55 identified VOCs in the 'FC50xFD54' population. Among them, 94 QTLs were highly stable for 43 VOCs. This low QTL stability is similar to previous studies in strawberry (Zorrilla-Fontanesi *et al.*, 2012 and Urrutia *et al.*, 2017). Focusing on QTLs for KVCs, 22 highly stable QTLs were detected for 16 compounds in our 'FC50xFD54' genetic map. However, only nine of the 51 detected QTLs were stable for seven KVCs in the '21AF' genetic map. Taking into account the great variation between

different seasons and populations, we assume that few QTLs will be really important for breeding.

In addition, some QTLs for the same trait are located in HGs as observed in most of the QTL analysis (Zorrilla-Fontanesi *et al.*, 2011, Zorrilla-Fontanesi *et al.*, 2012, Lerceteau-Köhler *et al.*, 2012, Cockerton *et al.*, 2018 and Cockerton *et al.*, 2019). In our population, six VOC QTLs were mapped in HGs being one more stable and significant than the others. This observation reinforces the hypothesis that similar genes located in HGs can contribute to the same trait, but one of them would have a predominant role and the others would modulate the accumulation of VOCs.

Some QTLs have been validated in different genetic backgrounds. The region responsible for the pleasant floral aroma provided by terpene compounds were mapped at the beginning of LG3B, namely *Terpenes_3B*. Since the '21AF' population presented the homozygous allele associated to a low terpene accumulation, its QTL showed a higher significance than the one found in the 'FC50xFD54' QTL analysis which lacks this allele in homozygosity (Fig. 2-11). Although this QTL is localised in the Fvb3-3 chromosome, some of its markers may correspond to the Fvb3-1 and Fvb3-2 chromosome, indicating a high synteny between the homeologous chromosomes. A terpene synthases cluster or a no-specific terpene substrate protein might explain this QTL. Another hypothesis is that some post-transcriptional reactions may alter the volatile terpene accumulation.

Similar to our finding, *F. x ananassa* nerolidol synthase, FaNES1, has been characterised as important for linalool and nerolidol synthesis (Aharoni *et al.*, 2004), mapped in LGIII-4 (Zorrilla-Fontanesi *et al.*, 2012) and annotated in Fvb3-1. Additionally, cultivated and wild octoploid strawberries having this functional protein produced higher linalool content (Chambers *et al.*, 2012). Other nerolidol synthases, namely FaNES1-like and FaNES2, are annotated in Fvb3-2 chromosome. These three genes show the same synteny position than our QTL located in Fvb3-3. Although a non-functional allele is described in *F. vesca* (Aharoni *et al.*, 2004 and Chambers *et al.*, 2012), terpene QTLs are also mapped in Fvb3 (Urrutia *et al.*, 2017) and a cluster of seven *FvNES1* genes can be found (Li *et al.*, 2019).

Since these results are based on the first version of the cultivated strawberry genome presenting high synteny with the homeologous chromosomes, mismatches may be likely to occur during assembling and therefore an homeologous *FaNES1* gene could also be located in the Fvb3-3. Another plausible explanation is that we are dealing with a new gene involved in these compounds' accumulation.

In cultivated strawberry, ester compounds are the most abundant family providing a wide range of fruity aroma. Several ester QTLs were found in different LGs such as LG1A, LG4B, LG6A and LG7D. According to our study, different ester compounds were mapped in the same QTLs depending on the mapping populations. This is the case of the QTLs from the LG1A, where methyl butanoate and methyl hexanoate mapped in the '21AF' genetic map while hexyl acetate, hexyl butanoate and octyl butanoate mapped in the 'FC50xFD54' genetic map. These '21AF' QTLs might be related to *Methyl_1A,* whereas the 'FC50xFD54' QTLs may be related to *Hexyl_1A*. In the literature, different ester QTLs, one of them for methyl hexanoate, were also mapped in LGI-1 (Zorrilla-Fontanesi *et al.*, 2012). Another example is the QTLs in LG6A mapped for different compounds in different populations, such as octyl group compounds, *Octyl_6A*, discovered in 'FC50xFD54' population and acetate related compounds, *Acetate_6A*, in '21AF'. In a previous study, the authors also found different compound QTLs without a direct relationship in one population (Zorrilla-Fontanesi *et al.*, 2012). Since we mapped the *BA_6A* in both populations, we expected to observed the same genotypic behaviour, however it was opposite (Fig. 2-12). We mapped another cluster QTLs in the LG4B only in the 'FC50xFD54' population, in this case related to hexanoate compounds namely *Hexanoate_4B*. Furthermore, our *MH_4B, MO_4B* and *MD_4B* QTLs might validate the ones mapped in LGIV-1 from the '1392x232' population (Zorrilla-Fontanesi *et al.*, 2012).

We localised a QTL for methyl hexanoate, isopropyl hexanoate and methyl octanoate in the LG7D. Since these three compounds share some methyl group, we called it *Methyl_7D*. The *MH_7D* presented a higher significance in the '21AF' genetic map compared to the 'FC50xFD54' genetic map. This difference in significance is due to the presence of homozygous allele linked to the low accumulation of methyl hexanoate in '21AF' and not in 'FC50xFD54' (Fig. 2-13).

The last step for ester biosynthesis is catalysed by *F. x ananassa alcohol acyl transferase* (*FaAAT*). This enzyme can use a wide variety of substrates (Aharoni *et al.*, 2000). Seven *AAT* genes were annotated in *F. x ananassa* genome, located in Fvb1-1 (2), Fvb7-1 (1), Fvb7-2 (3) and Fvb7-3 (1). These genes are mostly conserved in *Fragaria* genus, such as *FcAAT1* in *F. chiloensis* (Gonzalez, *et al.* 2009) and four *AAT* genes in diploid *F. vesca* genome annotated in LG6 (1) and LG7 (3) being one FvAAT enzyme already characterised (Beekwilder *et al.*, 2004). Trying to figure out a candidate gene for the *Methyl_7D* QTL*,* we underline the maker-Fvb7-3-snap-gene-89.68-mRNA-1 gene encoding an AAT.

Methyl anthranilate derived from the amino acid degradation is described as the wild strawberry aroma which is appreciated by consumers and, consequently, aroma breeding programs are interested. In the 'FC50xFD54' population, we mapped several methyl anthranilate QTLs in

different LGs, such as in LG1A, LG4D, three LGs of the HG5 and LG7A. Since methyl anthranilate is only present in a few cultivars (Olbricht *et al.*, 2008 and Schwieterman *et al.*, 2014), this explains that we were not able to detect it in the '21AF' population. The *anthranilate acid methyl transferase*, *FaAAMT*, which was mapped in the *F. vesca* chromosome 4 (Pillet *et al.*, 2017) may be responsible of our *MA_4D*. The HG5 and LG7A QTLs are probably related to the methyl anthranilate QTLs found in LG5 and LG7 in *F. vesca* (Urrutia *et al.*, 2017). We highlight that *MA_7A* was totally stable and a major QTL, and therefore appropriate for MAS.

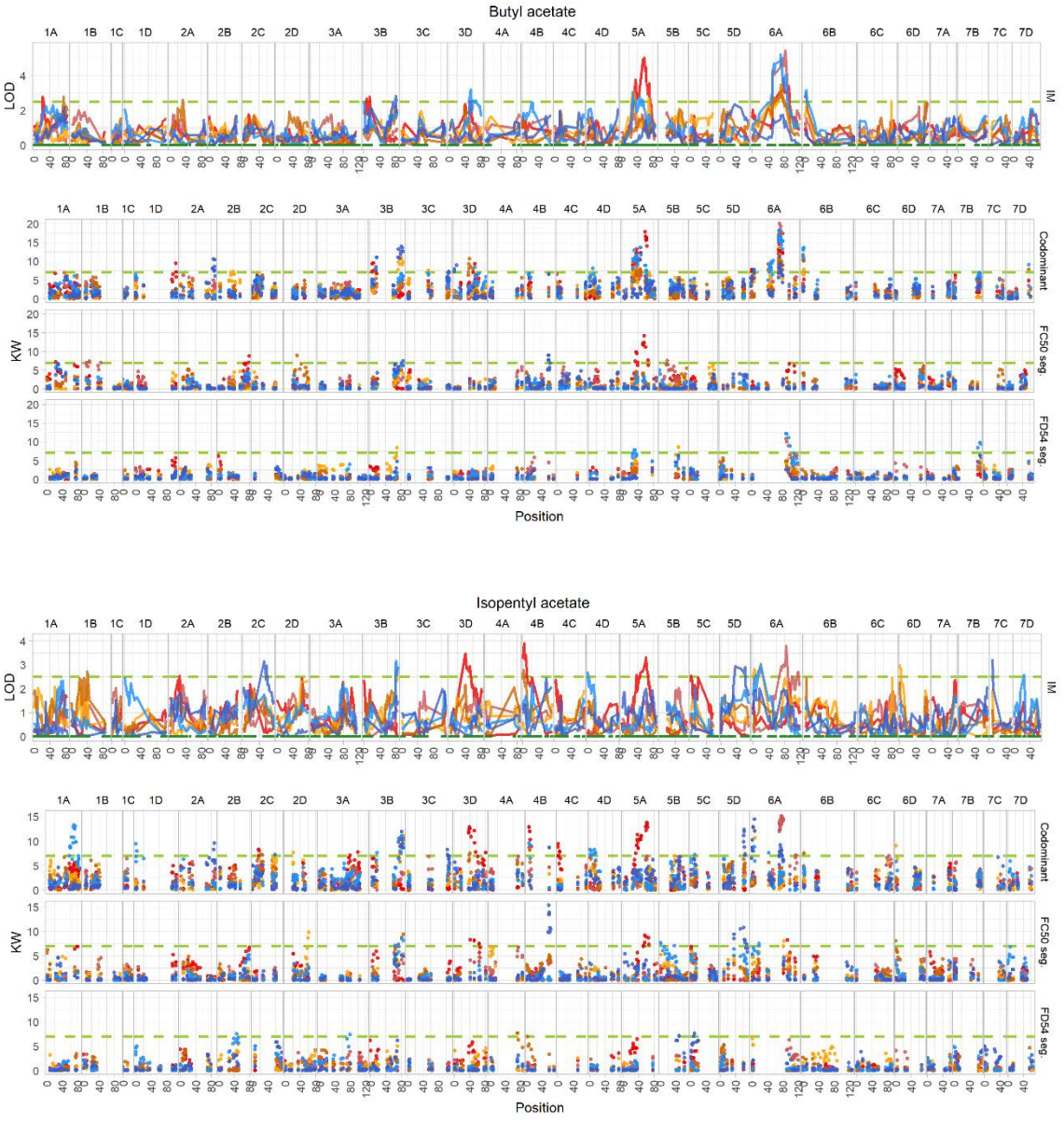
The peach aroma is attributed to γ-decalactone. A total of six QTLs for lactones compounds were detected in the two studied populations in LG2A, LG3D, LG4B, LG6B and LG6C. The major *GD_3D* QTL was only detected in '21AF' population. Although the allele linked to high γ-decalactone accumulation was present in both populations, only the '21AF' population exhibited the low γ-decalactone accumulation genotype (Fig. 2-14). This QTL is likely related to the *omega fatty acid desaturase* gene, *FaFAD1*, involved in the γ-decalatone synthesis and mapped in LGIII-2 (Zorrilla-Fontanesi *et al.*, 2012, Chambers *et al.*, 2014 and Sanchez-Sevilla *et al.*, 2014).

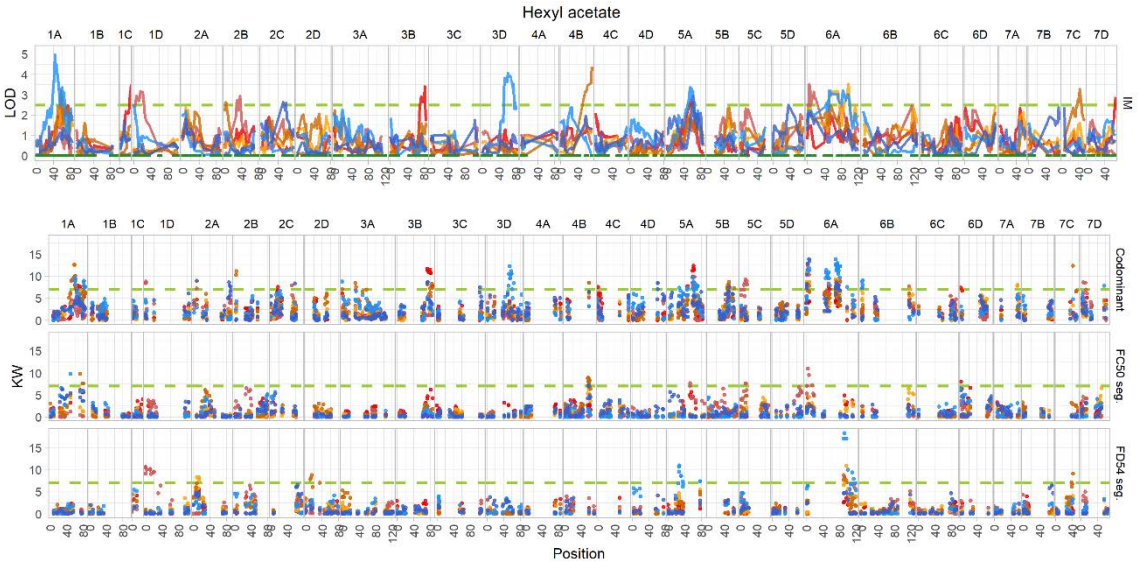
Additionally, the methyl butanoate QTL identified in our QTL analysis located in LG2C correspond, probably, to the FaP1D7 marker and the *GDo_6C* to the FaP1A7 marker identified in Gor *et al.* (2017).

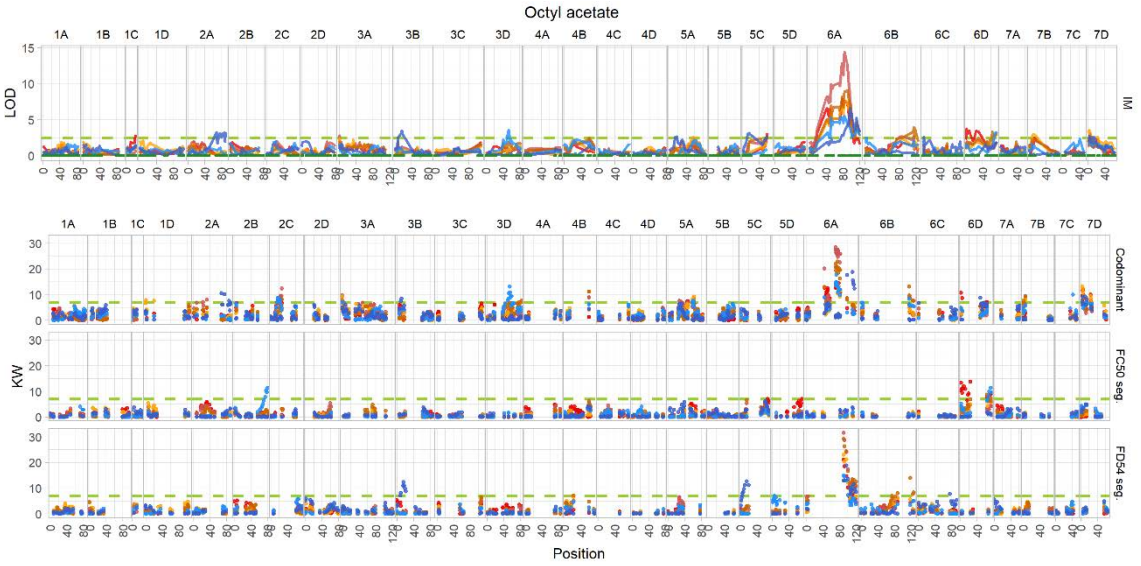
The 'FC50xFD54' volatilome revealed that this population is suitable for the discovery of volatile QTLs. Some of these QTLs are related to liking compounds (Schwieterman *et al.*, 2014 and Ulrich and Olbricht, 2016) and validate QTLs from previous studies. However, much more effort is needed to narrow down these regions in order to find a candidate gene that allows a better knowledge of how these compounds accumulate.

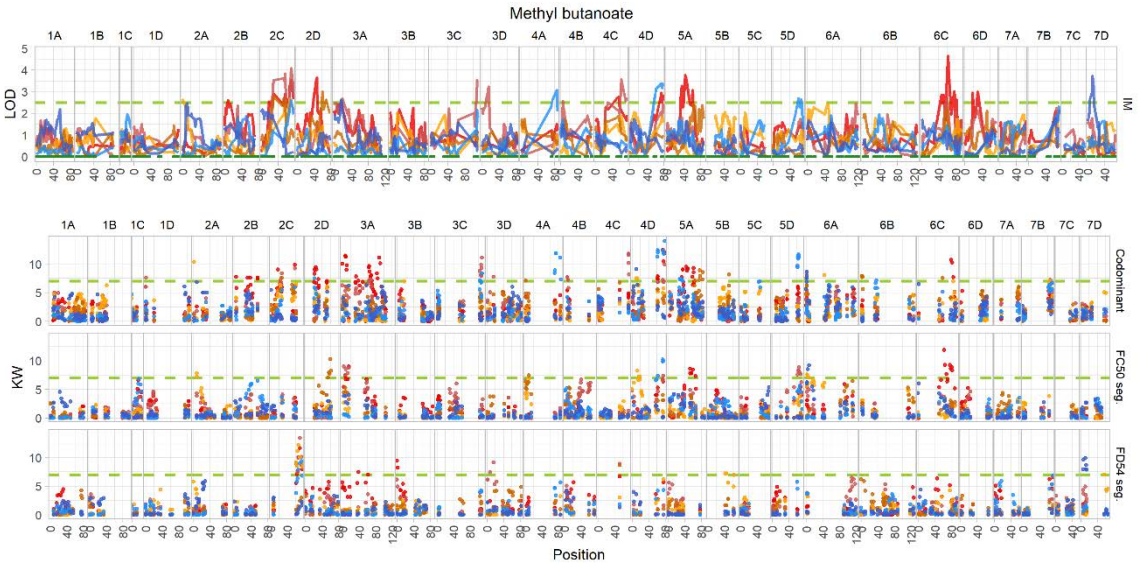
Supplementary figures

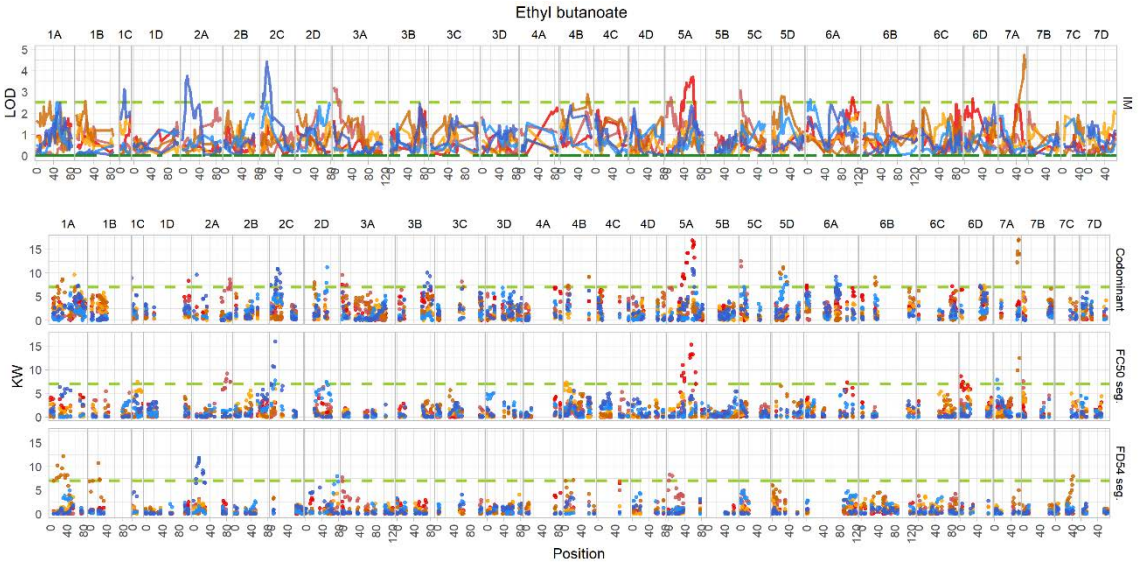
Suppl. figure 2-1. Graphical representation of whole genetic map QTLs for each VOCs in 'FC50xFD54' population. Upper: Interval mapping analysis. Lower: Kruskal-Wallis analysis separating for segregation: codominant, 'FC50' segregation and 'FD54' segregation markers. Harvest: 2016 (reddish), 2017 (orangish) and 2018 (bluish). Horizontal green line corresponds to the threshold of 2.5 LOD score

