Universitat Autònoma de Barcelona

Genetic analysis of fruit flavor and aroma volatile compounds in wild strawberry

Rong Zhang

UNIVERSITAT AUTÒNOMA DE BARCELONA

Facultat de Biociències

Dept. Biologia Animal, Biologia Vegetal i Ecologia

Genetic analysis of fruit flavor and aroma volatile compounds in wild strawberry

Dissertation presented by Rong Zhang for the degree of Doctor in Plant Biology and Biotechnology by the Universitat Autònoma de Barcelona (UAB).

This work was performed in the Centre for Research in Agricultural Genomics (CRAG).

Thesis directors	Tutor	PhD candidate
Dr. Amparo Monfort Vives	Dr. Amparo Monfort Vives	Rong Zhang

Barcelona, December 2019

Index

Index of Contents

Index of Figure

Index of Table

Supplementary Material

Summary

Summary

Strawberry, belonging to *Fragaria* genus, *Rosaceae* family, is the most commonly consumed berry fruit crop worldwide, with a production of around 9.2 million tons during 2017. Although traditionally breeding programs have been focused on improving agronomic traits, fruit quality has become a main goal recently. Fruit flavor is the main factor responsible for the fruit quality that is a direct factor attracting customers. The diploid strawberry (*Fragaria vesca*) serves as an important model plant for cultivated strawberry and *Rosaceae* family, due to its perennial life cycle, small genome, short generation time, a simple and efficient genetic transformation system and abundant genetic resources. The main goal of this work was to study the genetic basis of fruit flavor in diploid strawberry using a isogenic lines collection (NIL) developed by a cross between the recurrent parent *F. vesca* and the donor parent *F. bucharica*.

Firstly, we improved the previously developed NIL collection with 10 new lines added. Finally, this population consists of 49 lines with overlapping introgressions covering 94.8% of background of donor parental line. It is a highly relevant genetic tool for mapping a variety of traits for diploid strawberry.

Fruit flavor traits, including pH, citric acid and ^{o}Brix of ripe fruits of the collection were statistically analyzed for mapping as quantitative traits loci (QTL).One stable QTL was mapped for increasing pH and two major QTL for decreasing pH. Two major QTL were mapped for decreasing citric acid and one QTL for increasing ^{o}Brix.

Due to locating many major QTL for aroma volatile compounds in LG5, we deeply studied the aroma compounds in the new lines harboring introgressions in LG5. Seventeen key volatile compounds were identified and five QTL were mapped. The QTL for methyl 2-aminobenzoate and myrtenyl acetate were located in the same region LG5:20-35cM. The QTL for decreasing methyl butanoate content was located in LG5:11-20cM. Two QTL for green volatile compounds (Z)-3-hexenyl acetate and (E)-2-hexenyl acetate were located in LG5:50-76cM.

In order to investigate the major genes controlling aroma volatile compounds accumulation, eleven genes were selected as important candidate genes to analyze the transcription level in fully ripe strawberry fruits of *F. vesca* and NILs harboring *F. bucharica* alleles. Finally, three genes in LG5:0-35cM and three genes in LG5:50-76cM were verified to present significant differences in expression between the recurrent parent *F. vesca* and the NILs harboring *F. bucharica* alleles in these LG5 regions. The gene FvH4_5g29270 encoding a 3Z-2E-enal isomerase was selected as candidate gene for green leaf volatile compoundsin LG5:50-76cM. For verifying the function of gene FvH4_5g29270, the gene overexpression vector was constructed and transiently expressed in *F. vesca* and two NILs with the *F. bucharica* allele of

Summary

FvH4_5g29270 via agrobacterium mediated transfection. However, thegreen volatile compounds content results were unexpected, mainly due to the low transformation efficiency.

Attending to the important fruit appearance traits, the fruit color was analyzed with eight color parameters for three harvests. Orange fruit color was observed in some NILs and was mapped in LG5:35-39cM. Eighteen candidate genes in this region were analyzed for transcript expression level in fully ripen fruits from the recurrent parent *F. vesca* and NILs harboring the introgression containing LG5:35-39. The gene FvH4_5g14770 encoding light-harvesting complex chlorophyll A-B binding protein was selected as a good candidate gene since it showed significantly higher transcript level in orange colored fruits than in red fruits that can be related to chlorophyll content reduction. All these results provided fundamental basis for strawberry aroma and color breeding.

Resumen

Resumen

La fresa, perteneciente al género *Fragaria*, familia *Rosaceae*, es el cultivo de bayas más consumido en todo el mundo, con una producción de 9,2 millones de toneladas en el año 2017. Aunque tradicionalmente los programas de mejora se han centrado en los rasgos agronómicos, la calidad de la fruta se ha convertido en un objetivo principal en los últimos años. El sabor es el principal factor responsable de la calidad en la fruta, actuando directamente sobreelconsumidor. La fresa diploide (*Fragaria vesca*) es una planta modelo para la fresa cultivada y la familia *Rosaceae*, debido a su ciclo de vida perenne, pequeño genoma, corto periodo generacional, un sistema de transformación genética simple y eficiente y abundantes recursos genéticos. El objetivo principal de este trabajo es estudiar la base genética del sabor de la fruta en la fresa diploide utilizando una colección de líneas isogénicas (NIL) desarrollada por un cruce entre unparental recurrente *F. vesca* y un parental donante *F. bucharica*.

En primer lugar, incrementamos la colección NIL desarrollada previamente con 10 nuevas líneas. Actualmente, esta población consta de 49 líneas con introgresiones superpuestas que cubren el 94.8% del genoma de la línea parental donante. Es una herramienta genética relevante para mapar diferentes caracteres de la fresa diploide.

Para posicionarlos loci de los caracteres cuantitativos (QTL), se analizaron estadísticamente los parámetros indicativos de sabor en la fruta, incluyendo pH, ácido cítrico y ºBrix en fresas maduras de la colección. Se mapeó un QTL estable asociado al incremento del pH y dos QTL mayores asociados con su disminución. Se mapearon también, dos QTL mayores asociados con la reducción en ácido cítrico y un QTL relacionado con el incremento de los ºBrix.

Debido a la localización en LG5 de QTL asociados a compuestos volátiles, estudiamos la acumulación de dichos compuestos en las nuevas líneas que albergan introgresiones en LG5. Se identificaron diecisiete compuestos volátiles clave y se mapearon cinco QTL. Los QTL asociados al 2-aminobenzoato de metilo y al acetato de myrtenyl se localizan en la misma región LG5:20-35cM. El QTL asociado a la disminución de metil butanoato se localizó en LG5:11-20cM. Dos QTL ligados a la variación de compuestos volátiles verdes (Z)-3-acetato de hexenilo y (E)-2-acetato de hexenilo se localizan en LG5:50-76cM.

Para identificar los genes que controlan la acumulación de compuestos volátiles aromáticos en la región, se seleccionaron once genes candidatos y se analizó su nivel de transcripción en frutas completamente maduras de *F. vesca* y de las líneas NIL que albergan alelos de *F. bucharica* en la región LG5. Finalmente, se identificaron tres genes en LG5:0-35cM y tres genes en LG5:50-76cM que presentan diferencias significativas en la expresión entre el padre recurrente *F. vesca* y las NIL que albergan alelos de *F. bucharica*. El gen FvH4_5g29270 que codifica una isomerasa 3Z-2E-enal se seleccionó como gen

candidato para compuestos volátiles de hoja verde en LG5:50-76cM. Para verificar la función del gen FvH4_5g29270, se construyó el vector de sobreexpresión génica y se expresó de forma transitoria en *F. vesca* y en las NILs que contienen el alelo de FvH4_5g29270 en *F. bucharica* mediante transfección mediada por agrobacterium. Sin embargo, los resultados del contenido de compuestos volátiles verdes no fueron los esperados, principalmente debido a la baja eficiencia de transformación.

Respecto a los importantes rasgos de apariencia externa de la fruta, el color de la fresa se analizó aplicando ocho parámetros de color en tres cosechas distintas en la colección. Se observó el color del fruto anaranjado en algunas NILs y se mapeó en la región LG5:35-39cM. Se analizaron dieciocho genes candidatos en esta región para determinar el nivel de expresión de la transcripción en frutos completamente maduros en plantas de *F. vesca* y en las NILs que albergan la introgresión LG5:35-39. El gen FvH4_5g14770 que codifica para una proteína de unión al complejo clorofila A-B dependiente de luz se seleccionó como un buen gen candidato ya que mostró un nivel de transcripción significativamente mayor en las fresas de color anaranjado que en las fresas rojas, ydicho cambio puede estar relacionado con la reducción del contenido de clorofila. Todos estos resultados proporcionaron una base fundamental para la mejora del aroma y la variación del color de la fresa.

Resum

Resum

La maduixa, que pertany al gènere *Fragaria*, de la família *Rosaceae*, és el cultiu de fruita de baia més consumit a tot el món, amb una producció durant l'any 2017 d'uns 9,2 milions de tones. Tot i que els programes de millora tradicionalment s'han centrat en els trets agronòmics, la qualitat de la fruita s'ha convertit recentment en un objectiu principal. El sabor és el principal factor responsable de la qualitat de la fruita, que és un atractiu per el consumidor. La maduixa diploide (*Fragaria vesca*) es un model per a la maduixa conreada i les altres rosàcies, gràcies al seu cicle de vida perenne, el genoma reduït, el curt període generacional, al sistema de transformació genètica senzill i eficient, i als abundants recursos genètics. L'objectiu principal d'aquest treball ha estat estudiar la base genètica del sabor de la fruita en maduixa diploide mitjançant una col·lecció de línies isogéniques (NIL) desenvolupada per un encreuament entre un pare recurrent *F. vesca* i el progenitor donant *F. bucharica*.

En primer lloc, hem millorat la col·lecció NIL desenvolupada anteriorment amb deu noves línies. Actualment, aquesta població consta de 49 línies amb introgressions superposades que cobreixen el 94,8% de la línia parental donant. És una eina genètica altament rellevant per mapar diversos trets de maduixa diploides.

Per localitzar els loci delscaràcters quantitatius (QTL) es van analitzar estadísticament els trets indicatius del sabor a les maduixes madures de la col·lecció, inclosos el pH, l'àcid cítric i ºBrix. Es va mapar un QTL estable associat a l'augment del pH i dos QTL majors per la reducció del pH. Es van localitza dos QTL majors per la disminució dels nivells de l'àcid cítric i un QTL per l'augment dels ºBrix.

Degut a la localització de varius QTL majors per a compostos volàtils aromàtics trobats al LG5, hem estudiat en profunditat els compostos aromàtics a les línies que contenen noves introgressions al LG5. Es van identificar disset compostos volàtils clau i es van assignar cinc QTL. Els QTL de 2-aminobenzoat de metil i acetat de myrtenil es van localitzar a la mateixa regió LG5:20-35cM. El QTL per disminuir el contingut de butanoat de metil es va localitzar a LG5:11-20cM. Dos QTL per a compostos volàtils verds (Z) -3-hexenil acetat i acetat (E) -2-hexenil es van localitzar a LG5:50-76cM.

Per investigar els principals gens que controlen l'acumulació de compostos volàtils d'aroma, onze gens van ser seleccionats com a gen candidat important per analitzar el nivell de transcripció en maduixes de *F. vesca* i les NILs. Finalment, es van verificar tres gens en LG5:0-35cM i tres gens en LG5:50-76cM per presentar diferències significatives en l'expressió entre el pare recurrent i les NILs que albergaven al·lels de *F. bucharica* en aquestes regions LG5. El gen FvH4_5g29270 que codifica una isomerasa 3Z-2E-enal es va seleccionar com a gen candidat per a compostos volàtils de fulla verda en LG5:50-76cM. Per verificar la funció del gen FvH4_5g29270, es va construir el vector de sobreexpressió gènica i es va expressar de forma transitòria en *F. vesca* i en dos NILs amb l'al·lel *F. bucharica* de FvH4_5g29270

mitjançant transfecció mediada per agrobacterium. Tanmateix, els resultats de contingut de compostos volàtils verds no van ser els esperats, principalment a causa de la baixa eficiència de la transformació.

Observant un altre caràcter de qualitat important com es l'aparença del fruit, es va analitzar el color de la maduixa amb vuit paràmetres de color per a tres collites en la col·lecció. El color de la fruita taronja es va observar en algunes NILs i es va mapar en LG5:35-39cM. Es van analitzar divuit gens candidats en aquesta regió per obtenir el nivell d'expressió de la transcripció en fruites completament madures. El gen FvH4_5g14770 que codifica per una proteïna d'unió al complex de clorofil·la A-B dependent de la llum es va seleccionar com a gen candidat, ja que mostrava un nivell de transcripció significativament superior en fruites de color taronja que en fruites vermelles, i aquest canvi pot estar relacionat amb el procés de reducció del contingut de clorofil·la. Tots aquests resultats van proporcionar una base fonamental per a la millora de l'aroma i el color de la maduixa.

General introduction

1. *Fragaria* genus

1.1. Species and Distribution

Fragaria genus is a member of Rosaceae family, which contains around 100 genera with more than 3000 plant species including ornamental species such as rose, fruit species such as apple, pear, peach, cherry, almond and many berries (Buti *et al*. 2016). *Fragaria* genus has seven basic chromosomes (x = 7) with differ in ploidy level from diploid to decaploid. Thirteen diploids (2n = 2× = 14) includes *F. bucharica*, *F. vesca*, *F. chinensis*, etc. Five tetraploids (2n = 4× = 28) are *F. orientalis*, *F. corymbosa*, *F. gracilis*, *F. moupinensis* and *F. tibetica*. One hexaploid (2n = 6× = 42) is *F. moschata*. Four octoploids (2n = 8× = 56) are *F.* x *ananassa*, *F. chiloensis* (L.), *F. iturupensis* and *F. virginiana*. The only decaploid species (2n = 10× = 70) is *F. cascadensis* (Hummer *et al.* 2009; Liston *et al.* 2014). The genus distributed in the north temperate and holarctic zones (Table I-1), including Eurasia, North America and in western South America. Only one diploid species *Fragaria vesca*, differentiated into four regional subspecies, is found in both Eurasia and America. And also, it is the unique diploid species in North America. Combination with other factors and the broader distribution of *F. vesca*, it was suggested originated during the Cretaceous and the ancestor of *F. vesca* may be basal for *Fragaria* genus (Hummer and Hancock 2009).

Strawberry is one of the most important fruit consumed worldwide. It has not only appealing flavor and smell but is also beneficial for human health because of its remarkable nutritional composition of vitamin C, folates and phenolic compounds (Giampieri *et al.* 2014). Cultivated strawberries (*F.* ×*ananassa* notho subsp. *ananassa*) are octoploids, and derive from hybridization between *F. chiloensis* subsp. *chiloensis* forma *chiloensis* and *F. virginiana* subsp. *virginiana*. Because of the genomic complexity of octoploid strawberry, the diploid strawberry *F. vesca* has been mostly used as plant modelto study agronomically important traits of strawberry as well as other *Rosaceae* fruit.

Table I-1. *Fragaria* genes species, ploidy and distribution.

Species	Ploidy	Geographic distribution
F. bucarica	2*x*	Western Himalayas
F. daltoniana J. Gay		Himalayas
F. gracilis A. Los.		North China
F. innumae Makino		Japan
F. mandshurica Staudt		North China
F. nilgerrensis Schlect.		Southeastern Asia
F. nipponica Lindl.		Japan
F. nubicola Lindl.		Himalayas
F. pentaphylla Lozinsk		North China
F. vesca L.		Europe, Asia west of the Urals, North America
spp. *americana*		Eastern North America
spp. *bracteata*		Western North America
spp. *californica*		California
spp. *vesca*		Europe and temperate Asia
F. viridis Duch.		Europe and Asia
F. yezoensis		Japan
F. corymbosa	4*x*	Northern China
F. gracilis		Northwestern China
F. moupinensis (French.) Card		Northern China
F. orientalis Losinsk syn. = *F. corymbosa* Lozinsk		Russian Far East/ China
F. tibetica spec. nov.Staudt		China
F. x *bringhurstii* Staudt	5*x*	California
F. moschata Duch.	6*x*	Euro-Siberia
F. chiloensis (L.) Miller	8*x*	Western N. America, Hawaii and Chile
F. virginiana Miller		North America
F. iturupensis Staudt		Iturup Island, Kurile Islands
F. ×ananassa Duch. ex Lamarck		Cultivated worldwide

Note: From (Hummer and Hancock 2009)

1.2. Morphology

Strawberries are low-growing and perennial herbs. They are capable of sexual and vegetative reproduction. Vegetative propagation is via produce runners, which are trailing, above-ground stems that can take root at their nodes to establish new, clonal daughter plants (Figure I-1). *F. vesca* spp. *vesca*is a mutant of runnerless, it is self-fertile and can produce many seeds per fruit (Davis *et al.* 2007). Strawberries leaves were general trifoliate with long petioles that can be more than 20cm. The fleshy red and flavor strawberry ¨fruit¨, in the botanical sense, is actually the expanded receptacle of the strawberry flower. The true strawberry fruits are the achenes carrying seeds inside that dot the surface of receptacle (Hollender *et al.* 2012). For the purposes of this work, we will refer to consider the expanded receptacle with the achenes as fruit or berry. Ripe strawberry is non-climacteric and has been considered as the most

tractable non-climacteric model system although the influence of hormones including ethylene on the development of aroma volatiles, flavor and color has not been characterized unequivocally.