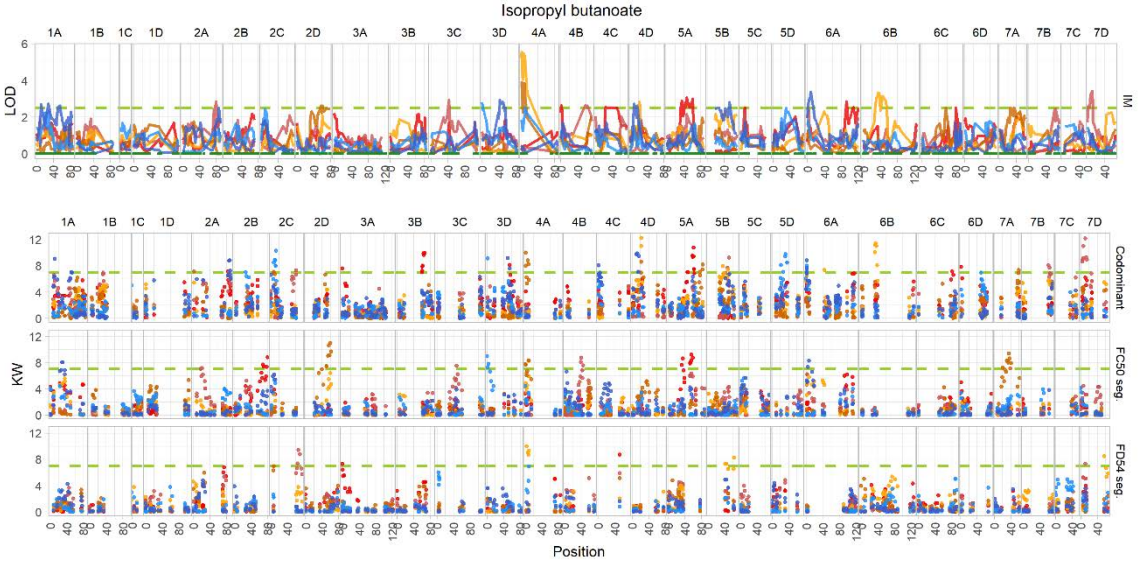


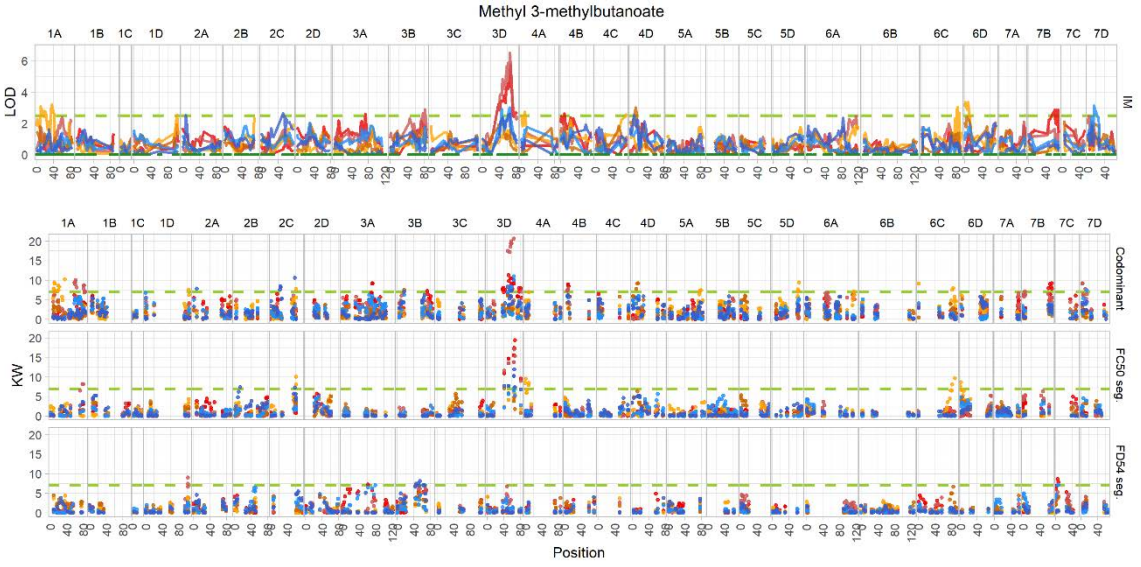

Hexyl acetate
LOD
IM
KW
Codominant
FC50 seg.
FD54 seg.
Position

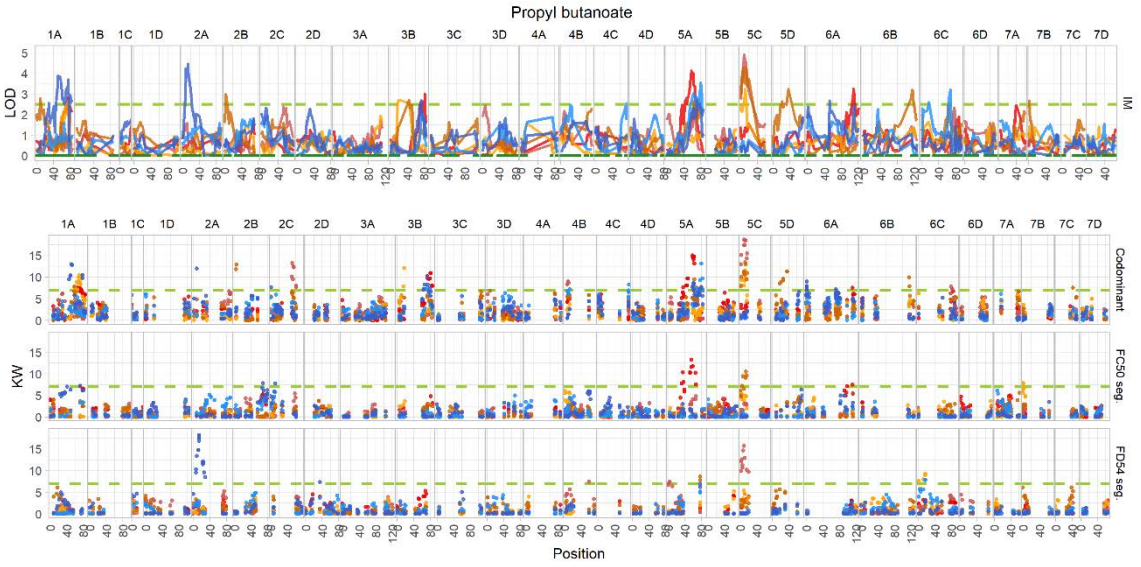

Octyl acetate
LOD
IM
KW
Codominant
FC50 seg.
FD54 seg.
Position

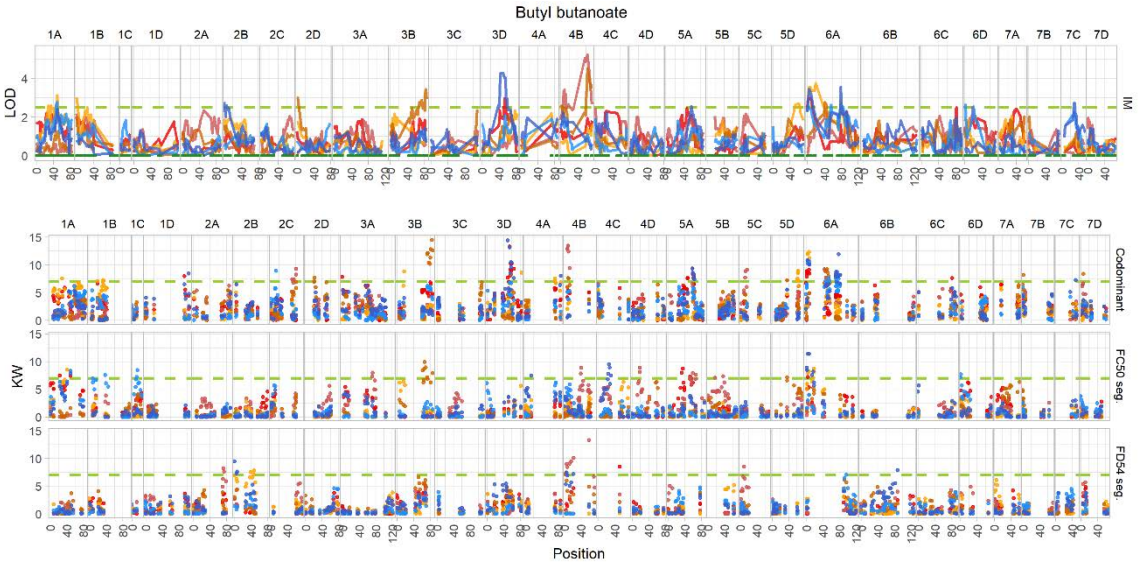

Methyl butanoate
LOD
IM
KW
Codominant
FC50 seg.
FD54 seg.
Position

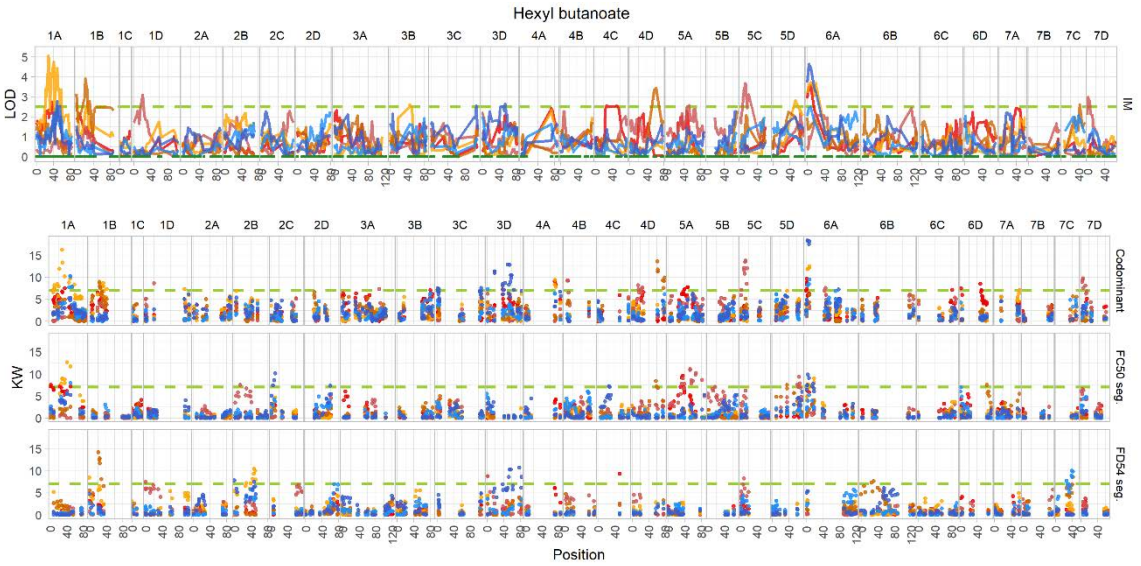

Ethyl butanoate
LOD
IM
KW
Codominant
FC50 seg.
FD54 seg.
Position

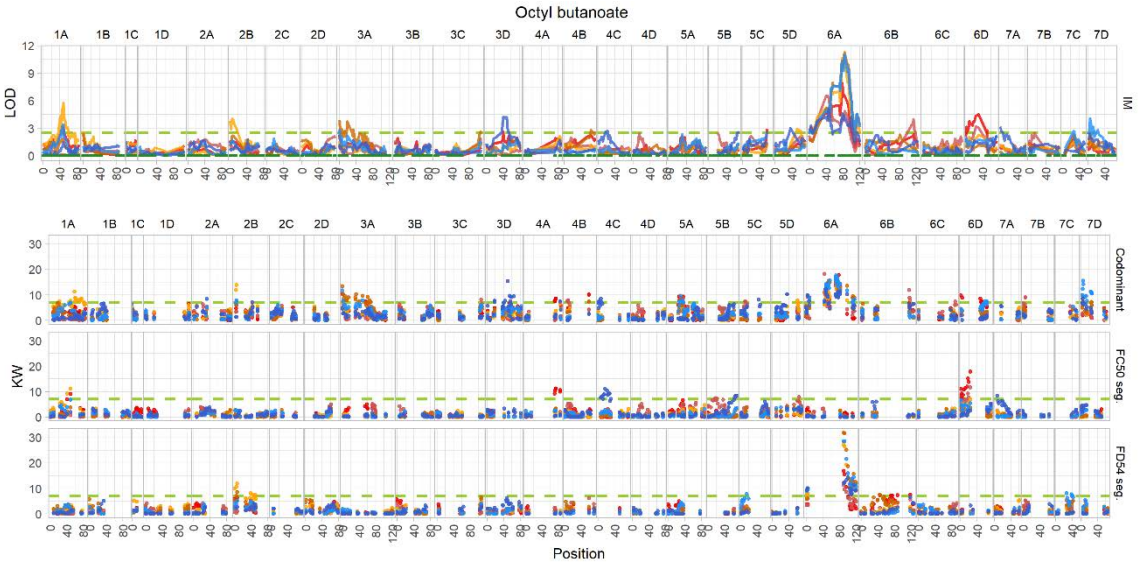

Isopropyl butanoate
LOD
IM
KW
Codominant
FC50 seg.
FD54 seg.
Position

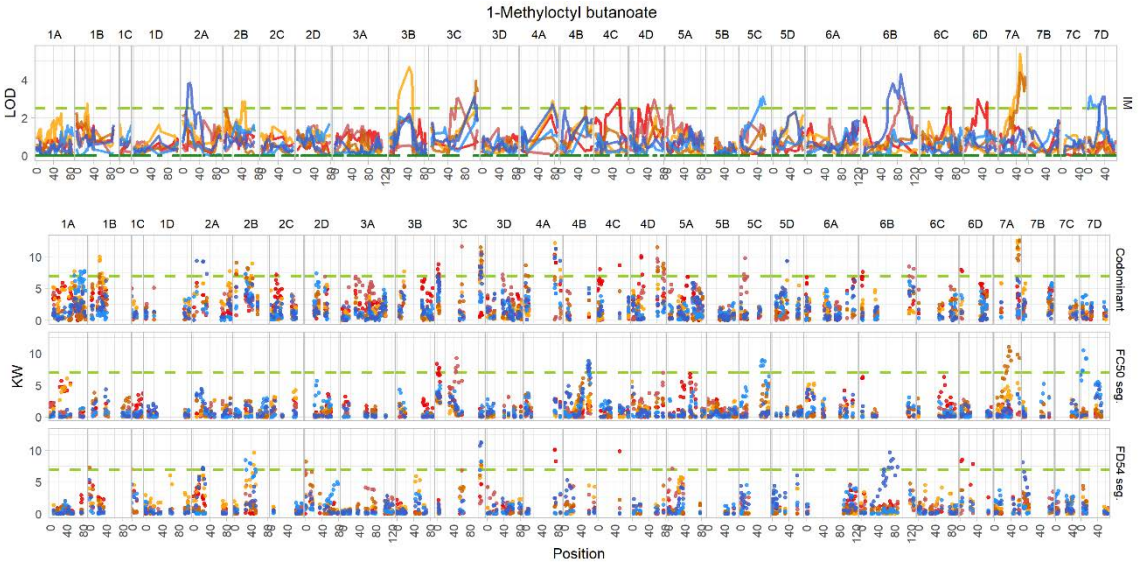

Methyl 3-methylbutanoate
LOD
IM
KW
Codominant
FC50 seg.
FD54 seg.
Position

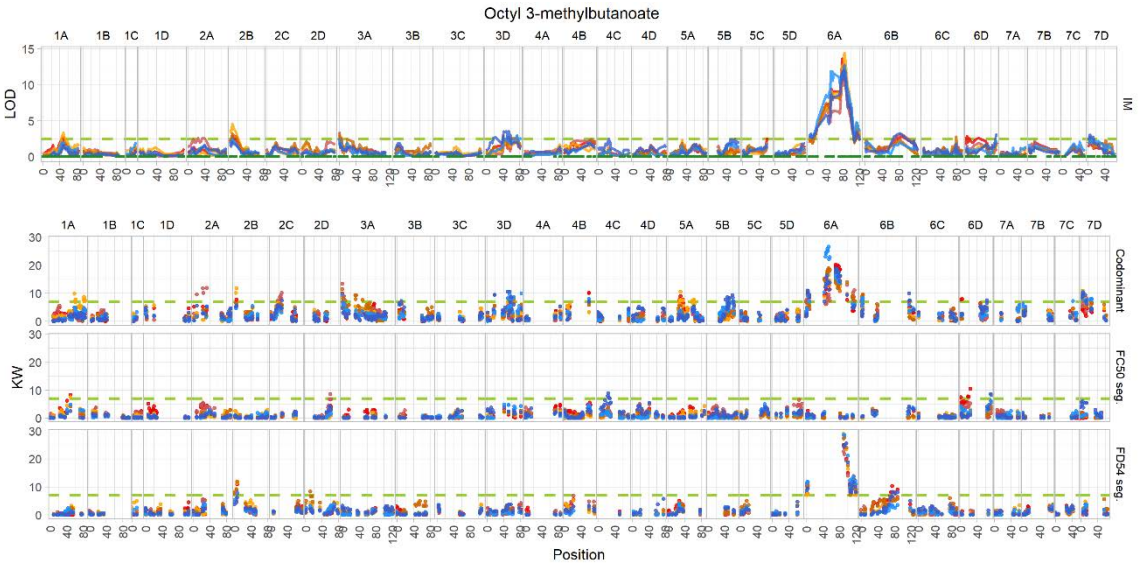

Propyl butanoate
LOD
KW
IM
Codominant
FC50 seg.
FD54 seg.
Position

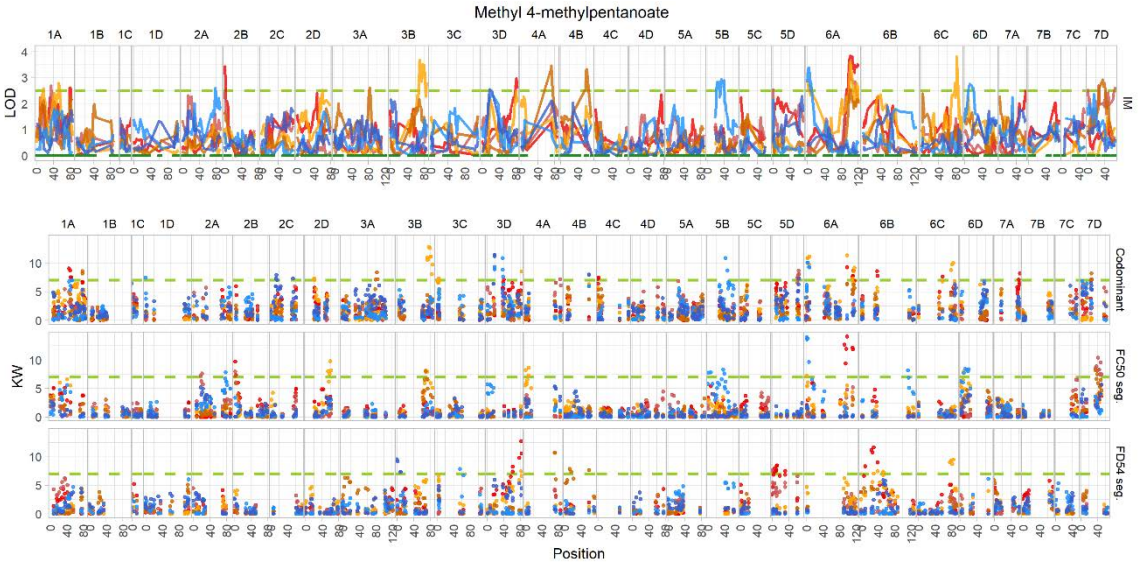

Butyl butanoate
LOD
KW
IM
Codominant
FC50 seg.
FD54 seg.
Position

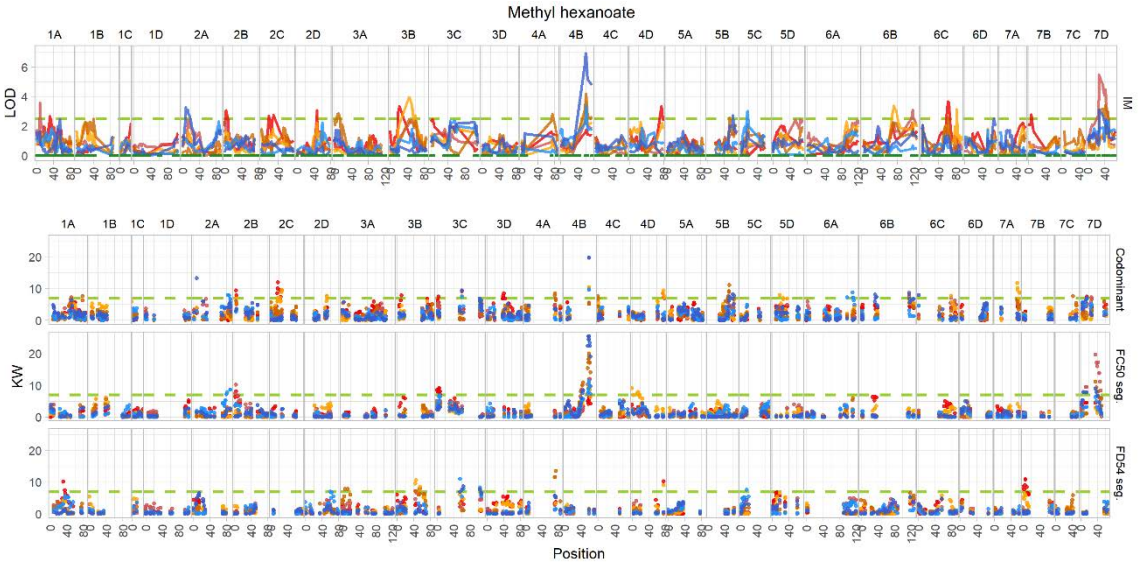

Hexyl butanoate
LOD
KW
IM
Codominant
FC50 seg.
FD54 seg.
Position

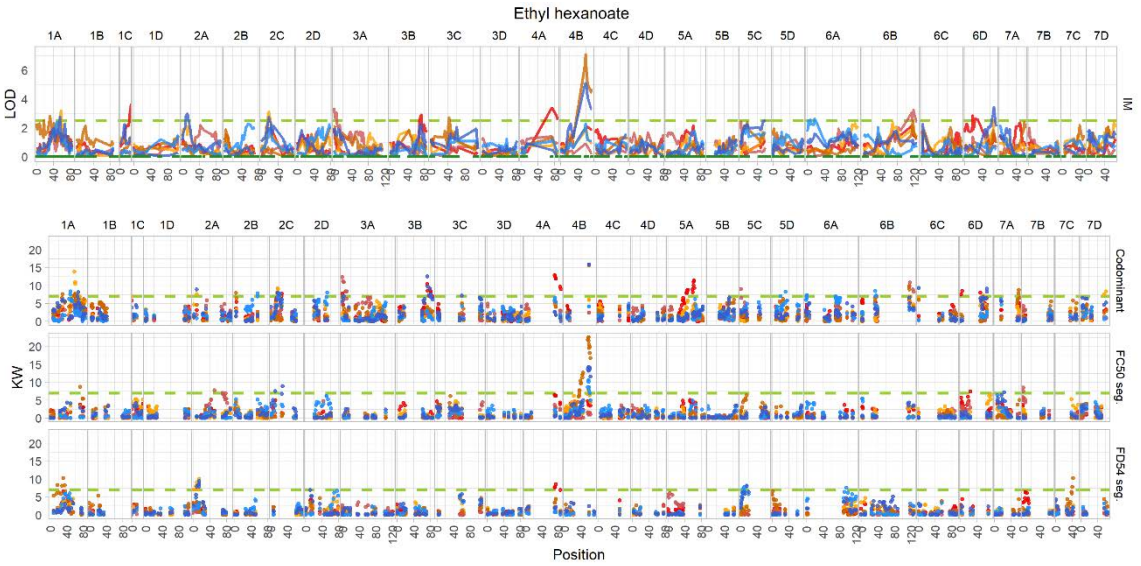

Octyl butanoate
LOD
IM
KW
Codominant
FC50 seg.
FD54 seg.
Position

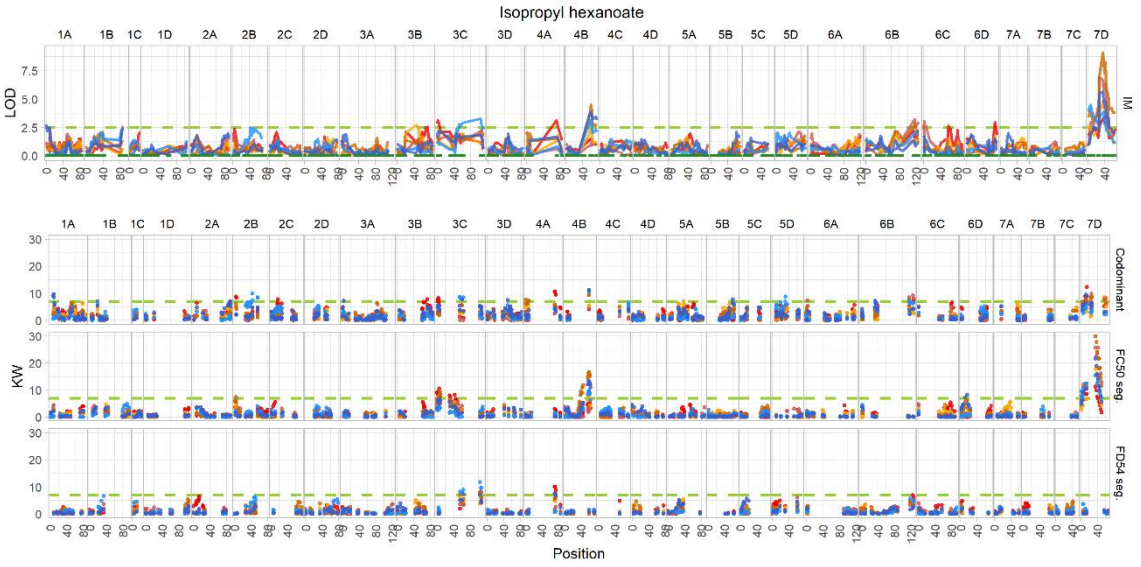

1-Methyloctyl butanoate
LOD
IM
KW
Codominant
FC50 seg.
FD54 seg.
Position

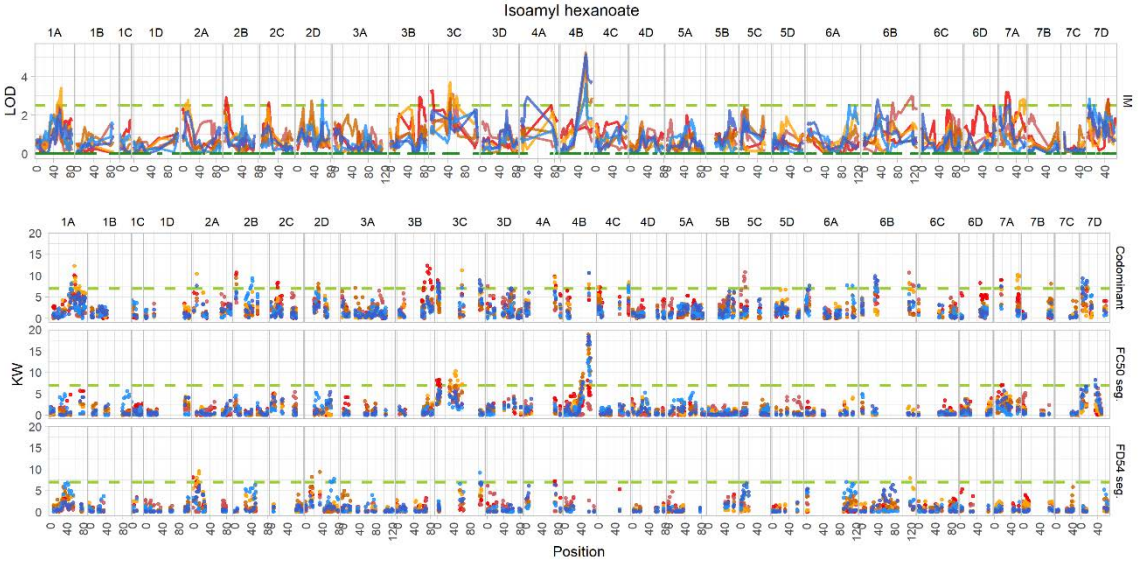

Octyl 3-methylbutanoate
LOD
IM
KW
Codominant
FC50 seg.
FD54 seg.
Position

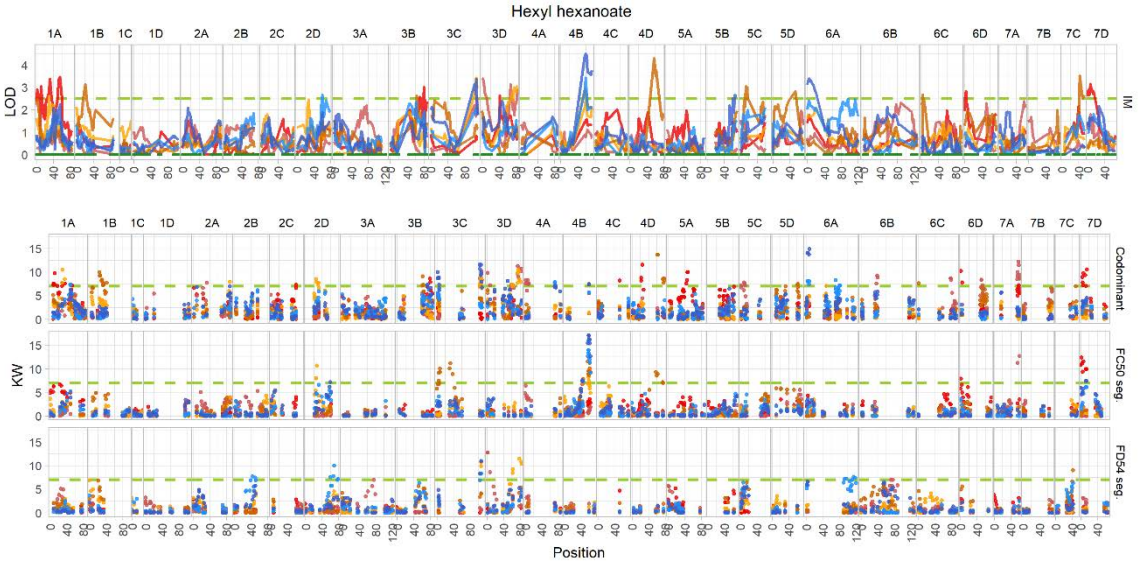

Methyl 4-methylpentanoate
LOD
IM
KW
Codominant
FC50 seg.
FD54 seg.
Position

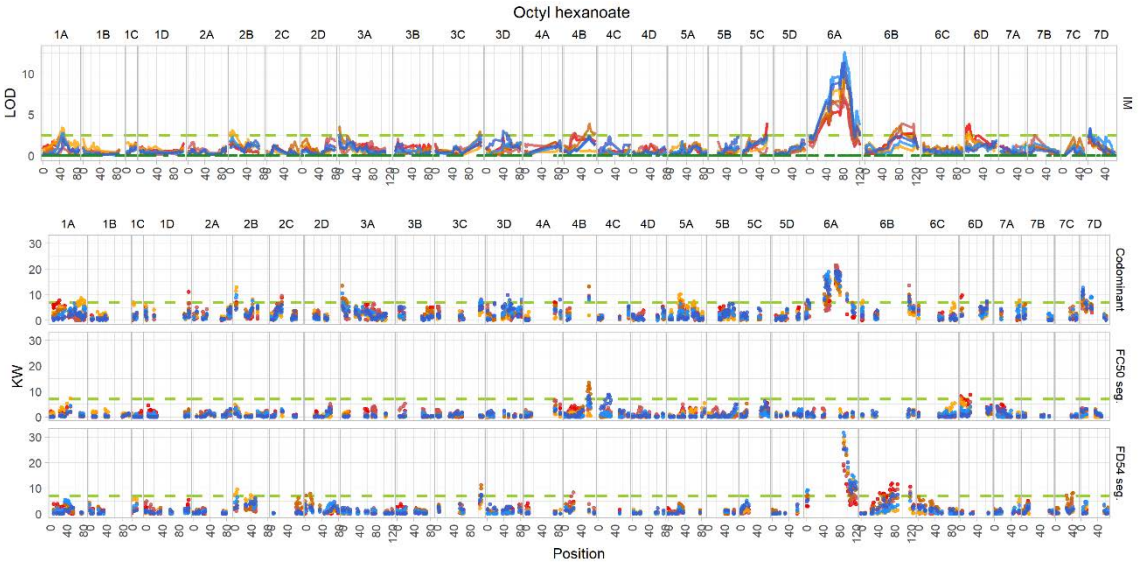

Methyl hexanoate
LOD
IM
KW
Codominant
FC50 seg.
FD54 seg.
Position

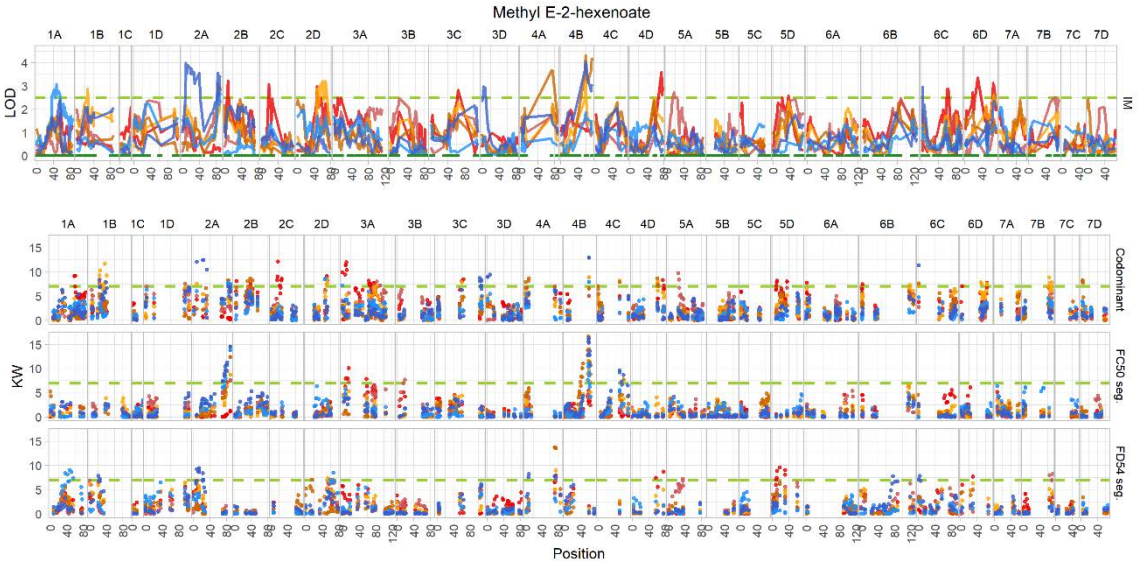

Ethyl hexanoate
LOD
IM
KW
Codominant
FC50 seg.
FD54 seg.
Position

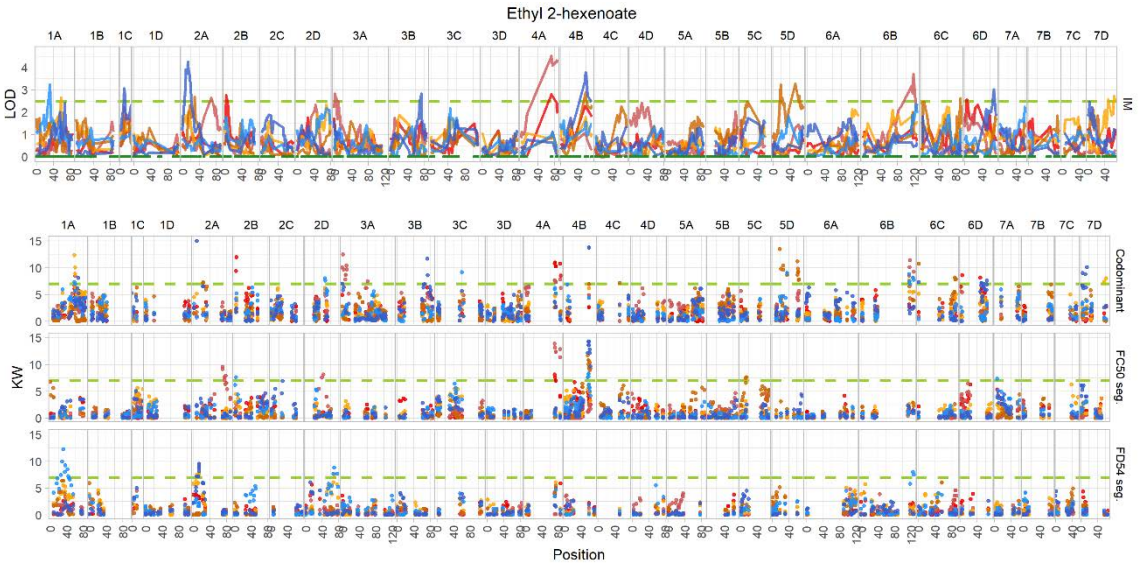

Isopropyl hexanoate
LOD
IM
KW
Codominant
FC50 seg.
FD54 seg.
Position

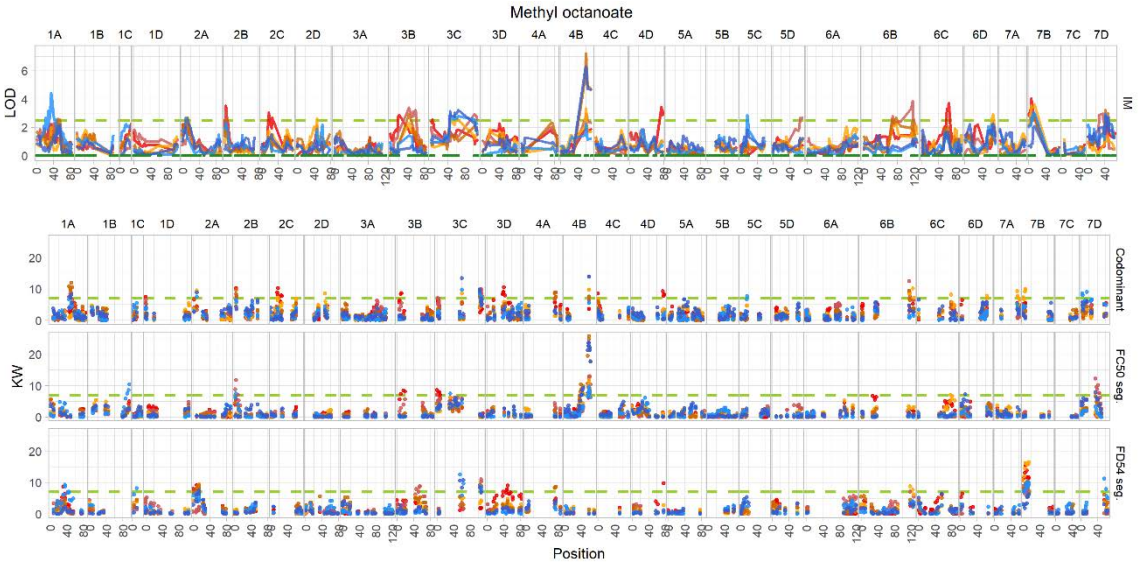

Isoamyl hexanoate
LOD
IM
KW
Codominant
FC50 seg.
FD54 seg.
Position

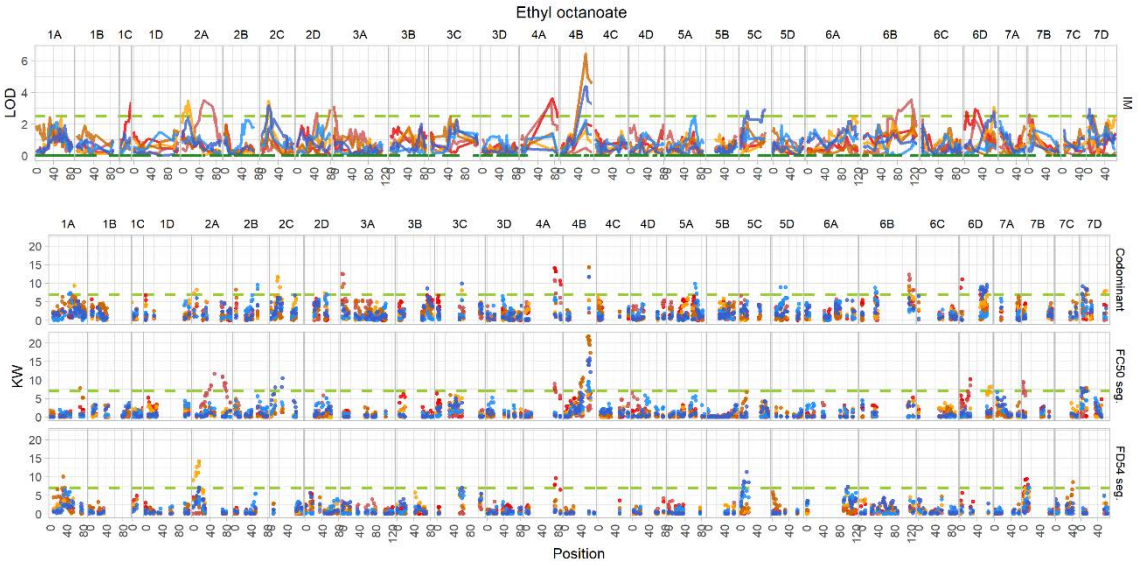

Hexyl hexanoate
LOD
IM
KW
Codominant
FC50 seg.
FD54 seg.
Position

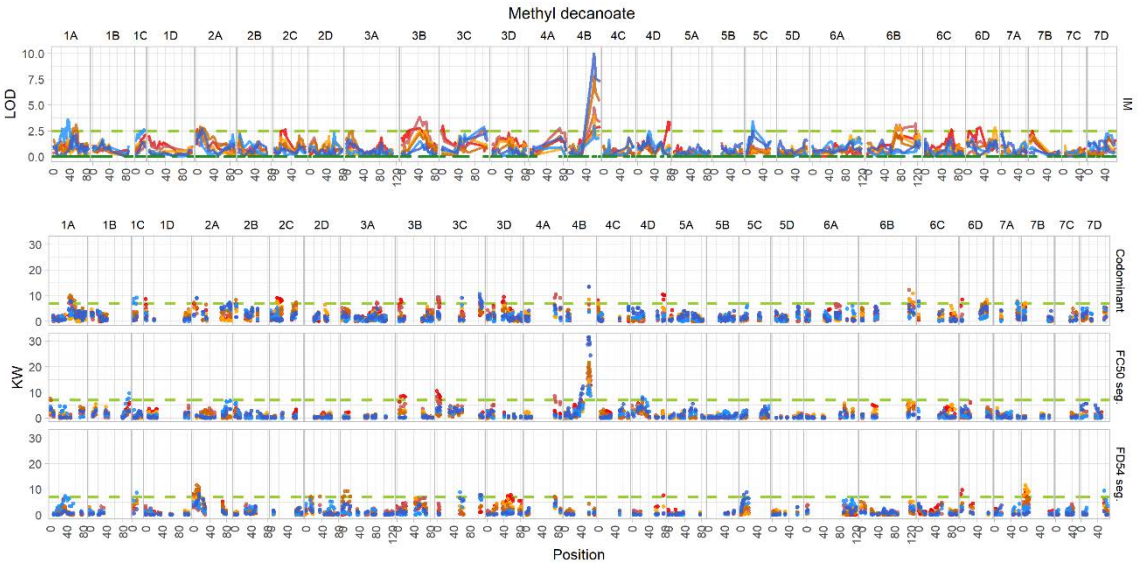

Octyl hexanoate
LOD
IM
KW
Codominant
FC50 seg.
FD54 seg.
Position

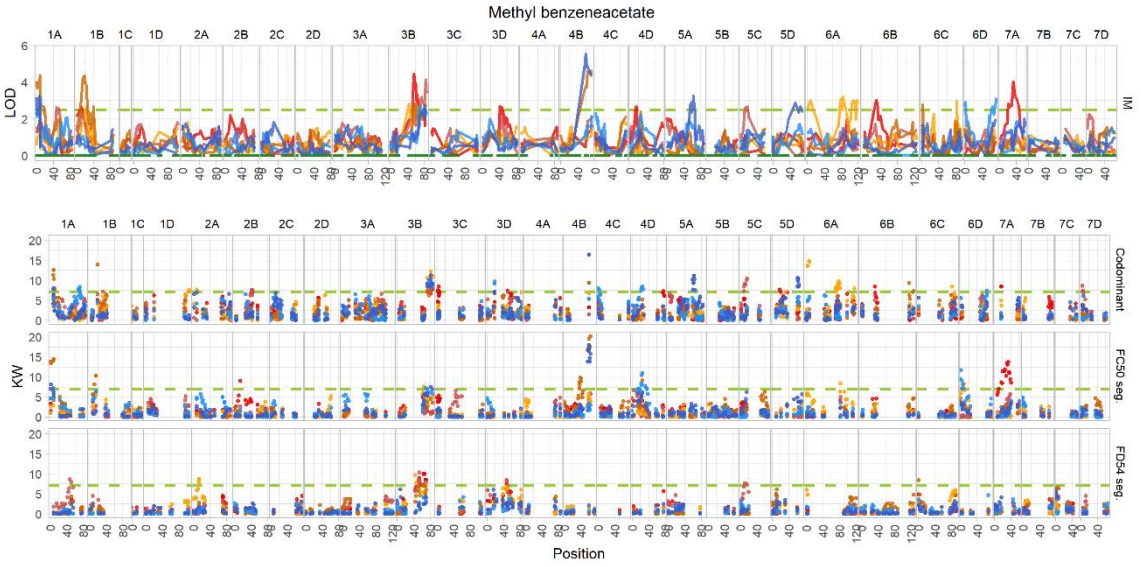

Methyl E-2-hexenoate
LOD
IM
KW
Codominant
FC50 seg.
FD54 seg.
Position

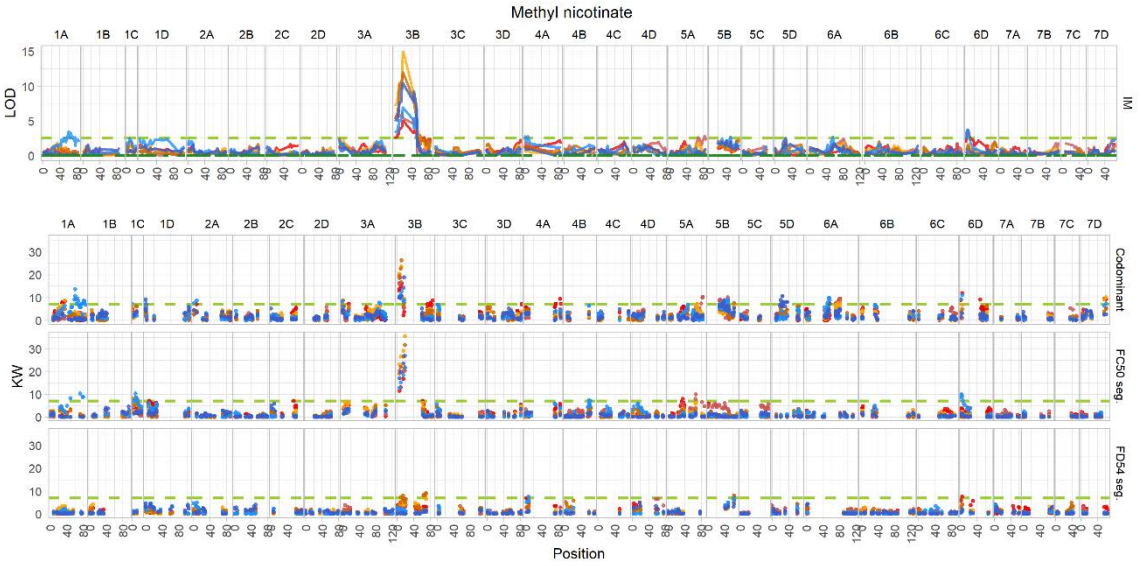

Ethyl 2-hexenoate
LOD
IM
KW
Codominant
FC50 seg.
FD54 seg.
Position

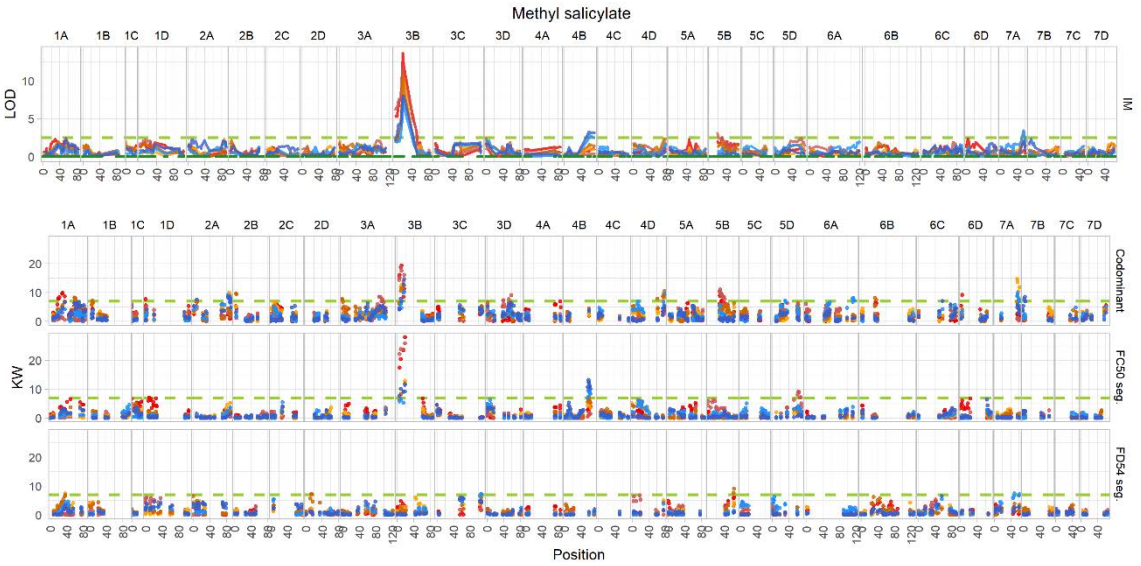

Methyl octanoate
LOD
KW
IM
Codominant
FC50 seg.
FD54 seg.
Position

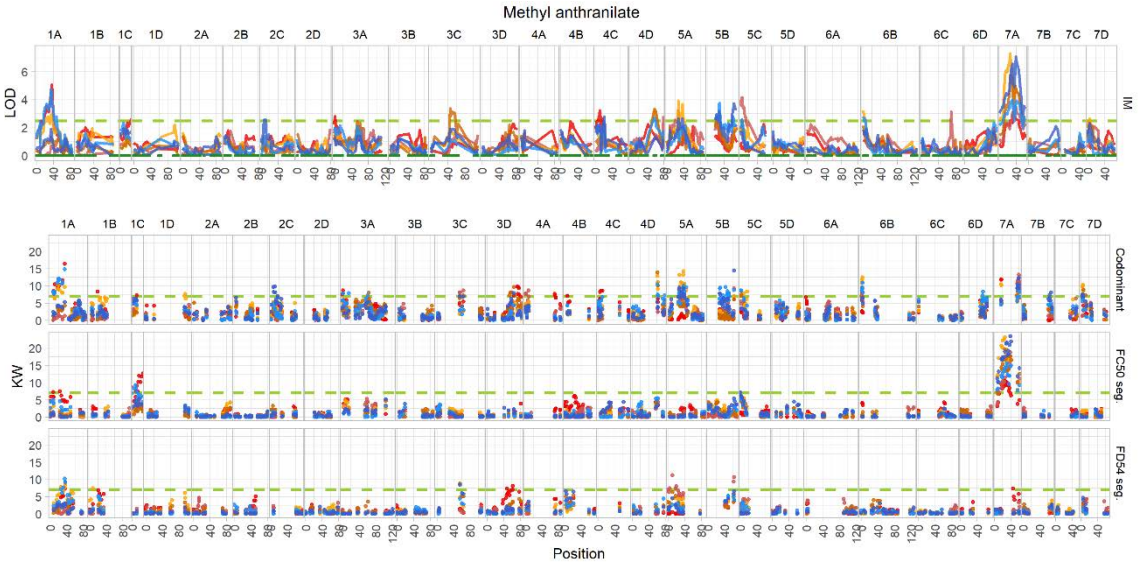

Ethyl octanoate
LOD
KW
IM
Codominant
FC50 seg.
FD54 seg.
Position

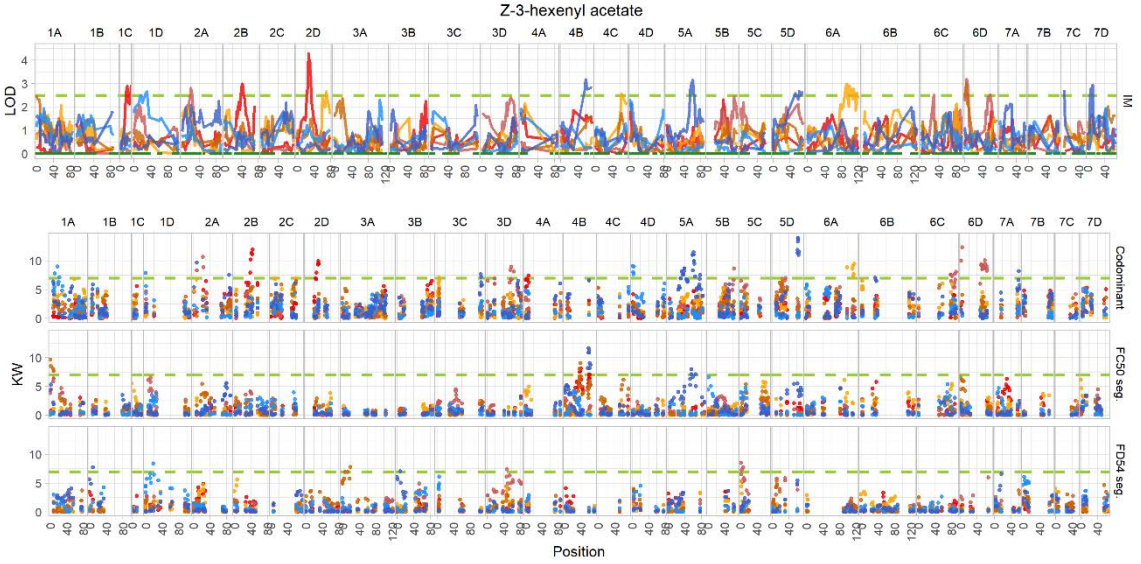

Methyl decanoate
LOD
KW
IM
Codominant
FC50 seg.
FD54 seg.
Position

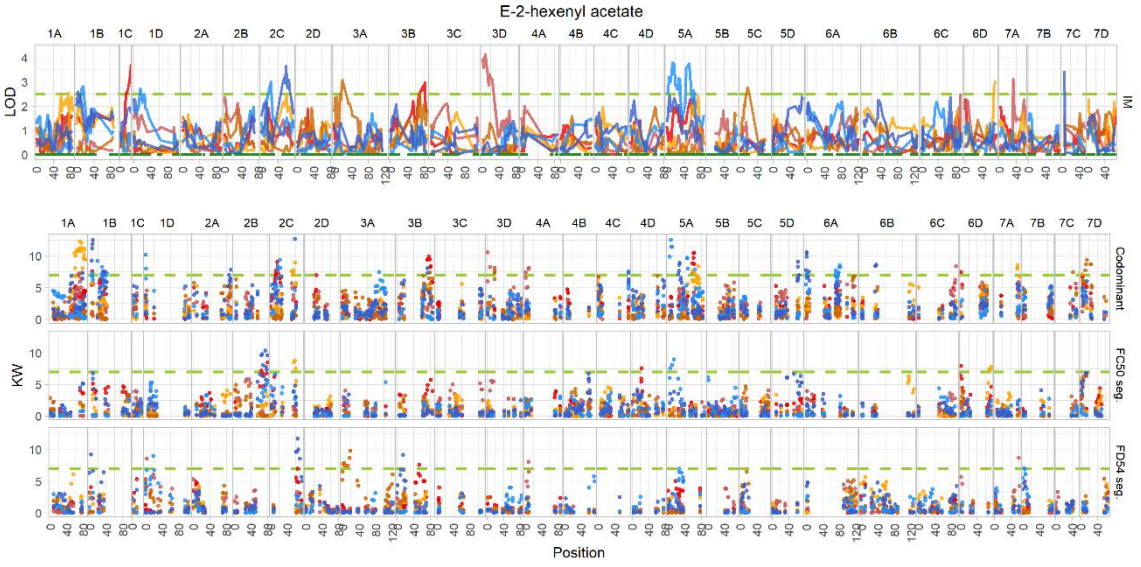

Methyl benzeneacetate
LOD
IM
KW
Codominant
FC50 seg.
FD54 seg.
Position

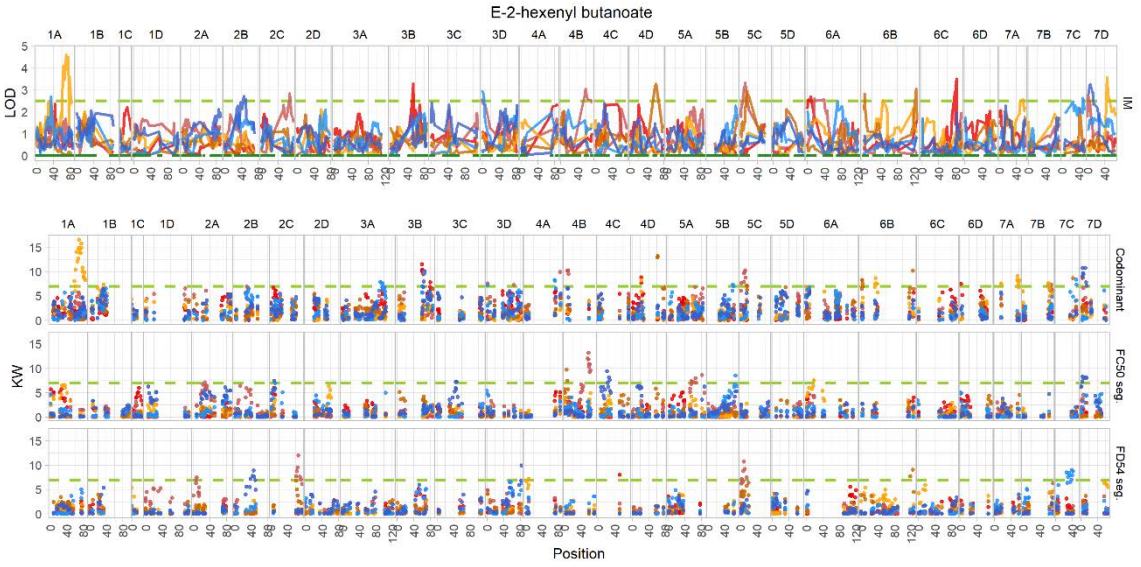

Methyl nicotinate
LOD
IM
KW
Codominant
FC50 seg.
FD54 seg.
Position

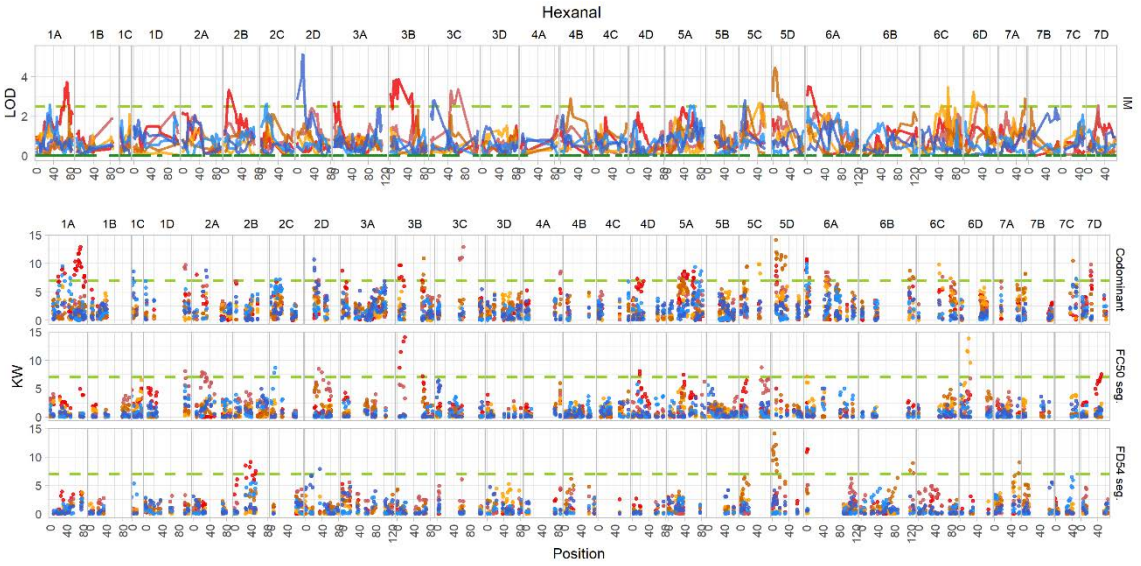

Methyl salicylate
LOD
IM
KW
Codominant
FC50 seg.
FD54 seg.
Position

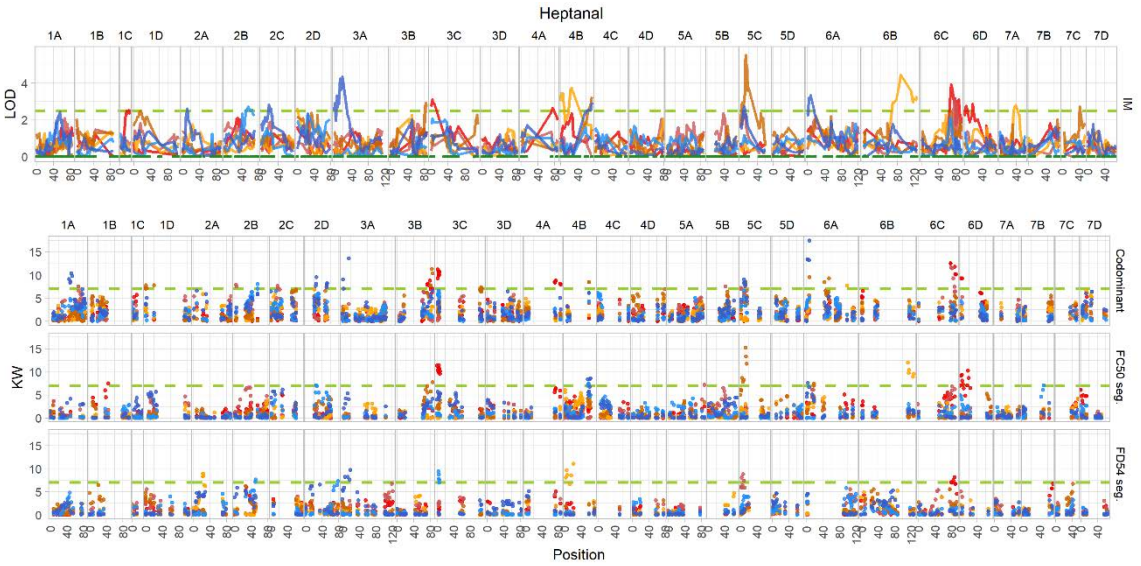

Methyl anthranilate
1A 1B 1C 1D 2A 2B 2C 2D 3A 3B 3C 3D 4A 4B 4C 4D 5A 5B 5C 5D 6A 6B 6C 6D 7A 7B 7C 7D
LOD
IM
KW
Codominant
FC50 seg.
FD54 seg.
Position

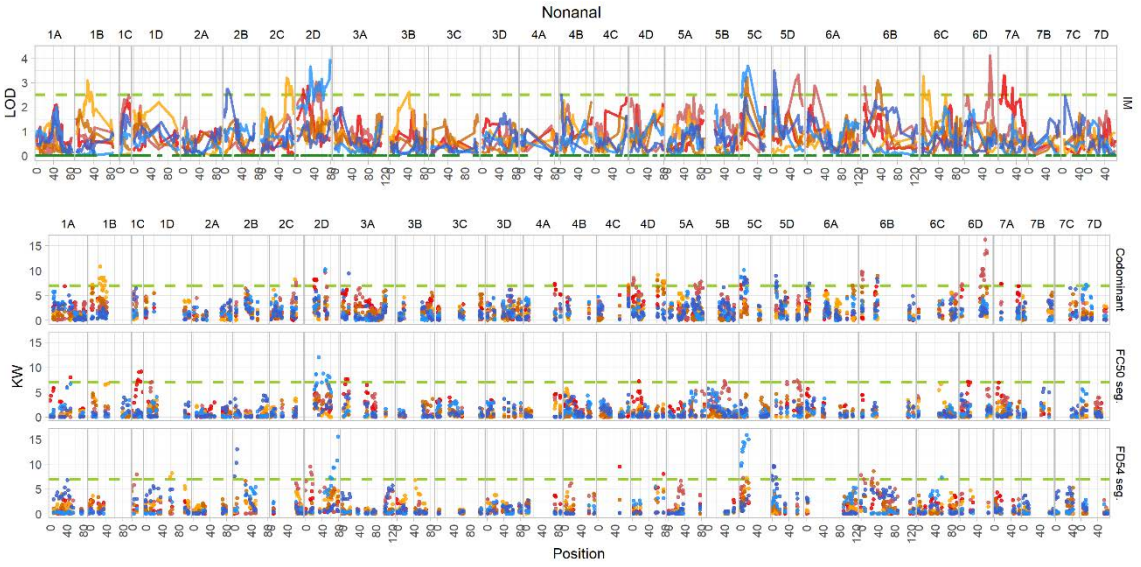

Z-3-hexenyl acetate
1A 1B 1C 1D 2A 2B 2C 2D 3A 3B 3C 3D 4A 4B 4C 4D 5A 5B 5C 5D 6A 6B 6C 6D 7A 7B 7C 7D
LOD
IM
KW
Codominant
FC50 seg.
FD54 seg.
Position

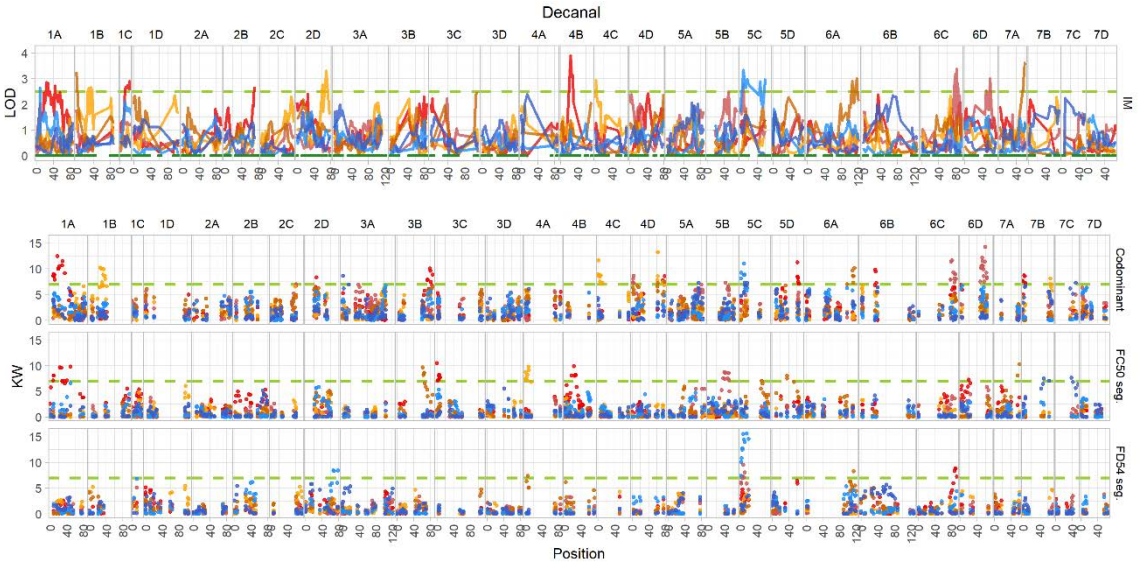

E-2-hexenyl acetate
1A 1B 1C 1D 2A 2B 2C 2D 3A 3B 3C 3D 4A 4B 4C 4D 5A 5B 5C 5D 6A 6B 6C 6D 7A 7B 7C 7D
LOD
IM
KW
Codominant
FC50 seg.
FD54 seg.
Position

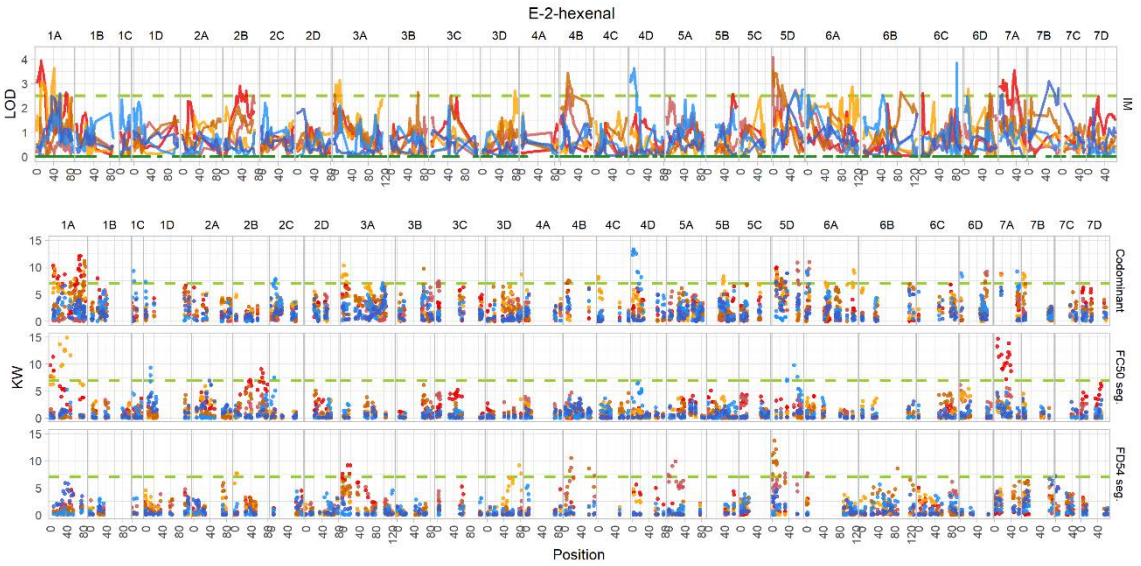

E-2-hexenyl butanoate
LOD
KW
IM
Codominant
FC50 seg.
FD54 seg.
Position

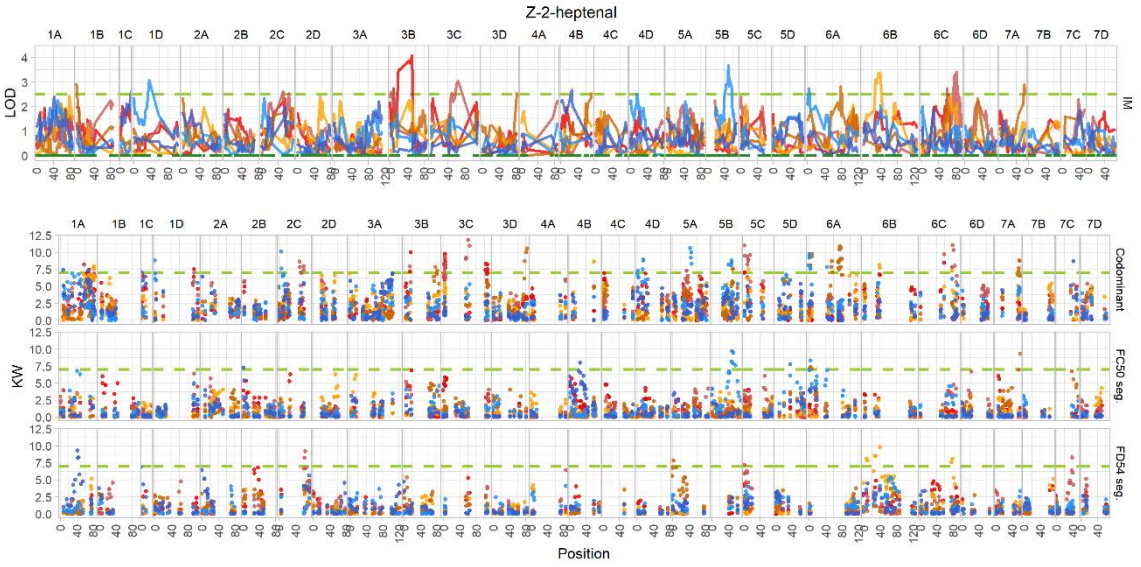

Hexanal
LOD
KW
IM
Codominant
FC50 seg.
FD54 seg.
Position

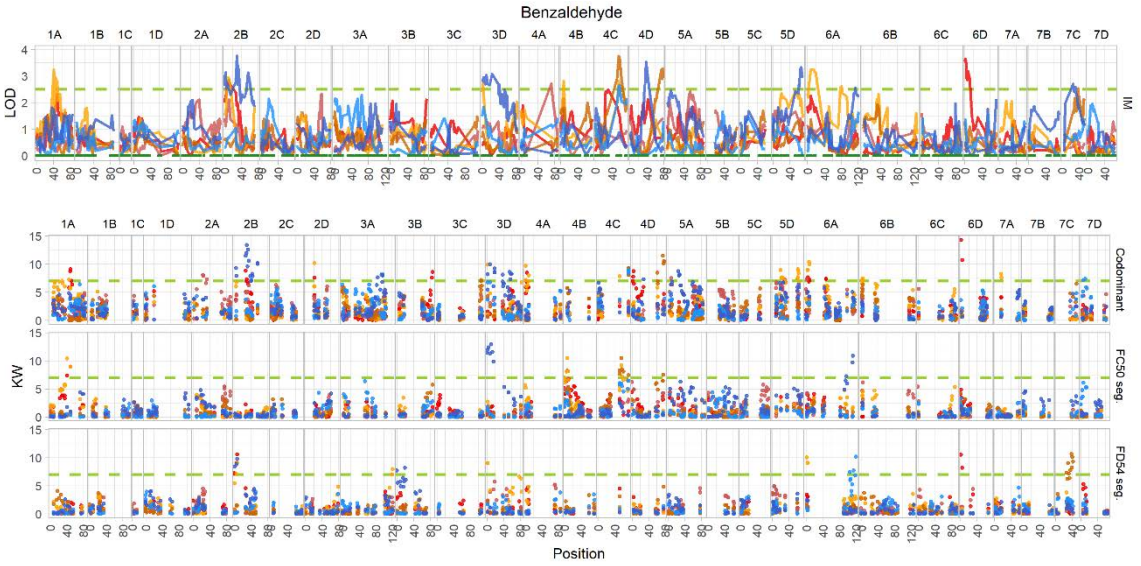

Heptanal
LOD
KW
IM
Codominant
FC50 seg.
FD54 seg.
Position

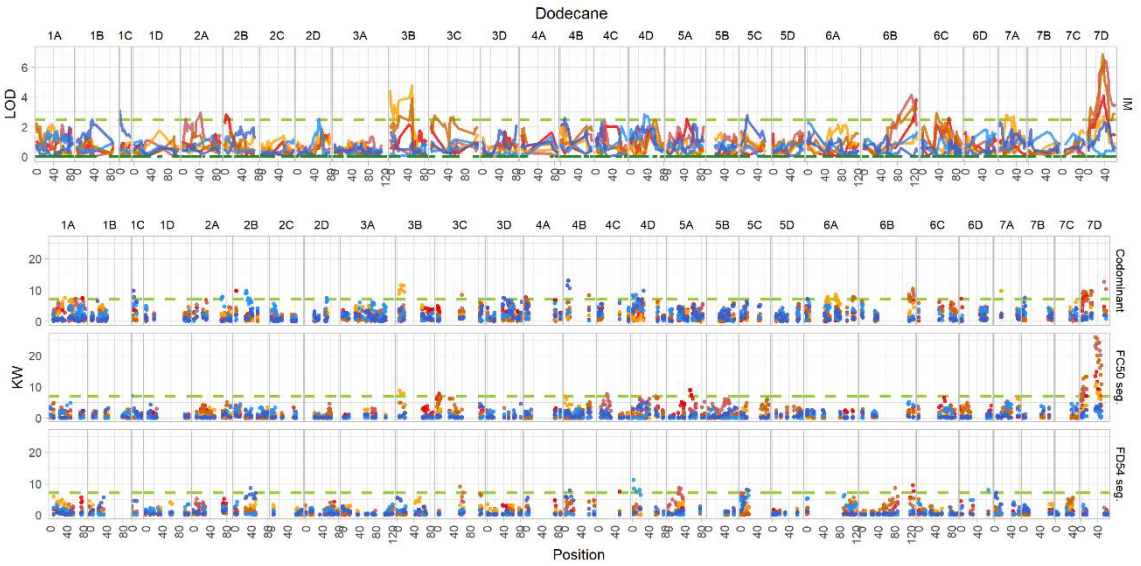

Nonanal
LOD
IM
KW
Codominant
FC50 seg.
FD54 seg.
Position

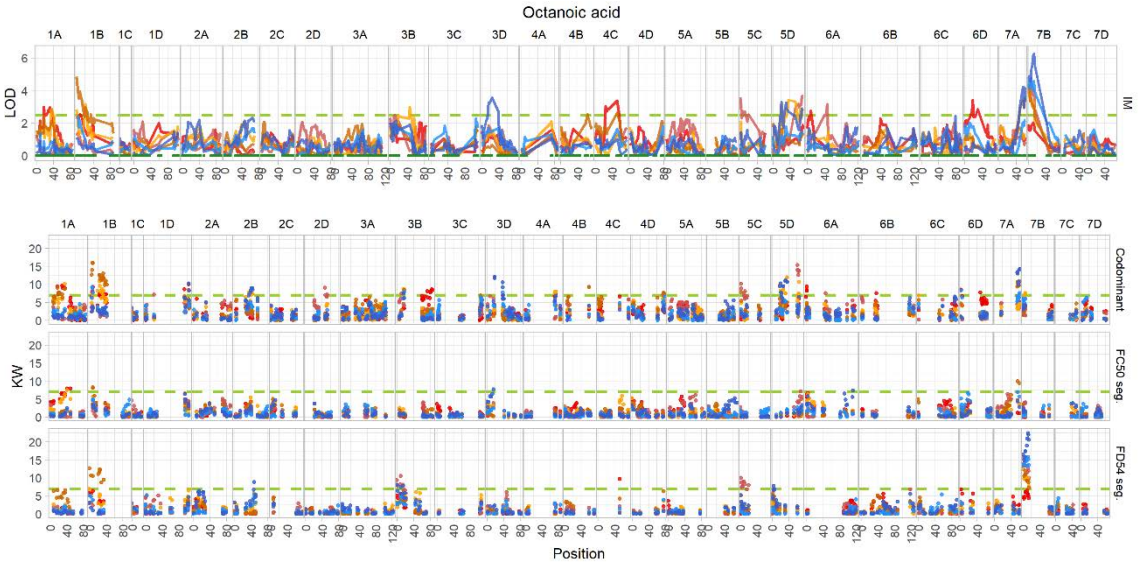

Decanal
LOD
IM
KW
Codominant
FC50 seg.
FD54 seg.
Position

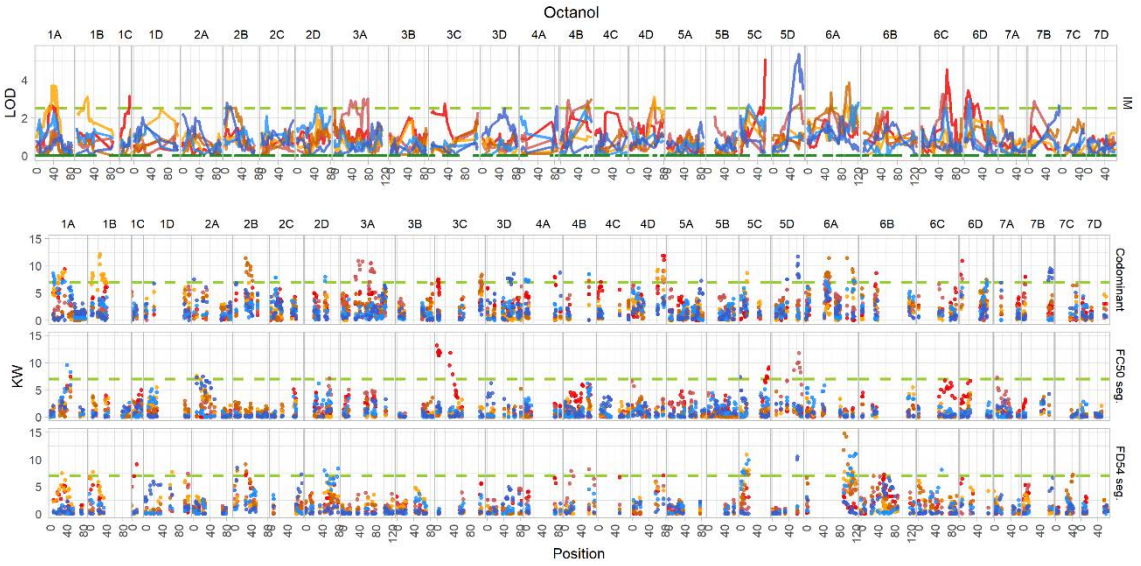

E-2-hexenal
LOD
IM
KW
Codominant
FC50 seg.
FD54 seg.
Position

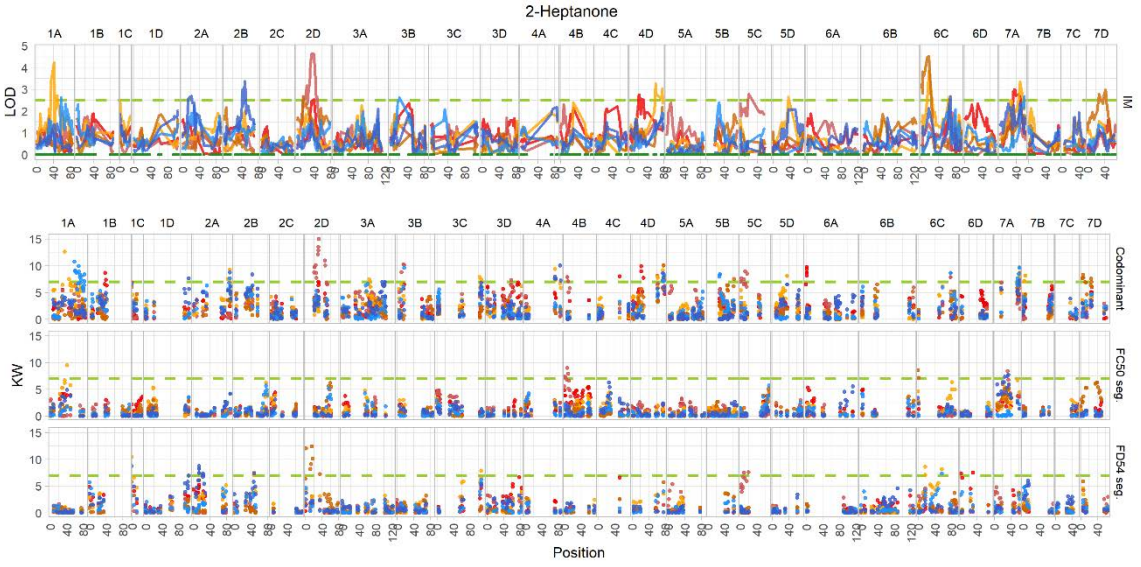

Z-2-heptenal
LOD
IM
KW
Codominant
FC50 seg.
FD54 seg.
Position

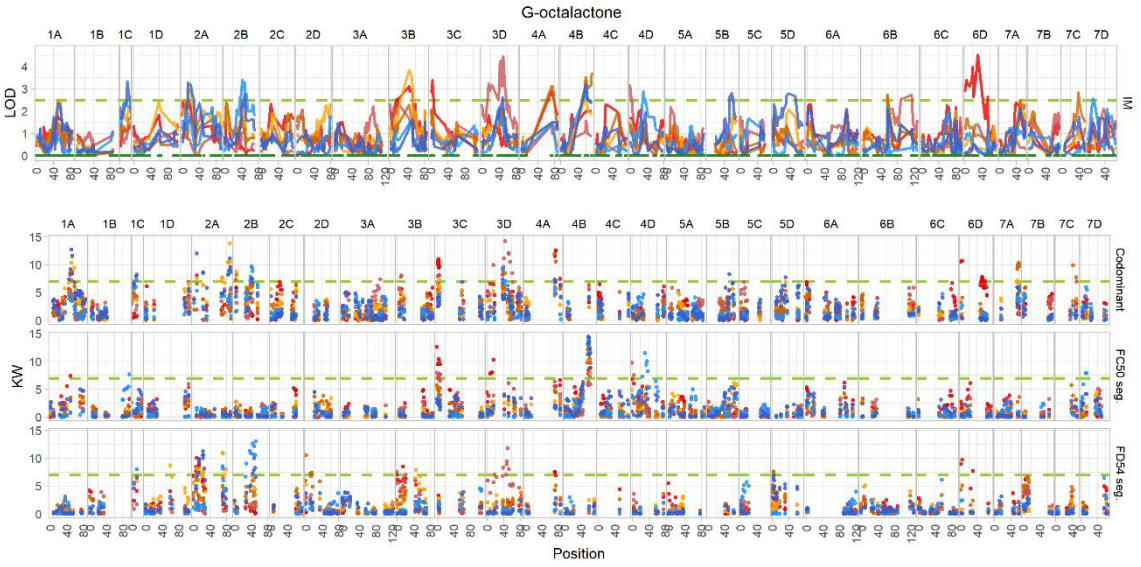

Benzaldehyde
LOD
IM
KW
Codominant
FC50 seg.
FD54 seg.
Position

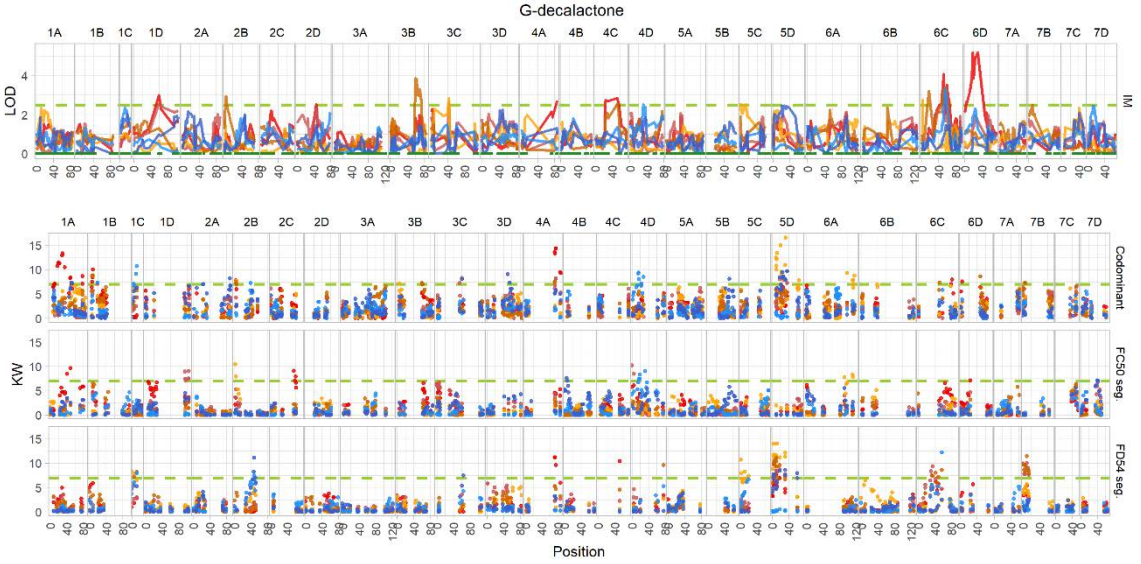

Dodecane
LOD
IM
KW
Codominant
FC50 seg.
FD54 seg.
Position

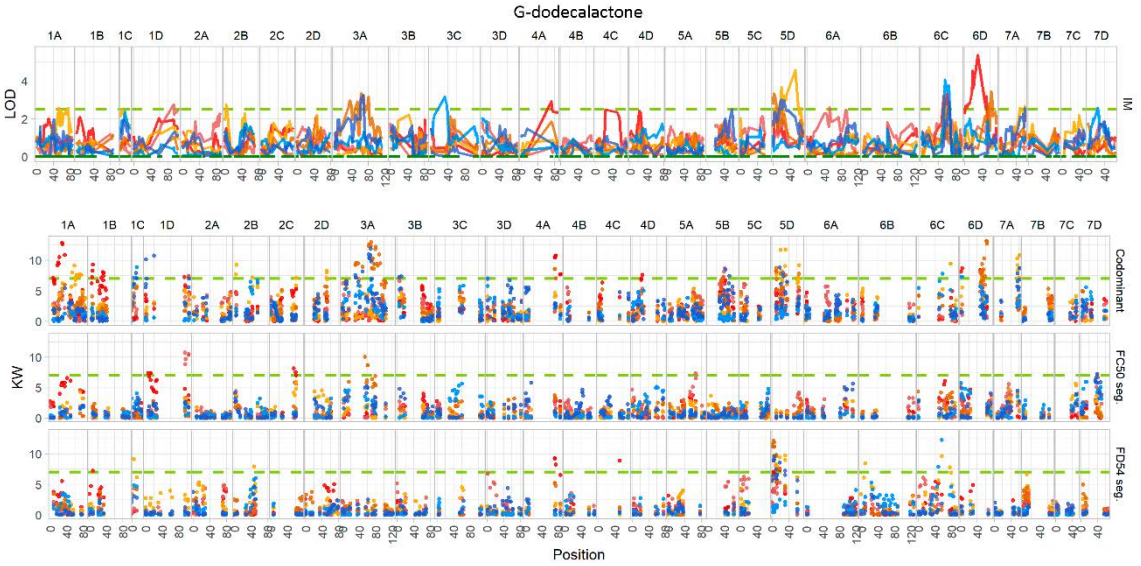

Octanoic acid
LOD
KW
IM
Codominant
FC50 seg.
FD54 seg.
Position

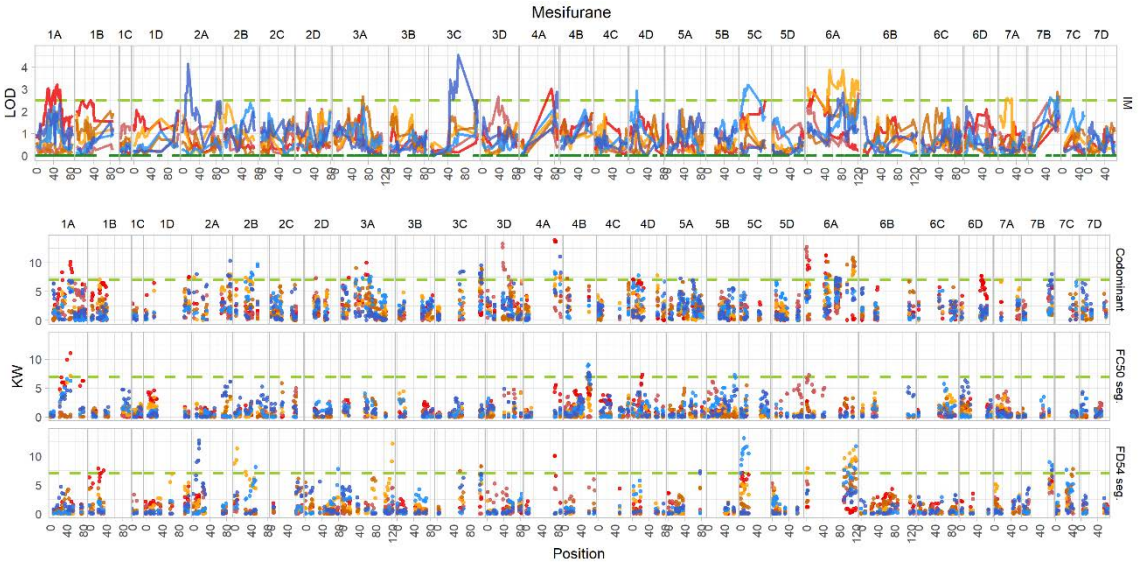

Octanol
LOD
KW
IM
Codominant
FC50 seg.
FD54 seg.
Position

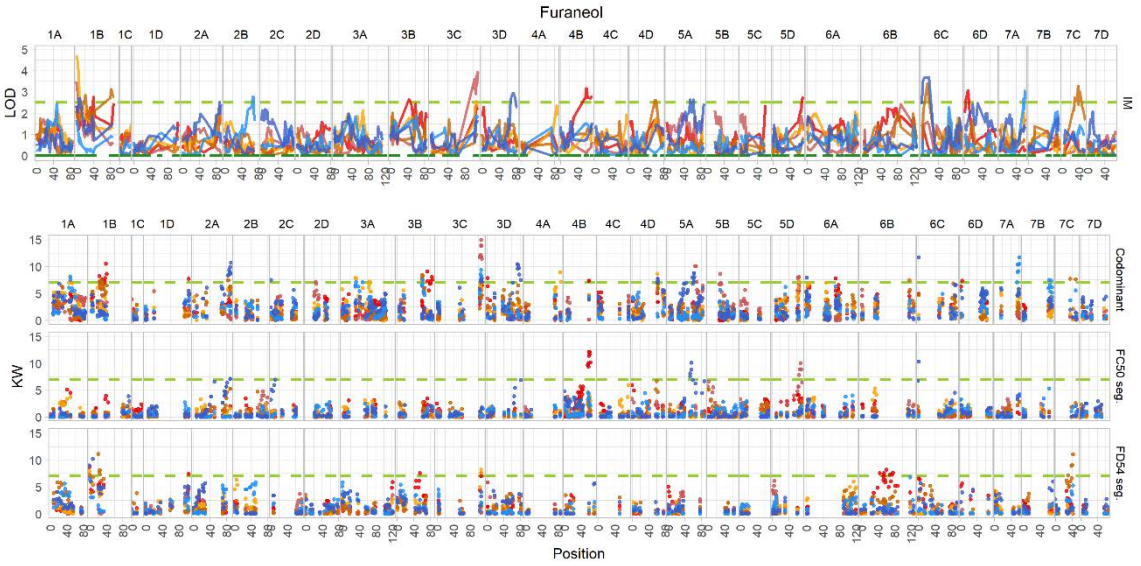

2-Heptanone
LOD
KW
IM
Codominant
FC50 seg.
FD54 seg.
Position

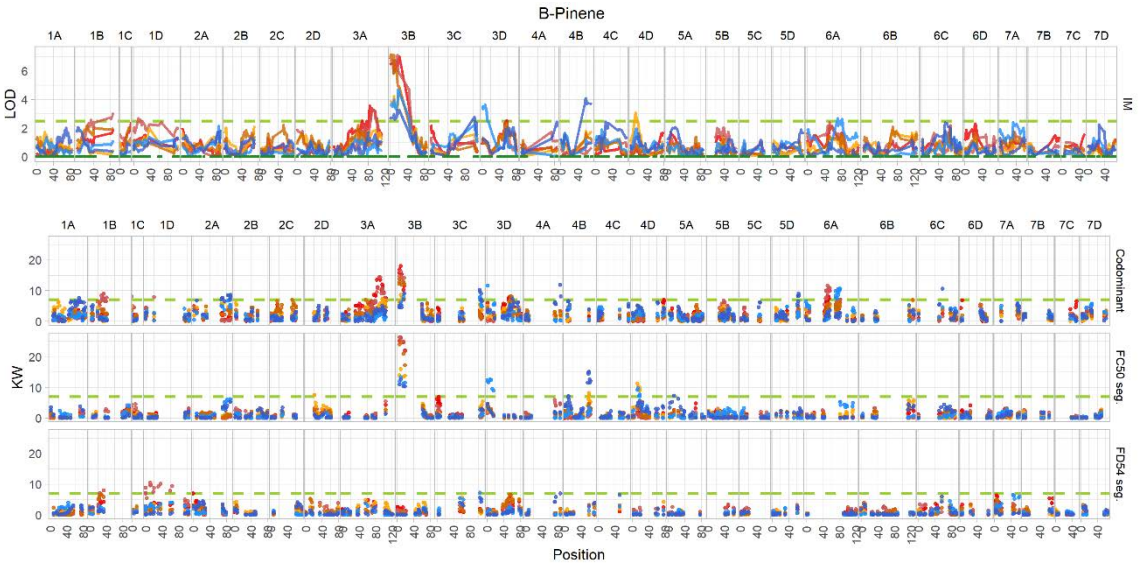

G-octalactone
LOD
IM
KW
Codominant
FC50 seg.
FD54 seg.
Position

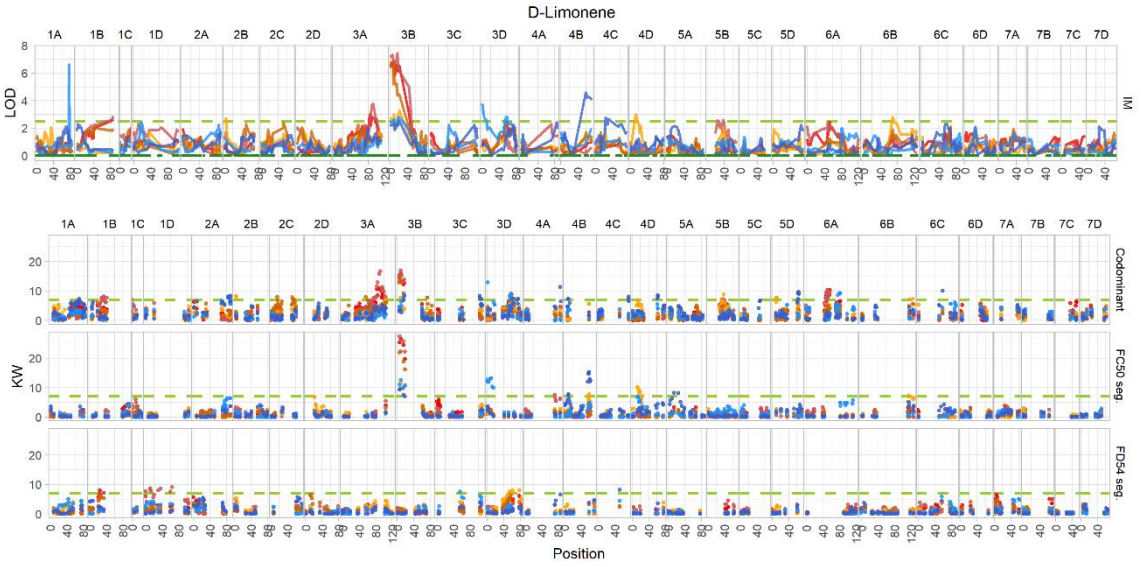

G-decalactone
LOD
IM
KW
Codominant
FC50 seg.
FD54 seg.
Position

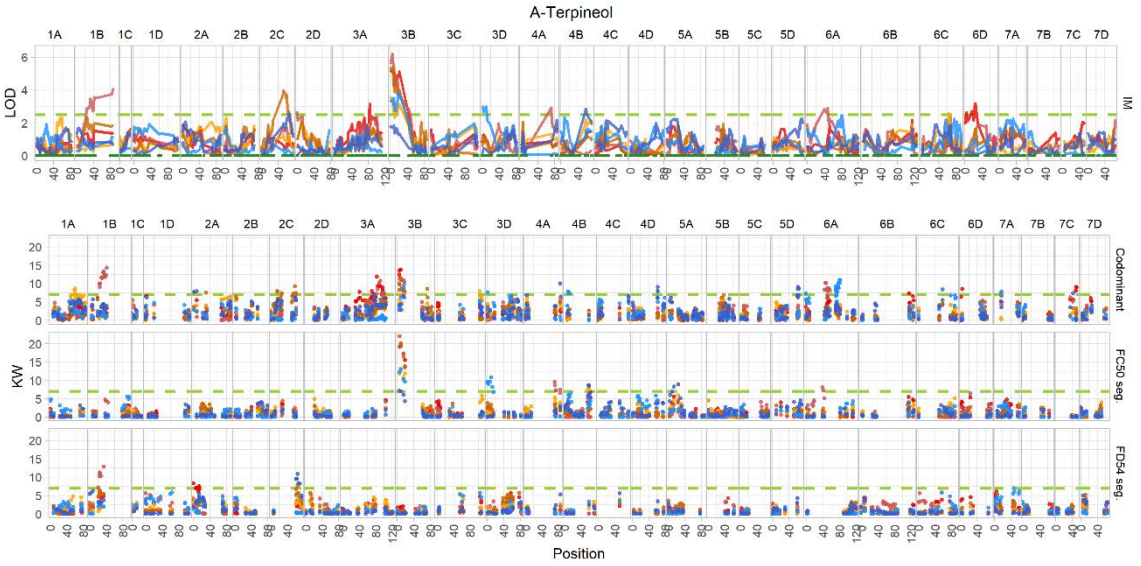

G-dodecalactone
LOD
IM
KW
Codominant
FC50 seg.
FD54 seg.
Position

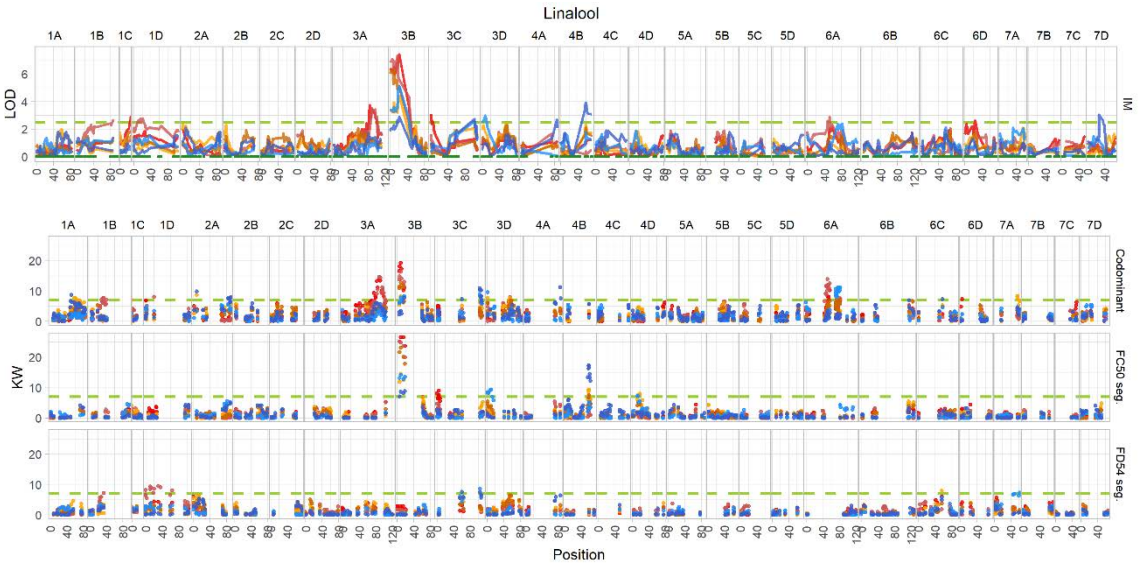

Mesifurane
LOD
IM
KW
Codominant
FC50 seg.
FD54 seg.
Position

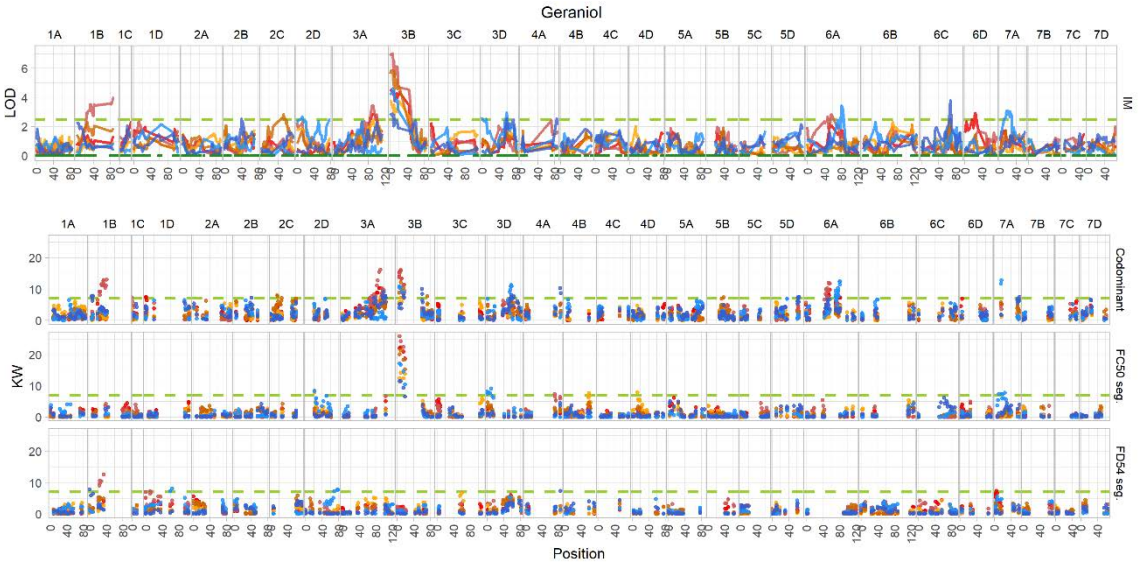

Furaneol
LOD
IM
KW
Codominant
FC50 seg.
FD54 seg.
Position

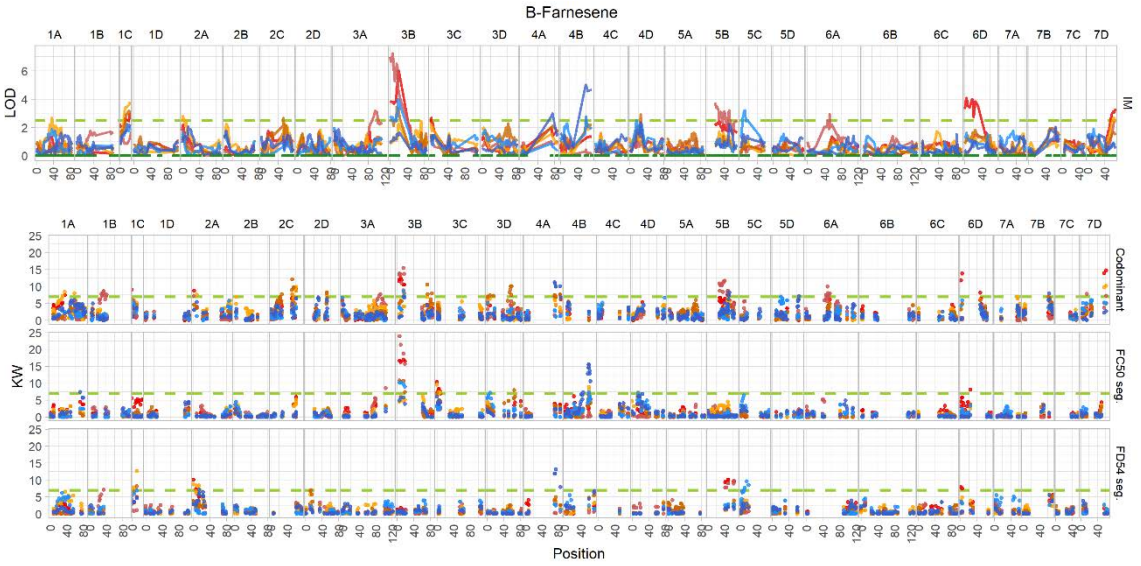

B-Pinene
LOD
IM
KW
Codominant
FC50 seg.
FD54 seg.
Position

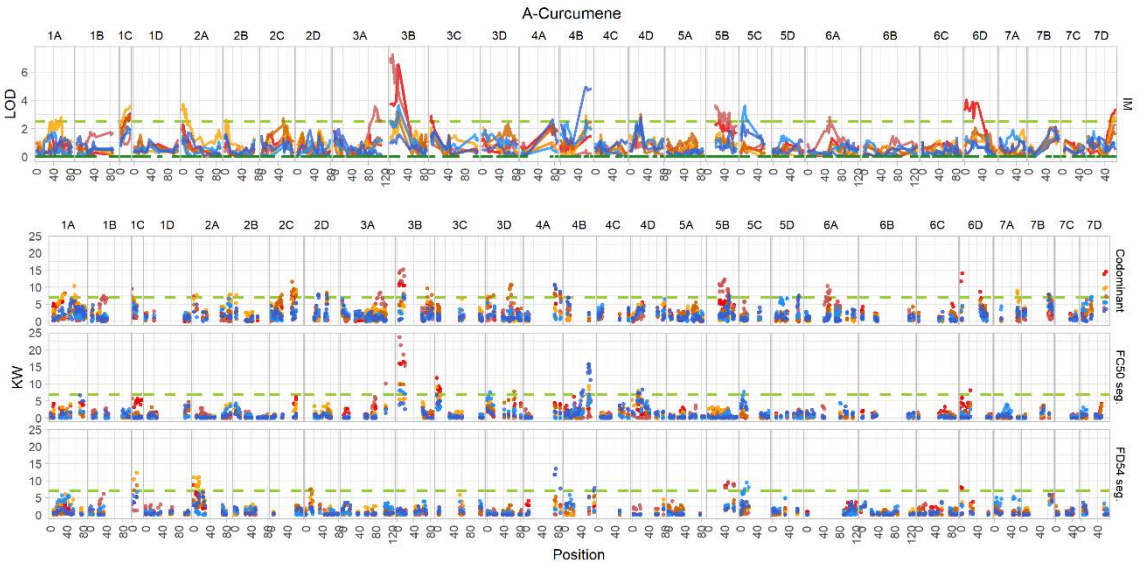

D-Limonene
LOD
KW
IM
Codominant
FC50 seg.
FD54 seg.
Position

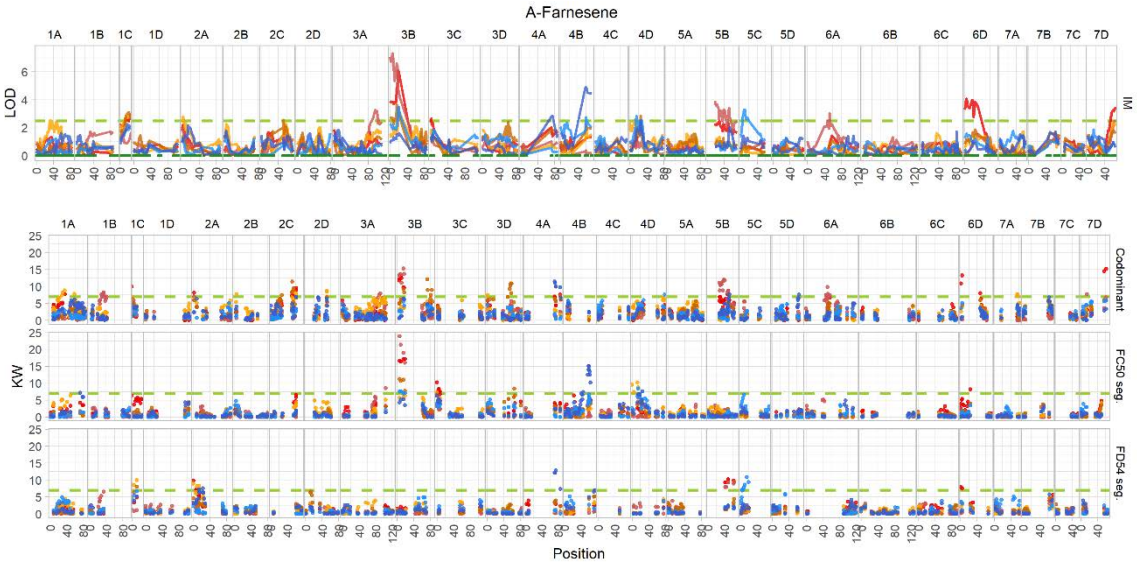

A-Terpineol
LOD
KW
IM
Codominant
FC50 seg.
FD54 seg.
Position

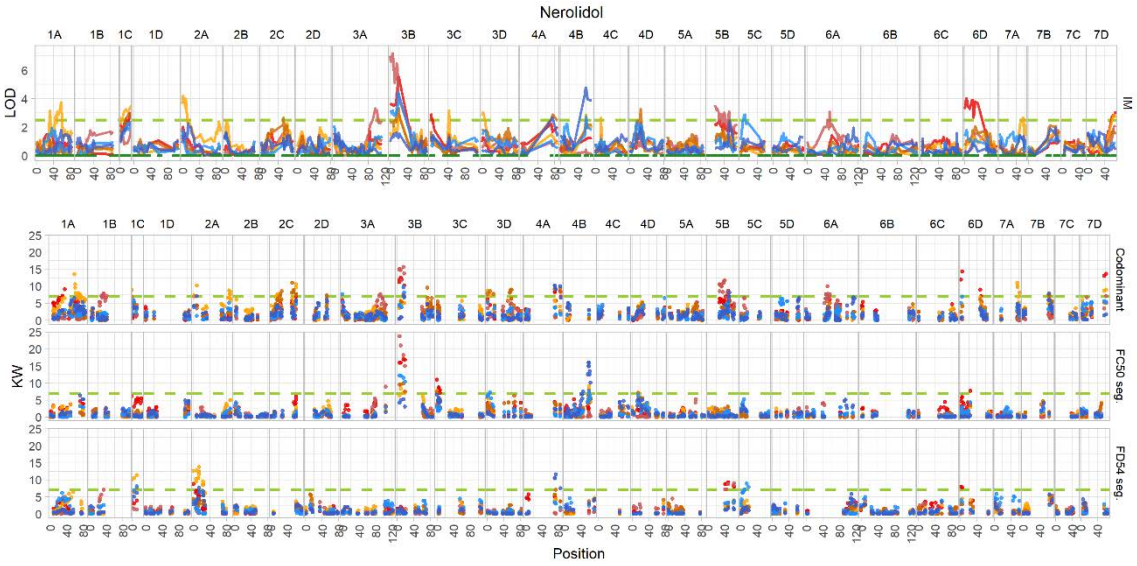

Linalool
LOD
KW
IM
Codominant
FC50 seg.
FD54 seg.
Position


Geraniol
LOD
IM
KW
Codominant
FC50 seg.
FD54 seg.
Position


B-Farnesene
LOD
IM
KW
Codominant
FC50 seg.
FD54 seg.
Position


A-Curcumene
LOD
IM
KW
Codominant
FC50 seg.
FD54 seg.
Position


A-Farnesene
LOD
KW
IM
Codominant
FC50 seg.
FD54 seg.
Position


Nerolidol
LOD
KW
IM
Codominant
FC50 seg.
FD54 seg.
Position

Annex chapter 2:

Development of a 48 SNP array linked to VOCs for MAS in strawberry aroma breeding programs

Introduction

Consumers are concerned about the strawberry flavour (Klee and Tieman, 2018). Since volatile compounds provide wide range of aromas contributing to the consumer acceptance, breeding programs are demanding new tools such MAS (Ulrich *et al.*, 1997, Ulrich and Olbricht, 2016 and Schwieterman *et al.*, 2014).

Some markers based on PCR are already available for MAS such as for γ-decalactone targeting *FaFAD1* in LG3 (Chambers *et al.*, 2014 and Cruz-Rus *et al.*, 2017), mesifurane targeting *FaOMT* in LG7 (Zorrilla-Fontanesi *et al.*, 2012 and Cruz-Rus *et al.*, 2017), methyl anthranilate targeting *FaAAMT* in LG4 (Pillet *et al.*, 2017) and linalool targeting *FaNES1* in LG3 (Chambers *et al.*, 2012).

As explained in the General introduction, there are different SNP genotyping platforms, which could be multiplexed and used for MAS being faster and more efficient than PCR based markers. From our QTL analysis, we discovered some promising SNPs linked to the accumulation of different volatiles (See Chapter 2) suitable to implement them in a breeding program.

Here, we present the preliminary results of the validation of an aroma SNP array in a new F1 population, derived from a cross between 'FC041' and 'FC084' breeding lines.

Materials and Methods

Plant material

An F1 population of 46 progenies (hereafter, 'C19' population) was obtained by crossing 'FC041' and 'FC084' breeding lines. Both parental lines were selected for having a special flavour. At least two clonal runner plants per progeny or parental lines were grown using standard cultivation practices in an open field of the south-west of Spain (Cartaya, latitude: 37°18'N, longitude 7°5'W) during 2018.

DNA extraction was performed as described in Material and Methods of Chapter 1. The DNAs were diluted to 20ng/µl as required for array hybridisation.

Design and analysis of the Fluidigm array

For the 48 selected SNPs, A1, A2 and C1 primers were designed using their flanking regions using the Primer Picker vs 0.26 software (KBioscience, UK). Primers sequence can be found in the Supplementary table A2-1. The genotyping was performed at the CRAG Genotyping facilities (Bellaterra, Spain) using 48x48 Dynamic Array™ (Fluidigm, CA, US). The results were assessed using the Fluidigm SNP Genotyping Analysis vs 4.1.3. software (Fluidigm, CA, US).

Volatile analysis

We phenotyped the KVCs from mature fruits of the 'C19' population collected in one harvest in 2018 using sampling and GC-MS protocol described in Chapter 2. KVCs were identified by comparing retention time and mass spectrum of each peak with those from the commercial standards and mass spectrum libraries, NIST08 and NIST11. We used the Enhanced ChemStation software (Agilent Technologies, CA, USA) for peak identification and area calculation. Peak areas were standardised with the internal standard, 3-hexanone, peak.

Results

A population from the cross of 'FC041' and 'FC084' cultivars selected for their special aroma was composed of 46 progenies and named 'C19' population. In the GC-MS analysis, we were able to detect 14 out of the 19 KVCs. The five KVCs that we could not detect were methyl cinnamate, methyl anthranilate, myrtenyl acetate, furaneol and Z-3-hexenal (Supplementary table A2-2).

We designed a 48 SNP Fluidigm® array composed by SNPs selected from the 'FC50xFD54' genetic map; ten SNPs from the terpene QTL (*Terpene_3B*, mainly from 13.39-29.17cM), ten from the hexanoate related esters (*Hexanoate_4B* mainly from 66.61-100.36cM), ten from the lactone QTL (*Loctone_6C* from 22.34-45.27cM), eleven from the methyl anthranilate QTL (*MA_7A* from 22.26-73.38cM) and the remaining seven were control markers.