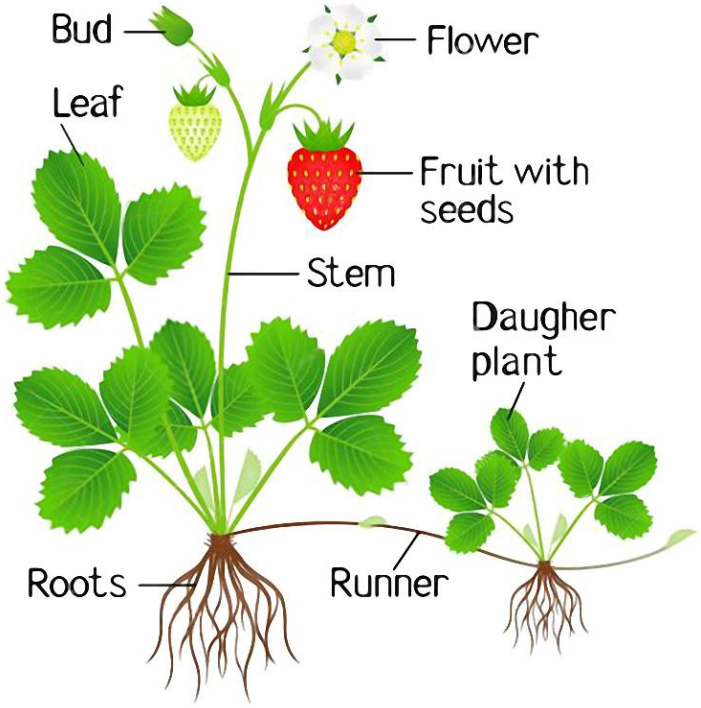

Figure I-1. Strawberry plant botanical drawing. (From https://cn.dreamstime.com)

1.3. *Fragaria vesca*

Fragaria vesca, one of the diploid ancestral subgenome donors for cultivated strawberry, has been used as an attractive model for the study of perennial plants and functional genomics research in *Rosaceae*. *Fragaria vesca* is the most widely distributed *Fragaria sp.* since it has diversified into four subspecies native from both Eurasia and America. Hence, it presents rich morphological diversity for fruits, floral and flowering habit, runnering and crown structure etc (Hadonou *et al.* 2004).

Over the past decades, evidences for *Fragaria vesca* as ideal model plant system have continued to increase. It has small size which caused it could be grown in 10cm pots in laboratory facilities total life and hundreds of plants were able to grown in a small greenhouse. It has a short life cycle (16-20 weeks) and could propagate sexually as well as vegetative reproduction. In addition, the stable and transient transformation could be used for gene functional studies in *F. vesca*. The most important advantage for *F. vesca* as a model species for *Fragaria* genus is it has a small sequenced genome (240Mb) that is amenable to genetic manipulations (Shulaev *et al.* 2011b). It shares a high degree of colinearity with

cultivated strawberry (*F. x ananassa*) and synteny with many commercially fruits crop species of *Rosaceae* family such as apple, pear, peach, plum, raspberry, etc (Illa *et al.* 2011).

F. vesca spp. *vesca* var. ¨Reine des vallées¨ is an alpine accession (PI 551824) which used as recurrent parent in this thesis. It is a runnerless variety with freshly red fruits. It is advantaged for its intense flavor and strong aroma of fruits.

2. Genomic resources in *Fragaria vesca*

2.1. Markers

Molecular markers offer means to detect DNA polymorphism, genetic diversity and population structure of a set of germplasm. DNA-based molecular markers include restriction fragment length polymorphism (RFLP), random amplified polymorphism (RAPD), simple sequence repeat (SSR), amplified fragment length polymorphisms (AFLP), single nucleotide polymorphisms (SNPs), and cleaved amplified polymorphic sequences (CAPS), among others. SSR markers also called microsatellites, is the main molecular markers used molecular markers in strawberry genetic studies and breeding since it has the characteristics of co-dominance, abundant polymorphism, high repeatability, and convenient experimental operation, etc.

Amount of SSRs have been developed for genotyping different accessions within the genus *Fragaria* including diploid strawberry *F. vesca* (James *et al.* 2003; Bassil *et al.* 2006; Monfort *et al.* 2006) and *F. viridis* (Sargent *et al.* 2003) and octoploid strawberry *F. x ananassa* (Govan *et al.* 2008; Rousseau-Gueutin *et al.* 2008).

SNPs are the most common form of genetic variation between individuals, making them very attractive for anchoring genome sequence contigs to linkage maps. In recent years, SNPs are developed based on available reference genome sequence of *F. vesca* and resequencing data of different octoploid species of strawberry (Shulaev *et al.* 2011b; Hirakawa *et al.* 2014; Tennessen *et al.* 2014) as well as abundance of SNPs information from other generated sequences. SNPs are advantageous over other molecular marker because they are bi-allelic and very frequent in genomes.

2.2. Genetic maps and mapping population

Genetic map presents the relative position of polymorphic markers in the whole genome scale, which could be useful in genetic analysis and gene mapping for many important traits. Genetic maps consisting microsatellite, gene-specific and morphological markers (Sargent *et al.* 2004; Sargent *et al.* 2006; Hadonou *et al.* 2004; Sargent *et al.* 2008) have been constructed in diploid strawberry. In octoploid

strawberry, high density genetic map constructed using microsatellite (Isobe *et al.* 2013; Nagano *et al.* 2017; Rousseau-Gueutin *et al.* 2008), IStraw90 Axiom® SNP array (Bassil *et al.* 2015) and double digest restriction-associated DNA sequencing (ddRAD) (Davik *et al.* 2015)were also performed. All these genetic maps would help to understand the genome structure of *F. x ananassa* and facilitate its molecular breeding progress.

The release of woodland strawberry draft genome using second-generation sequencing technology (Vladimir *et al.* 2011), improved genome V4 genome using single-molecule real-time sequencing from Pacific Biosciences (PacBio) (Figure I-2A) (Edger *et al.* 2018) and updated annotation of the V4 genome (Li *et al*. 2019) provide an important tool for researchers to study biologically relevant characters and for breeders to be used in breeding program. Apart from diploid strawberry, a dissection of octoploid strawberry (*Fragaria* × *ananassa*) genome was also performed on Illumina and Roche 454 platforms (Liu *et al.* 2016). Until recently, a near-complete chromosome-scale assembly for cultivated octoploid strawberry (*Fragaria* × *ananassa*) was reported and the phylogenetic analyses supported *Fragaria vesca* being one of the diploid progenitors (Figure I-2B). The release of the octoploid strawberry genome will facilitate cultivated strawberry research and enable molecular breeding in strawberry breeding programs (Edger *et al.* 2019).

Near isogenic lines (NIL) are a collection of lines covering the whole genome of a donor line, with each line containing a single DNA fragment insertion of the donor parental line at the background of a recurrent parental line. NIL is a powerful genetic tool for analysing different phenotypic traits, especially for those following Mendalian inheritance model. NIL has been developed in many important crops including rice (Maeda *et al.* 2014; Ding *et al.* 2011), wheat (Zhou *et al.* 2005; Botha *et al.* 2014; Xu *et al.* 2017), horticultural crops such as cucumber (Witkowicz *et al.* 2003), melon (Eduardo *et al.* 2005) as well as other plant species and has been used to detect QTL (Tanksley *et al.* 1996; Goodstal *et al.* 2005; Moreno *et al.* 2008; Vegas *et al.* 2013). In strawberry, a NIL collection containing 39 lines (Figure I-3) was obtained using *F. vesca* cv. Reine des Vallées and *F. bucharica* as recurrent parent and donor parent, respectively (Urrutia *et al.* 2015). This NIL collection has been applied in genetic QTL analysis of key volatile composition including esters, aldehydes, ketones, alcohols, terpenoids, furans and lactones (Urrutia *et al.* 2017), and in analysis of many stable QTLs related with the accumulation of (poly)phenols, including anthocyanins, flavonols, flavan-3-ols, flavanones, hydroxycinnamic acid derivatives, and ellagic acid (Urrutia *et al.* 2016).

A highly inbred line was developed in *F. vesca* ´Yellow Wonder´ providing an important tool for Rosaceae functional genomics analyses (Slovin *et al*. 2009).

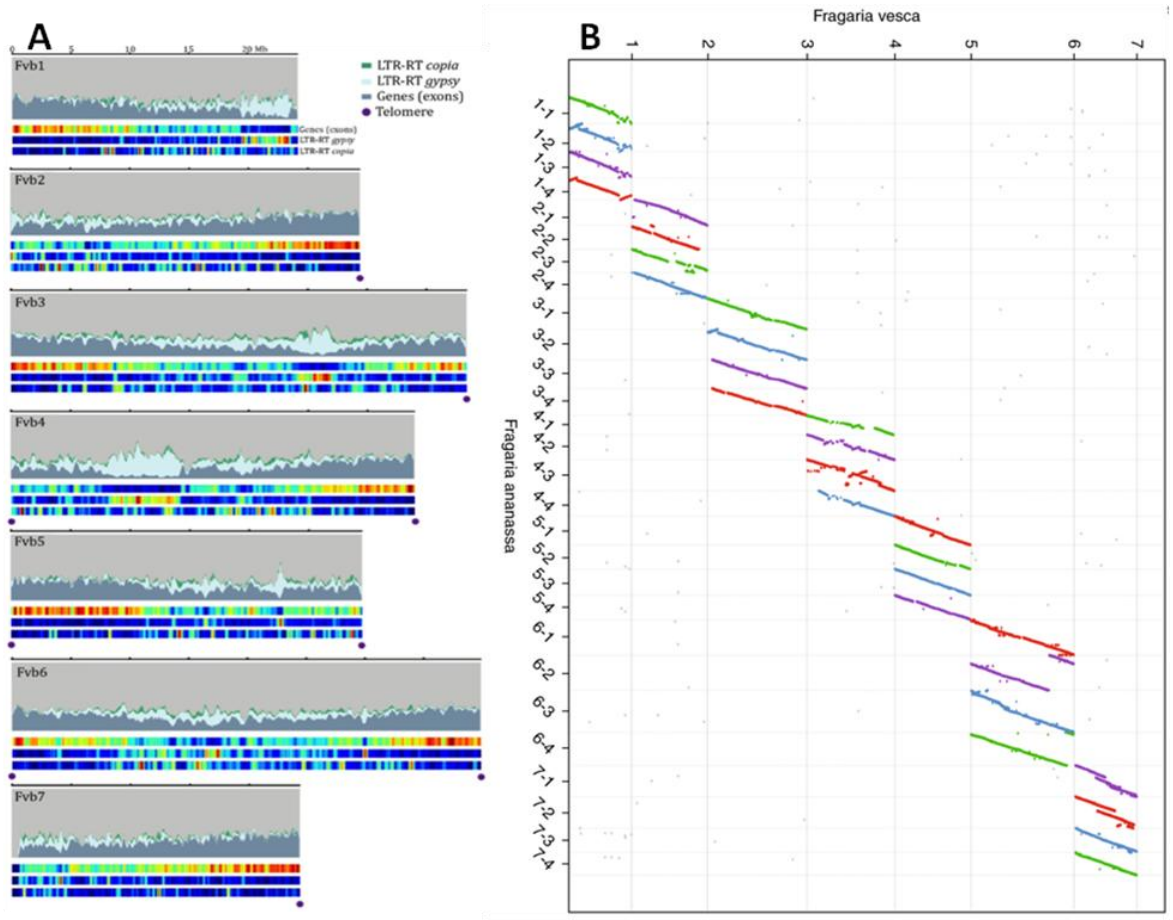

Figure I-2. Diploid and Octoploid Strawberry genomes.

A) Chromosome landscapes of the *F. vesca* V4 genome. The distribution of genes and long terminal repeat retrotransposons (LTR-RTs) are plotted for each of the 7 chromosomes (figure adopted from (Edger et al. 2018)); **B)** Macrosyntenic comparison of the entire *Fragaria* × *ananassa* and diploid *F. vesca* genome (*F. vesca* in red, *F. nipponica* in purple, *F. iinumae* in blue, and *F. viridis* in green) (figure modified from (Edger et al. 2019)).

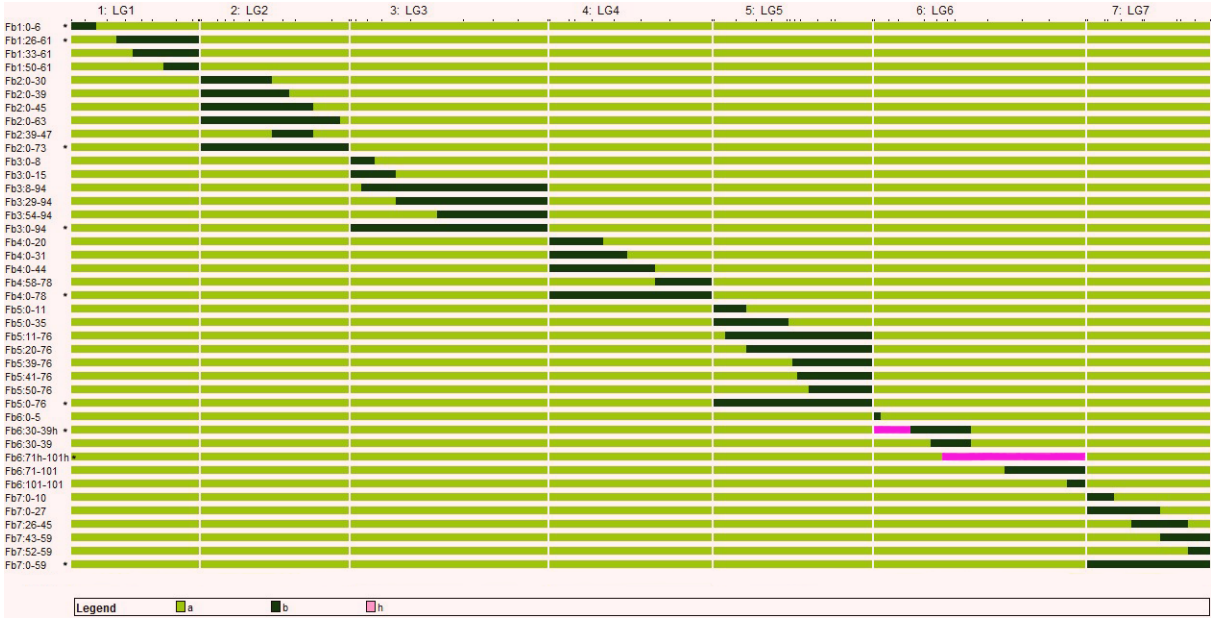

Figure I-3.Graphical genotypes of the *F. vesca* collection of 39 NILs and 2 heterozygous NILs.

(figure adopted from (Urrutia et al. 2015). *F. bucharica* homozygous introgressions are shown in black and heterozygous introgressions in pink. The *F. vesca* genetic background is shown in green. The NIL names are indicated on the right. The first number indicates the LG carrying the introgression. The following two numbers, separated with a hyphen, indicate the marker position at the start and end point of the introgression in centiMorgans, respectively. Dotted NILs indicate the minimal set of lines covering the entire *F. bucharica* genome.

3. Strawberry fruit quality traits

3.1. Strawberry fruit quality traits

Strawberries are a rich source of a variety of nutritive compounds including sugars, vitamins, and minerals as well as non-nutritive, bioactive compounds as flavonoids, anthocyanins and phenolic acids. All of these compounds exert a synergistic and cumulative effect on human health promotion and in disease prevention (Giampieri *et al.* 2015). Genetic factors as well as environmental conditions have an influence on the fruit quality and compound composition (Palmieri *et al.* 2017). Strawberry fruit quality traits are major factors to be considered in strawberry breeding programs, including all the compounds inferred as well as fruit appearance such as color and shape.

3.2. Fruit acidity and sugar content

Fruit sweetness is a major factor affecting strawberry commercial value. The ratio of sugar and organic acid in fleshy fruit has been considered as the major factor contributing to fruit sweetness (Qiao *et al.* 2017).

In most fruits, glucose, fructose and sucrose are the main forms of the soluble sugars, among which fructose is the sweetest sugar with almost twice the sweetness of glucose (Desnoues *et al.* 2014). Total sugar content in most plant species is evaluated with total soluble solid content (SSC) or ^{o}Brix degree measured using digital refractometer. Content of individual sugars has been mainly chromatographically analyzed using HPLC system. Many environmental factors such as culture condition, ions content in irrigation water were previously described affecting strawberry sugar content. Strawberry plants treated with appropriate iodine amount promoted plant growth as well as soluble sugar contents while $IO3^-$ increased fruit total acidity (Li *et al.* 2017a). Deficit irrigation decreased strawberry yield but increased the amount of sugar and organic acid (Weber *et al.* 2017).

Regarding genetic factors, many genes have also been found to have a great impact on sugar accumulation in different fruit species either by affecting sugar transport or sugar synthesis. Sour jujube (*Z. jujuba* Mill. *var. spinosa* Hu.) contained less sugar than cultivated jujube (*Ziziphus. jujuba* Mill.) and this sugar accumulation difference in fruit was mainly affected by sugar transport rather than sugar biosynthesis (Zhang *et al.* 2018). PbTMT4, a member of tonoplast monosaccharide transporter, was correlated with soluble sugar contents during the pear fruit development, mediated vacuolar sugar transport and affected sugar accumulation (Cheng *et al.* 2018b). Mineral element Mg concentration in the fruit of citrus was found positively correlated with organic acid while negatively correlated with sugar content, thus the balance of acids and sugars might change the flavor of citrus fruit (Zhou *et al.* 2018). CmTST2, a tonoplast sugar transporter, was highly expressed during melon fruit development, playing an important role in sugar accumulation (Cheng *et al*. 2018a). SlARF10, an auxin response factor, was involved in sugar accumulation during tomato fruit development (Yuan *et al.* 2018). In strawberry, characterization of a set of sucrose transporters revealed that FaSUT1 was a major component responsible for sucrose accumulation during fruit development (Jia *et al.* 2013). The interplay between FaMYB44.2 and FaMYB10 acted as a negative regulator in sucrose accumulation during strawberry fruit ripening (Wei *et al.* 2018). MicroRNA399 transgenic strawberry contained increased fructose, glucose and soluble solid contents compared with the wild type indicating that mRNA399 has a functional role in improving strawberry fruit quality (Wang *et al.* 2017b). A stable and major QTL for decreasing of fructose, glucose and total sugar content was mapped in LG2: 45-63cM in the *F. vesca* NIL collection and the QTL for decreasing fructose and glucose were detected in LG3 cM 54-94 and LG5 cM 41-50 respectively.