The hybridization of this array with the 'C19' (Supplementary table A2-3) showed segregation for 20 SNPs for aroma. Three of them were codominant, ten showed 'FC41' segregation and seven 'FC081' segregation. Most of the markers that did not segregate correspond to the lactone QTL.

Two major and stable QTLs detected in Chapter 2 were also significant in the 'C19' population. Four SNPs for linalool and three for methyl hexanoate were linked to their accumulation according to our genotyping array. The alleles that increase linalool and methyl hexanoate accumulation come from 'FC081' parental line and show 10-26% and 17-21% increase compared to the allele from 'FC041' parental line, respectively (Table A2-1). The XX genotype from Affx-88843366 marker is linked to linalool accumulation, while the XX genotype from Affx-88856956 marker is for methyl hexanoate accumulation (Fig. A2-1). However, we could not find any marker from the *Lactone_6C* associate with the γ-decalactone accumulation in this population.

Table A2-1. Summary 48 SNP array most linked to VOC content.

Marker	FC50xFD54 map		Genotype				Segregation	VOCs	Average content			Dif.
	LG	Pos	XX	XY	YY	NA	class.		XX	XY	YY	gen.
Affx-88834063	LG3B	13.84	0%	48%	46%	4%	FC081 seg	Linalool		**4.27**	3.78	11%
Affx-88836329	LG3B	25.87	0%	31%	60%	6%	FC081 seg			**4.38**	3.95	10%
Affx-88836006	LG3B	27.69	56%	31%	0%	6%	FC081 seg		3.42	**4.49**		24%
Affx-88843366	LG3B	29.17	60%	29%	0%	8%	FC041 seg		**4.46**	3.31		26%
Affx-88857536	LG4B	89.31	63%	31%	0%	4%	FC041 seg	Methyl hexanoate	**10.46**	8.38		20%
Affx-88856956	LG4B	93.67	67%	27%	0%	4%	FC041 seg		**10.47**	8.25		21%
Affx-88856914	LG4B	93.67	65%	29%	0%	2%	FC041 seg		**10.23**	8.51		17%

Bold numbers in average content indicate the genotype linked to the compound accumulation.

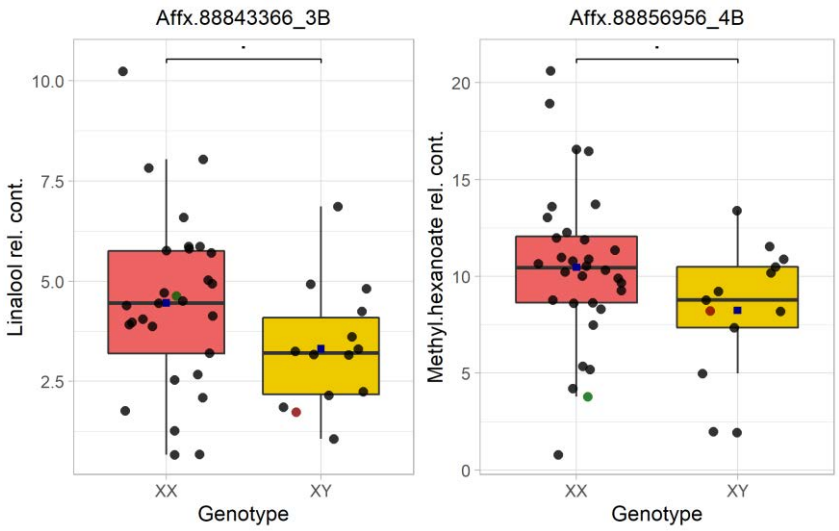


Figure A2-1. Linalool and methyl hexanoate boxplot for 'C19' population. Dot colours: progenies (black), 'FC041' (red) and 'FC081' (green). Blue squares represent means. Significant level <0.1 (·).

Discussion

Strawberry aroma is a key factor for consumer acceptance (Folta and Klee, 2016). Breeding for volatile compounds content is complex due to strawberry octoploid nature and the polygenic regulation of these traits. Hence, it is necessary to design markers to implement MAS and thus meet the consumer demand.

We provide preliminary genotyping results of the SNPs associated to four aroma QTLs in a 48x48 SNP array using the Fluidigm platform. Six markers did not amplify properly and 43% of remaining did not segregate in the 'C19' population. So, we ended with 20 segregating SNPs in this population associated with aroma traits.

For the *Terpene_3B* QTL*,* we found four markers linked to linalool accumulation and the selection of the genotype coming from the 'FC081' parental line might increase its content by 10-26%. Moreover, three other markers from *Hexanoate_4B* QTL selecting for the XX genotype may increase the methyl hexanoate content by 17-21%. The γ-decalactone variation that we found should be dependent on another QTL different from the *Lactone_6C* QTL. Therefore, using this population we could not validate neither the *Lactone_6C* nor the *MA_7A* linked markers.

Our final objective is to select markers with easy clustering genotype linked to most of vocQTLs discovered in Chapter 2 to apply in MAS. To do so, we can introduce some of the new subgenomic specific markers from the 850k and 50k SNP arrays (Hardigan *et al.*, 2020) which may be closely linked to the trait of interest and test them in a parental collection.

Chapter 3:

Genetic analysis of fruit quality traits in cultivated strawberry fruits

Introduction

Strawberries are an appreciated fruit used in fresh and processed market. Consumer acceptance lies on the balance between sweetness and acidity and the perception of some volatile compounds. However, other traits such as firmness, colour and shape are also important for the commercial value of these fruits.

Sweetness perception in fresh fruits is highly correlated to sugar content. The major compounds detected in strawberry fruits are glucose, fructose and sucrose. During fruit ripening, sugars and volatile compounds are accumulated in the receptacle. At subcellular level, sugars are stored in vacuoles through several transport mechanisms. Sucrose is also important for the regulation of fruit ripening and changing the activity of transporters such as *FaSUT1* gene in cultivated strawberry alters sucrose and ABA content (Jia *et al.*, 2013). Furthermore, a common transcription factor, regulated by sucrose and ABA, was found in tomato and strawberry fruits (Jia *et al.*, 2016).

Soluble solids content (SSC) is a good approximation to quantify sugar content, easily measured with a refractometer. In cultivated strawberry, several QTLs were detected in different homeologous groups (HGs) such as HGII, HGIII, HGV and HGVI (Zorrilla-Fontanesi *et al.*, 2011, Lerceteau-Köhler *et al.*, 2012 and Castro and Lewers, 2016). Moreover, metabolite QTLs were mapped for fructose in LGIVb and LGVc, for glucose in LGIb, LGIIIa, LGIIIc, LGIVa and LGVIa and for sucrose in LGVIa and LGVIIa (Lerceteau-Köhler *et al.*, 2012). However, metabolic QTLs in the diploid strawberry NIL population were located in different LGs, mapped in LG2 and LG3 for fructose, in LG2 and LG5 for glucose while no QTL were found for sucrose (Urrutia *et al.* 2015).

Another important trait involved in fruit flavour is acidity. Citric and malic acids are the main ones in strawberry (Reviewed in Perkins-Veazie, 1995). During fruit development, acids such as citric acid displayed different accumulation patterns in receptacle and achene, being constant in the former and dramatically declined in the latter (Moing *et al.*, 2001 and Fait *et al.*, 2008). For titratable acidity (TA), different tendencies were observed, starting with an increment during green and turning stages followed by a decrease until mature fruits (Moing *et al.*, 2001). Acid transport is important for their accumulations in the tonoplast (Jia *et al.*, 2018), but the mechanisms controlling these processes in strawberry are still unclear.

Several regions responsible for acidity traits, such as TA and pH, were mapped in HGI, HGIII, HGIV and HGV in different F1 populations (Zorrilla-Fontanesi *et al.*, 2011, Lerceteau-Köhler *et al.*, 2012 and Castro and Lewers, 2016). In addition, metabolite QTLs were localised for citric acid

in LGIa, LGIIc, LGIIIa, LGVb, LGVIa and LGVIIb and for malic acid in LGVd, LGVIa, LGVIb and LGVIId (Lerceteau-Köhler *et al.*, 2012). Two decreased acidity QTLs were found in LG4 and LG5 in a diploid NILs collection (Zhang *et al.*, 2017a). The sugar-acid ratio (SAR), which is used to quantify the balanced fruit taste, was mapped in LGVI3 (Castro and Lewers, 2016).

Under a commercial point of view, fruit firmness is another important trait. As explained in the General introduction, strawberries are soft fruits characterized by a very short shelf-life. Fruit softening is achieved through cell-wall degradation, cellulose and hemicellulose degradation and pectin release (Knee *et al.*, 1977 and Perkins-Veazie, 1995). Several cell-wall modifying enzymes involved in fruit firmness have been studied (Draye and Van Cutsem, 2008, Molina-Hidalgo *et al.*, 2013, Pose *et al.*, 2015, Paniagua *et al.*, 2017 and Mendez-Yanez *et al.*, 2020), however, little is known about its regulation.

As mentioned in the General introduction, many factors can alter the final fruit shape (Perkins-Veazie, 1995). A diploid NILs population mapped two round fruit QTLs in LG2 and LG5 and an elongated QTL in LG5 (Urrutia *et al.*, 2015). Fruit shape is an hormone-regulated process in which auxins and GA modulate fruit diameter and length, respectively, while ABA inhibits fruit growth and induces fruit ripening during the last stages (Symons *et al.*, 2012 and Liao *et al.*, 2018).

A relatively wide range of colours can be found in wild octoploid strawberry, going from light red in *Fragaria virginiana* to whitish in some *Fragaria chiloensis* accessions. In almost all breeding programs, strawberries were selected for the red fruit colour. Anthocyanins, regulated by ABA, are the major source of the red colour in strawberry. Moreover, the family of transcription factor R2R3 MYB regulates the anthocyanins pathway (Aharoni *et al.*, 2001, Schaart *et al.*, 2013 and Medina-Puche *et al.*, 2014). FaMYB10, controlled by hormones, is able to enhance the expression of genes involved in flavonoids/phenylpropanoids pathways (Medina-Puche *et al.*, 2014). An insertion that truncates the octoploid *FaMYB10* gene (Wang *et al.*, 2019) and an amino acid mutation in the diploid *FvMYB10* gene (Hawkins *et al.*, 2016) are responsible for the *Fragaria* white/yellow fruit phenotype. These genes are located in LG1. Other colour-related *MYB* genes are located in LG5 and LG6 (Wang *et al.*, 2019).

In the present study, we have evaluated during three years different fruit quality traits in an F1 population between ‘FC50’ and ‘FD54’ and an F2 population between ‘Camarosa’ and ‘Dover’ with the aim to compare these traits between both populations and define genetic regions responsible for their variability, focusing on the stable QTLs in different years and populations.

Materials and Methods

Plant material

Two PLANASA breeding lines, 'FD54' and 'FC50', were crossed to obtain an F1 population of 63 siblings (hereafter, 'FC50xFD54'). Parental characteristics and information about the cultivation of the population are described in Chapter 2, section Materials and Methods. Mature fruits were collected in six different harvests during three successive years at Le Barp, France.

Two commercial cultivars, 'Dover' and 'Camarosa', were crossed to obtain a hybrid ('H-21') which produced an F2 population of 117 progeny lines (hereafter, '21AF'). Mature fruits were picked at different times during three years between May and July at Caldes de Montbui, Spain. The fruits of each progeny were collected between three to five times per year.

All the fruit phenotypic data were recorded immediately or at least three days after their collection, keeping the fruits in a cold room. Data from 'FC50xFD54' population were averaged independently for each harvest (two harvests every year), whereas all the data from one year were averaged in the case of the '21AF' population (one harvest per year).

Phenotype analysis

Weight, firmness, SSC and acidity

Mature fruits from 63 progenies from 'FC50xFD54', parental lines ('FC50' and 'FD54') and grandparental lines ('FC030', 'FD197', 'FD016' and 'FD019') were collected in 2017 and 2018. Ripe fruits from 'Camarosa' and 'Dover' parental lines, hybrid '(H-21') and 117 progenies were collected in 2016 and 2017.

Fruit weight (FW) of five to seven fruits was measured with a balance and averaged (Ohaus Corp., Switzerland). Firmness (FIR) of three to four fruits was estimated with a penetrometer (Fruit Test^{TM}, Wagner Instruments) in gr force units. The juice of three to four strawberries was used to measure soluble solids content (SSC) with a digital hand refractometer (Atago Co. Ltd., Tokyo, Japan) and expressed in brix degrees. Five ml of juice diluted with 45ml deionised H_2O were measured in a HI 84532 titratable acidity Mini Titrator (Hanna instruments, Rhode Island, USA) reporting pH and titratable acidity (TA). TA is calculated by the quantity of NaOH (0.5M) needed to reach pH 8.1 and reporting citric acid content (g/100ml). Ratios between sweetness and acidity (SAR) were calculated by the SSC/TA index, SAR=10*SSC/TA.

Shape and colour

Mature fruits from 'FC50xFD54' and ancestors were collected in 2017, 2018 and 2019, whereas, mature fruits from '21AF' and ancestors were picked up in 2016, 2017 and 2019. Juice fruit colour was measured in 2017 and 2018 for 'FC50xFD54' population and 2016 and 2017 for '21AF' population.

Three to five fruits per progenies and ancestor lines were cut longitudinally and each half was scanned. Half fruit was faced up to measure shape and the internal colour and the other half was faced down to record the external fruit colour. Three to five fruits per progenies and ancestor lines were made juicy using a coffee grinder. All scanned fruits were analysed by Tomato Analyser v3.0 or v4.0 software (Darrigues *et al.*, 2008 and Gonzalo *et al.*, 2009). We selected some fruit shape measurements from the software such as fruit perimeter (FP), fruit area (FA), maximum fruit length (FL), maximum fruit diameter (FD) in cm units and fruit shape ratio (FS = FL/FD). Width widest position (WWP = y/FL) gives the ratio between length until width position divided by the total length (Fig. 3-1). Error ratios with the theoretical shapes such as elliptic (Ell), circular (Cir), rectangular (Rec) report values being further to 0 the less similar to each shape. Ovoid asymmetry (Ovo), vertical asymmetry (VA) and horizontal asymmetry (HA) give high values when they show high asymmetry. In strawberry fruit, ovoid shape (Ovo) is close to 0.4.

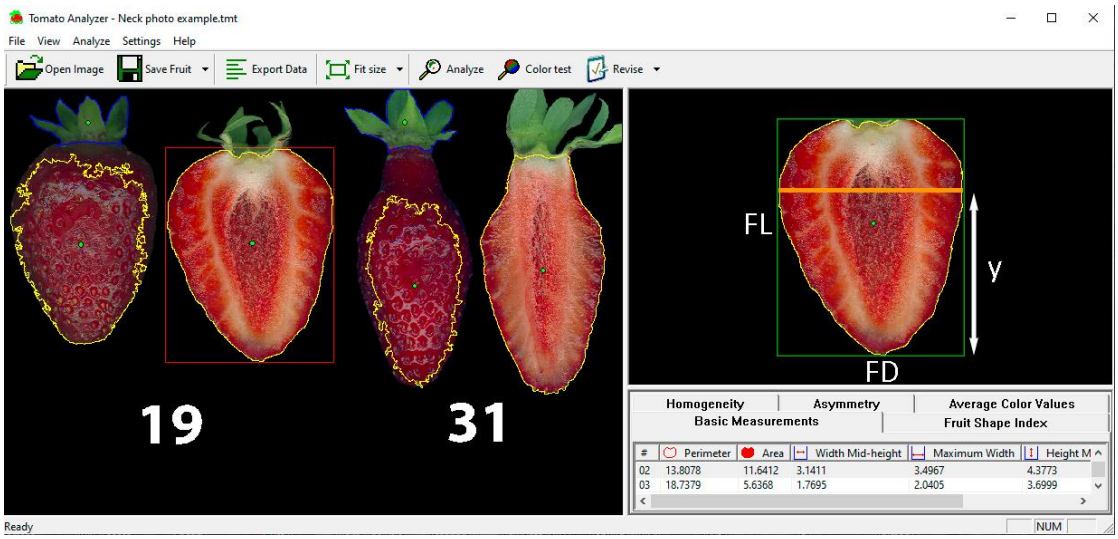


Figure 3-1. Example of strawberry cuts analysed with Tomato Analyser 4.0 software. Showing strawberry with neck and another without it. Parameter for FS (FL/FD) and WWP (y/FL) calculation.

In order to determine the internal (I), external (E) and puree (P) fruit colour, we chose the CIELab method, which is closer to human perception. CIELab colour detection consists of three axes: 'a', going from green (-a) to red (+a), 'b', going from yellow (-b) to blue (+b) and 'L', going from

darkness (-L) to lightness (+L). These axes range from -100 to 100 values. Hue (Hue) is the angle between 'a' and 'b' ranging from 0-360º and Chroma value (Chr) is also related with 'a' and 'b' values which represent the saturation or vividness of colour (Strecker *et al.*, 2010) (See Fig. I-9).

Visual neck phenotype

Since a neck phenotype was observed within 'FC50xFD54' progenies, a panel test characterised the scanned fruits from four harvests in 2017 and 2018 as presence or absence of neck. The proportion of fruits presenting neck in each progeny was used for QTL analysis.

Data analysis

The same R packages used in Chapter 2 were also used for fruit quality data analysis. In order to compare values of parental and 'H-21' lines in '21AF' population, we calculated the Tukey test using the 'HSD.test' function from 'agricolae' package (version 1.3-3).

QTL analysis

The reduced 'FC50xFD54' genetic map constructed with 1464 markers grouped in 28 LGs and covering a total size of 2273cM (see Chapter 1) and the '21AF' genetic map with 1457 markers grouped in 29 LGs and spanning 1808cM (see Annex 1) were used to perform Interval Mapping (IM) and non-parametric Kruskal-Wallis test (KW) using MapQTL®6 software (van Ooijen, 2009). As in Chapter 2, the QTLs mapped with a minimum LOD score of 2.5 in at least 2 harvests were considered as significant and stable. QTLs showing LOD score in IM higher than four were considered as major, whereas QTL stability was classified as highly stable if it was detected in three or four harvests and totally stable QTL if it was detected in five or six harvests for 'FC50xFD54' population. Since we phenotyped '21AF' in three harvests, stability was established as highly stable if QTLs were present in two years and totally stable if they were present in three years. Markers with highest KW test in different harvests were considered as the most linked to each trait.

Results

Fruit quality trait distribution

To explore the segregation of fruit quality traits, we phenotyped two biparental populations, an F1 population (named, 'FC50xFD54') and an F2 population (named, '21AF') during two or three years (Supplementary table 3-1). As far as we know, there is not pedigree relation between the parental lines.

The analysis of the parental lines of the 'FC50xFD54' population displayed significant differences (p-value <0.05) for fruit weight being 'FD54' fruits weightier than 'FC50' fruits. However, it was not significance for SSC and acidity traits. Even though fruit length and diameter were not significantly different comparing both parental lines, fruit shape ratio was significantly higher in 'FC50' compared to 'FD54' parental line. Additionally, other shape traits, such as circular, ovoid, width-widest position and vertical asymmetry traits, showed significant differences between the parental lines (Fig. 3-2).

Concerning the colour traits, we observed significant differences in the internal colour, but not for the external and puree colour. 'FD54' was significantly brighter (LI) than 'FC50' and 'FC50' showed significantly higher reddish (aI) and chroma values than 'FD54'.

Once we recorded the differences between the parental lines, we studied the segregation of all these traits in the 'FC50xFD54' population (Fig. 3-2 and Table 3-1). The most remarkable finding is that the progeny '46' is completely white inside and pink outside (Fig. 3-2).

To overview the stability of these traits under different environmental condition, we performed correlation analyses for each trait and we observed a high stability between the different harvests except for soluble solids content and vertical asymmetry (Table 3-1).

Once we compared the parental lines of the '21AF' population, 'Camarosa' and 'Dover', we observed significantly higher values in fruit weight, rectangular shape approximation, internal Hue value and external L value in 'Dover' compared to 'Camarosa'. On the other hand, titratable acidity and puree b value were higher in 'Camarosa' compared to 'Dover'. Despite 29 of the 33 studied traits did not significantly differ between the parental lines and 'H-21', the hybrid was closer to 'Camarosa' in the other four traits (fruit weight, pH, titratable acidity and internal Hue value) (Fig. 3-3).

Since eleven traits were measured only in two harvests in '21AF', we did not calculate the standard deviation and range of the correlation. Despite that, almost half (17) of the total traits showed an average correlation between harvests greater that 0.30 (Table 3-2).

Figure 3-2. Longitudinal fruit cuts showing internal, external and puree colour for all the individuals of the 'FC50xFD54' population, parental and grandparental lines from harvest 2019_1.

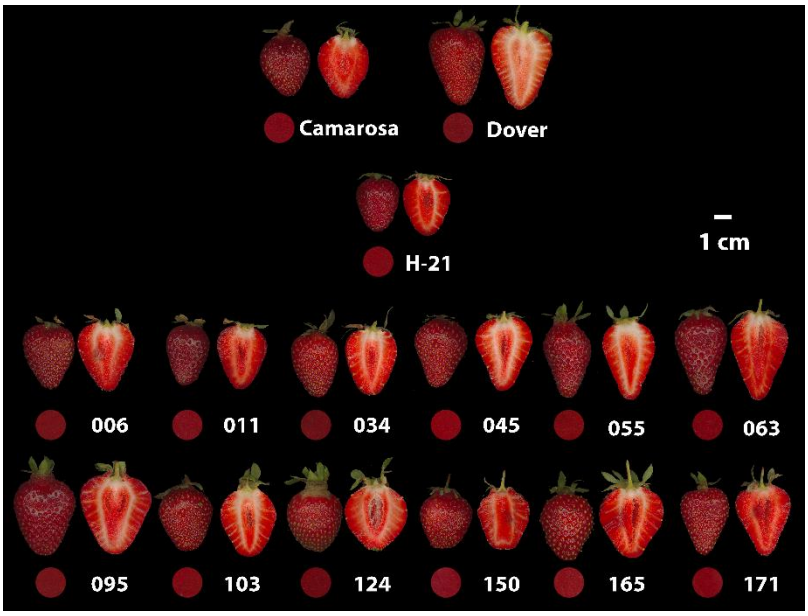


Figure 3-3. Longitudinal fruit cuts showing internal, external and puree colour for all the individuals of the '21AF' population, parental and 'H-21' lines from harvest 2016.

Table 3-1. Summary of 'FC50xFD54' fruit quality traits. Mean and standard deviation (SD) of each trait in 'FC50', 'FD54', and mean, SD and range of their F1 progeny and correlation between harvests in four or six harvests.

Traits	Abb.	FC50	FD54	FC50xFD54		Correlation	
		Mean ± SD	Mean ± SD	Mean ± SD	Range	Mean ± SD	Range
Fruit weight	FW	15.69 ± 1.50	**18.60 ± 1.50**	15.95 ± 3.95	5.29 - 29.36	0.45 ± 0.07	0.34 - 0.60
Firmness	FIR	298 ± 42	379 ± 87	356 ± 80	160 - 600	0.43 ± 0.11	0.24 - 0.62
pH	pH	3.38 ± 0.15	3.45 ± 0.13	3.48 ± 0.2	3.1 - 4.0	0.43 ± 0.07	0.35 - 0.56
Titratable acidity	TA	0.82 ± 0.13	0.77 ± 0.15	0.73 ± 0.14	0.31 - 1.11	0.47 ± 0.12	0.31 - 0.64
Soluble solids content	SSC	9.6 ± 0.6	9.7 ± 0.8	8.9 ± 1.3	5.3 - 13.1	0.20 ± 0.05	0.11 - 0.27
Sweetness-acidity ratio	SAR	119.08 ± 22.84	133.17 ± 42.17	126.24 ± 33.13	60.87 - 259.52	0.42 ± 0.11	0.27 - 0.60
Fruit perimeter	FP	12.96 ± 0.76	13.45 ± 2.09	13.4 ± 1.55	8.38 - 20.40	0.25 ± 0.08	0.08 - 0.51
Fruit area	FA	9.41 ± 1.38	9.78 ± 2.01	9.87 ± 1.70	4.40 - 16.78	0.32 ± 0.08	0.17 - 0.51
Fruit diameter	FD	2.97 ± 0.25	3.23 ± 0.33	3.10 ± 0.30	2.03 - 4.06	0.35 ± 0.09	0.17 - 0.52
Fruit length	FL	4.45 ± 0.38	4.29 ± 0.55	4.48 ± 0.52	2.96 - 6.14	0.49 ± 0.07	0.31 - 0.62
Fruit shape	FS	**1.51 ± 0.10**	1.33 ± 0.05	1.45 ± 0.17	1.00 - 2.01	0.70 ± 0.06	0.60 - 0.81
Width-widest position	WWP	0.36 ± 0.04	**0.40 ± 0.02**	0.37 ± 0.04	0.28 - 0.51	0.39 ± 0.06	0.28 - 0.54
Ellipsoid	Ell	0.06 ± 0.01	0.06 ± 0.01	0.06 ± 0.01	0.03 - 0.12	0.30 ± 0.11	0.08 - 0.61
Circular	Cir	**0.15 ± 0.03**	0.11 ± 0.01	0.14 ± 0.03	0.05 - 0.26	0.65 ± 0.05	0.55 - 0.77
Rectangular	Rec	0.42 ± 0.02	0.40 ± 0.01	0.41 ± 0.02	0.36 - 0.49	0.31 ± 0.09	0.11 - 0.53
Ovoid	Ovo	**0.28 ± 0.06**	0.22 ± 0.02	0.26 ± 0.06	0.00 - 0.41	0.45 ± 0.07	0.22 - 0.65
Vertical asymmetry	VA	0.04 ± 0.01	**0.05 ± 0.01**	0.05 ± 0.02	0.02 - 0.13	0.10 ± 0.08	-0.05 - 0.22
Horizonal asymmetry	HA	0.32 ± 0.12	0.23 ± 0.05	0.29 ± 0.10	0.00 - 0.62	0.38 ± 0.10	0.11 - 0.62
L internal	LI	43.52 ± 4.57	**52.26 ± 2.99**	46.13 ± 6.52	15.22 - 72.9	0.51 ± 0.12	0.27 - 0.79
a internal	aI	**33.88 ± 3.29**	25.33 ± 4.10	29.19 ± 5.11	4.55 - 41.05	0.68 ± 0.05	0.57 - 0.85
b internal	bI	24.06 ± 3.78	20.52 ± 5.44	22.42 ± 6.34	-0.01 - 38.61	0.23 ± 0.12	-0.05 - 0.55
Hue internal	HueI	35.38 ± 5.36	38.9 ± 8.28	37.51 ± 9.65	3.85 - 77.62	0.33 ± 0.19	0.01 - 0.86
Chroma internal	ChrI	**41.72 ± 3.12**	32.9 ± 4.86	37.32 ± 5.50	14.51 - 56.18	0.55 ± 0.08	0.30 - 0.79
L external	LE	26.15 ± 4.48	25.25 ± 3.70	24.76 ± 3.35	14.34 - 48.60	0.11 ± 0.23	-0.30 - 0.60
a external	aE	32.56 ± 4.02	29.86 ± 2.17	30.32 ± 3.23	14.72 - 41.16	0.36 ± 0.13	0.17 - 0.67
b external	bE	12.93 ± 4.14	9.74 ± 4.58	12.24 ± 5.33	0.33 - 32.37	0.19 ± 0.13	-0.03 - 0.42
Hue external	HueE	21.08 ± 5.38	47.58 ± 67.57	21.04 ± 8.24	0.64 - 42.86	0.08 ± 0.15	-0.16 - 0.39
Chroma external	ChrE	35.23 ± 4.81	31.70 ± 2.76	33.23 ± 3.95	15.85 - 50.29	0.37 ± 0.12	0.19 - 0.58

Traits	Abb.	FC50	FD54	FC50xFD54		Correlation	
		Mean ± SD	Mean ± SD	Mean ± SD	Range	Mean ± SD	Range
L puree	LP	25.29 ± 4.74	26.17 ± 2.87	27.50 ± 4.42	17.42 - 44.38	0.57 ± 0.07	0.45 - 0.75
a puree	aP	36.34 ± 2.09	36.86 ± 0.91	35.47 ± 5.12	9.13 - 43.22	0.73 ± 0.03	0.66 - 0.79
b puree	bP	24.12 ± 5.06	25.25 ± 3.64	24.55 ± 5.11	10.68 - 37.34	0.46 ± 0.12	0.33 - 0.83
Hue puree	HueP	33.35 ± 5.25	34.29 ± 4.07	34.61 ± 4.69	22.33 - 54.59	0.55 ± 0.12	0.36 - 0.90
Chroma puree	ChrP	43.76 ± 3.67	44.76 ± 1.96	43.26 ± 6.44	14.05 - 55.49	0.65 ± 0.06	0.55 - 0.77

Bold and underline numbers indicate the highest value in parental lines ‘FC50’ and ‘FD54’ when detected significant differences $p<0.05$ between them.

Table 3-2. Summary of '21AF' fruit quality traits. Mean and standard deviation (SD) of each trait in 'Camarosa', 'Dover and 'H-21', and mean, SD and range of their F2 progeny and correlation between harvests in two or three harvests.

Traits		Camarosa	Dover	H-21	21AF		Correlation	
		Mean ± SD	Mean ± SD	Mean ± SD	Mean ± SD	Range	Mean ± SD	Range
Fruit weight	FW	12.58 ± 1.15	**18.42 ± 0.68**	12.17 ± 2.17	10.80 ± 2.05	5.39 - 16.61	0.46	
Firmness	FIR	535 ± 67	467 ± 18	414 ± 19	451 ± 65	315 - 624	0.53	
pH	pH	3.6 ± 0.07	3.6 ± 0.0	3.4 ± 0.0	3.5 ± 0.1	3.3 - 3.7	0.38	
Titratable acidity	TA	**0.94 ± 0.04**	0.68 ± 0.06	1.03 ± 0.07	1.01 ± 0.15	0.59 - 1.52	0.34	
Soluble solids content	SSC	8.9 ± 0.9	7.3 ± 1.6	7.1 ± 0.5	8.2 ± 1.3	4.3 - 12.3	0.38	
Sweetness-acidity ratio	SAR	96.2 ± 15.42	110.16 ± 32.86	69.14 ± 9.67	82.99 ± 17.89	40.88 - 130.51	0.44	
Fruit perimeter	FP	12.29 ± 1.42	13.81 ± 0.61	12.27 ± 1.59	11.35 ± 1.13	6.25 - 14.35	0.14 ± 0.04	0.10 - 0.21
Fruit area	FA	8.85 ± 2.07	10.94 ± 1.16	8.71 ± 2.08	7.36 ± 1.37	2.60 - 11.44	0.15 ± 0.03	0.11 - 0.19
Fruit diameter	FD	3.21 ± 0.14	**3.93 ± 0.56**	3.26 ± 0.28	2.88 ± 0.27	1.84 - 3.66	0.19 ± 0.04	0.15 - 0.25
Fruit length	FL	3.76 ± 0.67	4.38 ± 0.25	3.86 ± 0.70	3.5 ± 0.44	2.24 - 4.90	0.32 ± 0.07	0.22 - 0.39
Fruit shape	FS	1.17 ± 0.16	1.28 ± 0.03	1.22 ± 0.12	1.23 ± 0.14	0.83 - 1.72	0.57 ± 0.02	0.53 - 0.58
Width-widest position	WWP	0.40 ± 0.05	0.35 ± 0.04	0.35 ± 0.03	0.36 ± 0.04	0.23 - 0.51	0.37 ± 0.06	0.30 - 0.45
Ellipsoid	Ell	0.06 ± 0.01	0.07 ± 0.01	0.07 ± 0.01	0.07 ± 0.01	0.03 - 0.10	0.39 ± 0.11	0.23 - 0.55
Circular	Cir	0.09 ± 0.03	0.11 ± 0.01	0.11 ± 0.02	0.10 ± 0.02	0.05 - 0.19	0.47 ± 0.04	0.41 - 0.53
Rectangular	Rec	0.43 ± 0.00	**0.44 ± 0.00**	0.43 ± 0.02	0.44 ± 0.03	0.37 - 0.51	0.35 ± 0.06	0.29 - 0.43
Ovoid	Ovo	0.24 ± 0.07	0.30 ± 0.05	0.29 ± 0.05	0.28 ± 0.06	0.06 - 0.49	0.4 ± 0.08	0.29 - 0.52
Vertical asymmetry	VA	0.07 ± 0.01	0.09 ± 0.03	0.08 ± 0.01	0.07 ± 0.02	0.02 - 0.13	-0.02 ± 0.09	-0.16 - 0.07
Horizonal asymmetry	HA	0.23 ± 0.11	0.34 ± 0.05	0.32 ± 0.10	0.26 ± 0.08	0.04 - 0.51	0.42 ± 0.12	0.25 - 0.56
L internal	LI	48.11 ± 7.84	51.93 ± 4.35	43.8 ± 7.05	44.20 ± 5.93	27.8 - 62.93	0.16 ± 0.04	0.09 - 0.21
a internal	aI	35.34 ± 5.20	28.18 ± 3.44	36.84 ± 2.33	33.21 ± 4.39	19.01 - 43.19	0.32 ± 0.11	0.20 - 0.48
b internal	bI	30.19 ± 3.69	27.53 ± 3.83	30.41 ± 4.24	28.32 ± 3.73	15.74 - 38.97	0.20 ± 0.09	0.08 - 0.34
Hue internal	HueI	40.84 ± 1.10	**44.72 ± 1.17**	39.50 ± 2.86	40.45 ± 3.34	26.28 - 53.88	0.18 ± 0.02	0.14 - 0.20
Chroma internal	ChrI	46.57 ± 6.27	39.47 ± 5.05	47.84 ± 4.22	43.77 ± 5.22	28.02 - 56.39	0.29 ± 0.10	0.15 - 0.42
L external	LE	24.10 ± 3.08	**28.03 ± 0.73**	24.39 ± 1.80	25.76 ± 3.00	18.44 - 39.64	0.16 ± 0.04	0.10 - 0.20
a external	aE	26.00 ± 2.44	29.04 ± 3.91	26.78 ± 1.95	24.98 ± 3.87	12.15 - 35.14	0.32 ± 0.07	0.21 - 0.43
b external	bE	13.91 ± 1.96	18.86 ± 4.83	13.90 ± 4.76	15.28 ± 5.23	-5.81 - 37.36	0.19 ± 0.02	0.17 - 0.22
Hue external	HueE	27.49 ± 4.35	31.95 ± 4.02	26.68 ± 6.88	37.68 ± 28.57	12.09 - 345.63	0.07 ± 0.11	-0.08 - 0.24
Chroma external	ChrE	29.66 ± 2.50	34.90 ± 5.58	30.42 ± 3.57	29.74 ± 5.39	15.26 - 47.42	0.3 ± 0.06	0.24 - 0.39

Traits		Camarosa	Dover	H-21	21AF		Correlation	
		Mean ± SD	Mean ± SD	Mean ± SD	Mean ± SD	Range	Mean ± SD	Range
L puree	LP	24.58 ± 2.56	24.65 ± 3.67	22.47 ± 1.93	22.72 ± 2.44	15.48 - 28.60	0.52	
a puree	aP	35.63 ± 3.80	31.97 ± 2.41	36.14 ± 0.49	36.09 ± 1.99	28.62 - 40.87	0.23	
b puree	bP	**22.52 ± 0.71**	19.73 ± 0.98	24.64 ± 4.91	23.25 ± 3.87	12.56 - 36.13	0.10	
Hue puree	HueP	32.27 ± 3.36	31.86 ± 3.39	33.88 ± 5.41	32.47 ± 3.54	23.69 - 43.47	0.01	
Chroma puree	ChrP	42.26 ± 2.73	37.60 ± 1.52	43.90 ± 2.41	43.05 ± 3.41	31.25 - 52.89	0.18	

Bold and underline numbers indicate the highest value in parental lines ‘Camarosa’ and ‘Dover’ when detected significant differences $p<0.05$ between them.

SD and range correlation calculated from two harvest could not be calculated.

To check if each trait is normal distributed in both populations, we applied the Shapiro-Wilk test. The results for the 'FC50xFD54' population indicated that most traits (26) were normally distributed with the exception of pH and colour traits (aI, aP, HueI, HueE, HueP and LP), whereas the '21AF' population showed that pH, vertical asymmetry (VA) and colour traits (LE, HueE, HueP and aP) did not have a normal distribution. Additionally, six colour traits (bI, bE, HueI, HueE, LE and LP) showed a skewed distribution between different harvests in both populations (Fig. 3-4).

Taking into account that both populations have totally different genetic background and were cultivated in different locations, 'FC50xFD54' population showed higher values in ten traits (FW, SSC, SAR, FP, FA, FD, FL, FS, WWP and Cir), but lower in other five (FIR, TA, Ell, Rec and VA) compared to '21AF' population (Fig. 3-4).

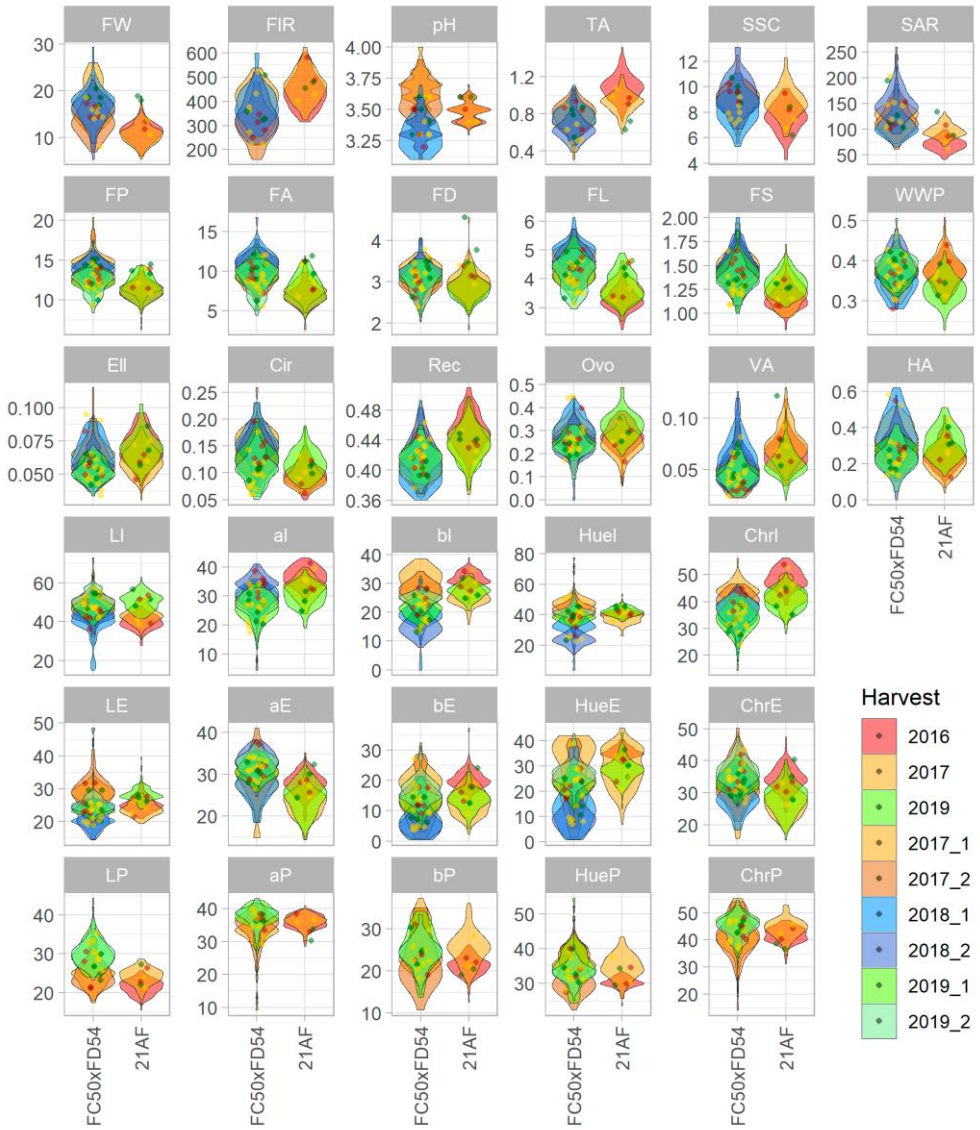


Figure 3-4. Violin plot of fruit quality trait distribution in 'FC50xFD54' and '21AF' populations. Harvest: 2016 (reddish), 2017 (orangish), 2018 (bluish), 2019 (greenish). Dots: mother (red), father (green) and grandparents in 'FC50xFD54' population and 'H-21' in '21AF' population (yellow).

Relations between fruit quality traits in each population

To explore the relationship between traits in different harvests, we performed a correlation analysis in both populations using Pearson coefficient (Supplementary table 3-2 and ST. 3-3). Looking at taste traits, we observed high correlation between the acidity traits (pH and TA) as we expected. Moreover, SAR, calculated from SSC and TA, also showed a strong correlation with these traits. Firmness is an important trait for strawberry commercial quality, but it was not correlated with any studied trait.

Concerning the shape analysis, FW was highly correlated with the direct shape measurements FP, FA, FD and FL, that also showed a high correlation with each other. Fruit length showed a stronger and positive correlation with fruit shape ratio than the fruit diameter, indicating that FL is the most important factor for FS. In addition, we observed a negative correlation between direct measurements and the approximate shape traits, such as WWP ratio with Ell and HA. Moreover, WWP ratio were also negatively correlated with Ovo. Since Cir showed a high correlation with FL and FS, the lengthier fruits are also the less circular in these two populations.

In relation to colour traits, internal reddish (aI) is negatively correlated with brightness (LI) and weakly correlated with yellowish (bI). HueI and ChrI, calculated using aI and bI values, showed a stronger correlation with aI than with bI. At the same time, LI was negatively correlated with ChrI and positively correlated with HueI. Each external trait was positively correlated with other external traits. The colour of the puree made with the whole fruit showed a behaviour similar to the one of the internal colour traits, explained by the higher abundancy of internal tissue compared with the external one.

Moreover, a high correlation was observed between acidity and internal colour in 'FC50xFD54' population, such as in the case of TA and aI, and also between acidity and puree colour, such as TA and aP, indicating that red colour is positively linked to acidity (Fig. 3-5 and Supplementary table 3-2 and ST. 3-3).

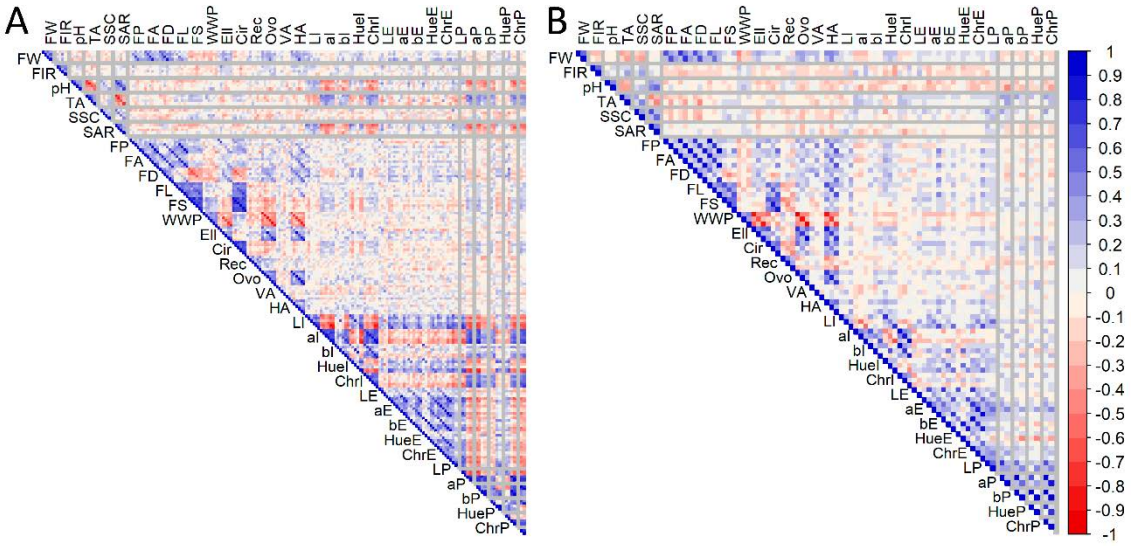


Figure 3-5. Heatmap visualization of Pearson correlation between fruit quality traits in different harvests. A. 'FC50xFD54' population data of six harvests (2017-19) B. '21AF' population data in three harvest (2016, 2017 and 2019).

To better visualize the relationship between traits, we performed a Cluster network analysis (CNA) which clearly divided fruit quality traits in two clusters, one for colour and acidity and the other for shape and weight. These two clusters exhibited a strong correlation inside each group. Since FIR is not related to any studied trait and VA and Rec are not really suitable for strawberry shape, they showed a weak correlation with other traits (Fig. 3-6).

High correlation, positive or negative, between each trait was similar than the ones seen in Fig. 3-5. Moreover, few trait correlations showed differences comparing to both populations. In 'FC50xFD54', a moderate correlation was observed between pH and colour traits (internal, puree, LE and bE) (Fig. 3-6A), whereas, in '21AF', pH is shown outside of the network (Fig. 3-6B). As expected, a high correlation between SSC and SAR was observed in 'FC50xFD54' population (Fig. 3-6A) and, in '21AF', SSC showed a positive correlation with SAR and LP (Supplementary table 3-4).

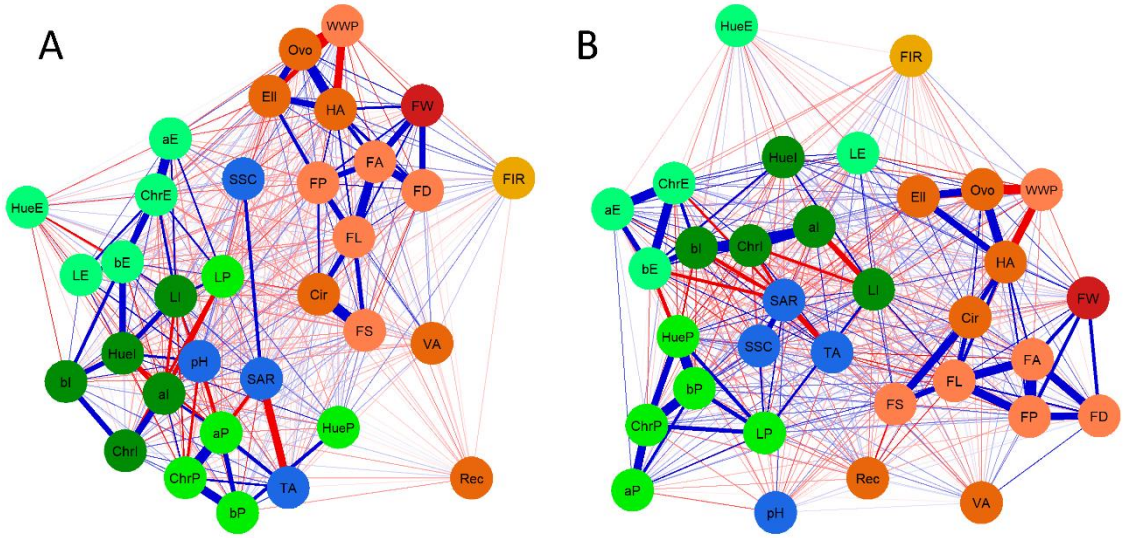


Figure 3-6. Cluster network analysis (CNA) for fruit quality traits. Graphical representation of the correlations between studied traits. A. 'FC50xFD54' and B. '21AF'. Traits are represented as nodes and coloured according to sweetness and acidity (blue), firmness (yellow), shape (reddish) and colour (greenish). Positive (blue) and negative (red) correlations are represented as link between traits, wider links represent stronger correlation.

With the aim to better understand the relationship between traits and individuals, a heatmap representation of hierarchical clustering analysis in 'FC50xFD54' population grouped the studied traits in four main groups. Cluster 1 grouped L and Hue from internal and puree colour, with external colour traits. Cluster 2 consisted of FIR and shape traits (FW, FD, Ovo, HA, Elli and Rec). Cluster 3 represented some of the shape traits (WWP and VA) and taste traits (pH, SSC and SAR). And finally, cluster 4 grouped a, b and Chr traits from internal and puree colour with other shape traits (FL, FS, Cir and FA) and the taste trait TA (Fig. 3-7).

To select the best progenies for future crosses, it is necessary to know the phenotypic relation between the lines. Four main clusters (A, B, C and D) were observed within individuals. Cluster A included 16 progenies characterised by having high values of cluster 3, corresponding to organoleptic traits, and some shape traits from cluster 4. Cluster B was the largest cluster and we divided it in two subgroups, B1 and B2. Cluster B1 was the biggest cluster and includes 'FC50', 'FD54', 'FD019' and 33 progenies showing high values of internal and puree colours. Cluster B2, comprises the 'FC030' grandparental line and ten progenies and have high values of cluster 2 and 4. Cluster C, with two grandparental lines, 'FD197' and 'FD016', and five progenies, was characterised by having high values of cluster 1 and 3 and low values of cluster 4. Finally, the '46' progeny clustered alone in cluster D at a huge distance from the other progenies. Since this progeny presents an inner white fruit colour and different shape (Fig. 3-2), it was characterised by having extreme colour values and low TA, FL, Cir and FS (Fig. 3-7).

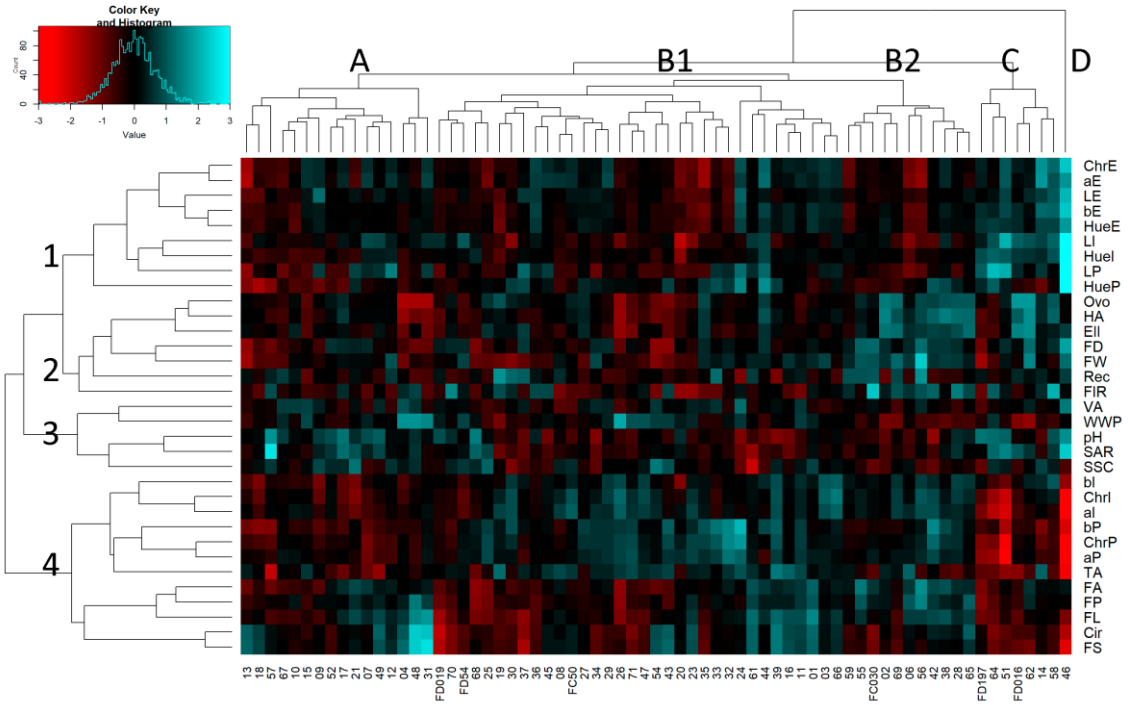


Figure 3-7. Hierarchical clustering analysis (HCA) and heatmap for 'FC50xFD54' population. HCA and dendrogram was calculated by Euclidean distance and complete clustering. Clusters are indicated by capital letters for individuals and by numbers for traits. 'FC50' and 'FD54' are parental lines and 'FC030', 'FD197', 'FD019' and 'FD016' are grandparental lines.

The phenotypic structure of the '21AF' population was explored performing a hierarchical clustering that provided three main groups for progeny lines (letters) and three clusters for traits (numbers). Cluster 1 was a diverse and large group consisting of shape traits (FA, FP, FL, FW, FD, Elli, Ovo, HA and VA), internal colour (HueI and LI) and external colour traits. Cluster 2 was the smallest group and comprised four traits, (FIR, TA and internal colour, aI, bI and ChrI). Finally, cluster 3 consisted of organoleptic traits (pH, SSC and SAR), fruit shape (FS, Rec, Cir and WWP) and puree colour traits. Furthermore, internal, external and puree colour traits were grouped in different clusters.

Concerning to the '21AF' progenies, we separated the two big clusters, A and B, in two subgroups each. Any ancestor lines were clustered in A. Cluster A1, consisted of four progenies, was characterised by high organoleptic and firmness traits and low values of traits from cluster 1, 2 and some from cluster 3. Cluster A2, including 30 progenies, had high values traits from cluster 2 and puree colour. Cluster B1, grouping 'Camarosa' and 35 progenies, was characterised by having low colour value traits and high FIR, pH, WWP and Rec values. The largest cluster B2, that comprised the hybrid 'H-21' and 48 progenies, had high values from cluster 1, low values from cluster 2 and different value degree from cluster 3. Finally, the parental line 'Dover' was

clustered alone in C, far from other lines, and was differentiated by having high values of shape traits such as FA, FP, FW and FD and low values for internal and puree colours (Fig. 3-8).

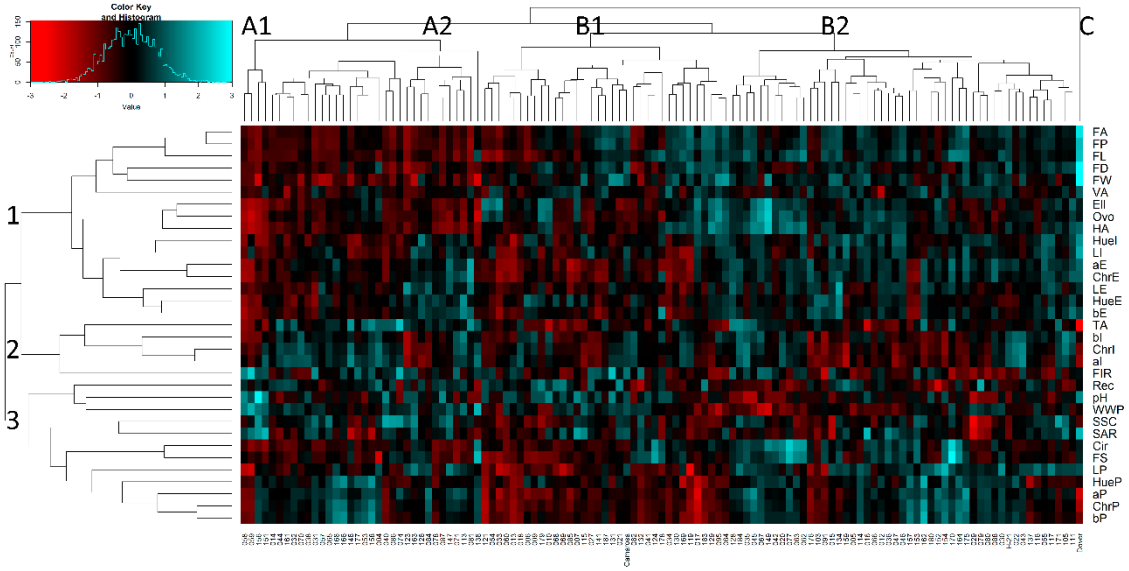


Figure 3-8. Hierarchical clustering analysis (HCA) and heatmap for '21AF' population. HCA and dendrogram were calculated by Euclidean distance and complete clustering. Clusters are indicated by capital letters for individuals and numbers for traits. 'Camarosa' and 'Dover' are parental lines and 'H-21' the hybrid.

Analyses of the environmental effect

Since plants were grown in open field and they are subject to environmental changes, a principal component analysis (PCA) and loading plot were performed in 'FC50xFD54' population to clarify which harvest showed the highest differences and which traits contributed the most (Fig. 3-9). Considering that we had different data from different years, we separated them in two groups, one group with FW, FIR, TA, pH, SSC and SAR from 2017 and 2018 (Fig. 3-9A1 and A2) and another group with colour and shape from 2017 and 2019 (Fig. 3-9B1 and B2). Since puree data from 2018 was not available, this year was discarded in the colour and shape analysis.

The three main dimensions explained 37.7% (Dim1), 21.6% (Dim2) and 15.6% (Dim3) of taste traits variance, whereas lower percentages of variance were observed in colour and shape dimensions being 24% (Dim1), 18% (Dim2) and 11% (Dim3). Different harvests in PCA plots were overlapping, indicating a little contribution of the environmental effect to each trait (Fig. 3-9 A1 and A2).

Regarding the trait contribution, FW and FIR presented the same direction in Dim2 but not observed in Dim3. A cluster with SSC, pH and SAR and almost opposite to TA was constant in both PCA plots (Fig. 9A1 and A2). Looking into the appearance traits, fruit shape traits were

located in one diagonal and colours traits in an opposite diagonal in Dim 2 (Fig. 9B1), while it was mixed in Dim 3 (Fig. 9B2).

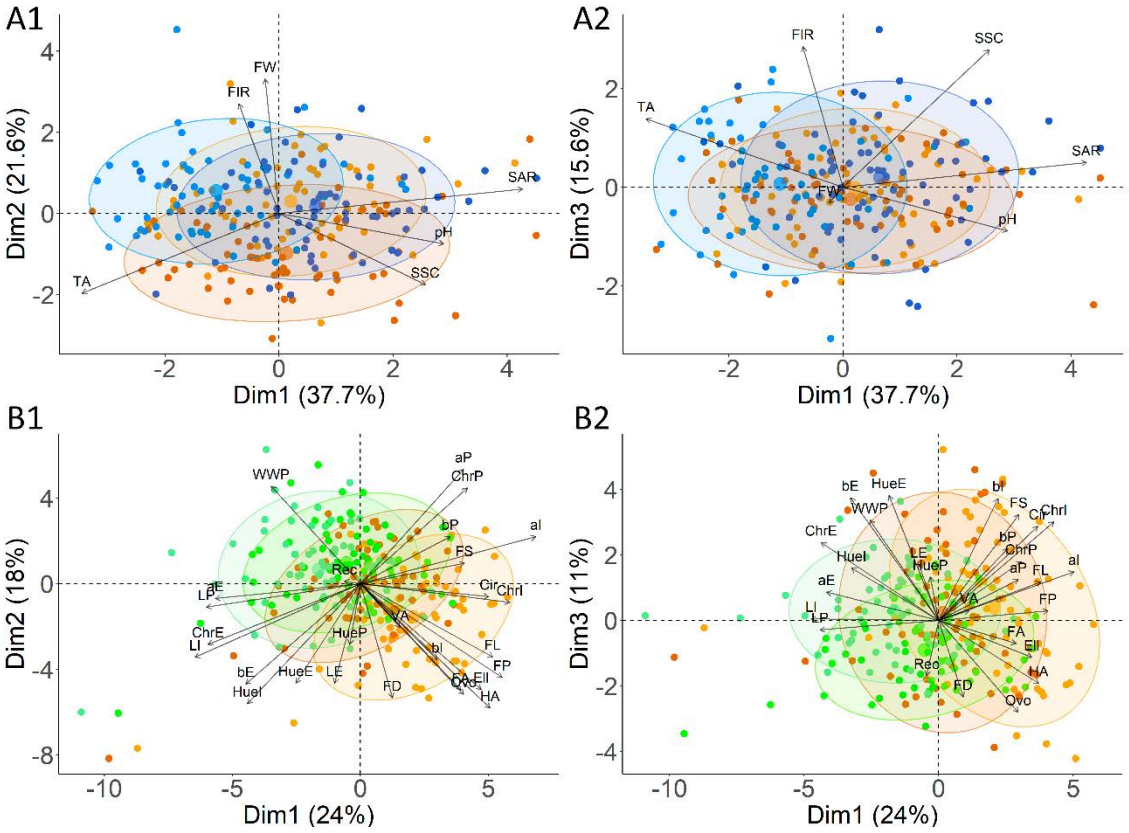


Figure 3-9. PCA and loading plot for fruit quality traits in 'FC50xFD54' population. A. PCA for fruit weight, firmness, acidity and sweetness traits: A1) Dim1 vs Dim2 and A2) Dim 1 vs Dim3. B. PCA for colour and shape traits B1) Dim1 vs Dim2 and B2) Dim 1 vs Dim3. Colours indicate harvests: 2017 (orangish), 2018 (bluish) and 2019 (greenish).

To further investigate the source of this phenotypic variance, we performed an analysis of variance (ANOVA) to reveal the Genetic factor (G), Environment factor (E) and interaction between them (GxE) in both populations.

Focusing on 'FC50xFD54 population, we observed that the G factor had a significant (p-value<0.05) contribution to the phenotypic variance in almost all trait (28), being the unique significant factor in SSC, FA, FD and Rec traits. Despite the high G factor found in observed variance, only eight traits did not showed significance in the environmental factor. Hence, twenty-three traits showed significant G and E factors. Only colour puree traits were significant in the GxE factor. Furthermore, we could not observe any significant factor in HueI, bE and HueE traits (Supplementary table 3-4).

To explain which percentage of the variability is caused by these factors, ω^2 test showed that G factor was the main contributor to the phenotypic variance over E and GxE factors. Specifically,

nineteen out of 33 traits showed that more than 25% of observed variance attributed to the G factor. However, we found especially low explanation of G factor in bI, HueI, LE, bE and HueE. Furthermore, we observed more than 30% of phenotypic variance was explained by E factor in pH and LP and GxE factor was nearly imperceptible in all traits. Puree colour traits, calculated in two years, had the highest G factor. However, a high proportion of the phenotypic variance cannot be explained by these three factors and we named 'error' (Fig. 3-10).

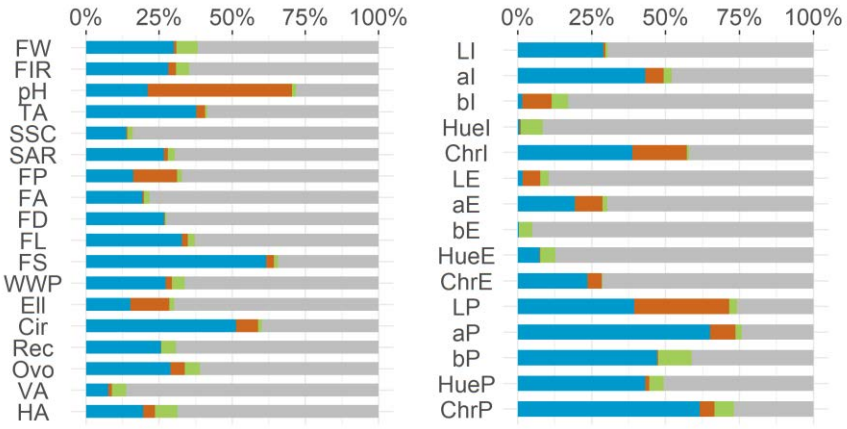


Figure 3-10. Omega square values (ω^2) for fruit quality traits in 'FC50xFD54' population: percentages of variance attributed to G factor (blue), E factor (orange), GXE factor (green) and error (grey).

Despite having few harvests to explore the environment effect in the '21AF' population, the three main dimensions from PCA explained similar phenotypic variance observed in 'FC50xFD54' population. However, a clear harvest dependence was showed in both PCAs (Data not shown). Studying the traits with three-harvest data, we found most of them being significant (p-value<0.05) environmental effect and 11 of them with significance for both G and E factors. In addition to the importance of the environmental effect, the G factor was high in some shape traits, especially in FS, Cir and Rec, explaining more than 35% of the observed variance (Fig. 3-11).

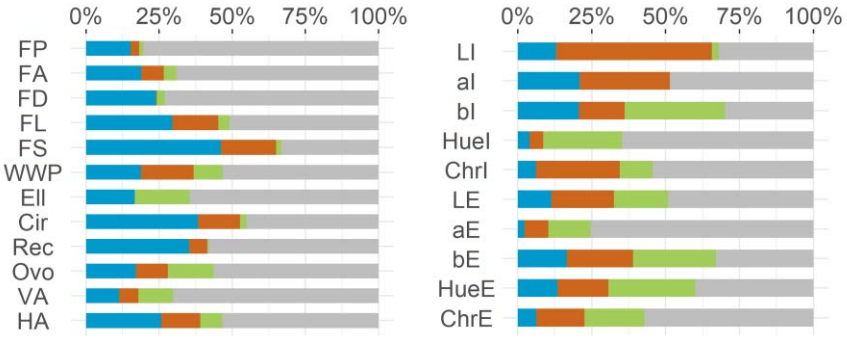


Figure 3-11. Omega square values (ω^2) for fruit quality traits in '21AF' population: percentages of variance attributed to G factor (blue), E factor (orange), GXE factor (green) and error (grey).

QTL analysis

'FC50xFD54' QTLs

In order to assess the genomic regions responsible for the phenotypic variability on the fruit quality traits, we performed a QTL analysis in both studied populations. Since fruits collected in different harvests are assumed to have completely different environmental characteristics, we have analysed each harvest independently to assess QTL stability.

The 'FC50xFD54' population was phenotyped during six different harvests and we were able to map a total of 77 QTLs (LOD score > 2.5 in at least two harvest). However, nearly one quarter (21) was highly stable and only two QTLs were stable in five different harvests (Table 3-3, Fig. 3-12 and Supplementary table 3-5). Additionally, 25 of the stable QTLs were considered as major QTLs (LOD>4).

Of the total stable QTLs, nine corresponded to weight, firmness and taste traits, being only one of them a major QTL. We could detect 28 different QTLs for shape traits being eight of them major QTLs. The majority of the QTLs (40) were mapped for colour traits, finding 16 major QTLs. Furthermore, QTLs were mainly located in only four different LGs, indicating that LG1A and LG1B are important for colour traits and LG4C and LG6C for shape traits.

As 'FC50xFD54' population showed a neck phenotype segregation within its progenies, a visual phenotype analysis was done by a panel test. Correlations between harvests ranged from 0.45 to 0.67 and between panellists were even higher (0.53 – 0.82). Moreover, this trait showed similar values of high positive correlation with FS and WWP ratios and negative correlation with Ovo, Ell, HA, Rec, FW and FD measures. QTL analysis revealed two stable QTLs, *Neck_3A* and *Neck_4B,* explaining a maximum of observed variance of 31.2% and 28.1%, respectively.

Of all the stable QTLs, only two were related to fruit weight and both of them were located at the beginning of LGs, such as *FW_1A* and *FW_3A* explaining 21.3-27.8% and 10-13% of the variance, respectively. The only QTL for firmness was detected in *FIR_7C*, being highly stable and explaining a range of variance between 17.8 and 26.9%.

Looking to QTLs for strawberry taste, a total of six QTLs related to acidity were mapped. TA and SAR QTLs were located in the middle of LG1A, being *SAR_1A*, a major QTL explaining a maximum of 33.4% of the phenotypic variance. From the three QTLs detected for pH, *pH_5C* was the most stable and explained 13.6-25.8% of the acidity variance (Table 3-3 and Fig. 3-12). Despite the SSC variability, any related genetic region could be detected.

Regarding the QTL analysis of the fruit shape traits, we identified one highly stable and significant QTL related to the diameter of the fruit in the middle of LG1A, *FD_1A*. Four different QTLs were mapped in the middle of LG2A related to horizontal axis, such as *FD_2A*, *Ovo_2A*, *Ell_2A* and *HA_2A*. Of them, *HA_2A* is a major QTL explaining a maximum of 29.3% of the observed variance. One of the two fruit area QTLs was a stable and major QTL, *FA_2D*, and explained between 13.0 and 34.4% of the variance. If we look at LG level, five different QTLs were mapped in three different positions along LG3A, with an accumulation of three of them at the beginning of the LG (*FW_3A* and two major QTLs *Neck_3A* and *WWP_3A*). In the middle of LG3A, we located a highly stable QTL for fruit shape ratio, *FS_3A*, explaining 10.9-24.8% of the variance and, at the distal part, one stable QTL for fruit diameter were detected. Furthermore, two QTLs for fruit shape were mapped in HGs, such as the stable *FS_4C* QTL (11.9-26.9% of phenotypic variance) and the major *FS_4D* QTL, that explains a maximum of 29% of the phenotypic variance. Two highly stable and significant fruit shape related QTLs were also mapped in the *FS_4C* position, the *WWP_4C* (10.9-28.3% of phenotypic variance) and the *HA_4C* (6.4-25.1% of variance), together with a highly stable and major QTL, *Ovo_4C* (11.6-36.8% of observed variance) and the *Neck_4C* QTL. Five different QTLs were mapped at the end of LG6D, where we also mapped the only fruit length QTL, *FL_6D*, being highly stable and contributing with 14.1-24.5% of the observed variance of vertical axis. Additionally, *FS_6D*, *Cir_6D, Rec_6D* mapped in the same place than *FL_6D.* Despite *Ovo_6D* was not exactly in the same place, it was close to them (Table 3-3 and Fig. 3-12).

Focusing on QTL analysis for colour traits, internal and puree QTLs were generally mapped in the same places. A cluster of eight QTLs for internal and puree traits were located in the first half of LG1A. They were highly stable and major QTLs and three of them were totally stable. Two different QTLs were mapped for LI in LG1A, *LI_1A.1* and *LI_1A.2* being the former highly significant but the latter highly stable. They explained 7.5-38.5 % and 8.3-30.6% of the brightness variance, respectively. Furthermore, a totally stable *aI_1A* and major *bI_1A* QTL, explaining a maximum of 37.8% and 34.8% of the observed variance respectively, were mapped at the same place than *ChrI_1A*. Regarding to puree QTLs in this region, we were able to detect the same traits found for internal colour, such as highly stable and major QTLs *LP_1A* (14.5-31.4%), *aP_1A* (14.7-41.7%) and a stable and major *ChrP_1A* QTL (maximum of 41.8% of variance). Additionally, some of these traits, four internal and one puree colour, were also mapped at the beginning of LG1B, being less stable than those in LG1A.

We were able to map only one cluster for external colour, *aE_3D, bE_3D* and *ChrE_3D*, located in LG3D and explaining a maximum of 22-29% of the observed variances.

Another cluster of four QTLs, one for the internal, two for puree and one for external colour traits, was mapped at the beginning of LG4B. Of them, two major QTLs were found for puree traits, *LP_4B* and *aP_4B,* explaining a maximum of 28.1% and 36% of the variance, respectively. Additionally, two internal QTLs, *b_4B* and *Chr_4B*, were mapped close to the previous cluster.

Furthermore, several major QTLs were mapped in different LGs, such as a highly stable and major *ChrI_2D* (7.4 - 36.2 %), a stable *bP_5B* (maximum of 31.5%) and stable *aE_5D* explaining a maximum of 30.4% of the variance (Table 3-3 and Fig. 3-12).

Table 3-3. List of fruit quality QTLs in 'FC50xFD54' population. Stable and significant QTLs ordered by traits and LG.

QTL ID	Harv.	LOD	% Expl.	KW	Genetic map		
					LG	QTL interval (cM)	Size (cM)
FW_1A	**4**	3.82	27.8	5.83	1A	0.00	22.31	22.31
FW_3A	2	3.12	23.8	11.31	3A	0.91	11.98	11.07
FIR_7C	**3**	3.68	26.9	5.51	7C	36.82	47.24	10.42
pH_5C	**3**	3.44	25.8	13.45	5C	0.00	2.73	2.73
pH_6C.1	2	3.46	25.5	15.82	6C	14.90	18.30	3.40
pH_6C.2	2	3.98	29.2	15.06	6C	91.81	93.63	1.82
TA_1A	2	3.83	27.8	3.97	1A	36.92	55.67	18.75
TA_5A	2	3.87	28.1	7.83	5A	52.62	57.79	5.17
SAR_1A	2	**4.77**	**34.4**	**17.60**	1A	39.92	49.17	9.24
FA_2D	**3**	**4.20**	**30.1**	**15.67**	2D	0.00	29.87	29.87
FA_5D	2	3.04	22.8	12.65	5D	57.15	72.24	15.09
FP_3B	2	3.28	24.0	14.13	3B	49.33	63.65	14.32
FL_6D	**4**	3.29	24.5	10.16	6D	52.85	68.53	15.68
FD_1A	**3**	2.97	22.4	7.14	1A	36.92	52.84	15.91
FD_2A	2	3.30	24.2	9.30	2A	21.21	57.33	36.11
FD_3A	2	3.05	22.9	12.21	3A	95.93	106.06	10.13
FS_3A	**3**	3.34	24.8	14.80	3A	31.75	62.32	30.57
FS_4C	**4**	3.68	26.9	11.81	4C	54.51	62.01	7.50
FS_4D	2	**4.15**	**29.3**	**14.66**	4D	8.57	21.42	12.85
FS_6A	2	2.96	22.0	9.93	6A	0.00	8.23	8.23
FS_6D	2	3.42	26.1	13.60	6D	49.22	64.83	15.61
WWP_3A	2	**4.10**	**29.1**	**14.53**	3A	4.88	18.36	13.48
WWP_4B	2	3.73	26.8	11.48	4B	56.49	61.90	5.41
WWP_4C	**3**	3.76	28.3	9.31	4C	54.51	62.01	7.50
Ell_2A	2	3.29	25.3	10.48	2A	25.77	48.27	22.50
Ell_4B	2	3.01	22.7	13.67	4B	56.49	63.77	7.28
Ell_4C	2	**4.22**	**31.2**	**8.80**	4C	55.51	61.20	5.68
Cir_6D	2	3.16	24.4	13.39	6D	49.22	63.83	14.61
Rec_6D	2	**4.84**	**33.8**	**13.67**	6D	57.39	61.15	3.76
Ovo_2A	2	3.08	22.7	11.16	2A	24.77	51.33	26.56
Ovo_4C	**4**	**5.19**	**36.8**	**10.10**	4C	54.51	63.01	8.50
Ovo_6D	2	3.77	27.1	4.92	6D	26.92	54.66	27.74
HA_2A	2	**4.14**	**29.3**	**12.55**	2A	14.05	51.33	37.28
HA_4A	2	3.76	27.4	10.03	4A	42.44	80.73	38.29
HA_4C	**3**	3.26	25.1	7.59	4C	54.51	62.01	7.50
Neck_3A	2	**4.47**	**31.2**	**14.41**	3A	0.00	19.16	19.16
Neck_4B	2	3.94	28.1	14.96	4B	57.48	73.96	16.48
LI_1A.1	4	**4.98**	**38.5**	**38.50**	1A	8.39	19.55	11.16
LI_1A.2	**5**	**4.29**	**30.6**	**30.60**	1A	47.35	56.67	9.32
LI_1B	2	**4.83**	**34.8**	**15.83**	1B	0.00	13.99	13.99
LI_3A	2	3.66	26.8	9.08	3A	1.82	19.16	17.34
LI_4B	2	3.14	24.3	7.63	4B	0.00	9.21	9.21

QTL ID	Harv.	LOD	% Expl.	KW	Genetic map			
					LG	QTL interval (cM)		Size (cM)
LI_6D	2	3.15	24.4	9.00	6D	60.15	64.83	4.67
aI_1A	**5**	**4.96**	**37.8**	**12.67**	1A	31.87	55.67	23.81
aI_1B	2	**4.99**	**36.1**	**18.70**	1B	0.00	11.28	11.28
bI_1A	2	**4.83**	**34.8**	**14.04**	1A	19.55	36.92	17.37
bI_1B	2	3.77	28.4	10.33	1B	0.00	22.93	22.93
bI_4B	2	3.54	26.0	8.52	4B	9.21	34.78	25.57
ChrI_1A	**5**	**4.96**	**38.4**	**14.10**	1A	16.10	56.67	40.57
ChrI_1B	**3**	**4.84**	**36.2**	**16.90**	1B	0.00	16.70	16.70
ChrI_2D	**3**	**4.25**	**29.9**	**16.43**	2D	51.36	64.16	12.80
ChrI_3A	2	3.09	23.9	9.13	3A	52.33	63.32	11.00
ChrI_4B	3	3.54	26.9	12.04	4B	9.21	34.78	25.57
ChrI_5C	2	3.66	26.8	18.73	5C	1.82	11.82	10.00
LE_4D	2	3.09	23.9	12.80	4D	67.15	78.91	11.77
aE_1D	2	3.05	22.5	11.34	1D	10.80	19.85	9.05
aE_3D	2	3.90	29.2	11.97	3D	10.33	41.01	30.68
aE_4B	2	2.91	22.7	10.76	4B	0.00	9.21	9.21
aE_5D	2	**4.10**	**30.4**	**15.43**	5D	6.44	11.87	5.43
bE_3D	2	2.84	21.2	12.31	3D	13.25	22.93	9.68
ChrE_1A	**3**	3.33	25.5	12.30	1A	75.08	85.27	10.19
ChrE_2A	2	2.81	22.0	11.25	2A	25.77	51.27	25.50
ChrE_3D	2	3.41	26.1	12.83	3D	14.25	22.93	8.68
ChrE_6D	2	3.39	24.7	12.12	6D	0.00	14.60	14.60
ChrE_7D	2	3.83	27.9	13.08	7D	47.69	56.92	9.23
LP_1A	**3**	**4.41**	**31.4**	**11.99**	1A	13.12	20.46	7.34
LP_1B	2	3.21	24.8	10.418	1B	0.00	12.99	12.99
LP_4B	2	**4.95**	**35.5**	**8.86**	4B	0.00	5.57	5.57
LP_6D	2	3.73	28.1	10.38	6D	60.15	66.64	6.49
aP_1A	**3**	**6.09**	**41.7**	**12.24**	1A	17.70	54.84	37.14
aP_4B	2	**5.29**	**36.3**	**16.30**	4B	0.00	9.21	9.21
bP_4C	2	2.81	21.3	11.10	4C	0.00	3.68	3.68
bP_5B	2	**4.43**	**31.5**	**9.75**	5B	25.09	36.95	11.86
bP_6C	2	2.79	21.9	5.694	6C	50.90	66.63	15.73
HueP_3C	2	2.98	22.4	13.232	3C	100.23	136.51	36.28
HueP_6B	2	3.13	24.2	11.038	6B	0.00	11.08	11.08
ChrP_1A	2	**6.11**	**41.8**	**12.975**	1A	17.70	54.84	37.14

Bold and underline numbers in harvest indicate highly stable QTLs and in LOD, % Expl. and KW test indicate major QTLs.

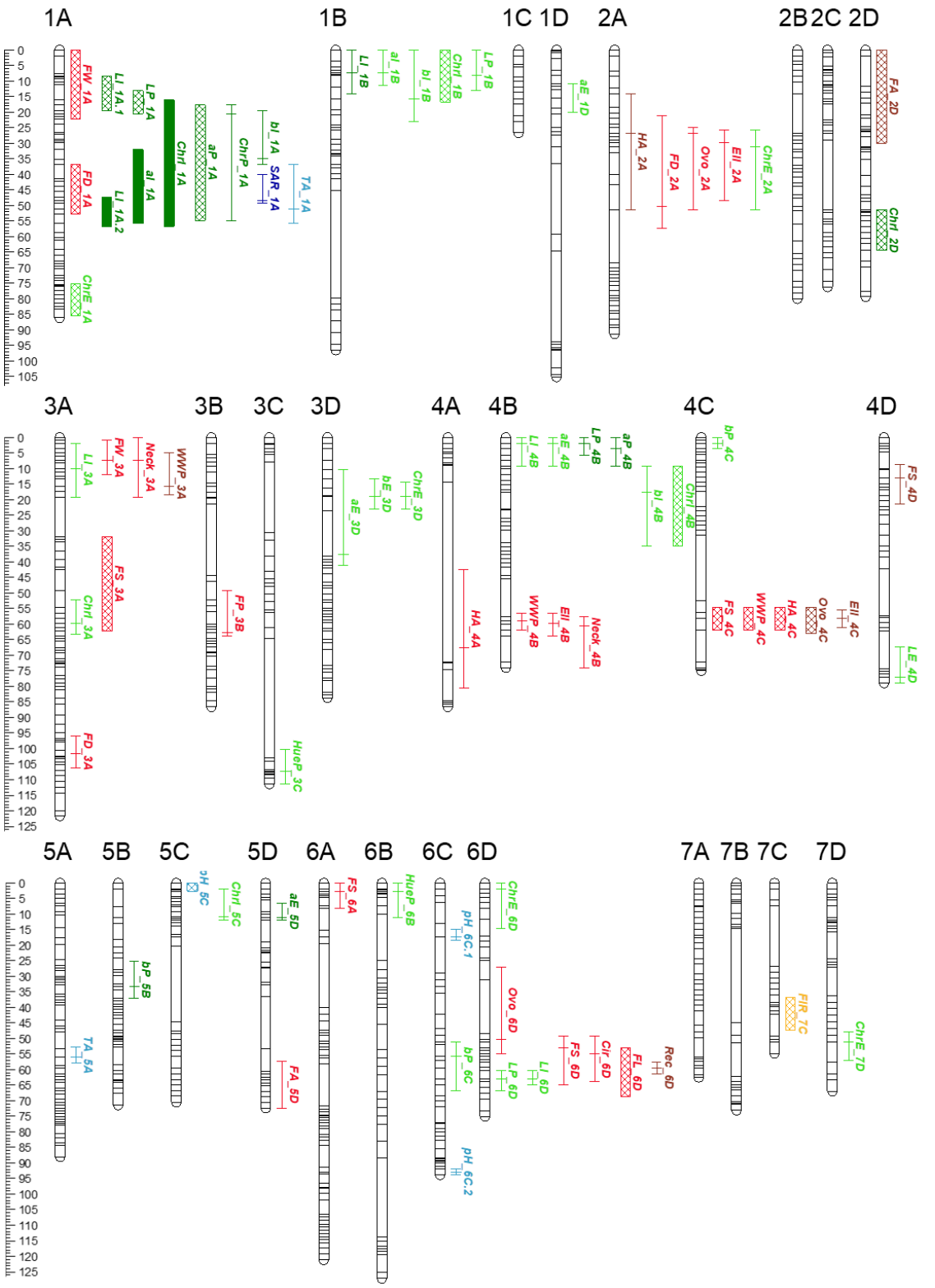


Figure 3-12. Fruit quality QTLs in 'FC50xFD54'. Colour indicates taste (blue), firmness (yellow), shape (red) and colour (green) QTLs. Dark colours indicate major QTLs. Degree of box filling means totally stable (filled), highly stable (semi-filled) and stable (lines).

The joint analysis of different traits mapped in the same genetic region could help to interpret the real effect of this QTL. Moreover, the study of trait variability within genotypes is also important to find which allele of a marker is linked to the high values of the trait.

Four different fruit shape QTLs were mapped in the same LG4C region, such as FS, WWP, Ovo and HA (Fig. 3-12). The selection of the Affx-88856304 marker located at 52.41cM of LG4C, which showed the highest Kruskal-Wallis values in different harvests, revealed that B allele was linked to higher values in fruit shape and WWP ratios and to less ovoid shape and horizontal asymmetry (Fig. 3-13).

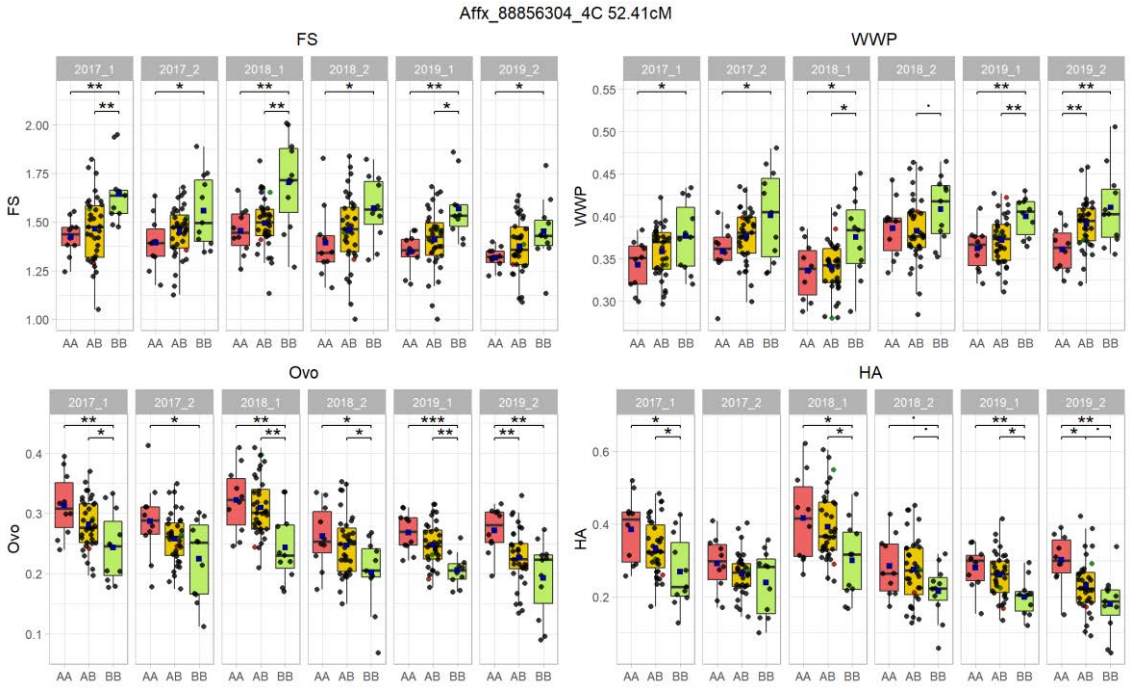


Figure 3-13. Boxplot for Affx-88856304 marker genotypes with shape traits values from LG4C QTLs. Boxplot colours indicate genotypes: AA (red), AB (yellow) and BB (green). Dot colours indicate: 'FC50xFD54' progenies (black), 'FC50' (green) and 'FD54' (red). Blue square is the average of each genotype. Significant levels <0.001 (***), <0.01 (**), <0.05 (*) and <0.1 (˙).

Other highly stable fruit shape QTLs were mapped in LG6D, being *FL_6D* the most stable (Fig. 3-12). The AA genotype from the Affx-88884749 marker located at 53,76cM in LG6D showed higher fruit length values compared to the BB genotype, which contribute to increase the fruit shape ratio and, therefore, the error fitting to the circular shape (Fig. 3-14).

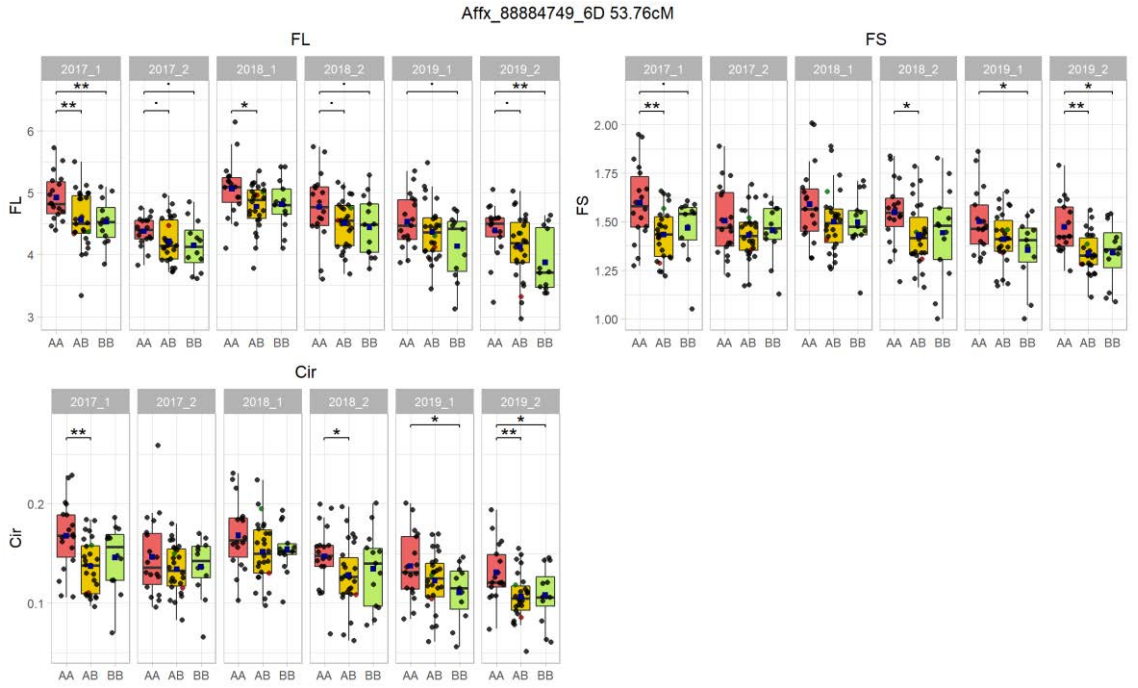


Figure 3-14. Boxplot for Affx-88884749 marker genotypes with shape traits values from LG6D QTLs. Boxplot colours indicate genotypes: AA (red), AB (yellow) and BB (green). Dot colours indicate: 'FC50xFD54' progenies (black), 'FC50' (green) and 'FD54' (red). Blue square is the average of each genotype. Significant levels <0.01 (**), <0.05 (*) and <0.1 (˙).

Focusing on the neck phenotype detected in 'FC50xFD54' population, we were able to find two QTLs. *Neck_3A* was mapped at the same place than *WWP_3A, FS_3A* and *FW_3A*. The presence of neck increased both fruit shape and WWP ratios, decreased the fruit weight. The B allele from the Affx-88833150 marker, mapped in the 8.217cM position of LG3A, was related to the presence of neck phenotype (Fig. 3-15).

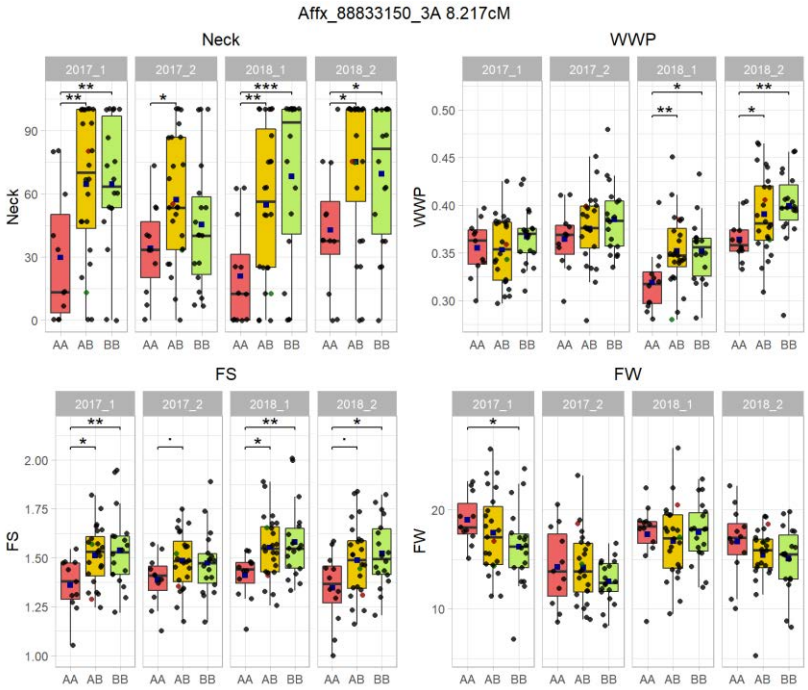


Figure 3-15. Boxplot for Affx-88833150 marker genotypes with neck and related traits values from LG3A QTLs. Boxplot colours indicate genotypes: AA (red), AB (yellow) and BB (green). Dot colours indicate: 'FC50xFD54' progenies (black), 'FC50' (green) and 'FD54' (red). Blue square is the average of each genotype. Significant levels <0.001 (***), <0.01 (**), <0.05 (*) and <0.1 (˙).

The other *Neck_4B* QTL is also related to WWP and FW. To deeply analyse all these QTLs, we selected the Affx-88857053 marker mapped in the 57.484cM position of LG4B and we revealed a pattern similar to previous QTL, even if less significant. Moreover, it indicated that the B allele in homozygosity is linked to the presence of the neck phenotype (Fig. 3-16).