In fruits, main organic acids included are citric, malic, tartaric, succinic and oxalic acids. Fleshy fruit acidity is measured by titratable acidity or pH. The predominant organic acids differ in ripe fruits between plant species. Malic acid is predominant in apple and pear while citric acid is predominant in citrus (a review see (Etienne *et al.* 2013)). The influence of malic acid is achieved through both genetic and

environmental factors in apple, for example. Several QTL in linkage group 8 and linkage group 16 were identified correlated with malic acid content of apple fruits (Zhang *et al*. 2012; Jia *et al.* 2018). Candidate genes *MdPP2CH*, *MdSAUR37* and *MdALMTII* were screened in major QTL region and validated to influence the malate content of apple fruits (Jia *et al.* 2018). Other genes such as MYB transcription factors *MdMYB1*, *MdSOS2L1* and *Ma10* were all found involved in regulating malic acid accumulation in apple fruit (Hu *et al.* 2016; Sun *et al.* 2016; Ma *et al.* 2019). Citric acid accumulation in citrus fruit is also affected by genetic and external conditions. External conditions such as γ-aminobutyric acid (GABA) treatment and ironshortage both could increase the content of citric acid *(Sheng et al. 2017; Shlizerman et al.* 2007). CsGAD1, a member of glutamate decarboxylases and a key regulator in the biosynthesis of GABA, had a strong correlation with citric acid utilization (Liu *et al.* 2014). *CsPH8*, a P-type proton pump gene and *CitERF13*, a transcription factor were found related with citric acid accumulation *(Shi et al*. 2019; Li *et al.* 2016) while genes *CitNAC62*, a transcription factor and two transport-related genes, *CitCHX* and *CitDIC* were involved in citric acid degradation (Li *et al*. 2017b; Lin *et al*. 2015).

Citric acid is the predominant organic acid in strawberry, however, most studies related with organic acid regulation were performed together with sugar, the ratio of ^{o}Brix and organic acid was a useful indictor to evaluated fruit flavor. Preharvest ultraviolet-C irradiation could slightly decrease titratable acidity and pH (Xie *et al.* 2016). Different genotypes of strawberry vary in acidity level (Mikulic-Petkovsek et al. 2012; Akhatou *et al.* 2016). One QTL for pH on LG 4CII and two QTLs for titratable acidity on LGs 2A and 5B were detected based on a pedigree-based QTL analysis in octoploid strawberry (Verma *et al.* 2017). QTL related with citrate and malate acid content were also detectedin different linkage groups but need to be further validated (Lerceteau-Kohler *et al.* 2012).

3.3. Volatile compounds

Even though volatile compounds make up for a minor proportion of strawberry weight, they could affect strawberry flavor with minor content modification. The most effective and frequently used volatile compound detection method gas is chromatography−mass spectrometry (GC-MS). More than 360 chemicals have been identified in strawberry flesh classified into groups of esters, alcohols, ketones, furans, terpenes, aldehydes, and sulfur compounds (Yan *et al.* 2018). Many factors such as genotype, environmental factors including cultivation practices, harvesting and storage conditions, as well as the analytical method all can influence the quantity and the quality of the identified substances (Ulrich *et al.* 2018). To date, the 20 most frequently identified volatile compounds in strawberry are methyl hexanoate, ethyl hexanoate, ethyl butanoate, methyl butanoate, linalool, γ-decalactone, hexyl acetate, γ-dodecalactone, DMMF (2,5-dimethyl-4methoxy-3(2H)-furanone, furaneol), (E)-2-hexenal, butyl acetate, DMHF (2,5-dimethyl-4-hydroxy-3(2H)-furanone, mesifurane), ethyl 3-methylbutanoate, ethyl 2-

methylbutanoate, hexanoaic acid, methyl octanoate, 2-methyl butanoic acid, ethyl acetate, hexanal and butanoic acid (Ulrich *et al.* 2018). Esters (ethyl butanoate, ethyl hexanoate, methyl butanoate, and methyl hexanoate, etc.), furanones (mainly DMHF and DMMF), terpenes (linalool and nerolidol), and sulfur compounds (methanethiol) are among the substances that account for the characteristic of strawberry aroma (Yan et al. 2018). Most of esters offer fruity odor for strawberry fruits and furanones contributed caramel-like, sweet, floral, and fruity aroma. However, the (E)-2-hexenal is the represent of green leaf compounds and hexanoaic acid and 2-methyl butanoic acid present sour flavor. Linalool as a flowery aroma contributor abundantly exists in octoploid species.

Flavor improvement is currently one of the most important but complex traits in strawberry breeding. For breeders, they could apply modern technology to utilize genotype with favourable flavor to improve the cultivars with high quality in other aspects. Diploid strawberry has more intensive volatile compounds conferring to pleasant aroma which is not common in cultivated octoploid strawberry. Using a diploid strawberry NIL collection, 50 major QTLs controlling volatile accumulation to increase wild strawberry flavor were identified (Urrutia *et al*. 2017) including the QTL for ¨green leaf volatile compounds¨ was mapped in LG5: 50-76cM, methyl 2-aminobenzoate decreasing QTL were detected in LG5:11-35cM and LG7:0-10cM and many stable QTL for key volatile compound like ethyl hexanoate, mesifurane, methyl butanoate, etc. These QTL might shed light on further investigation of gene identification for related volatiles.

In octoploid strawberry, a set of genes have been described regulating synthesis of some volatile compounds. DMHF is an important compound contributing to the flowery flavor of cultivated strawberry and is metabolized to the flavorless HDMF β-d-glucoside during fruit ripening. Functional characterization of ripening-related UGTs (UDP-glucosyltransferases) revealed that four Site-directed mutagenesissingle amino acid change in UGTs could increase HDMF glucosylation activity and provide the foundation for improvement of strawberry flavor (Song *et al.* 2016).

Transcription factor *FaDOF2* expression silencing could down regulate eugenol production in ripe fruit receptacles via decreasing the expression of the eugenol synthase gene *FaEGS2* and the *R2R3* MYB transcription factor (Molina-Hidalgo *et al.* 2017). γ-decalactone is a volatile compound conferring a peach flavor note to fruits and *FaFAD1* was identified as a possible gene controlling this compound's presence/absence using a combination of genetic, genomic and analytical chemistry methods (Sanchez-Sevilla *et al.* 2014; Chambers *et al.* 2014). Due to work in QTL analysis with related traits, molecular markers with high accuracy in predicting the presence of the important volatiles mesifurane and γ-

decalactone have been validated and could be used in breeding programs to specifically select cultivars with superior flavor (Cruz-Rus *et al.* 2017).

3.4. Fruit color

Fruit color is one of most important agronomical traits affecting the attractiveness of fruit for consumers, thus determining fruit commercial value. Most commercial cultivars (varieties) have appealing red fruits while still many white fruited strawberries can be found in the market or strawberry germplasm collections. For example, Yellow Wonder (YW) and Hawaii 4 (HW4) are different genotypes of diploid *F. vesca* with white fruits differing in metabolic and transcriptional level from red fruit cultivar. *Fragaria chiloensis* (L.) Mill spp. c*hiloensis* form *chiloensis*, is a strawberry that produces white fruits with unique aromas (Prat *et al.* 2014). In octoploid strawberry, there are also some varieties such as "Xiaobai", 'Snow Princess' and 'Tokun' (Zhao *et al.* 2018).

Spraying calcium was also reported to have the ability to enhance the accumulation of anthocyanins in strawberry fruit thus changing fruit color probably by regulating key structural genes related with synthesis of anthocyanins (Xu *et al.* 2014).

Strawberry is a non-climacteric horticultural plant. The development of strawberry fruit can be divided into several steps including small green, large green, degreening, white, initial red, partial red and full red. Apart from strawberry developmental color transition, strawberry has a wide range of diversity with varied fruit shape, fruit size as well as fruit color of ripe fruit, from white to red (Figure I-4). Polyamine spermine dominance, resulted from abundant transcripts of *S*-adenosyl-l-Met decarboxylase gene (*FaSAMDC*), regulate strawberry fruit ripening in an ABA-dominated, IAA-participating, and ethylene-coordinated manner (Guo *et al*. 2018). ABA receptors FaPYR1 (Chai *et al.* 2011), a highly conservative deduced protein, and FaRIPK1 (Hou *et al.* 2018), a leu-rich repeat receptor-like protein kinase, were also involved in strawberry fruit development and color transition. Gene *FaPAL6* with fruit-specific expression pattern showed increased transcription during fruit ripening along with the accumulation of anthocyanin. Also the expression of this gene was higher in an anthocyanin-rich strawberry cultivar than that in a cultivar with lower anthocyanin content (Pombo *et al.* 2011). RNA-seq and transcriptomic analyses revealed that the complex control of fruit color is mostly related with key genetic determinants of anthocyanin regulation and biosynthesis (Hossain *et al.* 2018; Zhang *et al.* 2015).

The quantity and variability of flavonoids, including proanthocyanidins (PAs), anthocyanins and flavonols control fruit coloration (Koes *et al.* 2005; Schaart *et al.* 2013).The syntheses of these components are catalyzed by a set of enzymes and could be illustrated in a biosynthetic pathway (Figure I-5). Key enzymes involved in the pathway are chalcone synthase (CHS), chalconeisomerase (CHI),

flavonoid 3',5'-hydroxylase (F3'5'H), flavonoid 3'-hydroxylase (F3'H), flavonoid 3-hydroxylase (F3H), dihydroflavonol-4-reductase (DFR), flavonol synthase (FLS), leucoanthocyanidinreductase (LAR), leucoanthocyanidin oxidase (LDOX), anthocyanidinreductase (ANR) and UDP-glucose:flavonoid 3-O-glucosyl transferase (UFGT). The concentration of flavonols, including myricetin, kaempferol and quercetin are mostly affected by enzyme activity of FLS. Expression of *leucoanthocyanidinreductase* (*LAR*) and *anthocyanidinreductase* (*ANR*) is required for the formation proanthocyanins. UFGTs determine the formation of anthocyanins (Bogs *et al.* 2006).

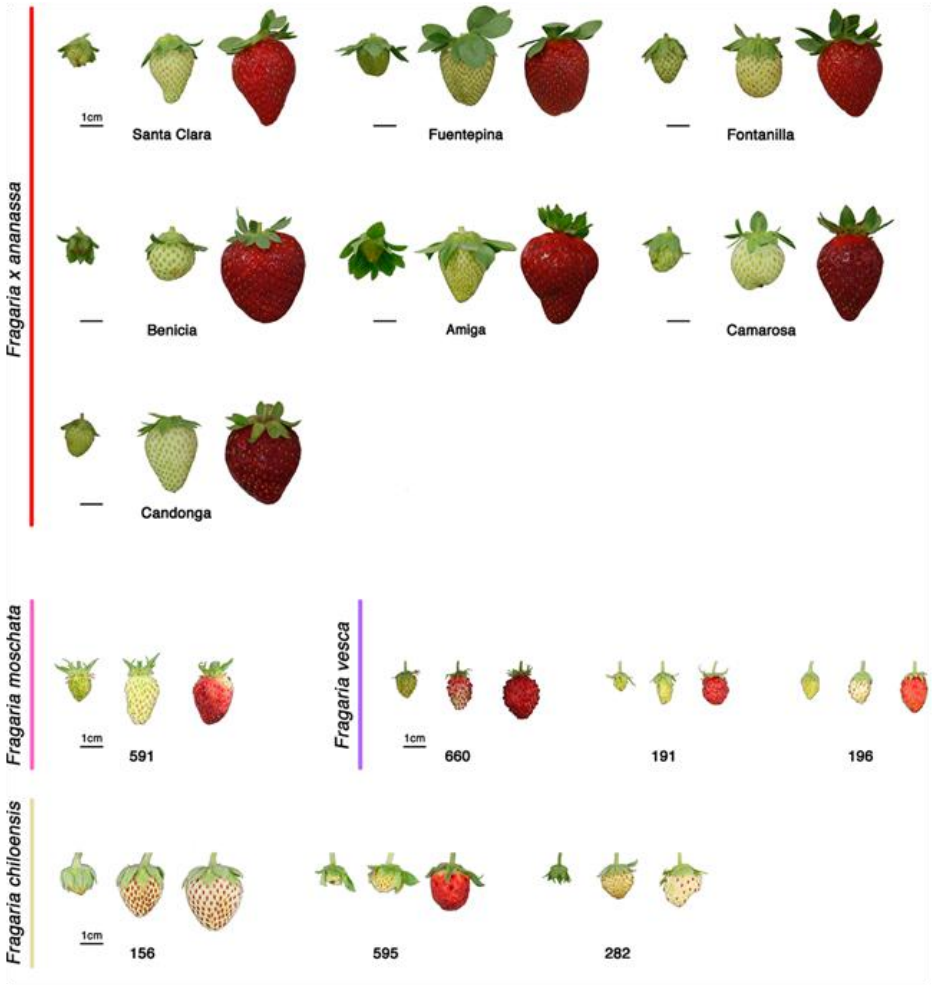

Figure I-4. Strawberry color and shape diversity. (Figure applied from (Vallarino *et al.* 2018)).

Transcription factors (TFs) have been described as key regulators controlling color differentiation from both genetic and transcriptomic aspects in many plant species, among which MYBs were mostly illustrated ones (Allan *et al.* 2008). For example, in sweet cherry, R2R3 MYB transcription factor PavMYB10.1 was identified involved in anthocyanin biosynthesis pathway and determines fruit skin color possibly by interacting with proteins PavbHLH and PavWD40, and binding to the promoter regions of the anthocyanin biosynthesis genes *PavANS* and *PavUFGT* (Jin *et al.* 2016). MYB family genes have been identified to affect fruit color differentiation also in strawberries. The diploid "Alpine" strawberry *F. vesca* ssp. *vesca* has red skin and light red flesh, whereas transgenic lines over-expressing *FvMYB10* showed purple skin and red flesh and RNAi line with silenced *FvMYB10* had white skin and white flesh (Lin-Wang *et al.* 2014). In cultivated octoploid strawberry (*Fragaria* x *ananassa* Duch.), an ACTTATAC insertion in the genomic region encoding the C terminus of the protein was found in *FaMYB10* and *FaMYB10-2* of white strawberry varieties, leading to premature termination of the protein, causing the inability to activate downstream flavonoid biosynthesis genes (Wang *et al.* 2019).

Other TFs like bHLH (Xie *et al.* 2012; Jo and Kim 2019), WRKY (Cheng *et al.* 2017; Wang *et al.* 2017a), APRR2 (Oren et al. 2019) were also found involved in plant fruit coloration. The R2R3 MYB TFs regulating anthocyanin biosynthesis have been shown to interact closely with TFs bHLH (Tohge *et al.* 2005). WRKY TFs have been found to regulate the production of a variety of phenolic compounds, resulting in altered biosynthesis of other phenolic-based compounds like flavonoids (Besseau *et al.* 2007). Ectopic expression of *APRR2* gene in tomato resulted in up-regulated several ripening-related genes with increasing plastid number and pigment content, enhancing the levels of both chlorophyll in immature unripe fruits, and carotenoids in red ripe fruits.

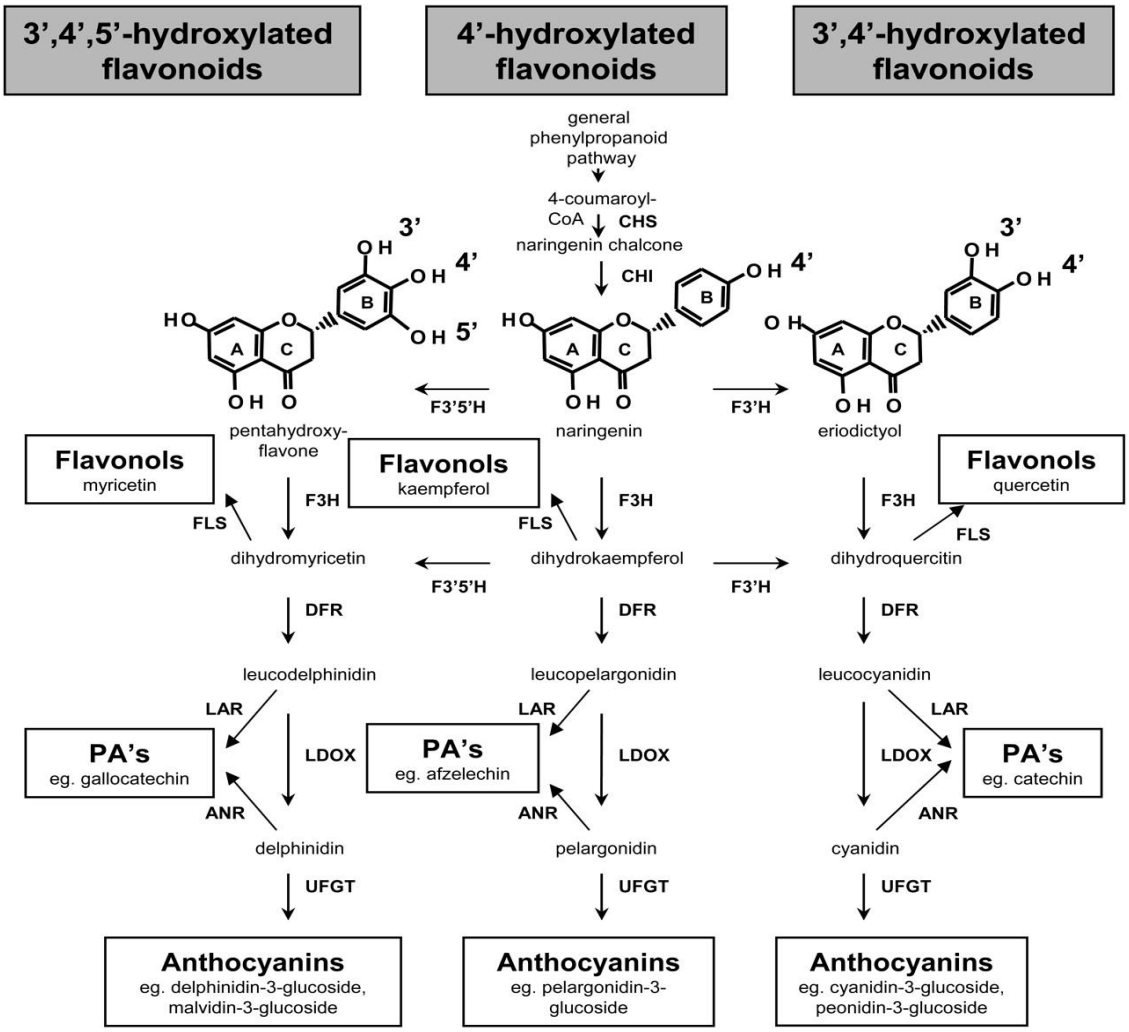

Figure I-5. Schematic representation of the flavonoid biosynthetic pathway. (Figure applied from Bogs *et al.*2006)

Objective

Objective

The general objective of this work is to detect QTL controlling traits related with strawberry fruit quality and flavor identifying genes responsible or regulators for the synthesis of compounds related with fruit flavor, taste and aroma, and appearance in diploid strawberry. To achieve this goal, three specific objectives are addressed:

1. Improvement of previously developed NIL collection of *F. vesca* and functional characterization of fruit quality taste, related to acid and sweet.
2. Genetic analysis of fruit aroma volatile compounds and functional verification of candidate genes related to green leaves aroma perception.
3. Genetic analysis of fruit appearance and identification of QTL and candidate gene for strawberry orange fruit color.

Chapter I

Improvement of *Fragaria vesca* near isogenic line (NIL) collection and functional characterization of fruit quality

Introduction

Strawberry, one of the most economically important fresh and processed fruit, is cultivated in all arable regions of the globe from the Arctic to the Tropics. The cultivated strawberry *F. x ananassa* is among the most complex of crop plants, harboring eight sets of chromosomes ($2n = 8x = 56$) derived from as many as four different diploid ancestors. Diploid strawberry *Fragaria vesca* ($2n= 2x=14$) is one of the putative diploid progenitors and donor of subgenome A in the octoploid genome (Edger *et al.* 2019). With its small genome size, short generation time, and well-established transformation system, *F. vesca* is an ideal model species for *Fragaria* and other *Rosaceae* species (Alger *et al.* 2018). The genome of *F. vesca* has been improved several times after the publication of the first genome in the year 2011 (Shulaev *et al.* 2011). The genome information has been an invaluable resource to the strawberry research and provides a powerful tool for gene mapping of agronomical important traits.

Near isogenic lines (NILs) are strains whose genetic content are identical except for a single DNA introgressed fragment from a donor line (Muehlbauer *et al.* 1988). NILs are developed through F_1 individual backcrossed with the recurrent parent, and the resulting BC_1 generation is recursively backcrossed for many (n) generations to obtain a single introgression of the desired genetic size. Then BC_n individuals are self-pollinated to fix homozygous lines. After each cross, progenies are selected by molecular markers to speed up selection process and reduce the number of generations needed to fully develop a NIL collection. After several generations, the genome of the selected individuals consists almost exclusively of background from recurrent parent and only one single introgression from donor parent (Young *et al.* 1988; Barrantes *et al.* 2014). NIL collections are excellent materials for genetic studies, including insertion of wild alleles and phenotypic variability into the elite germplasm (Zamir 2001; Zarouri *et al*. 2014; Merchuk-Ovnat *et al.* 2016), the exploration of gene effects (Tanksley *et al.* 1996; Brouwer and Clair 2004), screening of the molecular markers linked with the gene of interest and gene expression (Telebanco-Yanoria *et al.* 2011).

Over the past decades, NILs have been used extensively for mapping and tagging of both qualitative inherited genes and quantitative trait loci (QTL). If the phenotypes of a NIL and its recurrent parent present any obvious difference, this trait can be attributed to genetic factors in the introgressed fragment. In addition, NIL can be easily replicated via self-pollination, thus this kind of population has the advantage on increasing phenotyping accuracy as each line can be tested in different time points, decreasing environmental factor effect (Monforte *et al.* 2001). Since the first NIL collection developed in tomato (Eshed and Zamir 1994), dozens of QTL affecting yield, antioxidant capacity, lycopene content, and fruit-quality traits were mapped using NIL collection in tomato (Ashrafi *et al*. 2012; Barrantes *et*

*al.*2016; Eshed and Zamir 1995; Rousseaux *et al.* 2005). Several NIL collections have been developed as genomic resources in other species including Arabidopsis (Fletcher *et al.* 2013), tomato (Frary *et al.* 2000; Alpert and Tanksley 1996), rice (Zeng *et al.* 2009; Fukuoka *et al.* 2014; Khanna *et al.* 2015), wheat (Baek and Skinner 2003; Mago *et al.* 2005; Wang *et al.* 2019b), barley (Jiang *et al.* 2019; Chen *et al.* 2012) and melon (Eduardo *et al*. 2005b; Perpiñá *et al*. 2016).

The diploid strawberry NIL collection has been developed using two diploid species: *F. vesca*, as recurrent parent and *F. bucharica*, as donor parent (Urrutia *et al.* 2015). The NIL collection consisted of 39 homozygous lines and two heterozygous lines covering 522 cM and 192.7 Mb (96 and 92% of the genome in genetic and physical distance, respectively). Only a 19.3cM region on LG1 was not covered and lines covering a region of LG6 (Fb6**:**30-39h and Fb6**:**71h-101h) retained heterozygous introgressions. The NIL collection was phenotyped for a set of characters responsible for morphological and phenotypical characteristics of different parts of the plant and for characters related with the chemical composition of the fruits. A set of Mendelian inheritance genes and QTL was mapped. It is a new tool for in-depth study of the genetically important characters in strawberry and *Rosaceae* family.

In the recent years, fruit sensorial and nutritional traits have become major breeding targets in strawberry (Lerceteau-Köhler *et al.* 2012). Strawberries with intense flavor are characterised by their high titratable acidity, high soluble-solid content (Kader 1991) and high level of aromas (Aharoni *et al.* 2004). The main titratable acidity is citric acid in the strawberry, which accounts for 88% of the total acid content (Green 1971). The main soluble sugars represent more than 80% of the total sugars and 40% of the total dry weight (Wrolstad and Shallenberger 1981). Therefore, soluble sugars and citric acids of strawberries are regarded as significant quality factors.

In this study, in order to improve the NIL collection to be a more thorough and accurate genetic resource, selected heterozygous lines were backcrossed with the recurrent parent in order to obtain lines harbouring introgressions from *F. bucharica* that could cover the entire genome of *F. bucharica*. In addition, QTL analyses related tofruit quality traits like fruit pH, citric acid and soluble solid content were also performed, which will help to improve fruit quality in strawberry breeding.

Material and methods

Plant materials and sub-NIL development

Fragaria near isogenic line (NIL) collection with segments of *F. bucharica* (PI657844, also known as FDP601) (FB) introgressed into the background of *F. vesca* cv. 'Reine des Vallées' (RV), covering almost all seven linkage groups (Urrutia *et al.* 2015) and the parental lines of NIL collection FB and RV were used as materials in the study. The name of a NIL (for example Fb5:0-35) represents donor parent (Fb), the number of the linkage group (5), and the following two numbers, separated with a hyphen, indicate the marker position at the start (0) and end (35) of the introgression in centiMorgans. Since the NIL collection contained a gap region from 6cM to 26 cM in linkage group 1, and regions from 6cM to 30 cM and 39cM to 71cM in LG6 retained heterozygous introgressions, the heterozygous plants Fb1:0-49h, Fb1:6-61h, Fb1:0-61h, Fb6:6h-38 and Fb6:6h-101h (Figure1-1) were selfed to develop new NILs in homozygosis in LG1 and LG6. The mentioned lines only have heterozygous introgressions in LG1 and LG6 respectively with homozygous RV background in the other linkage groups. The lines Fb5:0-35 and Fb7:0-10 were backcrossed with *F. vesca* to produce sub-NILs with smaller introgressions in LG5 and LG7 because many stable and major QTL related to fruit flavor traits were mainly located on LG5 in the regions between 0-35cM and on LG7 between 0-10cM. They were backcrossed with the recurrent parent RV in 2016, and then the progenies were self-pollinated to produce sub-NILs. The entire NIL collection was used for phenotypic analysis of fruit quality related traits.

Figure 1-1. Graphical representation of the lines used to develop new lines and sub-NILs in the experiments. *F. bucharica* homozygous introgressions are shown in black. The *F. vesca* genetic background is shown in green. The heterozygous introgressions are shown in rose. The NIL names are indicated on the left. The first number indicates the LG carrying the introgression. The following two numbers, separated with a hyphen, indicate the marker position at the start and end point of the introgression in centiMorgans, respectively.

All seeds were germinated each year, from October 2015 to 2018 as described in Urrutia *et al* (2015). All the plants were grown under greenhouse conditions at day temperature between 22 to 24°C and 17°C during night without artificial lighting, relative humidity 40-50% at the Center for Research in Agricultural Genomics (CRAG) in Bellaterra (latitude:41° 29'N, longitude: 2° 06'E). The plants used for producing new NILs and sub-NILs were kept in greenhouse for selfing or backcrossing with the recurrent parent *F. vesca*.

For the full NIL collection, every line was represented by six plants when the seeds germinated, until March next year, when four individual plants of each genotype were transferred to a shaded greenhouse at Torre Marimon, Caldes de Montbui (latitude 41°36'N, longitude 2°10'E at 203 m of altitude from sea level) for fruit phenotypic evaluation. The remaining two plants were kept in CRAG for seed production. The plants in the shaded greenhouse grew under natural conditions and photoperiod without receiving additional lighting or heating from March to September. The agronomical practices were the usual for strawberry fruit production.

Genotyping

All individuals were genotyped by SSR (Single Sequence Repeat) markers to select candidates for selfing or backcrossing. Young leaf tissues were collected at three leaves stage and ground to fine powder in liquid nitrogen using 3-mm tungsten beads in a Retsch MM400 ball mill (Retch GmbH, Germany). DNA was extracted using the method of (Doyle *et al.* 1990) modified by the addition of 2% PVP-40. For PCR, the DNA was quantified and 20 ng per reaction were used. SSR markers were used to locate the introgressed regions. The NIL collection was genotyped by selecting three markers from each linkage group at the beginning, middle and end (Table1-1). In order to select new NILs in LG1, all the markers in LG1 and three markers from every other linkage group (markers in black) were used for genotyping. The same was done for the selection of new NILs in LG5, LG6 and LG7, respectively (Table 1-1). All markers have been described in (Urrutia *et al.* 2015).

PCR reactions were carried out in a final volume of 10 µl containing: 40 ng of DNA, 1X PCR buffer (50 mMKCl, 10 mM Tris-HCl pH 8.3, 0.001% gelatin), 1.5 mM $MgCl_2$, 0.2 mM dNTPs, 1U of DNA polymerase AmpliTaq (Perkin-Elmer, IL, USA), 0.15 µM forward primer, 0.2 µM reverse primer and 0.2 µM of M13 labelled primer, with an identical sequence added to the 5´-tail on the forward primers (Schuelke 2000). The choice of the dye label, 6-FAM, VIC, PET or NED (Applied Biosystems, CA, USA), depended on further multiloading capillary electrophoresis. PCR amplifications were run on a PE9600 thermal-cycler (Applied Biosystems, CA, USA) as follows: 2 min of initial denaturing at 94°C; 10 cycles of 15 sec at 94°C, 15 sec at annealing temperature and 30 sec at 72°C; 25 cycles of 15 sec at 94°C, 15 sec at 50°C and 30 sec at 72°C, followed by a final extension of 5 min at 72°C.

Amplicions were visualized by capillary electrophoresis in an ABI3130xl Genetic Analyzer (Applied Biosystems, CA, USA), run using 2 µl of a mix containing three differently labelled PCR products, 0.3 µl of LIZ-500 ladder and 12 µl of deionized formamide. Data generated by capillary electrophoresis were analysed using the GENEMAPPER software application (Applied Biosystems, CA, USA).

Table 1-1. SSR markers used for genotyping of NILs and selecting selfing or backcross progenies.

Linkage group	Marker*	Position (cM)	Linkage group	Marker*	Position (cM)
1	**EMFvi072**	0	5	UDF009	41.2
1	EMFn049	6.3	5	EMFv024	56.2
1	**EMFn136**	25.6	6	ARSFL007	0
1	UFF02F02	49.9	6	**CFaCT107**	5
1	**EMFv025**	60.5	6	**EMFn228**	11.4
1	CFV164	61.5	6	**FvH4123**	24
2	EMFv002	15.7	6	EMFn117	30.1
2	EMFv031	23.2	6	EMFn017	38.8
2	EMFv003	70.9	6	**CFVCT017**	54.3
3	EMFv029	2.4	6	**EMFv160AD**	70.9
3	CFVCT007	61.3	6	**EMFv160BC**	101
3	CFVCT012	94.3	7	**CFV3096**	0
4	UDF007	9.4	7	**FvH4152**	2.5
4	CFV3148	20.3	7	EMFv021	8.2
4	ChFaM23	77.5	7	**EMFn201**	9.5
5	EMFvi108	3.5	7	**CFV3896**	10.8
5	**CFV3132**	10.8	7	EMFv023	42.8
5	**EMFn110**	27.7	7	EMFV023	51.7
5	**FvH4095**	34.95			

*Markers in black were used to confirm the genotypes of all plants. Markers in **bold** were specifically used to select plants with small introgressions in LG1, LG5, LG6 and LG7, respectively.

Phenotyping and genetic analysis

For all the NILs, we collected fruits to measure fruit acidity and ° Brix.

Fruit samples from shaded greenhouse plants were collected with four harvests during May to July every year. Fully ripe red berries from the same line were pooled together and each harvest was analyzed as an independent biological replicate. Each biological replicate was a mix of at least 10 fruits (>10g). Fresh fruits were used to measure fruit acidity and °Brix.

Total fruit acidity was measured by citric acid titration (CA, expressed in g/100ml). Fruit samples of approximately 10 g (precise weight was recorded) were weighed and crushed using a handheld blender W112 (Dynamix, Greece) after adding 5 ml of MilliQ water. To make the measurement, MilliQ water was added to 5 ml of the puree until reaching a 50 ml volume. First pH was measured and scored and then the titration was measured, using automatic valuation system HI 84532 (Hanna Instruments (Pty) Ltd,

South Africa), by adding a solution of 0.1 M NaOH and the CA was recorded when the solution's pH reached 8.1. Each sample was measured using two technical replicates.

^{o} Brix values were measured using refractometer PAL-1 (Atago, Japan). The purée used was from the fruit acidity assays (10g fruit + 5 ml MilliQ water).

Statistical analyses were carried out using the JMP®8.0.1 statistical package (SAS Institute, NC, USA). The mean and ANOVA test were calculated for each trait. Means were compared with the recurrent parent by a Dunnett's test with $\alpha \leq 0.05$. QTL was determined when all significant lines shared the same interval.

Results

Accomplishment of NIL collection and development of sub-NILs

Linkage group 1

In the previous NIL collection created by Urrutia *et al* (2015), each NIL contained a single homozygous introgression and all seven linkage groups were covered by *F. bucharica* introgressions except a 19.3cM (6cM- 26cM) region on LG1 and a region of LG6 that both retained heterozygous introgressions.

To cover the gap in linkage group 1, two heterozygous plants (Fb1:6-61h and Fb1:0-61h) were selected to produce new NILs.

The lines Fb1:6-61h and Fb1:0-61hwere recovered from BC_2 generation due to being heterozygous in LG1 and homozygous for RV in other six linkage groups. In 2016, the progenies from selfings produced by Fb1:6-61h and Fb1:0-61h, were germinated and screened using SSR markers. In total, six and four lines were selected from 21 progenies of Fb1:6-61h and 40 progenies of Fb1:0-61h, respectively. Among them, one new NIL, Fb1:0-61 with an introgression covering the entire LG1 in homozygosis and nine different new heterozygous lines were obtained (Table1-2). Thereafter, all the selected ten individuals were maintained and selfed to produce offspring.

Table 1-2. Selected individuals from the selfing progenies of Fb1:6-61h and Fb1:0-61h.

primer	EMFvi072	EMFn049	EMFn136	UFF02F02	EMFv025	CFV164	name
LG	1	1	1	1	1	1	
Position (cM)	0	6.3	25.6	49.9	60.5	61.2	
1.6-61h	a	h	h	h	h	h	
1	a	h	h	h	h	h	1:6-61h
2	a	h	b	h	h	h	1:6h-25b
3	a	h	b	b	h	h	1:6h-49b
4	a	h	h	h	h	a	1:6-60h
5	a	h	a	a	a	a	1:6-6h
6	a	h	h	a	a	a	1:6-25h
1.0-61h	h	h	h	h	h	h	
1	a	a	a	a	h	a	1:60-60h
2	b	b	b	b	b	b	1:0-61
3	h	h	h	h	a	a	1:0-49h
4	a	h	h	h	h	h	1:0-61h

Yellow bar represents the parent plant and its genotype.

h = heterozygous, a= homozygous for the recurrent parent (RV) and b= homozygous for the donor parent (FB).

Name on the left indicates the number the seed was given after germination.

Name on the right are the names given after analysis according to their introgressions. The ‘h’ in the name designates that the introgression is still in heterozygosity, ‘b’ means the introgression is homozygous for FB at that position.

In 2017, the seeds of selected lines from 2016 were germinated and screened by SSR markers as shown in Table 1-1. In total, 20 out of 193 new plants were selected to produce more NILs harbouring small introgressions in LG1 (Table1-3). Three of them, Fb1:6-61, Fb1:60-61 and Fb1:0-50 were found with homozygous FB introgressions in LG1 and considered as new NILs.

Finally, four new NILs Fb1:0-50, Fb1:6-61, Fb1:61-61 and Fb1:0-61 were obtained after two years and they supplemented the introgression gap in LG1. Finally, eight NILs harbouring different introgressions could cover the entire LG1.All the seeds were stored in the fridge and the progress to produce new NILs or sub-NILs in linkage group 1 was completed.