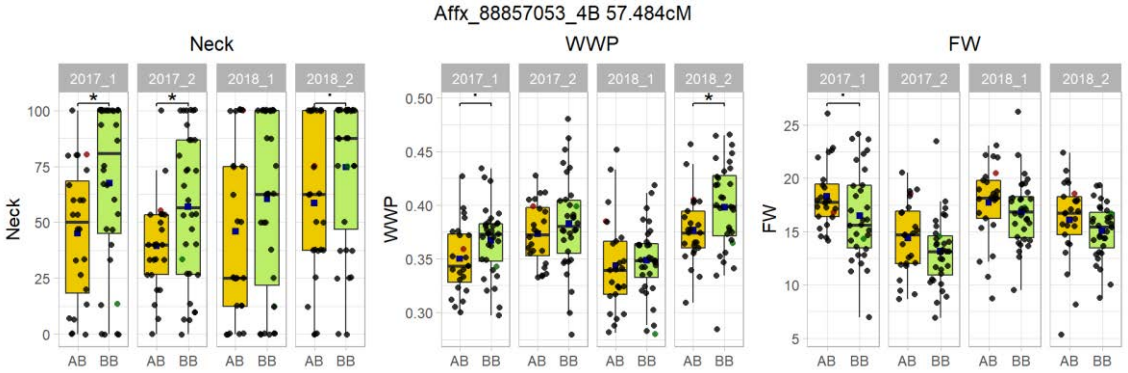


Figure 3-16. Boxplot for Affx-88857053 marker genotypes with neck and related traits values from LG4B QTLs. Boxplot colours indicate genotypes: AB (yellow) and BB (green). Dot colours indicate: 'FC50xFD54' progenies (black), 'FC50' (green) and 'FD54' (red). Blue square is the average of each genotype. Significant levels <0.05 (*) and <0.1 (˙).

The 'FC50xFD54' population showed a big range of internal colour and two regions in LG1A and LG1B explained its variance. The QTL located at the beginning of LG1A was important for L traits, followed by another region important for LI and aI traits. This QTL is related to the internal fruit brightness. To further study the relation between traits, we selected the Affx-88811723 marker located at 11.12cM of LG1A. This marker showed that the B allele with higher values of brightness (LI) compared to the A allele is related to lower values of red colour (aI). Furthermore, LI trait showed higher significant differences between genotypes compared with the other traits, indicating that it is more important for this QTL (Fig. 3-17). In addition to this QTL, other internal colour QTLs were found, but being less stable. For example, the Affx-88809848 marker located in LG1B at 6.456cM showed similar allele/phenotype behaviour for the LI, aI, bI and Chr traits than in LG1A.

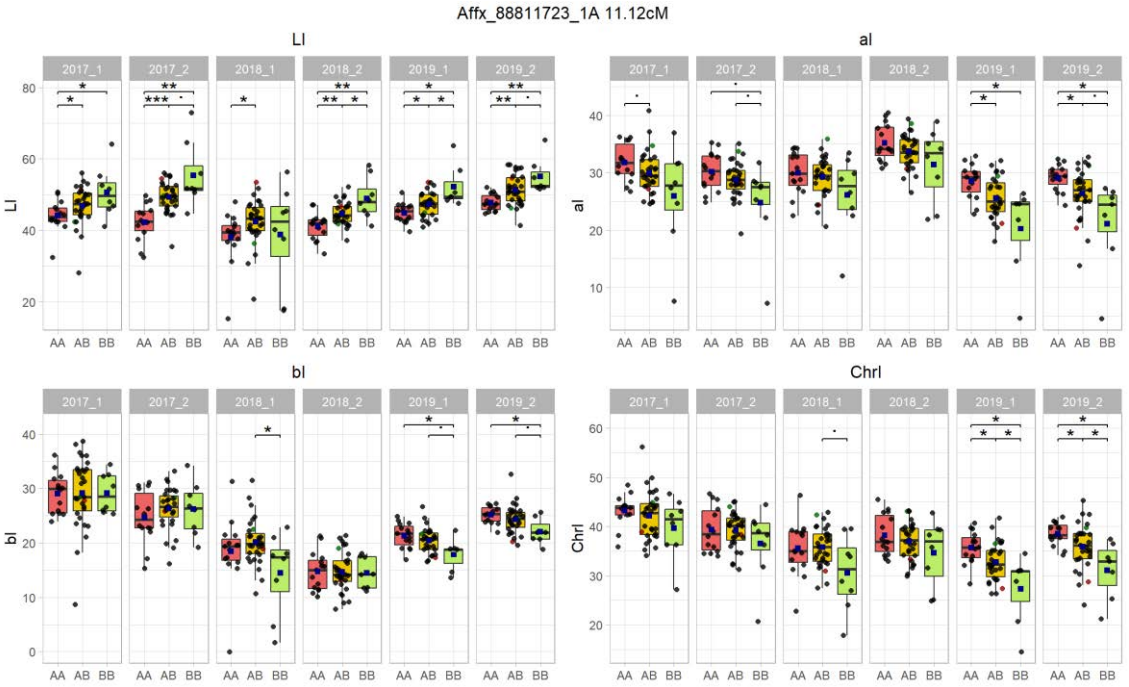


Figure 3-17. Boxplot for Affx-88811723 marker genotypes with colour traits values from LG1A QTLs. **B**oxplot colours indicate genotypes: AA (red), AB (yellow) and BB (green). Dot colours indicate: 'FC50xFD54' progenies (black), 'FC50' (green) and 'FD54' (red). Blue square is the average of each genotype. Significant levels <0.001 (***), <0.01 (**), <0.05 (*) and <0.1 (˙).

'21AF' QTLs

In order to validate some of the discovered QTLs, we performed a QTL analysis for the same traits in the '21AF' population and mapped 31 QTLs in at least one year. Seventeen of them were highly stable in at least two years and three QTLs were totally stable in all three years (Table 3-4 and Supplementary table 3-6). The majority of stable QTLs showed a highly significant LOD

score, indicating that they were major QTLs (LOD>4). In general, the explanations of the observed variance were lower than the ones detected in 'FC50xFD54' population.

The genetic analysis of firmness led to the identification of two QTLs. One of them, *FIR_7C*, was a major QTL explaining 14.5-17.8% of the observed variance. Regarding fruit shape analysis, three QTLs were mapped in the LG3A. The *FS_3A*, located in the middle of LG3A, was a major and totally stable QTL explaining between 11.1 and 16.2% of the fruit shape ratio variance. The two others, *Cir_3A* and *Rec_3A*, are QTLs related to shape approximation traits, that are overlapping and explained a maximum of 18.9 and 19.7% of the observed variance respectively. Furthermore, two QTLs were mapped at the beginning of LG3D, *FA_3D* and *FP_3D*, being the second one a major QTL explaining a maximum of 16.3% of the phenotype variance.

Concerning the colour traits, *aI_6B* was the only discovered QTL for the internal colour and explained a maximum of 20% of the observed variance. Contrarily to the internal colour traits, we could find five different QTLs for external traits, two in the LG3C and three in LG7B. In addition, the same QTLs observed for aE were also detected for ChrE. Finally, we were able to map two major QTLs for LP, such as *LP_6A* and *LP_7B*.

Table 3-4. Fruit quality QTLs in '21AF' population. List of stable and significant QTL ordered by traits and LG.

QTL ID	Harv.	LOD	% Expl.	KW	Genetic map		
					LG	QTL interval (cM)	Size (cM)
FIR_1A	2	3.48	13.1	15.07	1A	0.00 23.25	23.25
FIR_7C	2	**4.87**	**17.8**	**20.55**	7C	7.37 18.80	11.43
FA_3D	2	3.55	13.9	15.18	3D	4.80 13.44	8.64
FP_3D	2	**4.21**	**16.3**	**18.48**	3D	4.37 12.44	8.07
FS_3A	**3**	**4.37**	**16.2**	**20.25**	3A	19.74 40.35	20.62
Ell_7D	2	3.87	15.1	17.24	7D	13.43 28.15	14.72
Cir_2D	2	3.34	12.6	11.27	2D	4.85 13.54	8.69
Cir_3A	2	**5.19**	**18.9**	**16.71**	3A	60.17 88.17	28.00
Rec_3A	2	**5.43**	**19.7**	**21.36**	3A	73.17 101.50	28.33
aI_6B	2	**5.72**	**20.3**	**23.42**	6B	112.60 116.50	3.90
aE_3C	2	**4.88**	**18.6**	**19.03**	3C	25.44 34.60	9.16
aE_7B	**3**	**4.72**	**17.1**	**19.59**	7B	14.25 29.54	15.30
bE_7B	2	3.43	13.5	17.77	7B	23.88 34.15	10.27
ChrE_3C	2	**4.96**	**18.9**	**19.51**	3C	25.44 32.55	7.11
ChrE_7B	**3**	**4.54**	**16.5**	**18.25**	7B	20.11 34.15	14.04
LP_6A	2	**4.72**	**17.5**	**16.22**	6A	60.26 66.30	6.05
LP_7B	2	**4.73**	**17.1**	**20.95**	7B	18.11 29.54	11.43

Bold and underline numbers in harvest indicate totally stable QTLs and in LOD, % Expl. and KW test indicate major QTLs

Focusing on the overlapping QTLs, we detected two regions, LG3C and LG7B, with a cluster of external colour QTLs, such as *aE_3C* and *ChrE_3C,* and *aE_7B*, *bE_7B*, *ChrE_7B* and *LP_7B*, one region for shape traits QTLs, FA, FP and FW (the last in one harvest) in LG3D and one for acidity related traits QTLs, *pH_6A*, *TA_6A* and *SAR_6A* (all of them in one harvest). Moreover, some

QTLs were located in homeologous groups such as *TA_3B* and *TA_3D*, *SAR_6A* and *SAR_6C2* (both found only in one harvest) and *FIR_7C* and *FIR_7D* (the last one only in one harvest).

QTL comparison between 'FC50xFD54' and '21AF' populations

With the aim to detect robust QTLs in different genetic backgrounds, we focused on the QTLs that were identified in both populations and located in the same genetic regions. Two QTLs fulfilled these criteria, the *FS_3A* for fruit shape and *FIR_7C* for firmness. To work with the highest synteny, we have used the saturated genetic maps having the maximum segregating SNPs (See Chapter 1 and Annex 1).

The comparative analysis of fruit shape QTLs in LG3A revealed that the presence of a gap in the '21AF'population provide few common markers between both populations (Fig. 3-18). Concerning the allele segregation of the Affx-88834650 marker, the A allele in homozygosity showed high values for fruit shape ratio in both populations. Furthermore, little differences were observed between AB and BB alleles, revealing the importance of having the A allele in homozygosity to get a high FS ratio.

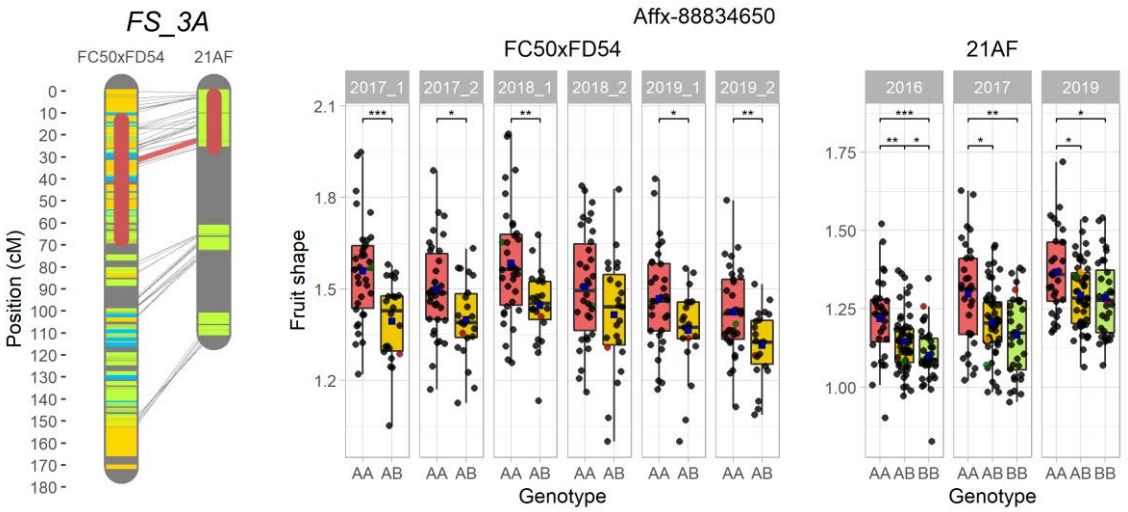


Figure 3-18. *FS_3A* QTL synteny and boxplot for 'FC50xFD54' and '21AF' populations. Left: LG3A synteny representation: codominant (green), 'FC50' segregated (blue) and 'FD54' segregated (yellow) marker class with red boxes represent the 1-LOD confidence interval. Grey lines indicate marker synteny and red line highlight the Affx-88834650 marker used for the boxplot. Right: boxplot from selected marker in both population and in different harvest. Significant levels <0.001 (***), <0.01 (**), <0.05 (*) and <0.1 (˙).

Concerning the firmness QTL, both *FIR_7C* were mapped at the end of LG7C. Although the LG7C from '21AF' is very short, it shows synteny with the end of the LG7C from the 'FC50xFD54' population. Regarding to the alleles of the Affx-88900969 marker, we could notice that B allele

in homozygosity was linked to a higher firmness whereas the presence of the A allele was to fruit softness (Fig. 3-19).

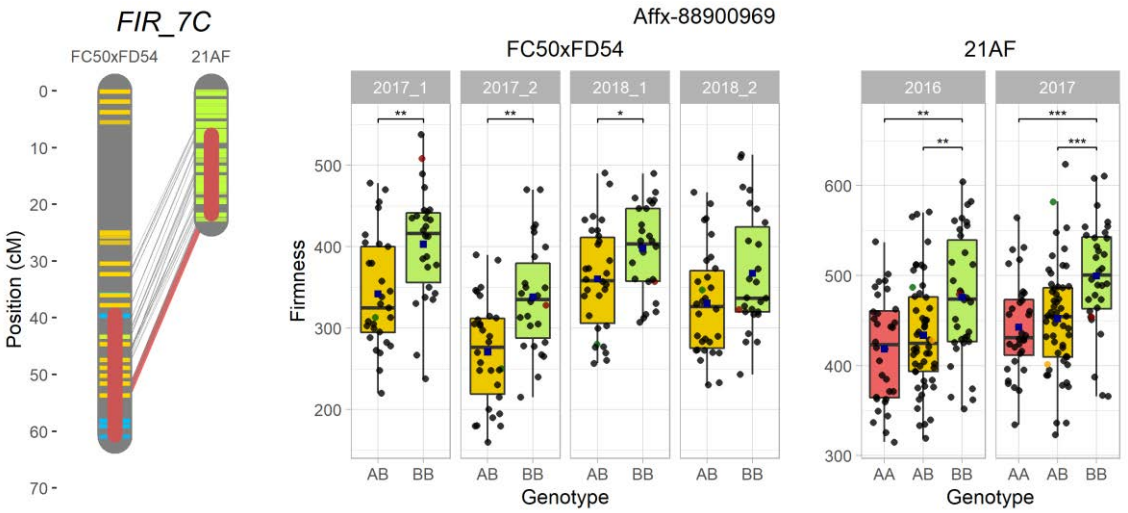


Figure 3-19. *FIR_7C* QTL synteny and boxplot for 'FC50xFD54' and '21AF' populations. Left: LG3A synteny representation: codominant (green), 'FC50' segregated (blue) and 'FD54' segregated (yellow) marker class with red boxes represent the 1-LOD confidence interval. Grey lines indicate marker synteny and red line highlight the Affx-888900969 marker used for the boxplot. Right: boxplot from selected marker in both population and in different harvest. Significant levels <0.001 (***), <0.01 (**), <0.05 (*) and <0.1 (˙).

Since both populations were cultivated in different environments, some QTLs could be expressed in homeologous groups, such as *pH_6C* ('FC50xFD54') and *pH_6A* ('21AF'), *FP_3B* ('FC50xFD54') and *FP_3D* ('21AF'). Several colour traits were mapped in HGs such as *aE_3D* ('FC50xFD54') and *aE_3C* ('21AF'), *ChrE_3D* ('FC50xFD54') and *ChrE_3C* ('21AF'), *LP_6D* ('FC50xFD54') and *LP_6A* ('21AF'), and *ChrE_7D* ('FC50xFD54') and *Chr_7B* ('21AF'). This suggested that colour traits are highly dependent on environment and each environment combined with different genetic background may express genes from one or other HGs.

Discussion

The major aim of this study is to give an insight into the genetic regions that control the principal fruit quality traits and to provide further knowledge and tools for the cultivated strawberry breeding programs. Therefore, four different parental lines, 'FC50', 'FD54', 'Camarosa' and 'Dover', without any family relationship, were chosen to develop an F1 ('FC50xFD54') and an F2 ('21AF') populations which may show different genetic background and heritability. To develop new genetic tools for these traits, we tried to detect stable QTLs in different environments. Therefore, we phenotyped 'FC50xFD54', cultivated in south of France, and '21AF', planted in north of Spain, during three consecutive years.

We studied the segregation of different organoleptic traits such as titratable acidity and soluble solids content, different parameters to decipher fruit shape, such as fruit length, fruit diameter and the presence of neck, and to evaluate fruit colour traits by the CIELab method for internal, external and puree surface. Almost every fruit quality trait showed a normal distribution. Moreover, similar correlation between traits in different harvests were observed in both populations clustering two big groups, one for shape traits and the other for colour and organoleptic traits (Fig. 3-6). In our populations, as previously seen in Lerceteau-Köhler *et al.* (2012), fruit weight correlates with other two shape traits, the fruit length and the diameter.

We have performed the first colour analysis using external, internal and whole fruit puree images. The relations between fruit surfaces were different in both studied populations. In the 'FC50xFD54' population, we observed a high correlation between internal and puree colour and of these colour traits with acidity values. However, in the '21AF' population colour traits were not correlated between them or with acidity (Fig. 3-5). The colour trait differences in both populations could be explained by the fact that some progenies from 'FC50xFD54' showed slightly internal whitish phenotype, whereas those of the '21AF' did not. The correlation between pH and colour traits was not consistent in both populations, as it is reported in other populations (Zorrilla-Fontanesi *et al.*, 2011 and Lerceteau-Köhler *et al.*, 2012).

The analysis of the phenotypic data can determine the parental and progenies relationship. Surprisingly, we found that the '46' individual from the 'FC50xFD54' population showed different colour and shape compared to its siblings and the parental lines (Fig. 3-7). These findings prompted to create a self-pollinated population whose phenotype and inheritance will be further analysed in the next seasons. Moreover, after the hierarchical clustering analysis of fruit quality traits and looking at the heatmap representation, we notice that 'Dover' clustered far from the '21AF' population (Fig. 3-8) reinforcing the idea that in some point a mislabelled

'Dover' occurred and, therefore, our phenotyped plant is not the real ancestor of the '21AF' population. Our genetic analysis also supports this hypothesis (see Annex Chapter 1).

Little is known about fruit quality QTLs in strawberry, either for being a highly environment dependent or for their extremely complex regulation in octoploid species.

Regarding the total number of QTLs, we observed big differences between both populations. A total of 77 stable QTLs were mapped in the 'FC50xFD54' population, however, only 31 stable QTLs were in '21AF'. The reason of this difference may be explained by the number of the analysed harvests (six against three). Moreover, the huge harvest variability found in the '21AF' population could complicate the analysis of QTL stability. Another reason could be the harvest-time methodology. We used the two-times harvest in the 'FC50xFD54' population which collected the differences between early and late season and the one-season harvest in the '21AF' population. Despite these methodological differences, we firmly think that the main reason for the differences in the number of QTLs detected in these populations is due to the lack of extreme phenotypes in the '21AF' population, as part of the allelic variability was reduced for being an F2 population compared to an F1 population.

A suitable population size is essential to perform a QTL analysis. We worked with a small 'FC50xFD54' population with 63 siblings and a large '21AF' population with 117 siblings. This variation in population size may explain the large percentage of the phenotypic variance explanation and low LOD score found in the 'FC50xFD54' population compared to the lower explanation of the phenotypic variance and the presence of major QTLs observed in the case of the '21AF' population. These differences were also observed in other strawberry populations when the same traits were analysed (Zorrilla-Fontanesi *et al.*, 2011 and Lerceteau-Köhler *et al.*, 2012).

Previous QTL analyses in fruit quality traits were done using genetic maps whose marker density are lower than the one of our map (Zorrilla-Fontanesi *et al.*, 2011, Lerceteau-Köhler *et al.*, 2012 and Castro and Lewers, 2016). Since these markers, SSRs and others, are not subgenome specific, it is not possible to distinguish between HGs. Hence, comparisons with previous studies are quite difficult, especially when QTLs for the same trait are mapped in different homeologous groups (HGs), such as colour traits in LG1A and LG1B or fruit shape traits in LG4B and LG4C. Regarding these two QTLs, we noticed that the stability of one homeologous QTL was higher than the other, remarking the possibility that some homeologous induced genes are environmental-dependent (Fig. 3-12). We also detected QTLs in HGs in different population whose different alleles or environmental conditions might trigger the expression of a gene from

one HG or another. As far as we know, only one QTL analysis using saturated SNP pedigree analysis revealed few QTLs but there were not stable in time or localization (Verma *et al.*, 2017b).

Concerning the genomic locations, the QTL analysis of fruit quality traits revealed that they are spread along the genome. However, in our populations there were two hotspots for fruit colour in LG1A and LG1B and three more for fruit shape detected in LG3A, LG4C and LG6D.

Throughout the bibliography of strawberry fruit weight, QTLs were reported in HG1 to HG5, being HG1 and HG5 important in more than one population (Fig. 3-20). In our populations, we were able to detect two fruit weight QTLs, *FW_1A* and *FW_3A*, reinforcing the QTL reported in HG1 and increasing the importance of HG3 for this trait. So, FW QTLs may be highly dependent on population or environment. Looking at the firmness trait, only the *FIR_7C* QTL was found both in 'FC50xFD54' and '21AF' populations also reported previously in LGVII-1 (Zorrilla-Fontanesi *et al.*, 2011). In the '21AF' population, we mapped another firmness QTL in LG1A. In the same population, the *FaRGlyase1* gene, encoding for an enzyme involved in cell-wall degradation, was linked to this trait and mapped in LG1B (Molina-Hidalgo, 2013).

Regarding the acidity QTL analysis, several QTLs have been mapped all over the genome, especially in HG4 and HG5 (Zorrilla-Fontanesi *et al.*, 2011, Lerceteau-Köhler *et al.*, 2012 and Castro and Lewers, 2016). *TA_5A* and *pH_5C* QTLs detected in this study could be related to the ones in HG5. Additionally, we found a common QTL for TA and SAR in LG1A. Although several SSC QTLs of minor significance have been described (Zorrilla-Fontanesi *et al.*, 2011, Lerceteau-Köhler *et al.*, 2012 and Castro and Lewers, 2016), we were not able to find any genomic regions explaining its variance, indicating the complexity of this trait.

Since a visual fruit shape classification could be imprecise, different softwares have been developed, such as the Tomato Analyser (Gonzalo *et al.*, 2009) designed for tomato and pepper fruit shapes. Very recently, a multi-dimensional machine learning approach was developed specifically for strawberry fruit shape (Feldmann *et al.*, 2020). This new tool will be very useful to further explore the genetics behind strawberry fruit morphology.

Lerceteau-Köhler *et al.* (2012) reported fruit shape QTLs spreading from HG2 to HG6, being the most stable in HG2 and HG3. In our study, a validated QTL, *FS_3A,* was detected in both populations. *FD_2A* QTL might be related to the round QTL detected in LG2 in a diploid NILs collection (Urrutia *et al.*, 2015). Additionally, we found new QTLs in HG4, such as *FS_4C* for ovoid shape and *FS_4D* for circular shape. It has been described that different hormones could be

related to some fruit shape QTLs, such as auxins, which are involved in the diameter, or gibberellic acids (GA), which are involved in the length of the fruit (Liao *et al.*, 2018).

For the neck phenotype, which is an undesirable trait for breeding programs, we mapped two QTLs, *Neck_3A* and *Neck_4B,* found in the 'FC50xFD54' populations. Since neck and WWP QTLs were overlapping, we concluded that WWP is appropriate, as an objective measurement, for fruit neck analysis.

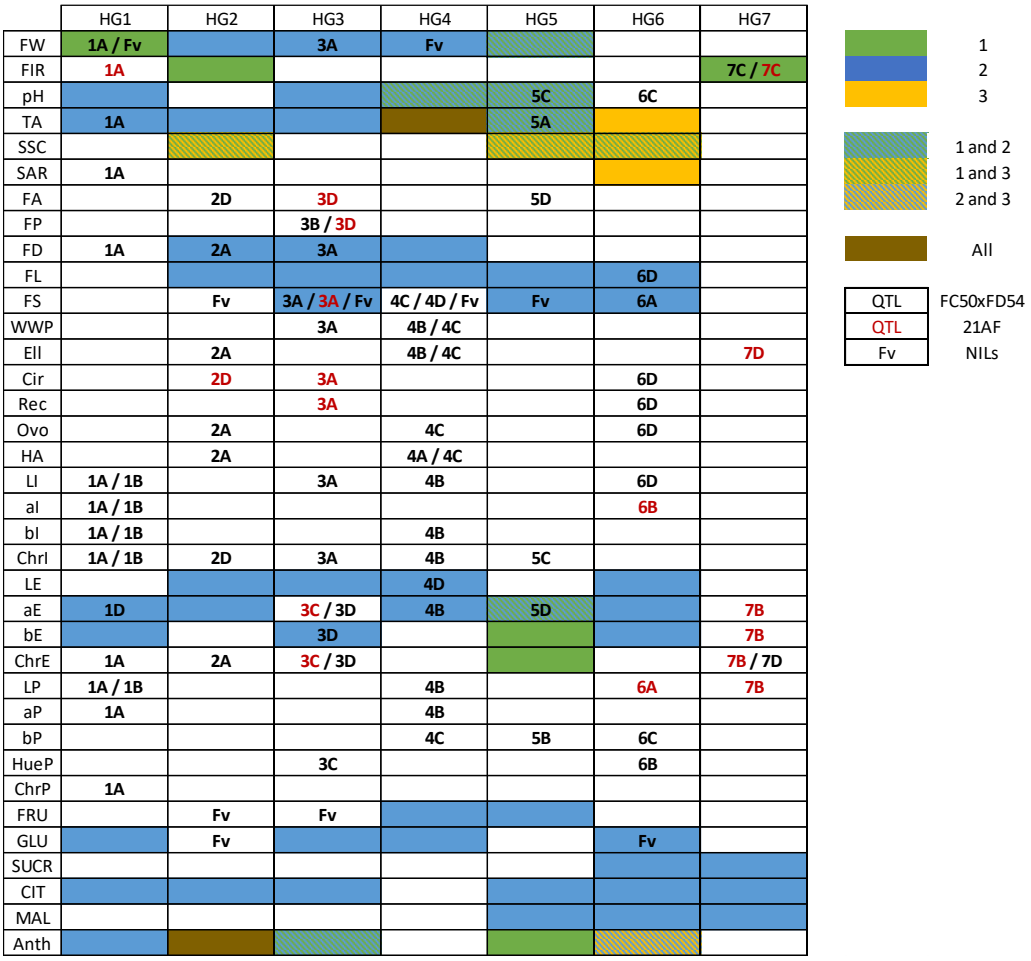


Figure 3-20. QTL summary in each HGs. Colour cells refer to different works: Zorrilla-Fontanesi *et al.*, 2011 (green, 1), Lerceteau-Köhler *et al.*, 2012 (blue, 2) and Castro and Lewers, 2016 (yellow, 3), stripped pattern for common QTLs and all shared in brown. QTLs detected in the 'FC50xFD54' population (black characters), '21AF' population (red characters) and in FvxFb NILs population (Urrutia *et al.*, 2015) (Fv mark). Trait abbreviations are presented as in Table 1 adding FRU (fructose), GLU (glucose), SUCR (sucrose), CIT, (citric acid), MAL (malic acid) and Anth (total anthocyanins).

External colour trait QTLs were reported in LGIa, LGIIa, LGIIIa, LGIVd, LGVb, LGVIa and LGVIb, being the one in LGIIIa the most stable (Lerceteau-Köhler *et al.*, 2012) and may co-localise with our hotspot QTLs in LG3D. Despite previous studies focused on skin colour, to our knowledge, our analysis is the first to identify internal and puree QTLs. We mainly mapped the highly

correlated internal and puree QTLs in HG1. Since colour QTLs in LG1A were much more significant and stable than QTLs in LG1B, we suggest that homeologous genes are expressed differently depending on the environment. A truncated transcription factor *FaMYB10* that regulates anthocyanins pathway is responsible for the white colour phenotype (Wang *et al.*, 2019) and it is annotated close to our LG1A colour QTLs. In addition, there are other *FaMYB* genes that regulate the anthocyanins content (Aharoni *et al.*, 2001, Schaart *et al.*, 2013 and Medina-Puche *et al.*, 2014).

Our in-depth characterization of fruit quality traits such as taste, firmness, colour and shape in two unrelated cultivated strawberry populations, 'FC50xFD54' and '21AF', has given insight to clarify the sub-genome LG of previously reported QTLs and added regions responsible for some fruit quality traits. In addition, we described the marker allele that are linked to some traits of interest, which could be used in breeding programs through MAS.

General discussion

Cultivated strawberry (*Fragaria x ananassa*) is an allo-octoploid (2n = 8x = 56) species. It appeared in the XVIII century from the hybridization of *Fragaria chiloensis* with *Fragaria virginiana*. Although it is a polyploid species, its genome is highly diploidised with four subgenomes. At present, there are two hypotheses explaining the origin of these four subgenomes, the first one is based only on two diploid species, *F. vesca* and *F. iinumae*, and a mixture of both (Tennessen *et al.*, 2014) while the most recent hypothesis suggested four different species, *F. vesca*, *F. iinumae*, *F. nipponica* and *F. viridis* (Edger *et al.*, 2019). Since some differences are detected between subgenome sequences, it is possible to genotype assuming markers with diploid segregation and considering the allelic dosage of the homeologous groups (Bassil *et al.*, 2015).

The Rosaceae family comprises several polyploid species with high economic, such as rose and blackberry. Moreover, other relevant polyploid crops can be found in other families, such as potato (Solanaceae), cotton (Malvaceae) and wheat and sugarcane (Poaceae).

Despite the big differences existing among the different polyploid species, we can compare some of their genetic knowledge. The genetic advancement of a given species will depend on the genetic and genomic tools already available. One of the most valuable tools that recently arisen are the SNP arrays. The number of SNPs, their distribution and density in the genome are key for the value of a SNP array. Fortunately, when this thesis begun, there were already two SNP arrays developed for cultivated strawberry, the 90k (Bassil *et al.*, 2015) and 35k (called IStraw35k) (Verma *et al.*, 2017a). Very recently, two new SNP arrays were released, the 850k (Hardigan *et al.*, 2020) and 50k (Hardigan *et al.*, 2020). We highlight that the 35k and the 50k SNP arrays are actually subsets of 90k and 850k, respectively. The other polyploid species from Rosaceae family with an available SNP array is the rose with the 68k SNP array (Koning-Boucoiran *et al.*, 2015).

Different SNP arrays have also been developed in other polyploid species belonging to other families, such as the 20k (Vos *et al.*, 2015) and 8.3k (Mengist *et al.*, 2018) arrays in autotetraploid potato, the 63k (Hulse-Kemp *et al.*, 2015) and 80k (Cai *et al.*, 2017) arrays in allotetraploid cotton, six different SNP arrays ranging from 9k (Cavanagh *et al.*, 2013) to 820k (Winfield *et al.*, 2016) in allohexaploid wheat and a very recent SNP array for an aneupolyploid sugarcane of 100k SNPs (You *et al.*, 2019). To sum up, a large diversity of SNP arrays has been developed for economically important polyploid species. A common issue for these SNP arrays is the necessity to know the allelic dosage of each marker and select, if possible, those that are subgenome-specific that could be evaluated as diploids.

Since cultivated strawberries are the result of a recent hybridization with any severe domestication or genetic drift, high genetic variability is maintained in their cultivars, which can be reproduced asexually. Thus, an F1 population can be considered a good mapping population segregating for many markers and traits.

For our study, we constructed a genetic map from an F1 population generated from two distant parental lines, 'FC50' selected for its wild strawberry aroma and 'FD54' for its fruity aroma. The resulted 'FC50xFD54' genetic map, developed using the IStraw35k array, contains 14695 SNPs, a number that is the highest compared to the publicly available genetic maps (Table 1-3). This high number of segregating SNPs confirms that our parental lines are genetically distant. Furthermore, comparing the relative markers position of our map with the *F. x ananassa* consensus map, the *F. vesca* and *F. x ananassa* genomes, we found a high synteny in all cases. However, some LGs in our map are inverted with the *F. x ananassa* chromosome orientation as observed in Fig. 1-8. We additionally constructed a reduced 'FC50xFD54' genetic map with shorten genetic distances for QTL analysis (Fig. 1-4).

Since the first SNP arrays in strawberry were constructed using the release diploid genome sequence, the IStraw35k array shows unbiased marker density for the *F. vesca-like* subgenome. The two new SNP arrays, 850k and 50k, were designed with subgenome specific markers and covering equally the whole genome (Hardigan *et al.*, 2020). Despite the high differences in marker density, we found similar predicted genetic maps using either the IStraw35k or the 850k SNP arrays (Fig. 1-10) confirming that both SNP arrays are appropriate for QTL analysis.

We also improved the F2 '21AF' genetic map derived from a cross between 'Dover' and 'Camarosa'. This population consisted of 117 progenies, a number that provides a highly accurate genetic map. Since the F2 population comes from a single hybrid, 'H-21', only 7977 SNPs segregate in its progenies using the IStraw90k array. We also obtained a good synteny between the '21AF' and both 'FC50xFD54' and *F. x ananassa* consensus maps (Fig. A1-2). The analysis of the genetic and phenotypic data revealed that the individual considered as the parental line of the '21AF' population, was neither the parent nor the 'Dover' variety. So, at some point, we must have lost the variety due to a mislabelling.

In order to compare the segregation ratio of some SNP arrays for different species, we focused on their consensus maps. In strawberry, 38% SNPs from the IStraw90k array (Bassil *et al.*, 2015) were selected to construct the *F. x ananassa* consensus map (Harrison, unpublished) and we mapped more than 14k from the IStraw35k array (39%) in the F1 population. In another Rosaceae specie, the same ratio (38%) from 63k SNP array were mapped in an F1 population of

Rosa hybrida (Bourke *et al.*, 2017). In general, higher segregation ratio was observed with other polyploid species than with *F. x ananassa*, such as the consensus maps in *Triticum aestivum* (Cavanagh *et al.*, 2013 and Wang *et al.*, 2014) or the GWAS analysis in cotton (Cai *et al.*, 2017).

The availability of whole genome sequences for polyploids or their related diploid species is a great advantage for genetic studies. In strawberry, the wild diploid genome (Shulaev *et al.*, 2011 and Edger *et al.*, 2018) was released much earlier than cultivated strawberry (Edger *et al.*, 2019), being the first Rosaceae polyploid crop having a whole genome sequence. In fact, the autotetraploid blackberry genetic studies are based on the closest published diploid genome, *Rubus occidentalis* (VanBuren *et al.*, 2016, Jibran *et al.*, 2018 and VanBuren *et al.*, 2018) and the same happen for the polyploid rose using the genomes of two diploid species, *Rosa chinensis* (Raymond *et al.*, 2018 and Hibrand Saint-Oyant *et al.*, 2018) and *Rosa multiflora* (Nakamura *et al.*, 2018).

Considering the publication time of the whole sequences of different polyploid genomes, wheat appears as the most advantageous species (Brenchley *et al.*, 2012 and Appels *et al.*, 2018), followed by the two cultivated species of cotton (Wang *et al.*, 2015, Li *et al.*, 2015, Liu *et al.*, 2015, Yuan *et al.*, 2015 and Hu *et al.*, 2019). By contrast, economically important polyploid species, such as potato or sugarcane, have no-complete genome sequences. Currently, only the diploid potato have been sequenced (Xu *et al.*, 2011 and Sharma *et al.*, 2013), while the polyploid genome is still ongoing (Kyriakidou *et al.*, 2020). The complex sugarcane genome is also in progress (Souza *et al.*, 2019). Hence, cultivated strawberry is one of the most advantageous polyploid crops (Fig. D-1).

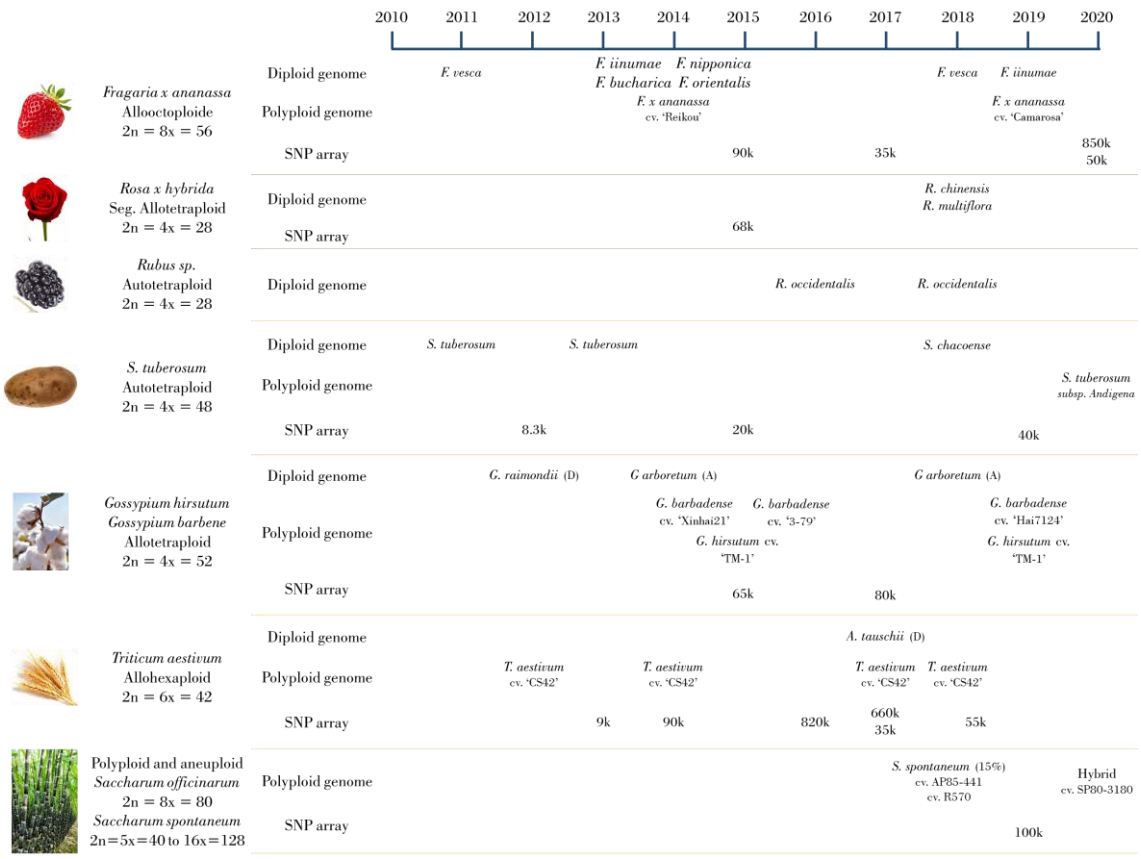


Figure D-1. Timeline of some polyploid genetic resource.

Strawberries are valued fruits but nowadays consumers are demanding tastier ones. Flavours is the admixture between taste, mainly characterised by the balanced sugar and acid content, and aroma, provided by more than 350 VOCs. Additionally, some volatile compounds can interfere with the sweetness perception (Schwieterman *et al.*, 2014 and Ulrich and Olbricht, 2016).

The accumulation of volatile compounds (VOCs) is characteristic for each kind of fruit. VOCs are classified as esters, alcohols, aldehydes, ketones, lactones, furans and terpenes and few of them confer the characteristic aroma of each fruit defined as key volatile compounds (KVCs). The most abundant KVCs in strawberry and apple are small straight-chain (C6) esters (Schieberle and Hofmann, 1997, Ulrich *et al.*, 1997 and Dixon and Hewett, 2000). Other compounds, such as C9 esters and lactones, are the most representative KVCs in peach (Eduardo *et al.*, 2010) and apricot (Greger and Schieberle, 2007). In the petals of roses, monoterpene alcohols and phenylpropanoids are the most released compounds (Spiller *et al.*, 2010). Despite these main characteristics, VOCs are highly dependent on genotype, environment and ripening level (Schwieterman *et al.*, 2014). In fact, environmental factors can activate one or more pathways to accumulate some compounds, such as the alternative pathway described under warm temperature for 2-phenyletanol in roses (Hirata *et al.*, 2016).

The polygenic inheritance of VOCs accumulation required a great effort to discover the major QTLs. For this reason, it is primordial to select the appropriate parental lines in order to generate a good segregating population. Since Rosaceae species are highly heterozygous, F1 populations are mainly used for QTL analysis.

We have characterised the aroma profile of an F1 population and the selected KVCs of an F2 population. We could detect high correlation between ester compounds which share the acyl-CoA or alcohol groups, indicating that they are synthetized by the same pathway (Fig. 2-2 and Fig. 2-3). Similarly, high correlation was observed within terpenes or lactones.

Regarding the floral aroma, we mapped a major and stable QTL in both populations for the terpene accumulation at the beginning of LG3, *Terpene_3B*, localised in the chromosome Fvb3-3 (Fig. 2-9 and Fig. 2-11). This QTL was also identified in a diploid NILs collection (Urrutia *et al.*, 2017). The blast analysis of some markers of the *Terpene_3B* QTL locates them in more than one homeologous group, such as Fvb3-1 and Fvb3-2. In these two regions, genes related to terpene synthesis such as FaNES1, FaNES1-like and FaNES2 are annotated. The FaNES1 catalyses the production of both the monoterpene linalool and the sesquiterpene nerolidol (Aharoni *et al.*, 2004). Moreover, this gene is present both in cultivated and wild octoploid terpene producer lines but not in diploid lines (Aharoni *et al.*, 2004 and Chambers *et al.*, 2012). In peach, two linalool synthase candidate genes were found in the terpene QTL located at the beginning of LG4 (Eduardo *et al.*, 2013 and Sánchez *et al.*, 2014). As the beginning of the Fragaria chromosome Fvb3 has high synteny with the beginning of the Prunus chromosome Pp4 (Jung *et al.*, 2012), it can be hypothesised that these genes might be present in the common ancestor. In addition to this QTL, we also mapped two more sesquiterpene QTLs in LG1C and LG4B. The latter QTL might be the one detected in LGIV-1 in '232x1392' population (Zorrilla-Fontanesi *et al.*, 2012).

More than half of the identified VOCs are ester compounds providing a large variety of fruity aroma notes. We localised several straight-chain ester QTLs in LG1A, LG4B, LG6A and LG7D. Focusing on the ones mapping in the LG1A, we found different compounds mapped in different populations, such as *Methyl_1A* in the '21AF' and *Hexyl_1A* in 'FC50xFD54'. Different ester QTLs, such as the ones for methyl hexanoate or benzyl acetate, were also found in LGI-1 in another F1 population (Zorrilla-Fontanesi *et al.*, 2012), indicating that these QTLs are in some way dependent on the genetic background or the environment.

We also mapped a major and stable QTL at the end of the LG4B that can be found only in the 'FC50xFD54' population. This QTL is related to hexanoate compounds, together with methyl

octanoate and methyl decanoate. A methyl decanoate QTL was also mapped at the end of LGIV-1 (Zorrilla-Fontanesi *et al.*, 2012). This kind of C6 compounds are synthetized through the lipoxygenase (LOX) pathway (Schwab *et al.*, 2008 and Granell and Rambla, 2013), but, the LOX gene was located in the middle of this LG (Zorrilla-Fontanesi *et al.*, 2012). The FaLOX has several isoforms which may have different physiological roles. One of them, active in the early stages of fruit development, produces the intermediaries of the C6 volatile compounds (Leone *et al.*, 2006). In apple, a cluster of LOX genes involved with hexanal production was located in LG7 (Vogt *et al.*, 2013). In tomato, a cluster of hexanoate-related QTLs was mapped at the top of chromosome 1 in the same position of *TomLoxC* (Rambla *et al.*, 2017). Of the different TomLox isoforms, TomLoxC is the only one which can use linoleic and linolenic acids to generate C6 flavour compounds (Chen *et al.*, 2004). Additionally, pan-genome analysis revealed that some *TomLoxC* promoter variants modified the C6 compound accumulation (Gao *et al.*, 2019). Since strawberry has C6 flavour compounds similar to the ones of apple and tomato, we hypothesise that LOX or hydroperoxide lyase (HPL) genes could be located in the *Hexanoate_4B* QTL.

The last step of ester biosynthesis is the esterification between alcohol and acyl-CoA groups catalysed by an alcohol acyltransferase (AAT). This enzyme can use a large variety of substrates (Aharoni *et al.*, 2000). A gene described as AAT gene, maker-Fvb7-3-snap-gene-89.68-mRNA-1, is located inside the *Methyl_7D* QTL detected in both 'FC50xFD54' and '21AF' populations (Fig. 2-13). In addition, four more AAT genes are annotated in HG7. This enzyme is essential to give the fruity aroma in *Fragaria* genus, thus, it was also characterised in *F. chiloensis* (Gonzalez *et al.*, 2009) and *F. vesca* (Beekwilder *et al.*, 2004). In apple, QTL analysis in different F1 populations revealed a cluster of ester and alcohol QTLs in LG2 that co-locate with MdAAT1. The MdAAT1 presented several allelic or paralogous variants accepting a wide range of substrates (Costa *et al.*, 2013, Souleyre *et al.*, 2014 and Yauk *et al.*, 2017). Since some regions of the chromosome Fvb7 are syntenic to Md02 (Jung *et al.*, 2012), this suggests a common gene localization.

Looking at the phylogenetic relationship between the cloned alcohol acyltransferase protein and the annotated ones in *F. x ananassa* genome (Supplementary table D-1), we were able to separate them in two groups. One of them groups FaAAT2 with other AATs from other species while the other groups mainly *Fragaria* proteins (Fig. D-2). In this second group, we can also find the maker-Fvb1-1-augustus-gene-211.40-mRNA-1 (FaAAT1_1-1_211.40) and the maker-Fvb7-3-snap-gene-89.68-mRNA-1 (FaAAT1_7-3_89.68) that are closely related to the FaSAAT described by Aharoni *et al.* (2000).