Table 1-3. Selected individuals from the selfing progenies of plants kept in 2016.

primer	EMFvi072	EMFn049	EMFn136	UFF02F02	EMFV025	CFV164	name
LG	1	1	1	1	1	1	
Position(cM)	0	6.3	25.6	49.9	60.5	61.5	
1.6-61h-1	a	b	b	h	h	h	
1.6-61h-10	a	b	b	b	b	b	1.6-61
1.6-61h-18	a	h	b	b	b	a	
1.6-61h-35	a	b	b	b	b	b	1.6-61
1.6-61h-39	a	b	h	h	h	h	
1.0-61h-13	a	a	h	b	b	b	
1.0-61h-18	a	a	a	b	b	b	1:50-61
1.0-61h-28	a	a	h	b	b	b	
1.0-61h-2	h	h	h	h	b	b	
1.0-61h-5	a	a	a	a	b	b	1:60-61
1.0-61h-7	h	h	h	b	b	b	
1.0-61h-6	b	b	h	h	h	h	
1.0-61h-11	h	b	h	h	a	a	
1.0-61h-16	h	b	a	h	a	a	
1.0-61h-21	h	b	b	h	h	h	
1.0-49h-2	h	h	h	a	a	a	
1.0-49h-9	h	h	a	a	a	a	
1.0-49h-15	b	b	b	b	a	a	1:0-50
1.0-49h-19	a	h	h	h	a	a	
1.0-49h-20	a	a	h	b	a	a	

H = heterozygous, a= homozygous for the recurrent parent (RV) and b= homozygous for the donor parent (FB). The names on the left indicate the number the seed was given after germination. The names on the right are the names given after analysis according to their introgressions.

Linkage group 6

A similar approach was taken for selecting candidates to improve the NIL definition in linkage group six.

To obtain the heterozygous regions of LG6 in homozygosis, the heterozygous lines Fb6:6h-38 and Fb6:6h-101h were chosen.

Twenty four progenies of Fb6:6h-38 and nineteen of Fb6:6h-101h lines were germinated and screened using SSR markers located in LG6 (Table1-1). Four and six plants respectively were chosen to go through another round of self-pollination due to containing smaller heterozygous regions in LG6 (Table1-4).

Table 1-4. Selected individuals from the selfing progenies ofFb6:6h-38 and Fb6:6h-101h.

primer	ARSFL007	CFaCT107	EMFn228	FvH4123	EMFn117	EMFn017	CFVCT017	EMFv160AD	EMFv010	EMFv160BC	name
LG	6	6	6	6	6	6	6	6	6	6	
Position (cM)	0	5	11.4	24	30.1	38.8	54.3	70.9	83.6	101	
6.6h-38	a	h	h	h	h	a	a	a	a	a	
5	a	a	a	a	h	-	a	a	a	a	6:30h-38
7	a	h	h	h	a	a	a	a	a	a	6:6h-24h
10	a	h	h	h	h	a	a	a	a	a	6:6h-30h
18	a	a	h	a	a	a	a	a	a	a	6:11h-11h
6.6h-101h	a	h	h	h	h	h	h	h	h	h	
9	a	a	a	h	h	h	h	h	h	h	6:24-202h
5	a	h	h	h	h	h	h	h	a	a	6:6-71h
10	a	h	h	h	h	a	a	a	a	a	6:6-30h
18	a	a	a	a	a	a	a	h	h	h	6:71h-101h
21	a	a	a	a	a	a	a	b	h	h	6:71b-101h
23	a	a	a	a	a	a	a	a	h	h	6:84-101h

The yellow bar represents the parent plant and its genotype.

h = heterozygous, a= homozygous for the recurrent parent (RV) and b= homozygous for the donor parent (FB). The names on the left indicate the number the seed was given after germination.

The names on the right are the names given after analysis according to their introgressions. The ‘h’ in the name designates that the introgression is still in heterozygosity, ‘b’ means the introgression is homozygous for FB at that position.

In 2017, the self-pollinated seeds of selected lines from 2016 were germinated and screened using SSR markers as previously described (Table 1-1). Based on the genotypes, seventeen out of 88 plants were selected. Out of the seventeen selected, different introgressions were observed, again providing further definition to the NIL collection. What’s more, three new lines Fb6:84-101, Fb6:24-30 and Fb6:30-30 all harboured homozygous introgressions of LG6 and could be considered as new NILs (Table1-5A). These new NILs reduced the region which did not cover regionsof the donor parent. In 2018, eighty nine plants were obtained from the self-pollinated individuals selected in the year 2017 and screened as described before. Three plants were selected to produce more progenies because they had the potential to add more

definition to the NIL collection (Table1-5B).

Table1-5. Genotypes of selected individuals in 2017 (A) and 2018 (B).

A

primer	ARSFL007	CFaCT107	EMFn228	FvH4123	EMFn117	EMFn017	CFVCT017	EMFv160AD	EMFv010	EMFv160BC	name
LG	6	6	6	6	6	6	6	6	6	6	
Position (cM)	0	5	11.4	24	30.1	38.8	54.3	70.9	83.6	101	
6.6h-38-2	a	a	h	a	a	a	a	a	a	a	6:11h-11h
6.6h-38-11	a	a	a	h	a	a	a	a	a	a	6:24h-24h
6.6h-38-12	a	h	h	a	a	a	a	a	a	a	6:6h-11h
6.6h-38-13	a	a	h	h	a	a	a	a	a	a	6:11-24h
6.6h-101h-8	a	h	h	b	b	h	-	h	h	h	6:24b-101h
6.6h-101h-11	a	a	a	a	h	h	h	h	h	h	6:30-101h
6.6h-101h-18	a	a	a	h	h	h	h	h	h	b	6:24-101b
6.71h-101h-1	a	a	a	a	a	a	a	a	b	b	**6:84-101**
6.6h-24h-1	a	a	h	a	a	a	a	a	a	a	6:11h-11h
6.6h-24h-2	a	h	h	h	a	a	a	a	a	a	6:6-24h
6.6h-24h-17	a	h	h	a	a	a	a	a	a	a	6:6-11h
6.6h-38-4	a	a	a	b	b	a	a	a	a	a	**6:24-30**
6.6h-30h-7	a	a	a	a	a	h	a	a	a	a	6:38h-38h
6.6h-30h-17	a	a	a	a	b	h	a	a	a	a	6:30-38h
6.6h-30h-18	a	a	a	a	b	a	a	a	a	a	**6:30-30**
6.6h-30h-27	a	h	b	b	a	a	a	a	a	a	6:6h-24

B

primer	ARSFL007	CFaCT107	EMFn228	FvH4123	EMFn117	EMFn017	CFVCT017	EMFv160AD	EMFv010	EMFv160BC	name
LG	6	6	6	6	6	6	6	6	6	6	
Position (cM)	0	5	11.4	24	30.1	38.8	54.3	70.9	83.6	101	
6:6h-24-1	a	h	b	a	a	a	a	a	a	a	6:6h-11
6:24-101-2	a	a	b	a	a	a	a	a	h	h	6:11b-84h
6:6h-24-11	h	h	b	a	a	a	a	a	a	a	6:0-11h

H = heterozygous, a= homozygous for the recurrent parent (RV) and b= homozygous for the donor parent (FB). The names on the left indicate the number the seed was given after germination.

The names on the right are the names given after analysis according to their introgressions; in bold we show homozygous lines. The 'h' in the name designates that the introgression is still in heterozygosity, 'b' means the introgression is homozygous for FB at that position.

Linkage group 5 and 7

Since many QTL related with fruit flavour traits were located inLG5:0-35cM and LG7:0-10cM, these lines were backcrossed with the current parent (RV) to narrow down these regions to obtain more precise QTL regions. The lines Fb5:0-35 and Fb7:0-10were backcrossed with the recurrent parent RV in 2015 and 5:0-35$BC_1$and 7:0-10BC_1 seeds were obtained. In 2016, the progenies of 5:0-35$BC_1$and 7:0-10BC_1 which contained heterozygous introgressions in LG5 region 0-35cM and LG7 region 0-10cM were selected to be self-pollinated and produce seeds. The seeds were germinated in 2017 and the plants were screened using SSR markers located in LG5 and LG7 (Table 1-1).

For LG5, twelve plants were chosen from eighty nine progenies of 5:0-35BC_1 since they had potential to produce the NILs desired. Among them, one line Fb5:11-35 was obtained with all introgressions in homozygosis and considered as new sub-NIL (Table 1-6). The other eleven lines were self-pollinated to obtain corresponding regions in homozygosis. And then, in total one hundred plants were germinated and also genotyped by SSR markers in 2018. Two new markers (CFV-3072 and CEL2) (Urrutia *et al.* 2015) in LG5 were used to get more precise genotypes and to select individuals. Three new lines (Fb5:0-4, Fb5:3-4 and Fb5:28-35) (Table1-7) were selected with small introgressions in homozygosis and to produce offspring for phenotype identification.

Table1-6. Selected individuals from the selfing progenies of 5:0-35BC1.

primer	EMFvi108	CFV3132	EMFn110	FvH4095	UDF009	EMFv024	name
LG	5	5	5	5	5	5	
Position(cM)	3.5	10.8	27.7	34.95	41.2	56.2	
5:0-35BC1	h	h	h	h	h	h	
5.0-35BC1-6-6	h	h	b	b	a	a	5:0h-35
5.0-35BC1-6-8	a	h	h	h	a	a	5:11h-35h
5.0-35BC1-6-10	b	b	h	h	a	a	5:0-35h
5.0-35BC1-6-17	a	h	b	h	a	a	5:11h-28-35h
5.0-35BC1-6-25	b	b	b	b	a	a	**5:0-35**
5.0-35BC1-6-27	b	b	b	h	a	a	5:0-28h
5.0-35BC1-2-4	a	h	b	b	a	a	5:11h-35
5.0-35BC1-2-10	a	a	h	h	a	a	5:28-35h
5.0-35BC1-2-13	b	h	a	a	a	a	5:0-11h
5.0-35BC1-4-6	b	h	a	a	a	a	5:0-11h
5.0-35BC1-4-12	a	b	b	b	a	a	**5:11-35**
5.0-35BC1-4-20	b	h	h	a	a	a	5:3-28h

The yellow bar represents the parent plant and its genotype.
h = heterozygous, a= homozygous for the recurrent parent (RV) and b= homozygous for the donor parent (FB). The names on the left indicate the number the seed was given after germination.
The names on the right are the names given after analysis according to their introgressions, homozygous lines are in **bold** letter.
The 'h' in the name designates that the introgression is still in heterozygosity, 'b' means the introgression is homozygous for FB at that position.

Table 1-7. New NILs harbouring introgressions in LG5 produced in 2018.

primer	CFV-3072	EMFvi108	CFV-3132	CEL2	EMFn110	FVH4095	name
LG	5	5	5	5	5	5	
Position(cM)	0	3.5	10.8	20.2	27.7	34.95	
5:0-28h-9	b	b	a	a	a	a	5:0-4
5:0-28h-11	b	b	b	a	a	a	5:0-11
5:3-28h-2	b	b	a	a	a	a	5:0-4
5:3-28h-3	b	b	a	a	a	a	5:0-4
5:3-28h-5	b	b	a	a	a	a	5:0-4
5:3-28h-7	b	b	a	a	a	a	5:0-4
5:3-28h-8	a	b	a	a	a	a	5:3-4
5:3-28h-9	b	b	a	a	a	a	5:0-4
5:11h-35h-3	a	a	b	b	b	b	5:11-35
5:28-35h-15	a	a	a	a	b	b	5:28-35
5:28-35h-22	a	a	a	a	b	b	5:28-35

a= homozygous for the recurrent parent (RV) and b= homozygous for the donor parent (FB).

The names on the left indicate the number the seed was given after germination.

The names on the right are the names given after analysis according to their introgressions.

Table 1-8. Genotypes of selected individuals in 2017 (A) and 2018 (B).

A

Primer	CFV3096	FvH4152	EMFv021	EMFn201	CFV3896	EMFv023	name
LG	7	7	7	7	7	7	
Position(cM)	0	2.5	8.2	9.5	10.8	42.8	
7.0-10BC1-5-1	b	h	h	h	h	a	7:0b-10h
7.0-10BC1-5-2	b	h	h	h	a	a	7:0b-9h
7.0-10BC1-5-3	a	a	a	a	h	a	7:10-10h
7.0-10BC1-5-4	a	h	h	h	a	a	7:3-9h
7.0-10BC1-5-12	b	h	b	b	b	a	7:3-3h
7.0-10BC1-4-17	h	h	b	b	b	a	7:0-3h
7.0-10BC1-4-27	h	h	h	h	a	a	7:0-9h

B

7:0b-10h-3	b	h	h	h	h	a	7:0b-10h
7:0b-9h-7	b	h	h	a	a	a	7:0b-8h
7:0-9h-13	h	h	h	b	a	a	7:0h-9b

h = heterozygous, a= homozygous for the recurrent parent (RV) and b= homozygous for the donor parent (FB). The names on the left indicate the number the seed was given after germination.

The names on the right are the names given after analysis according to their introgressions.

The ‘h’ in the name designates that the introgression is still in heterozygosity, ‘b’ means the introgression is homozygous for FB at that position.

Finally, four new NILsFb5:0-4, Fb5:3-4, Fb5:11-35 and Fb5:28-35 were obtained harboring smaller introgressions from FB in LG5, and the progress to produce new sub-NILs containing smaller introgressions in LG5 was completed.

A similar approach was taken in selecting candidates to improve the NIL definition in linkage group 7. Seven and three individuals were selected from sixty one and forty seven progenies of 7:0-10BC_1, respectively, based on the genotypes in 2017 and 2018 (Table 1-8). The selected plants will continue to provide further definition to the NIL collection.

Achievement of improved NIL collection

Finally, the strawberry NIL collection consisted of 49 homozygous lines (Figure1-2), each linkage group was represented by an average of seven NILs with overlapping introgressions，and only LG6 had two small regions (5cM-24cM and 39cM-71cM) that were not covered by *F. bucharica* homozygous introgressions.

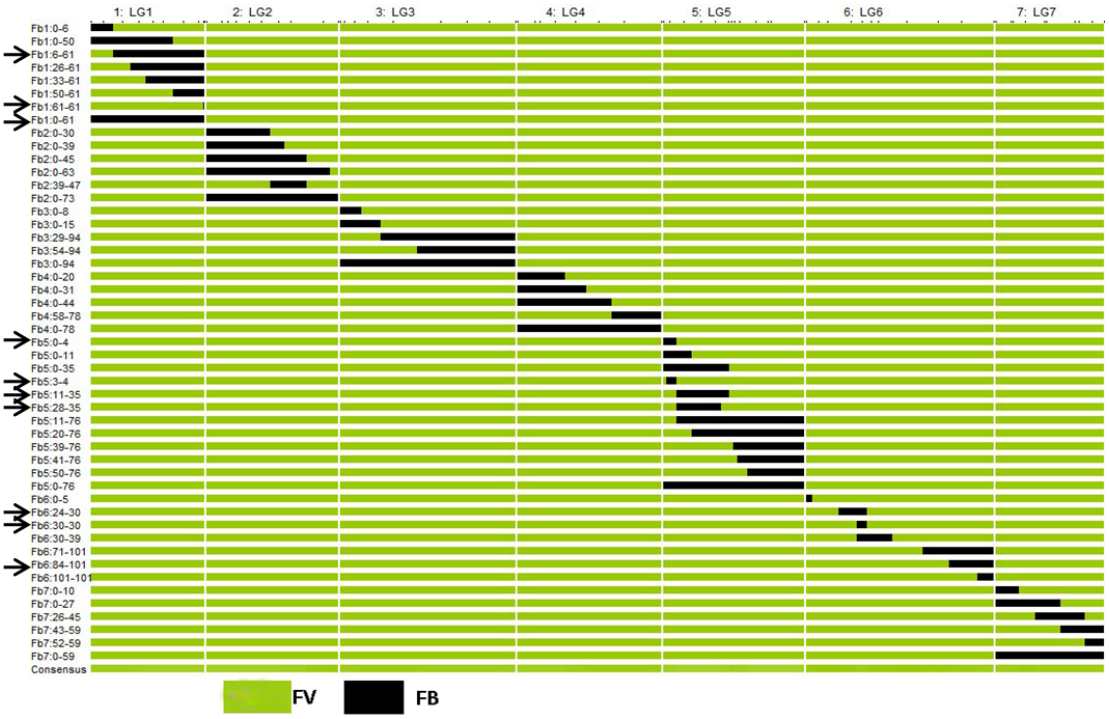

Figure 1-2.Graphical genotypes of the *F. vesca* collection of 49 NILs. *F. bucharica* homozygous introgressions are shown in black. The *F. vesca* genetic background is shown in green. The NIL names are indicated on the left. The first number indicates the LG carrying the introgression. The following two numbers, separated with a hyphen, indicate the marker position at the start and end point of the introgression in centiMorgans, respectively. New lines are marked by black arrow.

As the bins were defined by two consecutive breakpoints with no recombination between them in the NIL collection (Urrutia *et al.* 2015), six new bins were obtained in this research, two in LG1 and other four in

LG5, therefore this population had 43 bins (4-10 per linkage group) with an average size of 11.6 cM per bin. Apart from this, we also got some bins with much smaller regions in LG5 (1cM) (Table 1-9).

Table 1-9. Comparison of NIL collection characteristics from 2015 and 2018.

	2015	2018
Nb. NILs	39	49
Nb. Heterozygous NILs	2	0
Nb. NILs covering whole LG	5	6
Genome covered in homozygosity (cM)	479.3(88.5%)	495(94.8%)
Nb. Average NILs per LG	6	7
Nb. of BINs	37	43
Average BIN size (cM)	14.2	11.6
Smallest BIN (cM)	3.2	1
Largest BIN (cM)	65.7	65.7

Note: 2015 version from (Urrutia et al. 2015)

Fruit pH and citric acid (CA) content

The acidity content is one of the most important traits for flavor perception. This trait has been evaluated in NIL collection for three years. In 2016, the mean pH value for recurrent parent RV was 3.59 and the mean pH value of each NILs varied from 3.48 to 4.1, with the line Fb1:33-61 presenting the lowest value while line Fb7:0-59 was with the highest value. Among all the NILs, 10 NILS showed significant differences compared to that of RV (Table 1-10). In 2017, the average pH value for RV was 3.81 and the mean pH value of each NIL varied from 3.50 to 3.96, with the line Fb4:58-78 presenting the lowest value and line Fb4:0-31 with the highest value. Eleven lines showed significant differences compared to RV (Table 1-10). In 2018, the pH values for RV and the NIL collection were similar to that of the year 2017, and the Dunnett´s test result showed that 13 NILs presented significant differences compared to RV (Table 1-10).