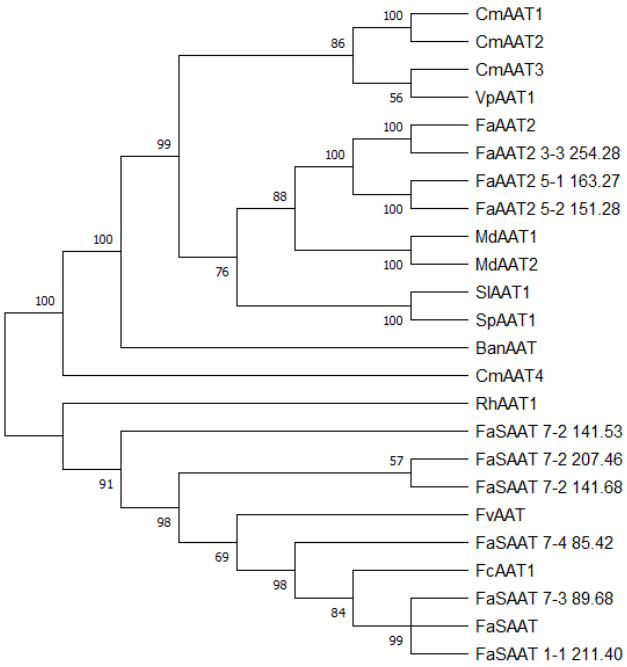


Figure D-2. Phylogenetic relationship of alcohol acyl transferases (AAT) of *Fragaria* and others genus. Values next to branching clades indicate the percentage of bootstrap support with 500 replications (MUSCLE alignment from MEGA6).

The last straight-chain ester QTLs were mapped in LG6A for different octyl related esters in 'FC50xFD54' and '21AF'. However, these two studied populations presented different allele behaviour, indicating that we are probably dealing with different QTLs that mapped in the same place (Fig. 2-12).

Esters with an aromatic group, such as methyl anthranilate, are highly important for strawberry breeding programs focused on the highly demanded wild strawberry flavour. These compounds are synthetized through the amino acid degradation pathway. However, methyl anthranilate is present in few cultivated strawberry lines (Olbricht *et al.*, 2008 and Schwieterman *et al.*, 2014), being one of them our 'FC50' parental line. We therefore mapped several QTLs for methyl anthranilate, being *MA_7A* the most significant and highly stable QTL and three more QTLs in the HG5. Methyl anthranilate QTLs were also detected in LG5 and LG7 in a diploid NILs collection (Urrutia *et al.*, 2017). Moreover, our *MA_4D* QTL might be related to the *anthranilate acid methyl transferase*, *FaAAMT*, which was located in *F. vesca* Fvb4, (Pillet *et al.*, 2017).

Regarding to lactone QTL, we could map a high significant QTL at the bottom of the LG3D, which might be related to *omega fatty acid desaturase* gene, *FaFAD1* (Chambers *et al.*, 2014 and Sanchez-Sevilla *et al.*, 2014), mapped at the end of LGIII-2 (Zorrilla-Fontanesi *et al.*, 2012).

QTL analysis should be directed to its implementation in breeding programs through molecular assisted selection (MAS). Volatile compounds are synthesised by few enzymes with a wide range of substrate affinity, such as FaNES1 (Aharoni *et al.*, 2004) or AAT (Aharoni *et al.*, 2000 and Yauk *et al.*, 2017). This fact makes more difficult to target a single compound. Furthermore, small changes in any of the VOCs, which are involved in the complex fruit aroma, could switch the pleasant perception to an undesirable flavour (Klee and Tieman, 2018). Hence, panel test should be in front of any breeding program, in order to evaluate the new varieties selected by MAS (Folta and Klee, 2016).

As a result of these genetic analyses of VOCs, we are able to provide some markers linked to the accumulation of some pleasant compounds, that can be easily implemented in breeding programs (Annex Chapter 2). By selecting the appropriate genotypes, the accumulation of one or more VOCs could be increased (Table D-1 and Figure D-3) and therefore, strawberries could gain the floral or fruity aroma which will satisfy the consumer demand.

Table D-1. Selected markers of some major and stable QTLs applicable in MAS and showing the genotype linked to the increment of each trait and population where was detected.

QTL	Marker	Genotype	Population
Methyl_1A	Affx-88812680	AA	21AF
Terpene_3B	Affx-88832495	AA	Both
Terpene_3B	TP_3.3_2.623_50k	XX	FC50xFD54
GD_3D	Affx-88845940	AA	21AF
Hexanoate_4B	Hex_4.3_4.204_50k	XX	FC50xFD54
MA_7A	Affx-88897207	AB	FC50xFD54
Methyl_7D	Affx-88897295	BB	Both
LI_1A	Affx-88811723	BB	FC50xFD54
FS_3A	Affx-88834650	AA	Both
FS_4C (Ovo)	Affx-88856304	BB	FC50xFD54
FL_6D	Affx-88884749	AA	FC50xFD54
FIR_7C	Affx-88900969	BB	Both

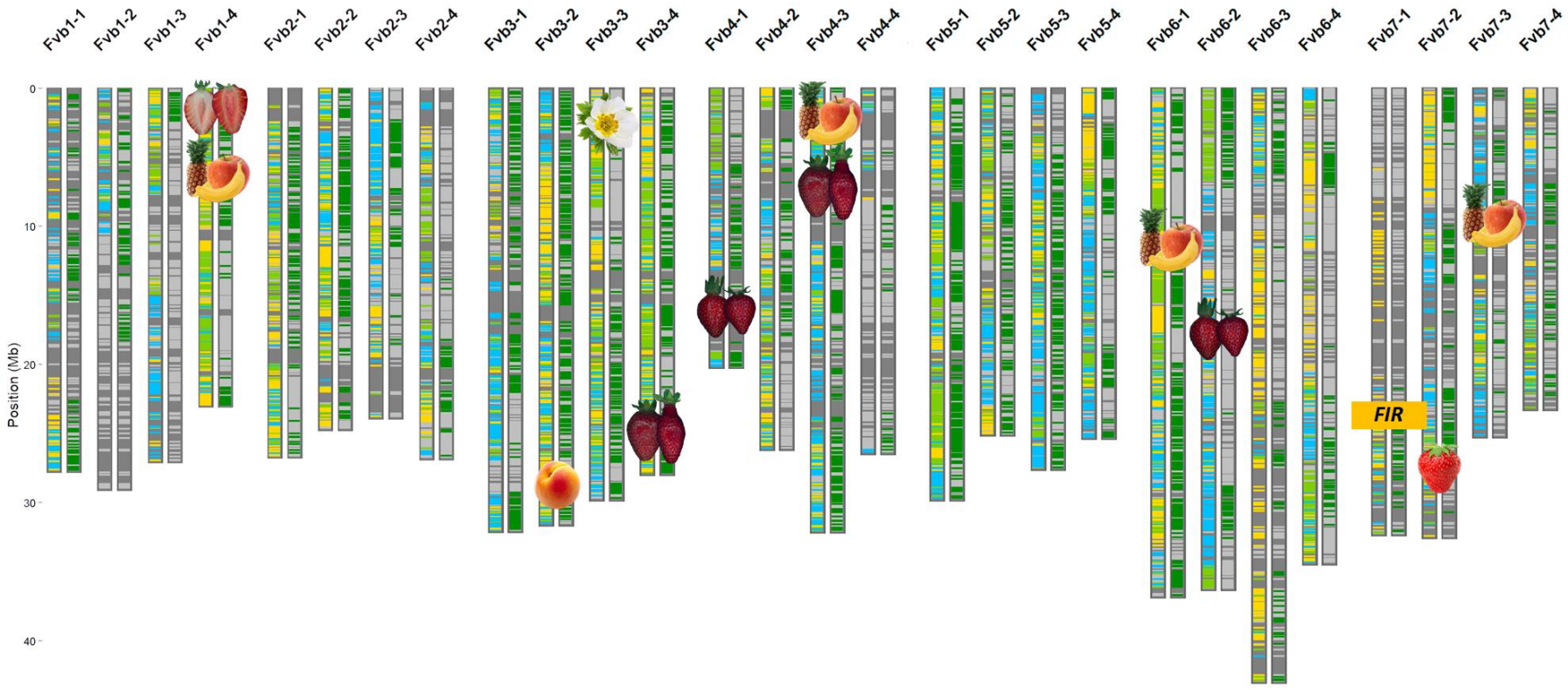


Figure D-3. Major QTLs in the *F. x ananassa* genome. Markers in 'FC50xFD54' (left) and '21F' (right) positioned in each chromosome. Marker colours indicate their segregation: co-dominant (green), 'FC50' segregated (blue), 'FD54' segregated (yellow), no segregated (light grey) and chromosome background (dark grey). Images representing each trait indicate QTL positions. QTLs from top to bottom and left to right: *Colour_1A* (Fvb1-4), *Methyl_1A* (Fvb1-4), *GD_3D* (Fvb3-2), *Terpene_3B* (Fvb3-3), *Neck_3A* (Fvb3-4), *FS_4C* (Fvb4-1), *Hexanoate_4B* (Fvb4-3), *Neck_4B* (Fvb4-3), *Acetate_6A* (Fvb6-1), *FL_6D* (Fvb6-2), *FIR_7C* (Fvb7-1), *MA_7A* (Fvb7-2) and *Methyl_7D* (Fvb7-3).

Not only volatile compounds are important for the strawberry flavour. Sweetness and acidity are also key traits for the consumer preferences. Despite the main sugar metabolites in strawberry are glucose, fructose and sucrose, SSC is the easy method to identify total sugar content. However, we were not able to detect any SSC QTL in our population. Sugar QTLs show a high polygenic behaviour with several minor QTLs spanning most of the HGs (Zorrilla-Fontanesi *et al.*, 2011, Lerceteau-Köhler *et al.*, 2012 and Castro and Lewers, 2016). This tendency is also observed in other species such as peach (Hernandez Mora *et al.*, 2017), apple and tomato. However, two major QTLs and candidate genes were identified, such as an apple vacuolar invertase gene for sucrose and fructose QTL located in LG1 (Sun *et al.*, 2015, Guan *et al.*, 2015 and Larsen *et al.*, 2019) and a tomato extracellular invertase (*LIN5*) responsible for sugar accumulation, that is located in chromosome 9 (Fridman *et al.*, 2004 and Tieman *et al.*, 2017). In addition, five more SSC QTLs have been mapped in different chromosomes (Tieman *et al.*, 2017 and Zhao *et al.*, 2019).

Citric and malic acid are the main responsible for strawberry acidity. Several minor QTLs were mapped in different HGs in previous studies (Zorrilla-Fontanesi *et al.*, 2011, Lerceteau-Köhler *et al.*, 2012 and Castro and Lewers, 2016). Some of these QTLs matches with our pH, TA and SAR QTLs in LG1A and LG5A. Besides the polygenic behaviour of acidity QTLs in apple, a major QTL was located in LG16 (Liebhard *et al.*, 2003, Khan *et al.*, 2013 and Ma *et al.*, 2016) caused by the transporter channel to vacuole, MA1 (Bai *et al.*, 2012 and Khan *et al.*, 2013). In peach, an auxin afflux carrier family protein was proposed as responsible for the sub-acid locus D (Cao *et al.*, 2016) located at the beginning of LG5 (Dirlewanger *et al.*, 2006 and Boudehri *et al.*, 2009).

Fruit firmness is an important trait related to shelf-life and an important target for the strawberry industry. In our firmness study, we were able to map two QTLs, namely *FIR_1A* and *FIR_7C.* In previous study in '21AF' population, *FaRGlyase1* gene was responsible for a QTL in LG1B (Molina-Hidalgo *et al.*, 2013) which might correspond to our *FIR_1A.* Our QTL in LG7C, *FIR_7C,* might be the one detected in LGVII-1 from '232x1392' population (Zorrilla-Fontanesi *et al.*, 2011).

Appearance is a key factor for strawberry fresh market. For this reason, we undertook a genetic study on the colour and shape of strawberries. Fruit shape is finely regulated during fruit development by hormones, such as auxin responsible for increasing fruit diameter and gibberellins for the fruit length (Liao *et al.*, 2018).

The 'FC50xFD54' population showed a fruit neck shape segregation and its QTL analysis identified two regions, LG3A and LG4B. Furthermore, in the same population, we were able to

map two main QTLs for fruit shape in LG4C, for the ovoid shape, and LG6D, for fruit length. Moreover, in both studied populations, we mapped the same fruit shape ratio QTL, *FS_3A* (Table D-1 and Fig. D-3). Looking at the candidate genes for fruit shape in other species, different candidate genes located at the beginning of LG6 (Dirlewanger *et al.*, 2006) have been proposed for the flat fruit shape of peach and nectarine (Cao *et al.*, 2016 and Lopez-Girona *et al.*, 2017), and in tomato, it has been shown that some genes, such as *sun, ovate, TRM5 and Sov1*, reorganize microtubules producing elongated tomato fruits (Lazzaro *et al.*, 2018 and Wu *et al.*, 2018). Moreover, fruit shape studies in other species have detected different QTLs such as *fsq8.1* in melon and *fs2-1* in cucumber (Monforte *et al.*, 2014 and Wu *et al.*, 2018).

Strawberry colour can vary from white to bright red. Reddish colours are provided by anthocyanins, mainly perlargonidin-3-glucoside. We mapped a major and totally stable QTL in LG1A (Table D-1 and Fig. D-3) and another minor and less stable QTL in the homeologous LG1B indicating that this QTL modulates colour trait in specific environments. White colour phenotype is caused by a truncated transcription factor, FaMYB10 which regulates anthocyanins pathway, (Wang *et al.*, 2019) mapped close to our LG1A colour QTL. Other FaMYB genes have been described to be involved in anthocyanin accumulation (Aharoni *et al.*, 2001, Schaart *et al.*, 2013 and Medina-Puche *et al.*, 2014). Flesh colour in peaches can vary from white, yellow or red. The white phenotype is controlled by a carotenoid dioxygenase, CCD4, (Brandi *et al.*, 2011) located in the middle of LG1 (Bliss *et al.*, 2002), while truncated alleles from this gene produced the yellow phenotype (Falchi *et al.*, 2013). Red fruit colour is controlled by a MYB10 gene located in LG3 (Bretó *et al.*, 2016). Similarly, the red colour QTL in pear was mapped in LG4 and LG5, with MYBs being the main candidate genes (Kumar *et al.*, 2019).

The results presented in this thesis could be followed by a fine mapping of some of the interested traits, genotyping the crosses that we have suggested to the company and whose seedlings are already growing. To discover some candidate genes for the VOC QTLs, the RNA sequencing pool of extreme phenotypes could detect differentially expressed genes. In addition, detecting variations in genomic sequences of the parental lines, such as 'FC50' and 'FD54' that we already have, would be useful to verify the causal mutations.

Fruit quality and appearance traits are polygenic and to enhance these traits several markers are required. With the applicability of these QTLs in mind, we are optimistic that the selected markers will work for MAS in other genetic backgrounds. This effort will surely pay off to satisfy the consumer demand.

Conclusions

1. An F1 population of 68 progenies, developed from a cross between two breeding cultivars, has been shown to be a valuable tool for the genetic characterization of fruit quality traits demanded by the strawberry consumers.

2. The 'FC50xFD54' genetic map constructed with 14595 SNPs grouping in 28 LGs and spanning 3451cM mapped the maximum number of SNPs compared to the published ones and showed high co-linearity with the *F. x ananassa* consensus map and the *F. vesca* and *F. x ananassa* genomes, last genome showing several LGs with opposite orientations and partial inversion in LG2D and LG6D.

3. The reduced genetic map constructed with 1461 SNPs and spanning 2332cM has the most appropriate genetic distances for QTL analysis.

4. The comparison between the predicted 'FC50xFD54' maps obtained using the IStraw35k and the 850k SNPs array showed similar parental segregating and no-segregating regions.

5. The improvement of the '21AF' genetic map, obtained from a F2 population of 'Camarosa' and 'Dover' cross, contains 7977 SNPs in 28 LGs and reduces to one third the total size of the previous genetic map being more convenient for QTL analysis

6. The volatile analysis in ripe fruits by GC-MS allowed the identification of a total of 58 VOCs, being 16 of them KVCs. High correlation between compounds from the same family was observed in several harvests.

7. The genetic analysis of VOCs segregation in the F1 population allowed to map 178 stable QTLs including 44 QTLs for 14 selected KVCs. Nine stable QTLs from the KVCs were mapped in the '21AF' population. The large number of QTL detected in the F1 population confirms its value for aroma analysis.

8. The QTL analysis in both populations were able to detect overlapping QTLs controlling terpenes accumulation at the beginning of LG3B, γ-decalactone QTL at the end of LG3D, methyl anthranilate QTL in the middle of LG7A and four major overlapping QTLs for ester accumulation in LG1A and LG7D for methyl group, LG4B for hexanoate related group and LG6A for octyl or acetate groups.

9. The candidate gene *FaAAT1* detected in the *Methyl*_7D region could be responsible for the methyl related esters variance.

10. The genetic analysis of fruit appearance and taste detected 77 stable QTLs in the 'FC50xFD54' population and 17 in the '21AF' population. The most important fruit quality QTLs found were the internal colour QTL in LG1A, the fruit shape QTLs in LG4C and LG6D, the neck QTLs in LG3A and LG4B and the firmness QTL in LG7C. Differences in shape and colour are more evident in the F1 population than in the F2.

11. Several SNPs linked to floral, fruity, peach and wild strawberry aroma have been identified and can be applied in MAS.

12. The analysis of VOCs composition allowed the selection of new parental lines for aroma breeding programs.

Bibliography

Aaby, K., Mazur, S., Nes, A., and Skrede, G. (2012). Phenolic compounds in strawberry (Fragaria x ananassa Duch.) fruits: Composition in 27 cultivars and changes during ripening. *Food Chemistry, 132*(1), 86-97. doi:10.1016/j.foodchem.2011.10.037

Agius, F., Gonzalez-Lamothe, R., Caballero, J. L., Munoz-Blanco, J., Botella, M. A., and Valpuesta, V. (2003). Engineering increased vitamin C levels in plants by overexpression of a D-galacturonic acid reductase. *Nature biotechnology, 21*(2), 177-181. doi:10.1038/nbt777

Aharoni, A., Giri, A. P., Verstappen, F. W., Bertea, C. M., Sevenier, R., Sun, Z., . . . Bouwmeester, H. J. (2004). Gain and loss of fruit flavor compounds produced by wild and cultivated strawberry species. *Plant Cell, 16*(11), 3110-3131. doi:10.1105/tpc.104.023895

Aharoni, A., De Vos, C. H. R., Wein, M., Sun, Z., Greco, R., Kroon, A., . . . O'Connell, A. P. (2001). The strawberry FaMYB1 transcription factor suppresses anthocyanin and flavonol accumulation in transgenic tobacco. *The Plant Journal, 28*(3), 319-332. doi:10.1046/j.1365-313X.2001.01154.x

Aharoni, A., Keizer, L. C. P., Bouwmeester, H. J., Sun, Z. K., Alvarez-Huerta, M., Verhoeven, H. A., . . . O'Connell, A. P. (2000). Identification of the SAAT gene involved in strawberry flavor biogenesis by use of DNA microarrays. *Plant Cell, 12*(5), 647-661. doi:Doi 10.1105/Tpc.12.5.647

Almeida, J. R., D'Amico, E., Preuss, A., Carbone, F., de Vos, C. H., Deiml, B., . . . Rosati, C. (2007). Characterization of major enzymes and genes involved in flavonoid and proanthocyanidin biosynthesis during fruit development in strawberry (Fragaria xananassa). *Archives of Biochemistry and Biophysics, 465*(1), 61-71. doi:10.1016/j.abb.2007.04.040

Anciro, A., Mangandi, J., Verma, S., Peres, N., Whitaker, V. M., and Lee, S. (2018). FaRCg1: a quantitative trait locus conferring resistance to Colletotrichum crown rot caused by Colletotrichum gloeosporioides in octoploid strawberry. *Theoretical and Applied Genetics, 131*(10), 2167-2177. doi:10.1007/s00122-018-3145-z

Appels, R., Eversole, K., Stein, N., Feuillet, C., Keller, B., Rogers, J., . . . Wang, L. (2018). Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science, 361*(6403), eaar7191. doi:10.1126/science.aar7191

Aragüez, I., Osorio, S., Hoffmann, T., Rambla, J. L., Medina-Escobar, N., Granell, A., . . . Valpuesta, V. (2013). Eugenol Production in Achenes and Receptacles of Strawberry Fruits Is Catalyzed by Synthases Exhibiting Distinct Kinetics. *Plant Physiology, 163*(2), 946-958. doi:10.1104/pp.113.224352

Ashley, M. V., Wilk, J. A., Styan, S. M. N., Craft, K. J., Jones, K. L., Feldheim, K. A., . . . Ashman, T. L. (2003). High variability and disomic segregation of microsatellites in the octoploid Fragaria virginiana Mill. (Rosaceae). *Theoretical and Applied Genetics, 107*(7), 1201-1207. doi:DOI 10.1007/s00122-003-1370-5

Bai, Y., Dougherty, L., Fazio, G., Li, M., Cheng, L., and Xu, K. (2012). A natural mutation-led truncation in one of the two aluminum-activated malate transporter-like genes at the Ma locus is associated with low fruit acidity in apple. *Molecular Genetics and Genomics, 287*, 663–678. doi:10.1007/s00438-012-0707-7

Baldi, P., Orsucci, S., Moser, M., Brilli, M., Giongo, L., and Si-Ammour, A. (2018). Gene expression and metabolite accumulation during strawberry (Fragaria x ananassa) fruit development and ripening. *Planta, 248*(5), 1143-1157. doi:10.1007/s00425-018-2962-2

Bassil, N. V., Davis, T. M., Zhang, H., Ficklin, S., Mittmann, M., Webster, T., . . . van de Weg, E. (2015). Development and preliminary evaluation of a 90 K Axiom® SNP array for the allo-octoploid cultivated strawberry Fragaria × ananassa. *BMC Genomics, 16*, 155-185. doi:doi:10.1186/s12864-015-1310-1

Bassil, N. V., Gunn, M., Folta, K., and Lewers, K. (2006). Microsatellite markers for Fragaria from 'Strawberry Festival' expressed sequence tags. *Molecular Ecology Notes, 6*(2), 473-476. doi:10.1111/j.1471-8286.2006.01278.x

Beekwilder, J., Alvarez-Huerta, M., Neef, E., Verstappen, F. W., Bouwmeester, H. J., and Aharoni, A. (2004). Functional characterization of enzymes forming volatile esters from strawberry and banana. *Plant Physiology, 135*(4), 1865-1878. doi:10.1104/pp.104.042580

Bekaert, M. (2016). *Genetic-Mapper: vectorial genetic map drawer.* Paper presented at the F1000Research.

Bianco, L., Cestaro, A., Linsmith, G., Muranty, H., Denance, C., Theron, A., . . . Troggio, M. (2016). Development and validation of the Axiom((R)) Apple480K SNP genotyping array. *The Plant Journal, 86*(1), 62-74. doi:10.1111/tpj.13145

Bliss, F. A., Arulsekar, S., Foolad, M. R., Becerra, V., Gillen, A. M., Warburton, M. L., . . . Mydin, K. K. (2002). An expanded genetic linkage map of Prunus based on an interspecific cross between almond and peach. *Genome, 45*(3), 520-529. doi:10.1139/g02-011

Boudehri, K., Bendahmane, A., Cardinet, G., Troadec, C., Moing, A., and Dirlewanger, E. (2009). Phenotypic and fine genetic characterization of the D locus controlling fruit acidity in peach. *BMC plant biology, 9*, 59. doi:10.1186/1471-2229-9-59

Bourke, P. M., Arens, P., Voorrips, R. E., Esselink, G. D., Koning-Boucoiran, C. F., Van't Westende, W. P., . . . Maliepaard, C. (2017). Partial preferential chromosome pairing is genotype dependent in tetraploid rose. *Plant Journal, 90*(2), 330-343. doi:10.1111/tpj.13496

Brandi, F., Bar, E., Mourgues, F., Horváth, G., Turcsi, E., Giuliano, G., . . . Rosati, C. (2011). Study of 'Redhaven' peach and its white-fleshed mutant suggests a key role of CCD4 carotenoid dioxygenase in carotenoid and norisoprenoid volatile metabolism. *BMC plant biology, 11*, 24. doi:10.1186/1471-2229-11-24

Brenchley, R., Spannagl, M., Pfeifer, M., Barker, G. L., D'Amore, R., Allen, A. M., . . . Hall, N. (2012). Analysis of the bread wheat genome using whole-genome shotgun sequencing. *Nature, 491*(7426), 705-710. doi:10.1038/nature11650

Bretó, M. P., Cantín, C. M., Iglesias, I., Arús, P., and Eduardo, I. (2016). Mapping a major gene for red skin color suppression (highlighter) in peach. *Euphytica, 213*(1). doi:10.1007/s10681-016-1812-1

Buendia, B., Gil, M. I., Tudela, J. A., Gady, A. L., Medina, J. J., Soria, C., . . . Tomas-Barberan, F. A. (2010). HPLC-MS analysis of proanthocyanidin oligomers and other phenolics in 15 strawberry cultivars. *Journal of agricultural and food chemistry, 58*(7), 3916-3926. doi:10.1021/jf9030597

Cai, C., Zhu, G., Zhang, T., and Guo, W. (2017). High-density 80 K SNP array is a powerful tool for genotyping G. hirsutum accessions and genome analysis. *BMC Genomics, 18*(1), 654. doi:10.1186/s12864-017-4062-2

Cao, K., Zhou, Z., Wang, Q., Guo, J., Zhao, P., Zhu, G., . . . Wang, L. (2016). Genome-wide association study of 12 agronomic traits in peach. *Nature Communications, 7*, 13246. doi:10.1038/ncomms13246

Carbone, F., Preuss, A., De Vos, R. C., D'Amico, E., Perrotta, G., Bovy, A. G., . . . Rosati, C. (2009). Developmental, genetic and environmental factors affect the expression of flavonoid genes, enzymes and metabolites in strawberry fruits. *Plant, Cell and Environment, 32*(8), 1117-1131. doi:10.1111/j.1365-3040.2009.01994.x

Castro, P., and Lewers, K. S. (2016). Identification of quantitative trait loci (QTL) for fruit-quality traits and number of weeks of flowering in the cultivated strawberry. *Molecular Breeding, 36*(10). doi:10.1007/s11032-016-0559-7

Castro, P., Weebadde, C. K., Finn, C. E., Bushakra, J. M., Wang, D., Luby, J. J., . . . Lewers, K. S. (2015). Genetic mapping of day-neutrality in cultivated strawberry. *Molecular Breeding, 35*(79), 1-16. doi:10.1007/s11032-015-0250-4

Cavanagh, C. R., Chao, S., Wang, S., Huang, B. E., Stephen, S., Kiani, S., . . . Akhunov, E. (2013). Genome-wide comparative diversity uncovers multiple targets of selection for improvement in hexaploid wheat landraces and cultivars. *Proceedings of the National Academy of Sciences, 110*(20), 8057-8062. doi:10.1073/pnas.1217133110

Chambers, A. H., Pillet, J., Plotto, A., Bai, J., Whitaker, V. M., and Folta, K. M. (2014). Identification of a strawberry flavor gene candidate using an integrated genetic-genomic-analytical chemistry approach. *BMC Genomics, 15*(217), 1-15.

Chambers, A., Whitaker, V. M., Gibbs, B., Plotto, A., and Folta, K. M. (2012). Detection of the linalool-producing NES1 variant across diverse strawberry (Fragaria spp.) accessions. *Plant Breeding, 131*(3), 437-443. doi:DOI 10.1111/j.1439-0523.2012.01959.x

Chen, G., Hackett, R., Walker, D., Taylor, A., Lin, Z., and 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*(1), 2641-2651. doi:10.1104/pp.104.041608

Cockerton, H. M., Li, B., Vickerstaff, R. J., Eyre, C. A., Sargent, D. J., Armitage, A. D., . . . Harrison, R. J. (2019). Identifying Verticillium dahliae Resistance in Strawberry Through Disease Screening of Multiple Populations and Image Based Phenotyping. *Frontiers in Plant Science, 10*, 924. doi:10.3389/fpls.2019.00924

Cockerton, H. M., Vickerstaff, R. J., Karlstrom, A., Wilson, F., Sobczyk, M., He, J. Q., . . . Harrison, R. J. (2018). Identification of powdery mildew resistance QTL in strawberry (Fragaria x ananassa). *Theoretical and Applied Genetics, 131*(9), 1995-2007. doi:10.1007/s00122-018-3128-0

Costa, F., Cappellin, L., Zini, E., Patocchi, A., Kellerhals, M., Komjanc, M., . . . Biasioli, F. (2013). QTL validation and stability for volatile organic compounds (VOCs) in apple. *Plant Science, 211*, 1-7. doi:10.1016/j.plantsci.2013.05.018

Cruz-Rus, E., Sesmero, R., Angel-Perez, J. A., Sanchez-Sevilla, J. F., Ulrich, D., and Amaya, I. (2017). Validation of a PCR test to predict the presence of flavor volatiles mesifurane and gamma-decalactone in fruits of cultivated strawberry (Fragaria x ananassa). *Molecular Breeding, 37*(10), 131. doi:10.1007/s11032-017-0732-7

Cruz-Rus, E., Amaya, I., Sanchez-Sevilla, J. F., Botella, M. A., and Valpuesta, V. (2011). Regulation of L-ascorbic acid content in strawberry fruits. *Journal of Experimental Botany, 62*(12), 4191-4201. doi:10.1093/jxb/err122

Darrigues, A., Hall, J., van der Knaap, E., Francis, D. M., Dujmovic, N., and Gray, S. (2008). Tomato Analyzer-color test: a new tool for efficient digital phenotyping. *Journal of the American Society for Horticultural Science, 133*(4), 579-586.

Darrow, G. (1966). *The Strawberry: History Breeding and Physiology*. New York, USA: Holt, Rinehart and Winston.

Davik, J., Sargent, D. J., Brurberg, M. B., Lien, S., Kent, M., and Alsheikh, M. (2015). A ddRAD Based Linkage Map of the Cultivated Strawberry, Fragaria xananassa. *PLoS One, 10*(9), e0137746. doi:10.1371/journal.pone.0137746

Davis, T. M., and Yu, H. (1997). A Linkage Map of the Diploid Strawberry, Fragaria vesca. *Journal of Heredity, 88*(3), 215-221.

Diamanti, J., Capocasa, F., Balducci, F., Battino, M., Hancock, J., and Mezzetti, B. (2012). Increasing strawberry fruit sensorial and nutritional quality using wild and cultivated germplasm. *PLoS One, 7*(10), e46470. doi:10.1371/journal.pone.0046470

DiMeglio, L. M., Staudt, G., Yu, H., and Davis, T. M. (2014). A phylogenetic analysis of the genus Fragaria (strawberry) using intron-containing sequence from the ADH-1 gene. *PLoS One, 9*(7), e102237. doi:10.1371/journal.pone.0102237

Dirlewanger, E., Cosson, P., Boudehri, K., Renaud, C., Capdeville, G., Tauzin, Y., . . . Moing, A. (2006). Development of a second-generation genetic linkage map for peach [Prunus persica (L.) Batsch] and characterization of morphological traits affecting flower and fruit. *Tree genetics & genomes, 3*(1), 1-13. doi:10.1007/s11295-006-0053-1

Dixon, J., and Hewett, E. W. (2000). Factors affecting apple aroma/flavour volatile concentration: A Review. *New Zealand Journal of Crop and Horticultural Science, 28*(3), 155-173. doi:10.1080/01140671.2000.9514136

Doyle, J. J., and Doyle, J. L. (1990). Isolation of plant DNA from fresh tissue. *Focus, 12*, 13–15.

Draye, M., and Van Cutsem, P. (2008). Pectin methylesterases induce an abrupt increase of acidic pectin during strawberry fruit ripening. *Journal of plant physiology, 165*(11), 1152-1160. doi:10.1016/j.jplph.2007.10.006

Edger, P. P., McKain, M. R., Yocca, A. E., Knapp, S. J., Qiao, Q., and Zhang, T. (2020). Reply to: Revisiting the origin of octoploid strawberry. *Nature Genetics, 52*(1), 5-7. doi:10.1038/s41588-019-0544-2

Edger, P. P., Poorten, T. J., VanBuren, R., Hardigan, M. A., Colle, M., McKain, M. R., . . . Knapp, S. J. (2019). Origin and evolution of the octoploid strawberry genome. *Nature Genetics, 51*(3), 541-547. doi:10.1038/s41588-019-0356-4

Edger, P. P., VanBuren, R., Colle, M., Poorten, T. J., Wai, C. M., Niederhuth, C. E., . . . Knapp, S. J. (2018). Single-molecule sequencing and optical mapping yields an improved genome of woodland strawberry (Fragaria vesca) with chromosome-scale contiguity. *Gigascience, 7*(2), 1-7. doi:10.1093/gigascience/gix124

Eduardo, I., Chietera, G., Pirona, R., Pacheco, I., Troggio, M., Banchi, E., . . . Pozzi, C. (2013). Genetic dissection of aroma volatile compounds from the essential oil of peach fruit: QTL analysis and identification of candidate genes using dense SNP maps. *Tree genetics & genomes, 9*, 189–204. doi:10.1007/s11295-012-0546-z

Eduardo, I., Chietera, G., Bassi, D., Rossini, L., and Vecchietti, A. (2010). Identification of key odor volatile compounds in the essential oil of nine peach accessions. *Journal of the Science of Food and Agriculture, 90*(7), 1146-1154. doi:10.1002/jsfa.3932

Fait, A., Hanhineva, K., Beleggia, R., Dai, N., Rogachev, I., Nikiforova, V. J., . . . Aharoni, A. (2008). Reconfiguration of the achene and receptacle metabolic networks during strawberry fruit development. *Plant Physiology, 148*(2), 730-750. doi:10.1104/pp.108.120691

Falchi, R., Vendramin, E., Zanon, L., Scalabrin, S., Cipriani, G., Verde, I., . . . Morgante, M. (2013). Three distinct mutational mechanisms acting on a single gene underpin the origin of yellow flesh in peach. *Plant Journal, 76*(2), 175-187. doi:10.1111/tpj.12283

Feldmann, M. J., Hardigan, M. A., Famula, R. A., Lopez, C. M., Tabb, A., Cole, G. S., and Knapp, S. J. (2020). Multi-dimensional machine learning approaches for fruit shape phenotyping in strawberry. *Gigascience, 9*(5). doi:10.1093/gigascience/giaa030

Figueroa, C. R., Rosli, H. G., Civello, P. M., Martínez, G. A., Herrera, R., and Moya-León, M. A. (2010). Changes in cell wall polysaccharides and cell wall degrading enzymes during ripening of Fragaria chiloensis and Fragaria ×ananassa fruits. *Scientia Horticulturae, 124*(4), 454-462. doi:10.1016/j.scienta.2010.02.003

Folta, K. M., and Klee, H. J. (2016). Sensory sacrifices when we mass-produce mass produce. *Horticulture research, 3*, 16032. doi:10.1038/hortres.2016.32

Folta, K. M., and Davis, T. M. (2006). Strawberry genes and genomics. *Critical Reviews in Plant Sciences, 25*(5), 399-415. doi:10.1080/07352680600824831

Folta, K. M., Staton, M., Stewart, P. J., Jung, S., Bies, D. H., Jesdurai, C., and Main, D. (2005). Expressed sequence tags (ESTs) and simple sequence repeat (SSR) markers from octoploid strawberry (Fragaria x ananassa). *BMC Plant Biol, 5*(1), 12. doi:10.1186/1471-2229-5-12

Forbes-Hernandez, T. Y., Gasparrini, M., Afrin, S., Bompadre, S., Mezzetti, B., Quiles, J. L., . . . Battino, M. (2016). The Healthy Effects of Strawberry Polyphenols: Which Strategy behind Antioxidant Capacity? *Crit Rev Food Sci Nutr, 56 Suppl 1*, S46-59. doi:10.1080/10408398.2015.1051919

Fridman, E., Carrari, F., Liu, Y., Fernie, A. R., and Zamir, D. (2004). Zooming in on a Quantitative Trait for tomato yield using interspecific introgressions. *Science, 305*(5691), 1786-1789. doi:10.1126/science.1101666

Gao, L., Gonda, I., Sun, H., Ma, Q., Bao, K., Tieman, D. M., . . . Fei, Z. (2019). The tomato pan-genome uncovers new genes and a rare allele regulating fruit flavor. *Nature Genetics, 51*(6), 1044-1051. doi:10.1038/s41588-019-0410-2

Giampieri, F., Alvarez-Suarez, J. M., and Battino, M. (2014). Strawberry and human health: effects beyond antioxidant activity. *Journal of agricultural and food chemistry, 62*(18), 3867-3876. doi:10.1021/jf405455n

Goff, S. A., and Klee, H. J. (2006). Plant volatile compounds: sensory cues for health and nutritional value? *Science, 311*(5762), 815-819. doi:10.1126/science.1112614

Gonzalez-Aguero, M., Troncoso, S., Gudenschwager, O., Campos-Vargas, R., Moya-Leon, M. A., and Defilippi, B. G. (2009). Differential expression levels of aroma-related genes during ripening of apricot (Prunus armeniaca L.). *Plant Physiology and Biochemistry, 47*(5), 435-440. doi:10.1016/j.plaphy.2009.01.002

Gonzalez, M., Gaete-Eastman, C., Valdenegro, M., Figueroa, C. R., Fuentes, L., Herrera, R., and Moya-Leon, M. A. (2009). Aroma development during ripening of Fragaria chiloensis fruit and participation of an alcohol acyltransferase (FcAAT1) gene. *Journal of agricultural and food chemistry, 57*(19), 9123-9132. doi:10.1021/jf901693j

Gonzalo, M. J., Brewer, M. T., Anderson, C., Sullivan, D., Gray, S., and van der Knaap, E. (2009). Tomato fruit shape analysis using morphometric and morphology attributes Implemented in Tomato Analyzer software program. *Journal of the American Society for Horticultural Science, 134*(1), 77-87.

Gor, M. C., Candappa, C., de Silva, T., Mantri, N., and Pang, E. (2017). Identification and validation of FaP1D7, a putative marker associated with the biosynthesis of methyl butanoate in cultivated strawberry (Fragaria x ananassa). *Scientific Reports, 7*(1), 17454. doi:10.1038/s41598-017-17448-1

Govan, C. L., Simpson, D. W., Johnson, A. W., Tobutt, K. R., and Sargent, D. J. (2008). A reliable multiplexed microsatellite set for genotyping Fragaria and its use in a survey of 60 F. x ananassa cultivars. *Molecular Breeding, 22*(4), 649-661. doi:10.1007/s11032-008-9206-2

Granell, A., and Rambla, J. L. (2013). *6. Biosynthesis of Volatile Compounds* (G. B. Seymour, Mervin Poole, J. J. Giovannoni, & G. A. Tucker. Eds.): JohnWiley & Sons, Inc.

Greger, V., and Schieberle, P. (2007). Characterization of the key aroma compounds in apricots (Prunus armeniaca) by application of the molecular sensory science concept. *Journal of agricultural and food chemistry, 55*(13), 5221-5228. doi:10.1021/jf0705015

Guan, Y., Peace, C., Rudell, D., Verma, S., and Evans, K. (2015). QTLs detected for individual sugars and soluble solids content in apple. *Molecular Breeding, 35*(6). doi:10.1007/s11032-015-0334-1

Halbwirth, H., Puhl, I., Haas, U., Jezik, K., Treutter, D., and Stich, K. (2006). Two-phase flavonoid formation in developing strawberry (Fragaria x ananassa) fruit. *Journal of agricultural and food chemistry, 54*(4), 1479-1485. doi:10.1021/jf0524170

Hamada, H., Petrino, M., and Kakunaga, T. (1982). A novel repeated element with Z-DNA-forming potential is widely found in evolutionarily diverse eukaryotic genomes. *Proceedings of the National Academy of Sciences, 79*, 6465-6469.

Hardigan, M. A., Feldmann, M. J., Lorant, A., Bird, K. A., Famula, R., Acharya, C., . . . Knapp, S. J. (2020). Genome synteny has been conserved among the octoploid progenitors of cultivated strawberry over millions of years of evolution. *Frontiers in Plant Science, 10*, 1789. doi:10.3389/fpls.2019.01789

Hardigan, M. A., Poorten, T. J., Acharya, C., Cole, G. S., Hummer, K. E., Bassil, N., . . . Knapp, S. J. (2018). Domestication of temperate and coastal hybrids with distinct ancestral gene selection in octoploid strawberry. *Plant Genome, 11*(3), 1-11. doi:10.3835/plantgenome2018.07.0049

Harrison, R. E., Luby, J. J., Furnier, G. R., and Hancock, J. F. (1997). Morphological and molecular variation among population of octoploid Fragaria virginiana and F. chiloensis (Rosaceae) from North America. *American journal of botany, 84*(5), 612-620.

Harrison, R. J., Gonzalo A. Accinelli Garcia, G. A., Yanes, L., Bates, H., Armitage, A., Clark, M., . . . Clavijo, B. J. (2019). *W1008: Haplotype Resolving the Genome of Fragaria x Ananassa*

Using the Basic Sequence Graph Framework. Paper presented at the XXVII Plant and Animal Genome Conference, San Diego, CA, US.

Hasin-Brumshtein, Y., Lancet, D., and Olender, T. (2009). Human olfaction: from genomic variation to phenotypic diversity. *Trends in Genetics, 25*(4), 178-184. doi:10.1016/j.tig.2009.02.002

Hawkins, C., Caruana, J., Schiksnis, E., and Liu, Z. (2016). Genome-scale DNA variant analysis and functional validation of a SNP underlying yellow fruit color in wild strawberry. *Scientific Reports, 6*, 29017. doi:10.1038/srep29017

Hernandez Mora, J. R., Micheletti, D., Bink, M., Van de Weg, E., Cantin, C., Nazzicari, N., . . . Aranzana, M. J. (2017). Integrated QTL detection for key breeding traits in multiple peach progenies. *BMC Genomics, 18*(1), 404. doi:10.1186/s12864-017-3783-6

Hibrand Saint-Oyant, L., Ruttink, T., Hamama, L., Kirov, I., Lakhwani, D., Zhou, N. N., . . . Foucher, F. (2018). A high-quality genome sequence of Rosa chinensis to elucidate ornamental traits. *Nature Plants, 4*(7), 473-484. doi:10.1038/s41477-018-0166-1

Hidalgo, J. M., Arús, P., and Monfort, A. (2014). *An improved version of cultivated strawberry, linkage map using the IStraw90 Axiom Array for QTL analysis*. Paper presented at the 7th International Rosaceae Genomics Conference, Seattle, WA, US.

Hirakawa, H., Shirasawa, K., Kosugi, S., Tashiro, K., Nakayama, S., Yamada, M., . . . Isobe, S. N. (2014). Dissection of the octoploid strawberry genome by deep sequencing of the genomes of Fragaria species. *DNA research, 21*(2), 169-181. doi:10.1093/dnares/dst049

Hirata, H., Ohnishi, T., Tomida, K., Ishida, H., Kanda, M., Sakai, M., . . . Watanabe, N. (2016). Seasonal induction of alternative principal pathway for rose flower scent. *Scientific Reports, 6*, 20234. doi:10.1038/srep20234

Honjo, M., Nunome, T., Kataoka, S., Yano, T., Hamano, M., Yamazaki, H., . . . Yui, S. (2015). Simple sequence repeat markers linked to the everbearing flowering gene in long-day and day-neutral cultivars of the octoploid cultivated strawberry Fragaria × ananassa. *Euphytica, 209*(2), 291-303. doi:10.1007/s10681-015-1626-6

Hu, Y., Chen, J., Fang, L., Zhang, Z., Ma, W., Niu, Y., . . . Zhang, T. (2019). Gossypium barbadense and Gossypium hirsutum genomes provide insights into the origin and evolution of allotetraploid cotton. *Nature Genetics, 51*(4), 739-748. doi:10.1038/s41588-019-0371-5

Hulse-Kemp, A. M., Lemm, J., Plieske, J., Ashrafi, H., Buyyarapu, R., Fang, D. D., . . . Stelly, D. M. (2015). Development of a 63K SNP Array for Cotton and High-Density Mapping of Intraspecific and Interspecific Populations of Gossypium spp. *GENES, GENOMES, GENETICS, 5*(6), 1187-1209. doi:10.1534/g3.115.018416

Hummer, K. E., and Hancock, J. (2009). *Strawberry genomics: botanical history, cultivation, traditional breeding and new technologies* (K. Folta & S. Gardiner Eds. Vol. 6): Springer-Verlag New York.

Isobe, S., Shirasawa, K., Chung, C., Boncan, D., Wada, T., Hirakawa, H., . . . Chan, T. (2020). *W1005 Strawberry Genomics: Whole Genome Assembly and Structure Analysis in Japanese Strawberry*. Paper presented at the XVIII Plant and Animal Genome Conference, San Diego, CA, US.