Taking three years' data together, the pH values showed high consistency in different years (Figure 1-3). The maximum correlation value (0.77) was obtained for pH value between 2017 and 2018, and citric acid data (0.56) between 2016 and 2017. Inverted correlation has been found between pH and citric acid content for each year ranking between 0.88 (2016) and 0.44 (2017).

Looking at pH values for each line separately, only one line, Fb4:0-31, shows a significantly higher pH value than RV in all three years. Seven lines were detected harboring different pH values from RV in harvests 2017 and 2018, and only one line in harvest 2016 and 2017. The lines harboring introgression of linkage group 1 present a significant difference with RV in two years (2017 and 2018). In 2016 only one

linkage group 1 line, Fb1:31-61, was analyzed and it also showed a lower pH than RV (Figure 1-4), therefore we considered that a QTL for decreased pH is located in LG1:50-61cM. A similar case also appeared in the line Fb2:0-45, showing a significant difference with RV in two years and a low average pH value with high variation in 2016 (Figure 1-4), so it contains the same QTL also. Overall, by comparing lines harboring the same region, one stable QTL for increased pH value could be mapped in LG4 in the interval 20 - 31 cM. Two QTL for decreased pH value were found based on two years´ measurements in the interval 50 -61 cM of LG1 and 39- 45 cM of LG2.

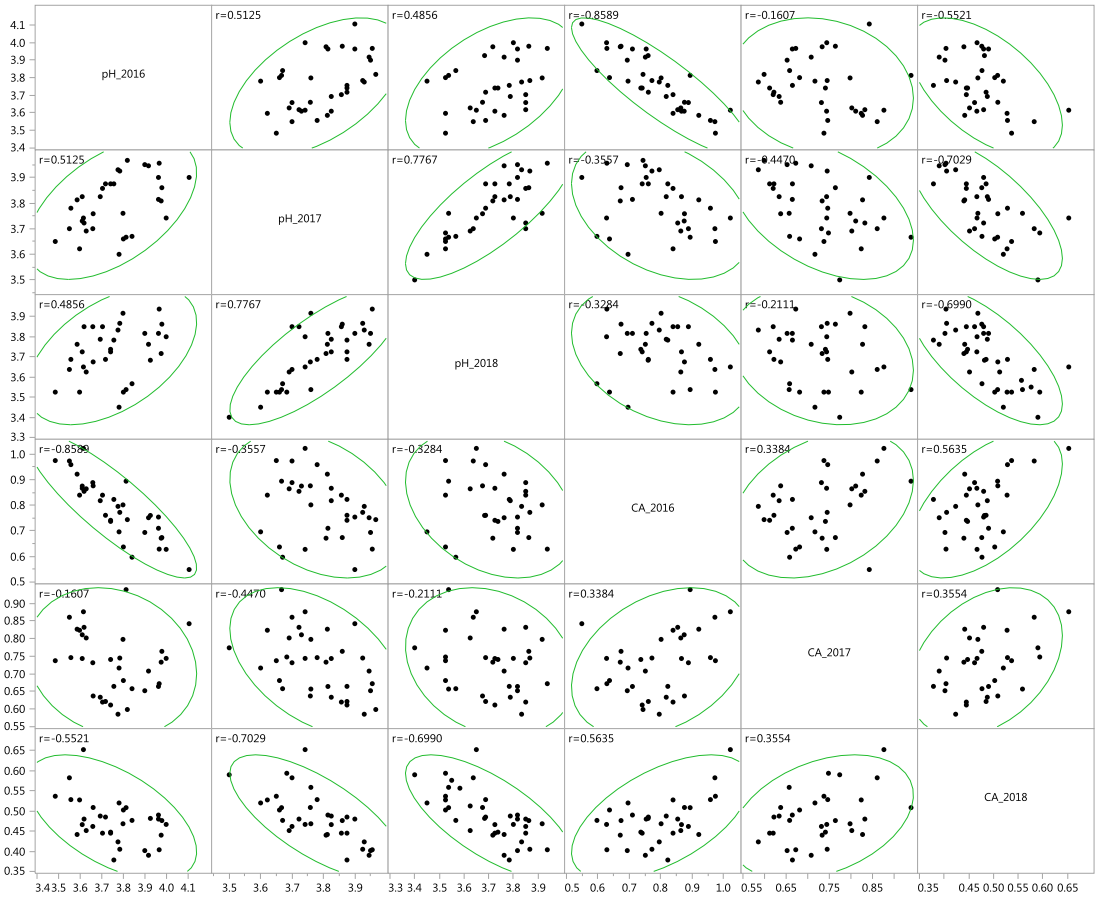

Figure 1-3. Correlation analysis of pH and citric acid (CAg/100ml) content in the years 2016, 2017 and 2018. Density ellipses α=0.95. r is correlation coefficients.

Citric acid titration is a quantification of fruit acidity content. This trait has been also evaluated in NIL collection for three years. The citric acid values in 2016 and 2018 showed a relatively high consistency but not in 2017. Also in 2016 and 2018, the citric acid values presented a negative correlation with pH values in the same year especially the citric acid values in 2018 showed high correlation with pH in all three years. The citric acid values in 2017 did not show significant correlation with other measurements (Figure 1-3).

The citric acid values measured in 2016 and 2017 showed a similar range of means in the NIL collection (0.60-1.02 g/100ml and 0.59-0.93 g/100ml, respectively); also in the recurrent parent the citric acid measurements were similar (0.92 and 0.83 g/100ml in 2016 and 2017, respectively). In 2018, both RV and the NIL collection showed a low range of means for citric acid values (0.44 and 0.38-0.59 g/100ml respectively). The result of Dunnett's test (Table 1-10) concludes that 12, 15 and 7 lines registered significant differences compared to RV in 2016, 2017 and 2018, respectively. Of these lines, six lines were significantly different from RV in both 2016 and 2017 and two lines in both 2017 and 2018. The lines Fb2:0-63, Fb2:0-73 and Fb2:39-47, all sharing an introgression in the region LG2:39-47cM, showed decreased acidity, therefore a QTL for decreased acidity in two years could be mapped in LG2:39-45cM. The line Fb4:0-30 presented a significant difference from RV in 2016 and 2017, and it also presented decreased acidity in 2018, although not significant (Figure 1-5). Hence, a QTL for decreased acidity could be mapped in the region between 20 and 30cM of LG4. The lines Fb5:11-76 and Fb5:41-76 showed a significant decreased acidity value as compared to RV and shared an introgression between 50 and 76 cM of LG5, but other lines which contained this introgression presented nearly no differences from RV. Therefore, two QTLs for decreased acidity were found in LG2 cM 39-45 and LG4 cM 20-31as based on measurements in2016 and 2017.

The results based on all three years' analysis proved the existence of one QTL that increase pH and decrease citric acid content on LG4 cM 20-31.

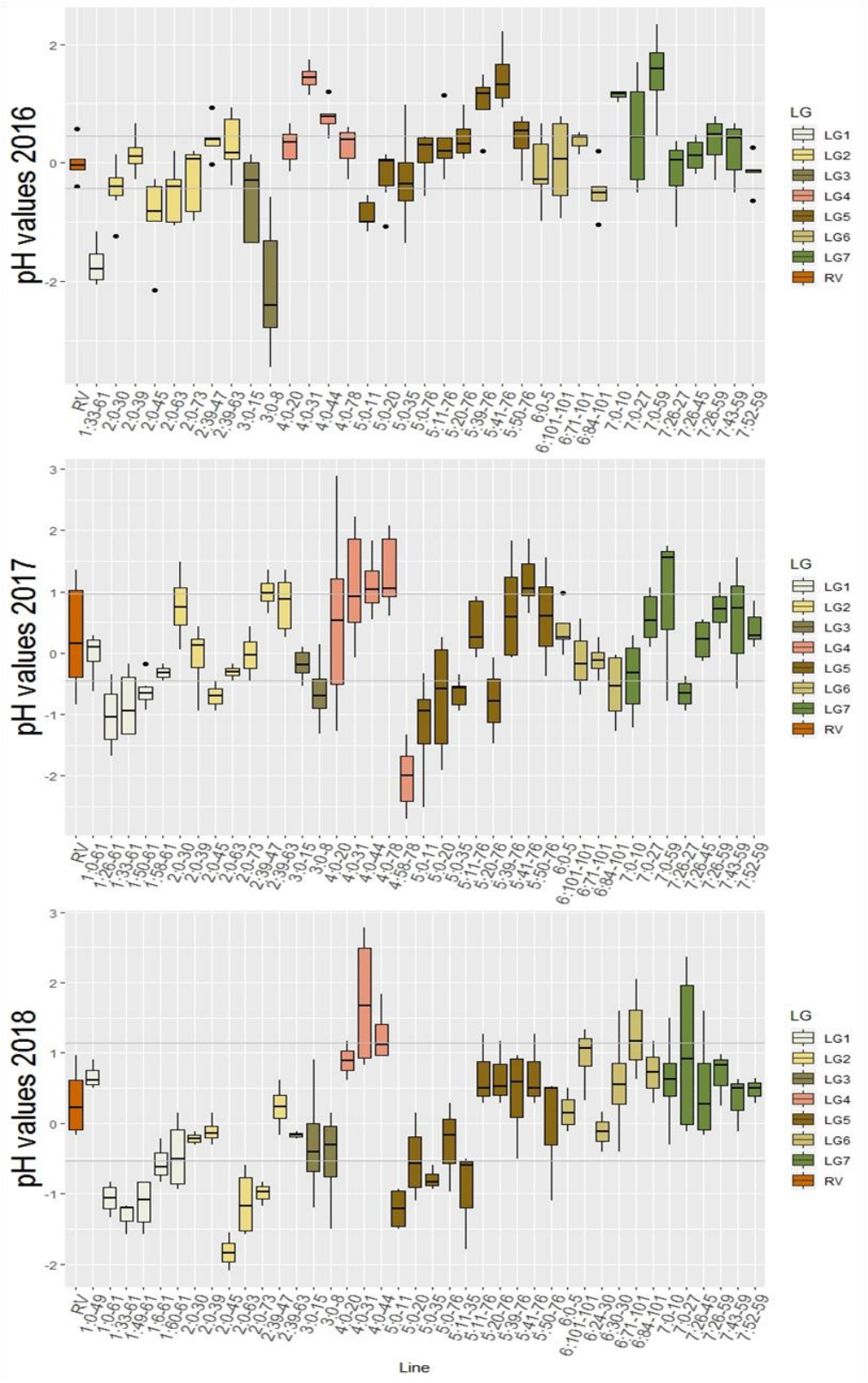

Figure 1-4. Boxplot representing pH values of RV and NILs in harvests in 2016, 2017 and 2018. The values shown as scaled for harvest weeks.

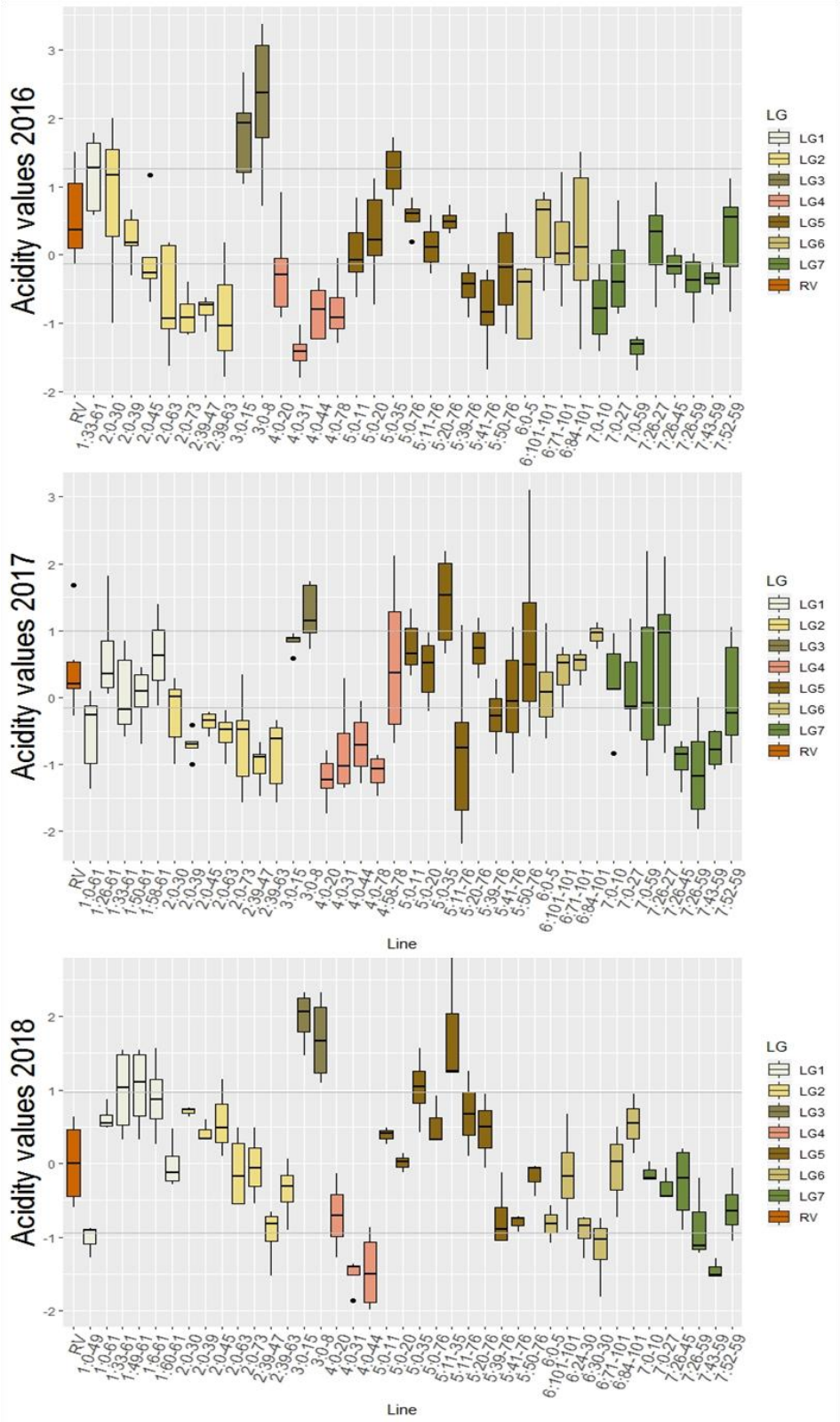

Figure 1-5. Boxplot representing acidity values of RV and NILs in harvests in 2016, 2017 and 2018. The values shown as scaled for harvest weeks.

Table 1-10. Dunnett's test results of fruit pH and citric acid content (CA g/100ml) in RV and NIL collection in the years 2016, 2017 and 2018.

	pH_2016		pH_2017		pH_2018		CA_2016		CA_2017		CA_2018	
Line	**Mean**	**p-Value**	**Mean**	**p-Value**	**Mean**	**p-Value**	**Mean**	**p-Value**	**Mean**	**p-Value**	**Mean**	**p-Value**
RV	3.59	1	3.81	1	3.76	1	0.92	1	0.83	1	0.44	1
1:0-61	-	-	3.76	0.9974	3.54	<.0001*	-	-	0.66	**0.0022***	0.56	**0.0175***
1.33-61	3.48	0.9992	3.65	**0.0029***	3.53	**<.0001***	0.97	1	0.74	0.4175	0.58	0.0027*
1:50-61	-	-	3.68	**0.048***	3.53	**<.0001***	-	-	0.75	0.7078	0.59	0.0004*
2.0-30	3.56	1	3.78	1	3.69	0.8641	0.96	1	0.75	0.7724	0.53	0.2702
2.0-39	3.66	1	3.76	0.9905	3.68	0.6556	0.88	1	0.64	0.0002*	0.51	0.5488
2.0-45	3.78	0.5514	3.60	**<.0001***	3.45	**<.0001***	0.70	0.0069*	0.72	0.2262	0.52	0.4118
2.0-63	3.80	0.3889	3.66	0.0545	3.53	<.0001*	0.64	**0.0001***	0.68	**0.0246***	0.50	0.6984
2.0-73	3.84	0.1605	3.67	0.051	3.57	0.0029*	0.60	**<.0001***	0.66	**0.0024***	0.48	0.9998
2.39-47	3.74	0.815	3.88	0.9467	3.73	1	0.74	**0.0475***	0.61	**<.0001***	0.45	1
2.39-63	3.72	0.9319	3.88	0.9467	3.69	0.8641	0.76	0.0896	0.62	<.0001*	0.49	0.9812
3.0-15	3.62	1	3.74	0.8413	3.65	0.2578	1.02	0.7932	0.88	0.9966	0.65	<.0001*
3.0-8	3.55	1	3.70	0.1729	3.64	0.1389	0.97	1	0.86	1	0.58	0.0015*
4.0-20	3.70	0.9875	3.86	0.9991	3.85	0.9958	0.84	0.9631	0.62	<.0001*	0.45	1
4.0-31	3.97	**0.001***	3.96	**0.0416***	3.94	**0.0048***	0.63	**<.0001***	0.67	**0.0125***	0.40	0.9971
4.0-44	3.78	0.4373	3.93	0.3046	3.87	0.4943	0.77	0.1941	0.75	0.8668	0.41	0.9995
4.0-78	3.82	0.2021	3.97	0.0188*	-	-	0.74	0.0544	0.60	<.0001*	-	-
4:58-78	-	-	3.50	**<.0001***	3.40	**<.0001***	-	-	0.77	0.9936	0.59	0.0248*
5.0-11	3.60	1	3.62	**<.0001***	3.53	**<.0001***	0.84	0.9446	0.82	1	0.53	0.2025
5.0-20	3.63	1	3.69	0.1	3.63	0.0683	0.86	0.9998	0.80	1	0.45	1
5.0-35	3.81	0.4072	3.67	**0.0129***	3.54	**<.0001***	0.89	1	0.93	**0.0323***	0.56	**0.0323***
5.0-76	3.93	0.0256*	-	-	3.68	0.8865	0.76	0.245	-	-	0.48	0.9977
5.11-76	3.96	0.0073*	3.81	1	3.82	0.9992	0.71	**0.0324***	0.66	**0.0174***	0.49	0.9692
5.20-76	3.96	0.0073*	-	-	3.82	0.9992	0.75	0.1981	-	-	0.48	0.9989
5.39-76	3.92	0.0115*	3.90	0.6077	3.76	1	0.75	0.0943	0.71	0.115	0.39	0.939
5.41-76	3.90	**0.0216***	3.95	**0.0464***	3.82	0.9992	0.69	**0.0041***	0.65	**0.0023***	0.40	0.9977
5.50-76	3.98	0.0047*	3.95	0.1729	3.72	1	0.67	0.0042*	0.73	0.9583	0.44	1
6.0-5	3.74	0.815	3.81	1	3.74	1	0.74	0.0379*	0.74	0.522	0.45	1
6.101-101	3.66	1	3.70	0.4692	3.85	0.6556	0.89	1	0.73	0.6633	0.46	1
6.71-101	3.80	0.3595	3.76	0.9974	3.92	0.0494*	0.80	0.558	0.80	1	0.47	1
6.84-101	3.62	1	3.72	0.6108	3.85	0.9084	0.85	0.9954	0.83	1	0.48	0.9999
7.0-10	4.00	0.0004*	3.74	0.8413	3.80	1	0.63	<.0001*	0.74	0.6285	0.47	1
7.0-27	3.98	0.0009*	3.86	0.9995	3.86	0.4343	0.67	0.0011*	0.76	0.9468	0.48	0.9994
7.0-59	4.10	<.0001*	3.90	0.967	-	-	0.55	<.0001*	0.84	1	-	-
7.26-45	3.69	0.9941	3.83	1	3.79	1	0.82	0.6926	0.63	0.0008*	0.49	0.9642
7.26-59	3.78	0.5092	3.93	0.1571	3.83	0.9603	0.80	0.4384	0.59	<.0001*	0.42	1
7.43-59	3.76	0.6851	3.88	0.9467	3.78	1	0.82	0.8173	0.66	0.003*	0.38	0.7359
7.52-59	3.61	1	3.83	1	3.73	1	0.87	0.9997	0.74	0.5769	0.47	1