Isobe, S. N., Hirakawa, H., Sato, S., Maeda, F., Ishikawa, M., Mori, T., . . . Tabata, S. (2013). Construction of an integrated high density simple sequence repeat linkage map in cultivated strawberry (Fragaria x ananassa) and its applicability. *DNA research, 20*(1), 79-92. doi:10.1093/dnares/dss035

Jetti, R. R., Yang, E., Kurnianta, A., Finn, C., and Qian, M. C. (2007). Quantification of selected aroma-active compounds in strawberries by headspace solid-phase microextraction gas chromatography and correlation with sensory descriptive analysis. *J Food Sci, 72*(7), S487-S496. doi:DOI 10.1111/j.1750-3841.2007.00445.x

Jia, D., Shen, F., Wang, Y., Wu, T., Xu, X., Zhang, X., and Han, Z. (2018). Apple fruit acidity is genetically diversified by natural variations in three hierarchical epistatic genes:

MdSAUR37, MdPP2CH and MdALMTII. *The Plant Journal, 95*(3), 427-443. doi:10.1111/tpj.13957

Jia, H., Jiu, S., Zhang, C., Wang, C., Tariq, P., Liu, Z., . . . Fang, J. (2016). Abscisic acid and sucrose regulate tomato and strawberry fruit ripening through the abscisic acid-stress-ripening transcription factor. *Plant Biotechnology Journal, 14*(10), 2045-2065. doi:10.1111/pbi.12563

Jia, H., Wang, Y., Sun, M., Li, B., Han, Y., Zhao, Y., . . . Jia, W. (2013). Sucrose functions as a signal involved in the regulation of strawberry fruit development and ripening. *New Phytologist, 198*(2), 453-465. doi:10.1111/nph.12176

Jibran, R., Dzierzon, H., Bassil, N., Bushakra, J. M., Edger, P. P., Sullivan, S., . . . Chagne, D. (2018). Chromosome-scale scaffolding of the black raspberry (Rubus occidentalis L.) genome based on chromatin interaction data. *Horticulture research, 5*, 8. doi:10.1038/s41438-017-0013-y

Jung, S., Lee, T., Cheng, C. H., Buble, K., Zheng, P., Yu, J., . . . Main, D. (2019). 15 years of GDR: New data and functionality in the Genome Database for Rosaceae. *Nucleic acids research, 47*(D1), D1137-d1145. doi:10.1093/nar/gky1000

Jung, S., Cestaro, A., Troggio, M., Main, D., Zheng, P., Cho, I., . . . Sargent, D. J. (2012). Whole genome comparisons of Fragaria, Prunus and Malus reveal different modes of evolution between Rosaceous subfamilies. *BMC Genomics, 13*(129). doi:10.1186/1471-2164-13-129

Khan, S. A., Beekwilder, J., Schaart, J. G., Mumm, R., Soriano, J. M., Jacobsen, E., and Schouten, H. J. (2013). Differences in acidity of apples are probably mainly caused by a malic acid transporter gene on LG16. *Tree genetics & genomes, 9*, 475–487. doi:10.1007/s11295-012-0571-y

Klee, H. J., and Tieman, D. M. (2018). The genetics of fruit flavour preferences. *Nature reviews genetics, 19*(6), 347-356. doi:10.1038/s41576-018-0002-5

Knee, M., Sargent, D. J., and Orborne, D. J. (1977). Cell wall metabolism in developing strawberry fruits. *Journal of Experimental Botany, 28*(108), 377-393, 395-396.

Koning-Boucoiran, C. F., Esselink, G. D., Vukosavljev, M., van 't Westende, W. P., Gitonga, V. W., Krens, F. A., . . . Smulders, M. J. (2015). Using RNA-Seq to assemble a rose transcriptome with more than 13,000 full-length expressed genes and to develop the WagRhSNP 68k Axiom SNP array for rose (Rosa L.). *Frontiers in Plant Science, 6*, 249. doi:10.3389/fpls.2015.00249

Kosambi, D. D. (1944). The Estimation of Map Distances from Recombination Values. *Annals of Eugenics, 12*, 172-175. doi:10.1111/j.1469-1809.1943.tb02321.x

Koskela, E. A., Sonsteby, A., Flachowsky, H., Heide, O. M., Hanke, M. V., Elomaa, P., and Hytonen, T. (2016). TERMINAL FLOWER1 is a breeding target for a novel everbearing trait and tailored flowering responses in cultivated strawberry (Fragaria x ananassa Duch.). *Plant Biotechnology Journal, 14*(9), 1852-1861. doi:10.1111/pbi.12545

Kovats, E. (1958). Gas chromatographische Charakterisierung organischer Verbindung. I. Retentions indices aliphatischer halogenide, alkohole, aldehyde und ketone. *Helv Chim Acta, 41*, 1915-1932.

Kumar, S., Kirk, C., Deng, C. H., Wiedow, C., Qin, M., Espley, R., . . . Brewer, L. (2019). Fine-mapping and validation of the genomic region underpinning pear red skin colour. *Horticulture research, 6*, 29. doi:10.1038/s41438-018-0112-4

Kyriakidou, M., Anglin, N. L., Ellis, D., Tai, H. H., and Stromvik, M. V. (2020). Genome assembly of six polyploid potato genomes. *Scientific Data, 7*(1), 88. doi:10.1038/s41597-020-0428-4

Larsen, B., Migicovsky, Z., Jeppesen, A. A., Gardner, K. M., Toldam-Andersen, T. B., Myles, S., . . . Pedersen, C. (2019). Genome-Wide Association Studies in Apple Reveal Loci for Aroma Volatiles, Sugar Composition, and Harvest Date. *Plant Genome, 12*(2). doi:10.3835/plantgenome2018.12.0104

Lazzaro, M. D., Wu, S., Snouffer, A., Wang, Y., and van der Knaap, E. (2018). Plant Organ Shapes Are Regulated by Protein Interactions and Associations With Microtubules. *Frontiers in Plant Science, 9*, 1766. doi:10.3389/fpls.2018.01766

Leone, A., Bleve-Zacheo, T., Gerardi, C., Melillo, M. T., Leo, L., and Zacheo, G. (2006). Lipoxygenase involvement in ripening strawberry. *Journal of agricultural and food chemistry, 54*(18), 6835-6844. doi:10.1021/jf061457g

Lerceteau-Köhler, E., Moing, A., Guérin, G., Renaud, C., Petit, A., Rothan, C., and Denoyes, B. (2012). Genetic dissection of fruit quality traits in the octoploid cultivated strawberry highlights the role of homoeo-QTL in their control. *Theoretical and Applied Genetics, 124*, 1059-1077. doi:10.1007/s00122-011-1769-3

Lerceteau-Köhler, E., Moing, A., Guérin, G., Renaud, C., Courlit, S., Camy, D., . . . Denoyes-Rothan, B. (2004). QTL Analysis for Fruit Quality Traits in Octoploid Strawberry (Fragaria x ananassa). *Acta Hort. 663, ISHS 2004, 663*, 331-336

Lerceteau-Kohler, E., Guerin, G., Laigret, F., and Denoyes-Rothan, B. (2003). Characterization of mixed disomic and polysomic inheritance in the octoploid strawberry (Fragaria x ananassa) using AFLP mapping. *Theoretical and Applied Genetics, 107*(4), 619-628. doi:10.1007/s00122-003-1300-6

Li, F., Fan, G., Lu, C., Xiao, G., Zou, C., Kohel, R. J., . . . Yu, S. (2015). Genome sequence of cultivated Upland cotton (Gossypium hirsutum TM-1) provides insights into genome evolution. *Nature biotechnology, 33*(5), 524-530. doi:10.1038/nbt.3208

Li, Y., Pi, M., Gao, Q., Liu, Z., and Kang, C. (2019). Updated annotation of the wild strawberry Fragaria vesca V4 genome. *Horticulture research, 6*, 61. doi:10.1038/s41438-019-0142-6

Liao, L., Li, M., Liu, B., Yan, M., Yu, X., Zi, H., . . . Yamamuro, C. (2018). Interlinked regulatory loops of ABA catabolism and biosynthesis coordinate fruit growth and ripening in woodland strawberry. *Proceedings of the National Academy of Sciences, 49*, 11542–11550. doi:10.1073/pnas.1812575115

Liebhard, R., Kellerhals, M., Pfammatter, W., Jertmini, M., and Gessler, C. (2003). Mapping quantitative physiological traits in apple (Malus × domestica Borkh.). *Plant Molecular Biology, 52*, 511-526. doi:10.1023/A:1024886500979

Lin-Wang, K., Bolitho, K., Grafton, K., Kortstee, A., Karunairetnam, S., McGhie, T. K., . . . Allan, A. C. (2010). An R2R3 MYB transcription factor associated with regulation of the anthocyanin biosynthetic pathway in Rosaceae. *BMC plant biology, 10*(50), 1-17.

Liston, A., Wei, N., Tennessen, J. A., Li, J., Dong, M., and Ashman, T. L. (2020). Revisiting the origin of octoploid strawberry. *Nature Genetics, 52*(1), 2-4. doi:10.1038/s41588-019-0543-3

Liu, X., Zhao, B., Zheng, H. J., Hu, Y., Lu, G., Yang, C. Q., . . . Chen, X. Y. (2015). Gossypium barbadense genome sequence provides insight into the evolution of extra-long staple fiber and specialized metabolites. *Scientific Reports, 5*, 14139. doi:10.1038/srep14139

Lopez-Girona, E., Zhang, Y., Eduardo, I., Mora, J. R. H., Alexiou, K. G., Arus, P., and Aranzana, M. J. (2017). A deletion affecting an LRR-RLK gene co-segregates with the fruit flat shape trait in peach. *Scientific Reports, 7*(1), 6714. doi:10.1038/s41598-017-07022-0

Ma, B., Zhao, S., Wu, B., Wang, D., Peng, Q., Owiti, A., . . . Han, Y. (2016). Construction of a high density linkage map and its application in the identification of QTLs for soluble sugar and organic acid components in apple. *Tree genetics & genomes, 12*, 1. doi:10.1007/s11295-015-0959-6

Mahoney, L. L., Sargent, D. J., Abebe-Akele, F., Wood, D. J., Ward, J. A., Bassil, N. V., . . . Davis, T. M. (2016). A high-density linkage map of the ancestral diploid strawberry, constructed with single nucleotide polymorphism markers from the IStraw90 array and genotyping by sequencing. *Plant Genome, 9*(2). doi:10.3835/plantgenome2015.08.0071

Mangandi, J., Verma, S., Osorio, L., Peres, N. A., van de Weg, E., and Whitaker, V. M. (2017). Pedigree-based analysis in a multiparental population of octoploid strawberry reveals

QTL alleles conferring resistance to Phytophthora cactorum. *GENES, GENOMES, GENETICS, 7*(6), 1707-1719. doi:10.1534/g3.117.042119

Medina-Puche, L., Molina-Hidalgo, F. J., Boersma, M. R., Schuurink, R. C., López-Vidriero, I., Solano, R., . . . Muñoz-Blanco, J. (2015). A R2R3-MYB transcription factor (FaEOBII) regulates eugenol production in ripe strawberry (*Fragaria × ananassa*) fruit receptacles. *Plant Physiology*. doi:10.1104/pp.114.252908

Medina-Puche, L., Cumplido-Laso, G., Amil-Ruiz, F., Hoffmann, T., Ring, L., Rodriguez-Franco, A., . . . Blanco-Portales, R. (2014). MYB10 plays a major role in the regulation of flavonoid/phenylpropanoid metabolism during ripening of Fragaria x ananassa fruits. *Journal of Experimental Botany, 65*(2), 401-417. doi:10.1093/jxb/ert377

Mendez-Yanez, A., Gonzalez, M., Carrasco-Orellana, C., Herrera, R., and Moya-Leon, M. A. (2020). Isolation of a rhamnogalacturonan lyase expressed during ripening of the Chilean strawberry fruit and its biochemical characterization. *Plant Physiology and Biochemistry, 146*, 411-419. doi:10.1016/j.plaphy.2019.11.041

Mengist, M. F., Alves, S., Griffin, D., Creedon, J., McLaughlin, M. J., Jones, P. W., and Milbourne, D. (2018). Genetic mapping of quantitative trait loci for tuber-cadmium and zinc concentration in potato reveals associations with maturity and both overlapping and independent components of genetic control. *Theoretical and Applied Genetics, 131*(4), 929-945. doi:10.1007/s00122-017-3048-4

Moing, A., Renaud, C., Gaudillère, M., Raymond, P., P., R., and Denoyes-Rothan, B. (2001). Biochemical changes during fruit development of four strawberry cultivars. *Journal of the American Society for Horticultural Science, 126*(4), 394–403.

Molina-Hidalgo, F. J., Franco, A. R., Villatoro, C., Medina-Puche, L., Mercado, J. A., Hidalgo, M. A., . . . Blanco-Portales, R. (2013). The strawberry *(Fragaria x ananassa*) fruit-specific *rhamnogalacturonate lyase 1* (*FaRGLyase1*) gene encodes an enzyme involved in the degradation of cell-wall middle lamellae. *Journal of Experimental Botany, 64*(6), 1471-1483. doi:10.1093/jxb/ers386

Monfort, A., Vilanova, S., Davis, T. M., and Arús, P. (2006). A new set of polymorphic simple sequence repeat (SSR) markers from a wild strawberry (*Fragaria vesca*) are transferable to other diploid Fragaria species and to *Fragaria × ananassa*. *Molecular Ecology Notes, 6*(1), 197-200.

Monforte, A. J., Diaz, A., Cano-Delgado, A., and van der Knaap, E. (2014). The genetic basis of fruit morphology in horticultural crops: lessons from tomato and melon. *Journal of Experimental Botany, 65*(16), 4625-4637. doi:10.1093/jxb/eru017

Montanari, S., Bianco, L., Allen, B. J., Martinez-Garcia, P. J., Bassil, N. V., Postman, J., . . . Neale, D. B. (2019). Development of a highly efficient Axiom 70 K SNP array for Pyrus and evaluation for high-density mapping and germplasm characterization. *BMC Genomics, 20*(1), 331. doi:10.1186/s12864-019-5712-3

Muñoz, C., Sanchez-Sevilla, J. F., Botella, M. A., Hoffmann, T., Schwab, W., and Valpuesta, V. (2011). Polyphenol composition in the ripe fruits of Fragaria species and transcriptional analyses of key genes in the pathway. *Journal of agricultural and food chemistry, 59*(23), 12598-12604. doi:10.1021/jf203965j

Nagano, S., Shirasawa, K., Hirakawa, H., Maeda, F., Ishikawa, M., and Isobe, S. N. (2017). Discrimination of candidate subgenome-specific loci by linkage map construction with an S1 population of octoploid strawberry (Fragaria x ananassa). *BMC Genomics, 18*(1), 374. doi:10.1186/s12864-017-3762-y

Nakamura, N., Hirakawa, H., Sato, S., Otagaki, S., Matsumoto, S., Tabata, S., and Tanaka, Y. (2018). Genome structure of Rosa multiflora, a wild ancestor of cultivated roses. *DNA research, 25*(2), 113-121. doi:10.1093/dnares/dsx042

Navarro-Retamal, C., Gaete-Eastman, C., Herrera, R., Caballero, J., and Alzate-Morales, J. H. (2016). Structural and Affinity Determinants in the Interaction between Alcohol

Acyltransferase from F. x ananassa and Several Alcohol Substrates: A Computational Study. *PLoS One, 11*(4), e0153057. doi:10.1371/journal.pone.0153057

Njuguna, W., Liston, A., Cronn, R., Ashman, T. L., and Bassil, N. (2013). Insights into phylogeny, sex function and age of Fragaria based on whole chloroplast genome sequencing. *Molecular Phylogenetics and Evolution, 66*(1), 17-29. doi:10.1016/j.ympev.2012.08.026

Njuguna, W., Hummer, K. E., Richards, C. M., Davis, T. M., and Bassil, N. V. (2011). Genetic diversity of diploid Japanese strawberry species based on microsatellite markers. *Genetic Resources and Crop Evolution, 58*(8), 1187-1198. doi:10.1007/s10722-010-9652-7

Nuzzi, M., Lo Scalzo, R., Testoni, A., and Rizzolo, A. (2008). Evaluation of Fruit Aroma Quality: Comparison Between Gas Chromatography–Olfactometry (GC–O) and Odour Activity Value (OAV) Aroma Patterns of Strawberries. *Food Analytical Methods, 1*(4), 270-282. doi:10.1007/s12161-008-9039-y

Olbricht, K., Grafe, C., Weiss, K., and Ulrich, D. (2008). Inheritance of aroma compounds in a model population of Fragaria x ananassa Duch. *Plant Breeding, 127*(1), 87-93. doi:DOI 10.1111/j.1439-0523.2007.01422.x

Paniagua, C., Santiago-Domenech, N., Kirby, A. R., Gunning, A. P., Morris, V. J., Quesada, M. A., . . . Mercado, J. A. (2017). Structural changes in cell wall pectins during strawberry fruit development. *Plant Physiology and Biochemistry, 118*, 55-63. doi:10.1016/j.plaphy.2017.06.001

Peace, C., Bassil, N., Main, D., Ficklin, S., Rosyara, U. R., Stegmeir, T., . . . Iezzoni, A. (2012). Development and evaluation of a genome-wide 6K SNP array for diploid sweet cherry and tetraploid sour cherry. *PLoS One, 7*(12), e48305. doi:10.1371/journal.pone.0048305

Perkins-Veazie, P. (1995). Growth and Ripening of Strawberry Fruit. *Horticultural Reviews: Volume, 17*, 267-298.

Perrotte, J., Gaston, A., Potier, A., Petit, A., Rothan, C., and Denoyes, B. (2016). Narrowing down the single homoeologous FaPFRU locus controlling flowering in cultivated octoploid strawberry using a selective mapping strategy. *Plant Biotechnology Journal, 14*(11), 2176-2189. doi:10.1111/pbi.12574

Pillet, J., Chambers, A. H., Barbey, C., Bao, Z., Plotto, A., Bai, J., . . . Folta, K. M. (2017). Identification of a methyltransferase catalyzing the final step of methyl anthranilate synthesis in cultivated strawberry. *BMC plant biology, 17*(1), 147. doi:10.1186/s12870-017-1088-1

Pose, S., Kirby, A. R., Paniagua, C., Waldron, K. W., Morris, V. J., Quesada, M. A., and Mercado, J. A. (2015). The nanostructural characterization of strawberry pectins in pectate lyase or polygalacturonase silenced fruits elucidates their role in softening. *Carbohydrate Polymers, 132*, 134-145. doi:10.1016/j.carbpol.2015.06.018

Potter, D., Eriksson, T., Evans, C., Oh, S., Smedmark, J. E. E., Morgan, D. R., . . . Campbell, C. S. (2007). Phylogeny and classification of Rosaceae. *Plant Systematics and Evolution, 266*, 5-43. doi:10.1007/s00606-007-0539-9

Potter, D., Luby, J. J., and Harrison, R. E. (2000). Phylogenetic Relationships Among Species of Fragaria (Rosaceae) Inferred from Non-coding Nuclear and Chloroplast DNA Sequences. *Systematic Botany, 25*(2), 337-348. doi:10.2307/2666646

Prat, L., Espinoza, M. I., Agosin, E., and Silva, H. (2014). Identification of volatile compounds associated with the aroma of white strawberries (Fragaria chiloensis). *Journal of the Science of Food and Agriculture, 94*(4), 752-759. doi:10.1002/jsfa.6412

Raab, T., Lopez-Raez, J. A., Klein, D., Caballero, J. L., Moyano, E., Schwab, W., and Munoz-Blanco, J. (2006). FaQR, required for the biosynthesis of the strawberry flavor compound 4-hydroxy-2,5-dimethyl-3(2H)-furanone, encodes an enone oxidoreductase. *Plant Cell, 18*(4), 1023-1037. doi:10.1105/tpc.105.039784

Rambla, J. L., Medina, A., Fernandez-Del-Carmen, A., Barrantes, W., Grandillo, S., Cammareri, M., . . . Granell, A. (2017). Identification, introgression, and validation of fruit volatile

QTLs from a red-fruited wild tomato species. *Journal of Experimental Botany, 68*(3), 429-442. doi:10.1093/jxb/erw455

Raymond, O., Gouzy, J., Just, J., Badouin, H., Verdenaud, M., Lemainque, A., . . . Bendahmane, M. (2018). The Rosa genome provides new insights into the domestication of modern roses. *Nature Genetics, 50*(6), 772-777. doi:10.1038/s41588-018-0110-3

RCoreTeam. (2019). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org/.

Roach, J. A., Verma, S., Peres, N. A., Jamieson, A. R., van de Weg, W. E., Bink, M. C., . . . Whitaker, V. M. (2016). FaRXf1: a locus conferring resistance to angular leaf spot caused by Xanthomonas fragariae in octoploid strawberry. *Theoretical and Applied Genetics, 129*(6), 1191-1201. doi:10.1007/s00122-016-2695-1

Rousseau-Gueutin, M., Gaston, A., Aïnouche, A., Aïnouche, M. L., Olbricht, K., Staudt, G., . . . Denoyes-Rothan, B. (2009). Tracking the evolutionary history of polyploidy in Fragaria L. (strawberry): New insights from phylogenetic analyses of low-copy nuclear genes. *Molecular Phylogenetics and Evolution, 51*(3), 515-530. doi:10.1016/j.ympev.2008.12.024

RStudio. (2019). RStudio: Integrated development environment for R (Version Version 0.98.1049). Boston, MA, USA: RStudio. Retrieved from Available from http://www.rstudio.org/

Salinas, N., Verma, S., Peres, N., and Whitaker, V. M. (2019). FaRCa1: a major subgenome-specific locus conferring resistance to Colletotrichum acutatum in strawberry. *Theoretical and Applied Genetics, 132*(4), 1109-1120. doi:10.1007/s00122-018-3263-7

Salvatierra, A., Pimentel, P., Moya-Leon, M. A., and Herrera, R. (2013). Increased accumulation of anthocyanins in Fragaria chiloensis fruits by transient suppression of FcMYB1 gene. *Phytochemistry, 90*, 25-36. doi:10.1016/j.phytochem.2013.02.016

Samykanno, K., Pang, E., and Marriott, P. J. (2013). Genotypic and environmental effects on flavor attributes of 'Albion' and 'Juliette' strawberry fruits. *Scientia Horticulturae, 164*, 633-642. doi:10.1016/j.scienta.2013.09.001

Sanchez-Sevilla, J. F., Cruz-Rus, E., Valpuesta, V., Botella, M. A., and Amaya, I. (2014). Deciphering gamma-decalactone biosynthesis in strawberry fruit using a combination of genetic mapping, RNA-Seq and eQTL analyses. *BMC Genomics, 15*(218). doi:10.1186/1471-2164-15-218

Sánchez, G., Martínez, J., Romeu, J., García, J., Monforte, A. J., Badenes, M. L., and Granell, A. (2014). The peach volatilome modularity is reflected at the genetic and environmental response levels in a QTL mapping population. *BMC plant biology, 14*, 137. doi:10.1186/1471-2229-14-137

Sargent, D. J., Buti, M., Surbanovski, N., Brurberg, M. B., Alsheikh, M., Kent, M. P., and Davik, J. (2019). Identification of QTLs for powdery mildew (Podosphaera aphanis; syn. Sphaerotheca macularis f. sp. fragariae) susceptibility in cultivated strawberry (Fragaria xananassa). *PLoS One, 14*(9), e0222829. doi:10.1371/journal.pone.0222829

Sargent, D. J., Yang, Y., Surbanovski, N., Bianco, L., Buti, M., Velasco, R., . . . Davis, T. M. (2016). HaploSNP affinities and linkage map positions illuminate subgenome composition in the octoploid, cultivated strawberry (Fragariaxananassa). *Plant Science, 242*, 140-150. doi:10.1016/j.plantsci.2015.07.004

Sargent, D. J., Passey, T., Surbanovski, N., Lopez Girona, E., Kuchta, P., Davik, J., . . . Simpson, D. W. (2012). A microsatellite linkage map for the cultivated strawberry (Fragaria x ananassa) suggests extensive regions of homozygosity in the genome that may have resulted from breeding and selection. *Theoretical and Applied Genetics, 124*, 1229-1240. doi:10.1007/s00122-011-1782-6

Sargent, D. J., Kuchta, P., Lopez-Girona, E., Zhang, H., Davis, T. M., Celton, J., . . . Simpson, D. W. (2011). Simple sequence sepeat marker development and mapping targeted to

previously unmapped regions of the strawberry genome sequence. *Plant Genome, 4*(3), 165-177. doi:10.3835/plantgenome2011.05.0014

Sargent, D. J., Clarke, J., Simpson, D. W., Tobutt, K. R., Arús, P., Monfort, A., . . . Battey, N. H. (2006). An enhanced microsatellite map of diploid *Fragaria*. *Theoretical and Applied Genetics, 112*(7), 1349-1359.

Sargent, D. J., Davis, T. M., Tobutt, K. R., Wilkinson, M. J., Battey, N. H., and Simpson, D. W. (2004). A genetic linkage map of microsatellite, gene-specific and morphological markers in diploid Fragaria. *Theoretical and Applied Genetics, 109*(7), 1385-1391. doi:10.1007/s00122-004-1767-9

Schaart, J. G., Dubos, C., Romero De La Fuente, I., van Houwelingen, A. M. M. L., de Vos, R. C. H., Jonker, H. H., . . . Bovy, A. G. (2013). Identification and characterization of MYB-bHLH-WD40 regulatory complexes controlling proanthocyanidin biosynthesis in strawberry (Fragaria × ananassa) fruits. *New Phytologist, 197*(2), 454-467. doi:10.1111/nph.12017

Schieberle, P., and Hofmann, T. (1997). Evaluation of the character impact odorants in fresh strawberry juice by quantitative measurements and sensory studies on model mixtures. *Journal of agricultural and food chemistry, 45*(1), 227-232. doi:10.1021/Jf960366o

Schwab, W., Davidovich-Rikanati, R., and Lewinsohn, E. (2008). Biosynthesis of plant-derived flavor compounds. *The Plant Journal, 54*(4), 712-732. doi:10.1111/j.1365-313X.2008.03446.x

Schwieterman, M. L., Colquhoun, T. A., Jaworski, E. A., Bartoshuk, L. M., Gilbert, J. L., Tieman, D. M., . . . Clark, D. G. (2014). Strawberry Flavor: Diverse Chemical Compositions, a Seasonal Influence, and Effects on Sensory Perception. *PLoS One, 9*(2), e88446. doi:10.1371/journal.pone.0088446

Sharma, S. K., Bolser, D., de Boer, J., Sonderkaer, M., Amoros, W., Carboni, M. F., . . . Bryan, G. J. (2013). Construction of reference chromosome-scale pseudomolecules for potato: integrating the potato genome with genetic and physical maps. *GENES, GENOMES, GENETICS, 3*(11), 2031-2047. doi:10.1534/g3.113.007153

Shulaev, V., Sargent, D. J., Crowhurst, R. N., Mockler, T. C., Folkerts, O., Delcher, A. L., . . . Folta, K. M. (2011). The genome of woodland strawberry (Fragaria vesca). *Nature Genetics, 43*(2), 109-116. doi:10.1038/ng.740

Souleyre, E. J., Chagne, D., Chen, X., Tomes, S., Turner, R. M., Wang, M. Y., . . . Atkinson, R. G. (2014). The AAT1 locus is critical for the biosynthesis of esters contributing to 'ripe apple' flavour in 'Royal Gala' and 'Granny Smith' apples. *Plant Journal, 78*(6), 903-915. doi:10.1111/tpj.12518

Souza, G. M., Van Sluys, M. A., Lembke, C. G., Lee, H., Margarido, G. R. A., Hotta, C. T., . . . Heckerman, D. (2019). Assembly of the 373k gene space of the polyploid sugarcane genome reveals reservoirs of functional diversity in the world's leading biomass crop. *Gigascience, 8*(12). doi:10.1093/gigascience/giz129

Spigler, R. B., Lewers, K. S., and Ashman, T. L. (2011). Genetic architecture of sexual dimorphism in a subdioecious plant with a proto-sex chromosome. *Evolution, 65*(4), 1114-1126. doi:10.1111/j.1558-5646.2010.01189.x

Spigler, R. B., Lewers, K. S., Main, D. S., and Ashman, T. L. (2008). Genetic mapping of sex determination in a wild strawberry, Fragaria virginiana, reveals earliest form of sex chromosome. *Heredity, 101*(6), 507-517. doi:10.1038/hdy.2008.100

Spiller, M., Berger, R. G., and Debener, T. (2010). Genetic dissection of scent metabolic profiles in diploid rose populations. *Theoretical and Applied Genetics, 120*, 1461–1471. doi:10.1007/s00122-010-1268-y

Spyropoulou, E. A., Dekker, H. L., Steemers, L., van Maarseveen, J. H., de Koster, C. G., Haring, M. A., . . . Allmann, S. (2017). Identification and characterization of (3Z):(2E)-hexenal isomerases from cucumber. *Frontiers in Plant Science, 8*, 1342. doi:10.3389/fpls.2017.01342

Staudt, G. (1962). Taxanomic studies in the genus Fragaria. Typification of Fragaria species known at the time of Linnaeus. *Canadian Journal d Botany, 48*, 869-886.

Strecker, J., Rodríguez, G., Njanji, I., Thomas, J., Jack, A., Darrigues, A., . . . van der Knaap, E. (2010). Tomato analyzer color test user manual version 3.

Sturtervant, A. H. (1913). The linear arrangement of six-linked factor in Drosophila, as shown by their mode of association. *Journal of Experimental Zoology, 14*, 43-59.

Sun, R., Chang, Y., Yang, F., Wang, Y., Li, H., Zhao, Y., . . . Han, Z. (2015). A dense SNP genetic map constructed using restriction site-associated DNA sequencing enables detection of QTLs controlling apple fruit quality. *BMC Genomics, 16*, 747. doi:10.1186/s12864-015-1946-x

Symons, G. M., Chua, Y.-J., Ross, J. J., Quittenden, L. J., Davies, N. W., and Reid, J. B. (2012). Hormonal changes during non-climacteric ripening in strawberry. *Journal of Experimental Botany, 63*(13), 4741-4750. doi:10.1093/jxb/ers147

Tennessen, J. A., Govindarajulu, R., Ashman, T. L., and Liston, A. (2014). Evolutionary origins and dynamics of octoploid strawberry subgenomes revealed by dense targeted capture linkage maps. *Genome biology and evolution, 6*(12), 3295-3313. doi:10.1093/gbe/evu261

Thill, J., Miosic, S., Gotame, T. P., Mikulic-Petkovsek, M., Gosch, C., Veberic, R., . . . Halbwirth, H. (2013). Differential expression of flavonoid 3'-hydroxylase during fruit development establishes the different B-ring hydroxylation patterns of flavonoids in Fragaria x ananassa and Fragaria vesca. *Plant Physiology and Biochemistry, 72*, 72-78. doi:10.1016/j.plaphy.2013.03.019

Tieman, D., Zhu, G., Resende Jr, M. F. R., Lin, T., Nguyen, C., Bies, D., . . . Klee, H. (2017). A chemical genetic roadmap to improved tomato flavor. *Science, 355*, 391-394. doi:10.1126/science.aal1556

Tieman, D., Bliss, P., McIntyre, L. M., Blandon-Ubeda, A., Bies, D., Odabasi, A. Z., . . . Klee, H. J. (2012). The chemical interactions underlying tomato flavor preferences. *Current Biology, 22*(11), 1035-1039. doi:10.1016/j.cub.2012.04.016

Ulrich, D., Kecke, S., and Olbricht, K. (2018). What Do We Know about the Chemistry of Strawberry Aroma? *Journal of agricultural and food chemistry, 66*(13), 3291-3301. doi:10.1021/acs.jafc.8b01115

Ulrich, D., and Olbricht, K. (2016). A search for the ideal flavor of strawberry − Comparison of consumer acceptance and metabolite patterns in Fragaria × ananassa Duch. *Journal of applied botany and food quality, 89*, 223-234. doi:10.5073/JABFQ.2016.089.029

Ulrich, D., Komes, D., Olbricht, K., and Hoberg, E. (2007). Diversity of aroma patterns in wild and cultivated Fragaria accessions. *Genetic Resources and Crop Evolution, 54*(6), 1185-1196. doi:10.1007/s10722-006-9009-4

Ulrich, D., Hoberg, E., Rapp, A., and Kecke, S. (1997). Analysis of strawberry flavour - discrimination of aroma types by quantification of volatile compounds. *Zeitschrift Fur Lebensmittel-Untersuchung Und-Forschung a-Food Research and Technology, 205*(3), 218-223. doi:10.1007/s002170050154

Urrutia, M., Rambla, J. L., Alexiou, K. G., Granell, A., and Monfort, A. (2017). Genetic analysis of the wild strawberry (Fragaria vesca) volatile composition. *Plant Physiol Biochem, 121*, 99-117. doi:10.1016/j.plaphy.2017.10.015

Urrutia, M., Schwab, W., Hoffmann, T., and Monfort, A. (2016). Genetic dissection of the (poly)phenol profile of diploid strawberry (Fragaria vesca) fruits using a NIL collection. *Plant Science, 242*, 151-168. doi:10.1016/j.plantsci.2015.07.019

Urrutia, M., Bonet, J., Arús, P., and Monfort, A. (2015). A near-isogenic line (NIL) collection in diploid strawberry and its use in the genetic analysis of morphologic, phenotypic and nutritional characters. *Theoretical and Applied Genetics, 128*, 1261-1275. doi:10.1007/s00122-015-2503-3

Vallarino, J. G., Pott, D. M., Cruz-Rus, E., Miranda, L., Medina-Minguez, J. J., Valpuesta, V., . . . Amaya, I. (2019). Identification of quantitative trait loci and candidate genes for primary

metabolite content in strawberry fruit. *Horticulture research, 6*, 4. doi:10.1038/s41438-018-0077-3

van Dijk, T., Pagliarani, G., Pikunova, A., Noordijk, Y., Yilmaz-Temel, H., Meulenbroek, B., . . . van de Weg, E. (2014). Genomic rearrangements and signatures of breeding in the allo-octoploid strawberry as revealed through an allele dose based SSR linkage map. *BMC plant biology, 14*, 55. doi:10.1186/1471-2229-14-55

van Ooijen, J. W. (2018). JoinMap ® 5, Software for the calculation of genetic linkage maps in experimental population of diploid species. Kyazma B.V., Wageningen, Netherlands

van Ooijen, J. (2009). MapQTL 6. *Software for the mapping of quantitative trait loci in experimental populations of diploid species. Kyazma BV: Wageningen, Netherlands*

VanBuren, R., Wai, C. M., Colle, M., Wang, J., Sullivan, S., Bushakra, J. M., . . . Bassil, N. V. (2018). A near complete, chromosome-scale assembly of the black raspberry (Rubus occidentalis) genome. *Gigascience, 7*(8). doi:10.1093/gigascience/giy094

VanBuren, R., Bryant, D., Bushakra, J. M., Vining, K. J., Edger, P. P., Rowley, E. R., . . . Mockler, T. C. (2016). The genome of black raspberry (Rubus occidentalis). *Plant Journal, 87*(6), 535-547. doi:10.1111/tpj.13215

Vandendriessche, T., Vermeir, S., Mayayo Martinez, C., Hendrickx, Y., Lammertyn, J., Nicolaï, B. M., and Hertog, M. L. A. T. M. (2013). Effect of ripening and inter-cultivar differences on strawberry quality. *LWT - Food Science and Technology, 52*(2), 62-70. doi:10.1016/j.lwt.2011.12.037

Verde, I., Bassil, N., Scalabrin, S., Gilmore, B., Lawley, C. T., Gasic, K., . . . Peace, C. (2012). Development and evaluation of a 9K SNP array for peach by internationally coordinated SNP detection and validation in breeding germplasm. *PLoS One, 7*(4), e35668. doi:10.1371/journal.pone.0035668

Verma, S., Bassil, N. V., van de Weg, E., Harrison, R. J., Monfort, A., Hidalgo, J. M., . . . Whitaker, V. M. (2017a). Development and evaluation of the Axiom® IStraw35 384HT array for the allo-octoploid cultivated strawberry Fragaria × ananassa. *Acta Horticulturae*(1156), 75-82. doi:10.17660/ActaHortic.2017.1156.10

Verma, S., Zurn, J. D., Salinas, N., Mathey, M. M., Denoyes, B., Hancock, J. F., . . . Whitaker, V. M. (2017b). Clarifying sub-genomic positions of QTLs for flowering habit and fruit quality in U.S. strawberry (Fragariaxananassa) breeding populations using pedigree-based QTL analysis. *Horticulture research, 4*, 17062. doi:10.1038/hortres.2017.62

Vogt, J., Schiller, D., Ulrich, D., Schwab, W., and Dunemann, F. (2013). Identification of lipoxygenase (LOX) genes putatively involved in fruit flavour formation in apple (Malus × domestica). *Tree genetics & genomes, 9*, 1493–1511. doi:10.1007/s11295-013-0653-5

Voorrips, R. E. (2002). MapChart: software for the graphical presentation of linkage maps and QTLs. *Journal of Heredity, 93*(1), 77-78.

Vos, P. G., Uitdewilligen, J. G., Voorrips, R. E., Visser, R. G., and van Eck, H. J. (2015). Development and analysis of a 20K SNP array for potato (Solanum tuberosum): an insight into the breeding history. *Theoretical and Applied Genetics, 128*(12), 2387-2401. doi:10.1007/s00122-015-2593-y

Wang, D. G., Fan, J., Siao, C., Berno, A., Young, P., Sapolsky, R., . . . Lander, E. S. (1998). Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. *Science, 280*, 1077-1082.

Wang, H., Zhang, H., Yang, Y., Li, M., Zhang, Y., Liu, J., . . . Jin, W. (2019). The control of red color by a family of MYB transcription factors in octoploid strawberry (Fragaria x ananassa) fruits. *Plant Biotechnology Journal*. doi:10.1111/pbi.13282

Wang, S., Chen, J., Zhang, W., Hu, Y., Chang, L., Fang, L., . . . Zhang, T. (2015). Sequence-based ultra-dense genetic and physical maps reveal structural variations of allopolyploid cotton genomes. *Genome biology, 16*, 108. doi:10.1186/s13059-015-0678-1

Wang, S., Wong, D., Forrest, K., Allen, A., Chao, S., Huang, B. E., . . . Akhunov, E. (2014). Characterization of polyploid wheat genomic diversity using a high-density 90,000 single

nucleotide polymorphism array. *Plant Biotechnol Journal, 12*(6), 787-796. doi:10.1111/pbi.12183

Weber, J. L., and May, P. E. (1989). Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. *American Journal of Human Genetics, 44*, 388-396.

Wein, M., Lavid, N., Stefan Lunkenbein, S., Lewinsohn, E., Schwab, W., and Kaldenhoff, R. (2002). Isolation, cloning and expression of a multifunctional O-methyltransferase capable of forming 2,5-dimethyl-4- methoxy-3(2H)-furanone, one of the key aroma compounds in strawberry fruits. *The Plant Journal, 31*(6), 755-765.

Winfield, M. O., Allen, A. M., Burridge, A. J., Barker, G. L., Benbow, H. R., Wilkinson, P. A., . . . Edwards, K. J. (2016). High-density SNP genotyping array for hexaploid wheat and its secondary and tertiary gene pool. *Plant Biotechnol Journal, 14*(5), 1195-1206. doi:10.1111/pbi.12485

Wu, S., Zhang, B., Keyhaninejad, N., Rodriguez, G. R., Kim, H. J., Chakrabarti, M., . . . van der Knaap, E. (2018). A common genetic mechanism underlies morphological diversity in fruits and other plant organs. *Nature Communications, 9*(1), 4734. doi:10.1038/s41467-018-07216-8

Xiang, Y., Huang, C. H., Hu, Y., Wen, J., Li, S., Yi, T., . . . Ma, H. (2017). Evolution of Rosaceae Fruit Types Based on Nuclear Phylogeny in the Context of Geological Times and Genome Duplication. *Mol Biol Evol, 34*(2), 262-281. doi:10.1093/molbev/msw242

Xu, W., Dubos, C., and Lepiniec, L. (2015). Transcriptional control of flavonoid biosynthesis by MYB-bHLH-WDR complexes. *Trends in plant science, 20*(3), 176-185. doi:10.1016/j.tplants.2014.12.001

Xu, X., Pan, S., Cheng, S., Zhang, B., Mu, D., Ni, P., . . . Visser, R. G. (2011). Genome sequence and analysis of the tuber crop potato. *Nature, 475*(7355), 189-195. doi:10.1038/nature10158

Yang, Y., and Davis, T. M. (2017). A New Perspective on Polyploid Fragaria (Strawberry) Genome Composition Based on Large-Scale, Multi-Locus Phylogenetic Analysis. *Genome biology and evolution, 9*(12), 3433-3448. doi:10.1093/gbe/evx214

Yauk, Y. K., Souleyre, E. J. F., Matich, A. J., Chen, X., Wang, M. Y., Plunkett, B., . . . Atkinson, R. G. (2017). Alcohol acyl transferase 1 links two distinct volatile pathways that produce esters and phenylpropenes in apple fruit. *Plant Journal, 91*(2), 292-305. doi:10.1111/tpj.13564

You, Q., Yang, X., Peng, Z., Islam, M. S., Sood, S., Luo, Z., . . . Wang, J. (2019). Development of an Axiom Sugarcane100K SNP array for genetic map construction and QTL identification. *Theoretical and Applied Genetics, 132*(10), 2829-2845. doi:10.1007/s00122-019-03391-4

Yuan, D., Tang, Z., Wang, M., Gao, W., Tu, L., Jin, X., . . . Zhang, X. (2015). The genome sequence of Sea-Island cotton (Gossypium barbadense) provides insights into the allopolyploidization and development of superior spinnable fibres. *Scientific Reports, 5*, 17662. doi:10.1038/srep17662

Zhang, B., Shen, J. Y., Wei, W. W., Xi, W. P., Xu, C. J., Ferguson, I., and Chen, K. (2010). Expression of genes associated with aroma formation derived from the fatty acid pathway during peach fruit ripening. *Journal of agricultural and food chemistry, 58*(10), 6157-6165. doi:10.1021/jf100172e

Zhang, J., Lei, Y., Wang, B., Li, S., Yu, S., Wang, Y., . . . Zhang, Z. (2020). The high-quality genome of diploid strawberry (Fragaria nilgerrensis) provides new insights into anthocyanin accumulation. *Plant Biotechnology Journal*. doi:10.1111/pbi.13351

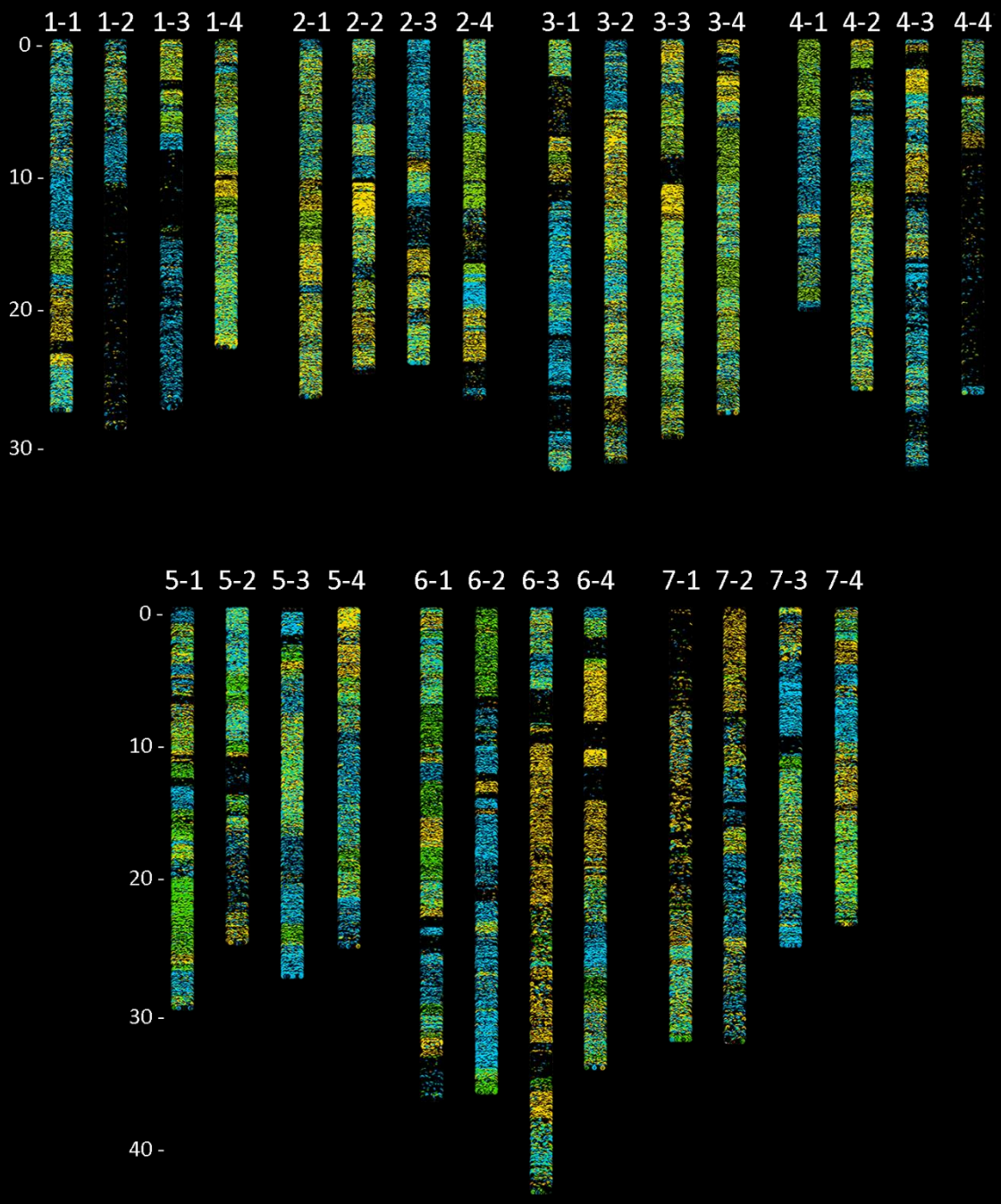
Zhang, R., Taris, K. E., Urrutia, M., Mnejja, M., and Monfort, A. (2017a). *Charactering agronomical traïts in a diploid strawberry NIL collection*. Paper presented at the International Strawberry Congress.

Zhang, S. D., Jin, J. J., Chen, S. Y., Chase, M. W., Soltis, D. E., Li, H. T., . . . Yi, T. S. (2017b). Diversification of Rosaceae since the Late Cretaceous based on plastid phylogenomics. *New Phytologist, 214*(3), 1355-1367. doi:10.1111/nph.14461

Zhao, J., Sauvage, C., Zhao, J., Bitton, F., Bauchet, G., Liu, D., . . . Causse, M. (2019). Meta-analysis of genome-wide association studies provides insights into genetic control of tomato flavor. *Nature Communications, 10*(1), 1534. doi:10.1038/s41467-019-09462-w

Zorrilla-Fontanesi, Y., Rambla, J. L., Cabeza, A., Medina, J. J., Sánchez-Sevilla, J. F., Valpuesta, V., . . . 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*(2), 851-870. doi:10.1104/pp.111.188318

Zorrilla-Fontanesi, Y., Cabeza, A., Domínguez, P., Medina, J. J., Valpuesta, V., Denoyes-Rothan, B., . . . Amaya, I. (2011). Quantitative trait loci and underlying candidate genes controlling agronomical and fruit quality traits in octoploid strawberry (Fragaria x ananassa). *Theoretical and Applied Genetics, 123*, 755-778. doi:10.1007/s00122-011-1624-6


1-1
1-2
1-3
1-4
2-1
2-2
2-3
2-4
3-1
3-2
3-3
3-4
4-1
4-2
4-3
4-4
0
10
20
30
5-1
5-2
5-3
5-4
6-1
6-2
6-3
6-4
7-1
7-2
7-3
7-4
0
10
20
30
40