"*" presents a significant p-value($p<0.05$),**bold** p-values represent lines with2 or 3 years significance compared toRV

°Brix content in NIL collection

The °Brix as a measure of Solid Soluble Content (SSC) is associated with sweetness related to flavor perception. This trait has been evaluated by refractometer in NIL collection for three years. The °Brix value showed high correlation between three years ranking between 0.6 to 0.7 (Figure 1-6). The °Brix

value measured in 2016 and 2017 showed similar means in collection (7.11 and 7.37 respectively); and also similar values in the recurrent parent (7.72 and 7.86 respectively). However, in 2018, both RV and the means of NIL collection showed low °Brix values (5.95 and 6.30 respectively) compared to those in the years 2016 and 2017. The result of Dunnett's test (Table 1-11) concluded that seven, two and four lines registered significant differences compared to RV in 2016, 2017 and 2018, respectively. Among these lines, only two lines showed significant differences compared to RV in two harvest years. The lines harboring introgression region from FB in linkage group 1 all present an increased °Brix value in both 2016 and 2018. In 2017, these lines did not show significant difference with RV but °Brix is higher than RV and most of other lines in NIL collection (Figure1-7). Therefore, considering the same region is included in all these lines, a QTL for increased soluble solid content were mapped in LG1:50-61cM. In contrast to these, the line Fb5:11-76 showed decreased °Brix values in both 2016 and 2017, but the line Fb5:0-76, also harboring LG5:50-61cM, was not obviously different from RV. Therefore, taking together our three years' °Brix measurements for the entire NIL collection, only one QTL was concluded responsible for increased °Brix, which is LG1:50-61cM.

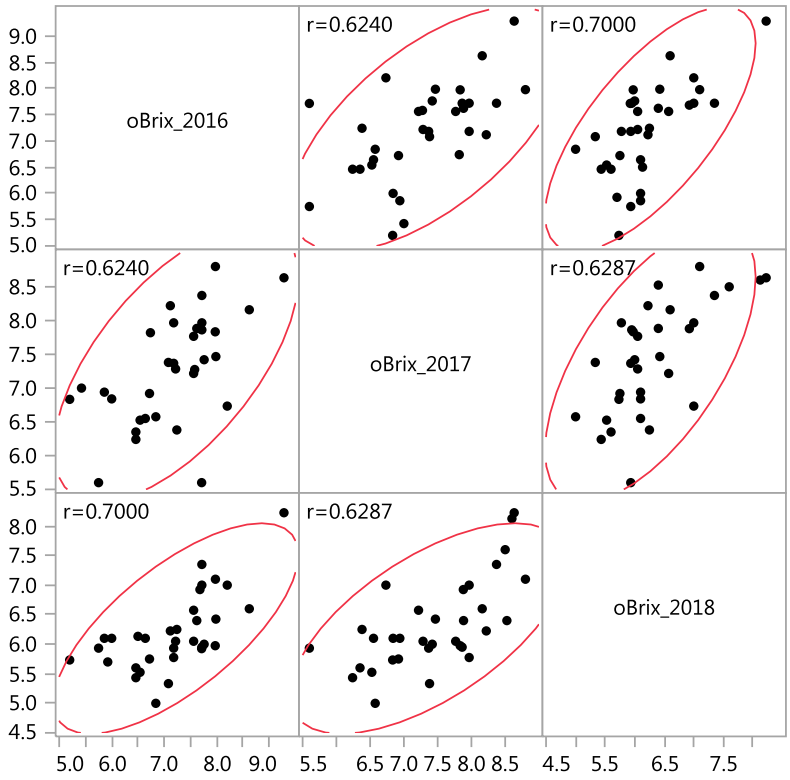

Figure 1-6. Correlation analysis of oBrix in the years 2016, 2017 and 2018. Density ellipses α=0.95. r is correlation coefficients.

Table 1-11. Dunnett's test results of oBrix in fruits of RV and NIL collection in the years 2016, 2017 and 2018.

	°Brix_2016		°Brix_2017		°Brix_2018	
Line	**Mean**	**p-Value**	**Mean**	**p-Value**	**Mean**	**p-Value**
RV	7.72	1	7.86	1	5.95	1
1:0-61	-	-	8.60	0.9947	8.13	<.0001*
1:33-61	9.28	**0.0346***	8.63	0.9803	8.23	**<.0001***
1:50-61	-	-	8.50	0.9995	7.60	0.0015*
2:0-30	7.71	1	5.60	0.013*	5.93	1
2.0-39	7.24	0.9997	6.38	0.2479	6.25	1
2.0-45	6.00	0.013*	6.84	0.8289	6.10	1
2.0-63	6.46	0.179	6.35	0.3218	5.60	1
2.0-73	6.64	0.3784	6.55	0.5582	6.10	1
2.39-47	8.62	0.6546	8.16	1	6.60	0.8285
2.39-63	7.98	1	7.47	1	6.43	0.9927
3.0-15	7.56	1	7.22	0.9985	6.58	0.8696
3.0-8	7.68	1	7.88	1	6.93	0.2362
4.0-20	7.56	1	7.77	1	6.05	1
4.0-31	6.54	0.2549	6.53	0.5256	5.53	0.9987
4.0-44	6.84	0.6995	6.58	0.5912	5.00	0.269
4.0-78	7.58	1	7.28	1	-	-
4:58-78	-	-	8.53	0.9997	6.40	0.9999
5.0-11	7.72	1	8.37	1	7.35	0.0133*
5.0-20	7.72	1	7.97	1	7.00	0.1556
5.0-35	7.98	1	8.80	0.8651	7.10	0.0837
5.0-76	6.50	0.2948	-	-	6.13	1
5.11-76	5.75	**0.0053***	5.60	**0.0052***	5.93	1
5.20-76	5.93	0.0165*	-	-	5.70	1
5.39-76	6.72	0.4989	6.92	0.9093	5.75	1
5.41-76	5.20	<.0001*	6.83	0.7487	5.73	1
5.50-76	8.20	0.9999	6.73	0.9021	7.00	0.2398
6.0-5	7.76	1	7.42	1	6.00	1
6.101-101	7.18	0.9975	7.97	1	5.78	1
6.71-101	7.08	0.9755	7.38	1	5.33	0.9387
6.84-101	7.22	0.9982	7.28	0.9998	6.05	1
7.0-10	7.62	1	7.88	1	6.40	0.9967
7.0-27	5.86	0.0049*	6.94	0.925	6.10	1
7.0-59	5.43	0.0005*	7.00	0.9958	-	-
7.26-27	6.74	0.3496	7.82	1	-	-
7.26-45	7.12	0.9785	8.22	1	6.23	1
7.26-59	6.46	0.179	6.24	0.1434	5.43	0.992
7.43-59	7.18	0.9975	7.37	1	5.93	1
7.52-59	7.97	1	7.83	1	5.98	1

"*" presents a significant p-value($p<0.05$), **bold** p-values represent lines with 2 or 3 years' significance compared with RV

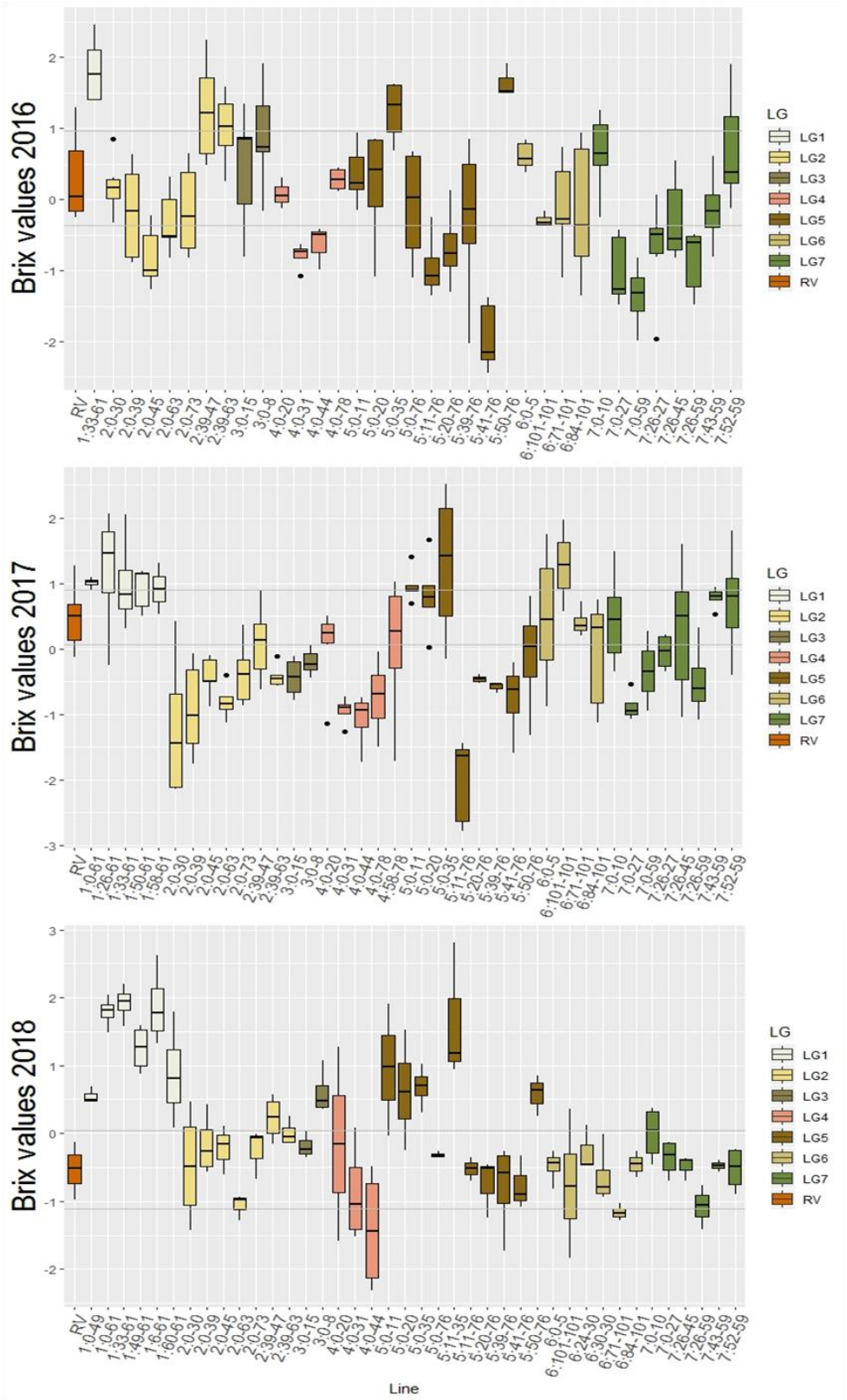

Figure 1-7. Boxplot representing oBrix value of RV and NIL in harvests in 2016, 2017 and 2018. The values shown as scaled for harvest weeks.

Discussion

NIL collections have been applied in many crops to study various aspects of agronomical, nutritional and organoleptic traits. The NILs were developedto study the purple pericarp trait in wheat and mapped a new weak effect gene (Gordeeva *et al.* 2015). A set of 25 NILs was developed for genetic analysis of brown plant hopper resistance genes in rice (Jena *et al.* 2017). An allele was mapped to confer resistance to both southern leaf blight and gray leaf spot in maize (Lennon *et al.* 2017). And some QTLs affecting yield, antioxidant capacity, lycopene content, and fruit-quality traitswere mapped using NILs in tomato (Ashrafi *et al.* 2012; Barrantes *et al.* 2014; Eshed and Zamir 1994; Rousseaux *et al.* 2005). In strawberry, *Fragaria vesca* serves as an ideal model plant for cultivated strawberry (*Fragaria* × *ananassa*) and the Rosaceae family. A *Fragaria vesca* NIL collection was built on *F. vesca* background with *F. bucharica* introgressions (Urrutia *et al.* 2015). However, in this previous NIL collection, the *F. bucharica* genome was not well represented with some regions still missing (Figure I-3). A NIL collection with introgressions covering the entire *F. bucharica* genome with higher resolution (i.e. by smaller introgressions) will be a powerful tool for more accurate analyses of QTL controlling various agronomic traits.

In this work, ten new lines have been made to improve the definition within the collection, covering nearly the entire genome of *F. bucharica* with only 4.9% of *F. bucharica* genome missing. Genome covered in homozygosity has increased by 6.3% covering 94.8%. Six new bins have been added and the average bin size has been reduced by 2.6 cM. Finally, this population consists of 49 NILs and all lines have single homozygous introgressions from *F. bucharica*. The lines added in linkage group 1 supply the gap of 6-26cM which was not covered by *F. bucharica* genome before. More importantly, we acquired the line Fb1:0-61 which harbours the introgression covering whole LG 1. The whole-chromosome coverage line is an essential material for backcrossing with recurrent parent to produce a series of recombinant individuals with diverse coverage of this chromosome. They will be useful tools for QTL analysis and fine mapping genes in LG1. Also, the line Fb1:0-61 is a significant reference for phenotypic comparisons of other lines containing introgressions in LG1. Through backcrossing a previous line Fb5:0-35 with RV, four new sub-NILs harbouring smaller introgressions of *F.bucharica* were obtained. Since the line Fb5:0-35 was mapped with some stable and strong QTL linked to the accumulation of several volatiles of alcohols and esters (Urrutia *et al.* 2017), the new sub-NILs will provide us better opportunities to narrow down the QTL region and making it easier to select candidate genes for related traits.

Fleshy fruit acidity is an important agronomic trait affecting strawberry commodity value since individual persons differ in the acidity sensitivity and tolerance. Fleshy fruit acidity, as measured by titratable acidity and pH, is an important component of fruit organoleptic quality (Esti *et al.* 2002; Harker *et al.* 2002; Bugaud *et al.* 2011). Citric acid is the predominant organic acid in ripe strawberry fruit (Kallio *et al.* 2000). Fruits from the NIL collection were assessed for acidity content and pH. In this study, increased pH value could be observed in the line Fb4:0-31 in all three years and also decreased citric acid content in two years, so one stable QTL for decreased fruit acidity content could be mapped in LG4 interval 20 - 31 cM. Fruits from line Fb4:0-44 and Fb4:0-78 didn't show any difference compared to RV, and in the line Fb4:58-78, a significantly lower pH compared to RV was observed in two years. Therefore, we speculate that maybe a QTL exists in 31-78cM in LG4 for increasing the fruit acidity content. A major effect QTL for pH was detected in LG4 CII in cultivated strawberry (Verma *et al.* 2017) and a QTL for pH has also been previously reported on HG 4 in different cultivated varieties (Lerceteau-Köhler *et al.* 2012; Zorrilla-Fontanesi *et al.* 2012), but it cannot be confirmed that the QTL reported here are in the same regions as found in the earlier studies.

Many lines were detected to have significant differences from RV only in two years. For instance, the lines Fb1:33-61 and Fb1:50-61 both showed a significant p-value in both 2017 and 2018 and shared an introgression between 50 and 61cM of LG1 matching the low ph QTL. But in 2016, the lines containing this introgression didn't present any obvious differences from RV. In a previous study, a QTL for pH was detected for one year in LG1C (Lerceteau-Köhler *et al.* 2012). We suppose that this variation occurs because the processes involved in the metabolism and accumulation of organic acid in mesocarp cells are under both genetic and environmental control (Etienne et al. 2013a), causing the pH to have high variation between years. The literature shows that the plant source: sink ratio, mineral fertilization, water supply, and temperature are the agro-environmental factors that have the highest impact on fruit acidity (Souty *et al.* 1999; Wu *et al.* 2002; Lechaudel *et al.* 2005; Lobit *et al.* 2003; Spironello *et al.* 2004; Alva *et al.* 2006; Wang and Camp 2000; Gautier *et al.* 2005).

The main acids in strawberry are citric and malic acids; glycolic and shikimic acids are also present but in lesser quantities (Woodward, 1972). Earlier studies show that high temperature growing conditions significantly reduced citric acid in strawberry fruit and increased malic acid content (Wang and Camp 2000). This could explain why one QTL for decreased pH in LG1:50-61 cM did not show increased citric acid content, and the lines harboring introgressions 39-45cM in LG2 presented decreased pH and also decreased citric acid, we speculated that the mentioned lines may harbor QTL for increased malic acid content.

Finally, one QTL that decrease fruit organic acid content on LG4 cM 20-31 was mapped and genetic markers developed from this QTL will be a great tool to breeders for selecting cultivars with low acid content.

Strawberry flavor is determined in part by the balance between sugar content and titratable acid content in ripe fruits. The main sugar components to be investigated were fructose, glucose and sucrose in strawberry fruit. Based on Kallio's result (Kallio *et al.* 2000), soluble solid content (^{O}Brix) could be considered a rather good indicator for the evaluation of sugar content of ripe fruits. In this study, soluble solid content (^{O}Brix) of fruit from NIL collection was measured in three years (2016, 2017 and 2018). The average ^{O}Brix values of RV and all NILs in the year 2016 were similar to those in 2017, while a big difference was detected in the year 2018, with low ^{O}Brix values for both RV and NIL collection. In the earlier study by Urruria *et al*. 2015, a stable QTL for decreasing fructose, glucose and total sugar content was mapped in LG2:45-63 cM. In our experiments, the lines harboring introgression of this region showed lower ^{O}Brix values than RV in 2016 and 2017 although the differences were not statistically significant. We consider LG2:45-63 cM as a candidate QTL for decreasing sugar content. Many previous studies present that soluble solid content is influenced by a number of factors, including genetics, climate, water management and other cultivation practices (Shaw 1990; Prange and DeEll 1997; Neuweiler *et al.* 2003; Wold and Opstad 2007; Cao *et al.* 2015). The WMO Statement on the State of the Global Climate in 2018 said Europe experienced exceptional heat and drought through the late spring and summer of 2018. Temperatures were well above average and rainfall well below average from April onwards in much of northern and western Europe, particularly in Spain and Portugal. We assume that the climate has a significant impact during the ripening offruit. However, the ^{O}Brix values of total NILs showed similar trends in three years. For instance, the lines harboring introgressions in LG1 presented a bit higher values than the other lines, although in 2017 they didn't have significant difference with RV. In 2016, only one line in LG1 was analyzed. However, now we have more lines containing introgressions from *F. bucharica* in LG1 and will collect more ^{O}Brix data from them to confirm whether there is a QTL for increased soluble solids content. In previous reports, a moderate effect QTL for soluble solids content was detected on LG6A in cultivated strawberry (Verma *et al.* 2017), and also this QTL was under large environmental influence. Environmental influences may have prevented the detection of some smaller effect QTLs in metabolites as sugars and acids related to flavor perception.

Bibliography

Aharoni A, Giri AP, Verstappen FW, Bertea CM, Sevenier R, Sun Z, Jongsma MA, Schwab W, Bouwmeester HJ (2004) Gain and loss of fruit flavor compounds produced by wild and cultivated strawberry species. The Plant Cell 16 (11):3110-3131

Alger EI, Colle M, Edger PP (2018) Genomic Resources for the Woodland Strawberry (Fragaria vesca). In: The Genomes of Rosaceous Berries and Their Wild Relatives. Springer, pp 25-33

Alpert KB, Tanksley SD (1996) High-resolution mapping and isolation of a yeast artificial chromosome contig containing fw2. 2: a major fruit weight quantitative trait locus in tomato. Proceedings of the National Academy of Sciences 93 (26):15503-15507

Alva AK, Mattos Jr D, Paramasivam S, Patil B, Dou H, Sajwan KS (2006) Potassium management for optimizing citrus production and quality. International Journal of Fruit Science 6 (1):3-43

Ashrafi H, Kinkade MP, Merk HL, Foolad MR (2012) Identification of novel quantitative trait loci for increased lycopene content and other fruit quality traits in a tomato recombinant inbred line population. Molecular Breeding 30 (1):549-567. doi:10.1007/s11032-011-9643-1

Baek K-H, Skinner DZ (2003) Alteration of antioxidant enzyme gene expression during cold acclimation of near-isogenic wheat lines. Plant Science 165 (6):1221-1227

Barrantes W, Fernández-del-Carmen A, López-Casado G, González-Sánchez MÁ, Fernández-Muñoz R, Granell A, Monforte AJ (2014) Highly efficient genomics-assisted development of a library of introgression lines of Solanum pimpinellifolium. Molecular breeding 34 (4):1817-1831

Barrantes W, López-Casado G, García-Martínez S, Alonso A, Rubio F, Ruiz JJ, Fernández-Muñoz R, Granell A, Monforte AJ (2016) Exploring new alleles involved in tomato fruit quality in an introgression line library of Solanum pimpinellifolium. Frontiers in plant science 7:1172

Brouwer D, Clair DS (2004) Fine mapping of three quantitative trait loci for late blight resistance in tomato using near isogenic lines (NILs) and sub-NILs. Theoretical and Applied Genetics 108 (4):628-638

Bugaud C, Deverge E, Daribo MO, Ribeyre F, Fils- Lycaon B, Mbéguié- A- Mbéguié D (2011) Sensory characterisation enabled the first classification of dessert bananas. Journal of the Science of Food and Agriculture 91 (6):992-1000

Cao F, Guan C, Dai H, Li X, Zhang Z (2015) Soluble solids content is positively correlated with phosphorus content in ripening strawberry fruits. Scientia Horticulturae 195:183-187

Chen G, Li H, Zheng Z, Wei Y, Zheng Y, McIntyre C, Zhou M, Liu C (2012) Characterization of a QTL affecting spike morphology on the long arm of chromosome 3H in barley (Hordeum vulgare L.) based on near isogenic lines and a NIL-derived population. Theoretical and applied genetics 125

(7):1385-1392

Doyle JJ, Doyle JL, Brown A, Grace J (1990) Multiple origins of polyploids in the Glycine tabacina complex inferred from chloroplast DNA polymorphism. Proceedings of the National Academy of Sciences 87 (2):714-717

Edger PP, Poorten TJ, VanBuren R, Hardigan MA, Colle M, McKain MR, Smith RD, Teresi SJ, Nelson AD, Wai CM (2019) Origin and evolution of the octoploid strawberry genome. Nature genetics 51 (3):541

Eduardo I, Arús P, Monforte AJ (2005) Development of a genomic library of near isogenic lines (NILs) in melon (Cucumis melo L.) from the exotic accession PI161375. Theoretical and applied genetics 112 (1):139-148

Eshed Y, Zamir D (1994) A genomic library of Lycopersicon pennellii in L. esculentum: a tool for fine mapping of genes. Euphytica 79 (3):175-179

Eshed Y, Zamir D (1995) An introgression line population of Lycopersicon pennellii in the cultivated tomato enables the identification and fine mapping of yield-associated QTL. Genetics 141 (3):1147-1162

Esti M, Cinquanta L, Sinesio F, Moneta E, Di Matteo M (2002) Physicochemical and sensory fruit characteristics of two sweet cherry cultivars after cool storage. Food Chemistry 76 (4):399-405

Etienne A, Génard M, Lobit P, Mbeguié-A-Mbéguié D, Bugaud C (2013) What controls fleshy fruit acidity? A review of malate and citrate accumulation in fruit cells. Journal of experimental botany 64 (6):1451-1469

Fletcher RS, Mullen JL, Yoder S, Bauerle WL, Reuning G, Sen S, Meyer E, Juenger TE, McKay JK (2013) Development of a next-generation NIL library in Arabidopsis thaliana for dissecting complex traits. BMC genomics 14 (1):655

Frary A, Nesbitt TC, Frary A, Grandillo S, Van Der Knaap E, Cong B, Liu J, Meller J, Elber R, Alpert KB (2000) fw2. 2: a quantitative trait locus key to the evolution of tomato fruit size. Science 289 (5476):85-88

Fukuoka S, Yamamoto S-I, Mizobuchi R, Yamanouchi U, Ono K, Kitazawa N, Yasuda N, Fujita Y, Nguyen TTT, Koizumi S (2014) Multiple functional polymorphisms in a single disease resistance gene in rice enhance durable resistance to blast. Scientific Reports 4:4550

Gautier H, Rocci A, Buret M, Grasselly D, Causse M (2005) Fruit load or fruit position alters response to temperature and subsequently cherry tomato quality. Journal of the Science of Food and Agriculture 85 (6):1009-1016

Gordeeva E, Shoeva O, Khlestkina E (2015) Marker-assisted development of bread wheat near-isogenic lines carrying various combinations of purple pericarp (Pp) alleles. Euphytica 203 (2):469-476

Green A (1971) Soft fruits. The biochemistry of fruits and their products 2:375-410

Harker F, Marsh K, Young H, Murray S, Gunson F, Walker S (2002) Sensory interpretation of instrumental measurements 2: sweet and acid taste of apple fruit. Postharvest Biology and Technology 24 (3):241-250

Jena KK, Hechanova SL, Verdeprado H, Prahalada G, Kim S-R (2017) Development of 25 near-isogenic lines (NILs) with ten BPH resistance genes in rice (Oryza sativa L.): production, resistance spectrum, and molecular analysis. Theoretical and applied genetics 130 (11):2345-2360

Jiang Y, Habib A, Zheng Z, Zhou M, Wei Y, Zheng Y-L, Liu C (2019) Development of tightly linked markers and identification of candidate genes for Fusarium crown rot resistance in barley by exploiting a near-isogenic line-derived population. Theoretical and applied genetics 132 (1):217-225

Kader AA (1991) Quality and its maintenance in relation to the postharvest physiology of strawberry. The strawberry into the 21st century Timber Press, Portland, OR:145-152

Kallio H, Hakala M, Pelkkikangas A-M, Lapveteläinen A (2000) Sugars and acids of strawberry varieties. European Food Research and Technology 212 (1):81-85

Khanna A, Sharma V, Ellur RK, Shikari AB, Krishnan SG, Singh U, Prakash G, Sharma T, Rathour R, Variar M (2015) Development and evaluation of near-isogenic lines for major blast resistance gene (s) in Basmati rice. Theoretical and applied genetics 128 (7):1243-1259

Lechaudel M, Joas J, Caro Y, Genard M, Jannoyer M (2005) Leaf: fruit ratio and irrigation supply affect seasonal changes in minerals, organic acids and sugars of mango fruit. Journal of the Science of Food and Agriculture 85 (2):251-260

Lennon JR, Krakowsky M, Goodman M, Flint-Garcia S, Balint-Kurti PJ (2017) Identification of teosinte alleles for resistance to southern leaf blight in near isogenic maize lines. Crop Science 57 (4):1973-1983

Lerceteau-Köhler E, Moing A, Guérin G, Renaud C, Petit A, Rothan C, 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 (6):1059-1077

Lobit P, Génard M, Wu B, Soing P, Habib R (2003) Modelling citrate metabolism in fruits: responses to growth and temperature. Journal of Experimental Botany 54 (392):2489-2501

Mago R, Bariana H, Dundas I, Spielmeyer W, Lawrence G, Pryor A, Ellis J (2005) Development of PCR markers for the selection of wheat stem rust resistance genes Sr24 and Sr26 in diverse wheat germplasm. Theoretical and Applied Genetics 111 (3):496-504

Merchuk-Ovnat L, Barak V, Fahima T, Ordon F, Lidzbarsky GA, Krugman T, Saranga Y (2016) Ancestral QTL alleles from wild emmer wheat improve drought resistance and productivity in modern

wheat cultivars. Frontiers in plant science 7:452

Monforte A, Friedman E, Zamir D, Tanksley S (2001) Comparison of a set of allelic QTL-NILs for chromosome 4 of tomato: deductions about natural variation and implications for germplasm utilization. Theoretical and Applied Genetics 102 (4):572-590

Muehlbauer G, Specht J, Thomas-Compton M, Staswick P, Bernard R (1988) Near-isogenic lines—a potential resource in the integration of conventional and molecular marker linkage maps. Crop Science 28 (5):729-735

Neuweiler R, Bertschinger L, Stamp P, Feil B (2003) The impact of ground cover management on soil nitrogen levels, parameters of vegetative crop development, yield and fruit quality of strawberries. European Journal of Horticultural Science 68 (4):183-191

Perpiñá G, Esteras C, Gibon Y, Monforte AJ, Picó B (2016) A new genomic library of melon introgression lines in a cantaloupe genetic background for dissecting desirable agronomical traits. BMC plant biology 16 (1):154

Prange RK, DeEll JR (1997) Preharvest factors affecting postharvest quality of berry crops. HortScience 32 (5):824-830

Rousseaux MC, Jones CM, Adams D, Chetelat R, Bennett A, Powell A (2005) QTL analysis of fruit antioxidants in tomato using Lycopersicon pennellii introgression lines. Theoretical and Applied Genetics 111 (7):1396-1408

Schuelke M (2000) An economic method for the fluorescent labeling of PCR fragments. Nature biotechnology 18 (2):233

Shaw DV (1990) Response to selection and associated changes in genetic variance for soluble solids and titratable acids contents in strawberries. Journal of the American Society for Horticultural Science 115 (5):839-843

Shulaev V, Sargent DJ, Crowhurst RN, Mockler TC, Folkerts O, Delcher AL, Jaiswal P, Mockaitis K, Liston A, Mane SP (2011) The genome of woodland strawberry (Fragaria vesca). Nature genetics 43 (2):109

Souty M, Génard M, Reich M, Albagnac G (1999) Effect of assimilate supply on peach fruit maturation and quality. Canadian Journal of Plant Science 79 (2):259-268

Spironello A, Quaggio JA, Teixeira LAJ, Furlani PR, Sigrist JMM (2004) Pineapple yield and fruit quality effected by NPK fertilization in a tropical soil. Revista brasileira de fruticultura 26 (1):155-159

Tanksley S, Grandillo S, Fulton T, Zamir D, Eshed Y, Petiard V, Lopez J, Beck-Bunn T (1996) Advanced backcross QTL analysis in a cross between an elite processing line of tomato and its wild relative L. pimpinellifolium. Theoretical and applied genetics 92 (2):213-224

Telebanco-Yanoria MJ, Koide Y, Fukuta Y, Imbe T, Tsunematsu H, Kato H, Ebron LA, Nguyen TMN, Kobayashi N (2011) A set of near-isogenic lines of Indica-type rice variety CO 39 as differential varieties for blast resistance. Molecular breeding 27 (3):357-373

Urrutia M, Bonet J, Arus P, 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. TAG Theoretical and applied genetics Theoretische und angewandte Genetik 128 (7):1261-1275.

Urrutia M, Rambla JL, Alexiou KG, Granell A, Monfort A (2017) Genetic analysis of the wild strawberry (Fragaria vesca) volatile composition. Plant physiology and biochemistry : PPB 121:99-117.

Verma S, Zurn JD, Salinas N, Mathey MM, Denoyes B, Hancock JF, Finn CE, Bassil NV, Whitaker VM (2017) Clarifying sub-genomic positions of QTLs for flowering habit and fruit quality in US strawberry (*Fragaria× ananassa*) breeding populations using pedigree-based QTL analysis. Horticulture research 4:17062

Wang SY, Camp MJ (2000) Temperatures after bloom affect plant growth and fruit quality of strawberry. Scientia Horticulturae 85 (3):183-199

Wang X, Liu H, Liu G, Mia MS, Siddique KH, Yan G (2019) Phenotypic and genotypic characterization of near-isogenic lines targeting a major 4BL QTL responsible for pre-harvest sprouting in wheat. BMC plant biology 19 (1):1-10

Wold A-B, Opstad N (2007) Fruit quality in strawberry (Fragaria x ananassa Duch. cv. Korona) at three times during the season and with two fertilizer strategies. Journal of applied botany and food quality 81 (1):36-40

Wrolstad R, Shallenberger R (1981) Free sugars and sorbitol in fruits--a complication from the literature. Journal-Association of Official Analytical Chemists 64 (1):91-103

Wu BH, Genard M, Lescourret F, Gomez L, Li SH (2002) Influence of assimilate and water supply on seasonal variation of acids in peach (cv Suncrest). Journal of the Science of Food and Agriculture 82 (15):1829-1836

Young ND, Zamir D, Ganal MW, Tanksley SD (1988) Use of isogenic lines and simultaneous probing to identify DNA markers tightly linked to the tm-2a gene in tomato. Genetics 120 (2):579-585

Zamir D (2001) Improving plant breeding with exotic genetic libraries. Nature reviews genetics 2 (12):983

Zarouri B, Fergany M, Eduardo I, Alvarez Alvarez JM, Picó B, Díaz Bermúdez A (2014) Mapping and Introgression of QTL Involved in Fruit Shape Transgressive Segregation into 'Piel de Sapo'Melon (Cucucumis melo L.).

Zeng Y, Yang S, Cui H, Yang X, Xu L, Du J, Pu X, Li Z, Cheng Z, Huang X (2009) QTLs of cold tolerance-related traits at the booting stage for NIL-RILs in rice revealed by SSR. Genes

&Genomics 31 (2):143-154

Zorrilla-Fontanesi Y, Rambla JL, Cabeza A, Medina JJ, Sanchez-Sevilla JF, Valpuesta V, Botella MA, Granell A, Amaya I (2012) Genetic analysis of strawberry fruit aroma and identification of O-methyltransferase FaOMT as the locus controlling natural variation in mesifurane content. Plant physiology 159 (2):851-870.

Chapter II

Genetic analysis of aroma volatile compounds and the role of a (3Z):(2E)-hexenalisomerase in the development of fruit aroma

Introduction

Over the last 10 years, customers' demand for better-tasting fruits and berries has increased dramatically, and strawberries are not an exception. Strawberry taste is a complex trait formed by an interaction of fruit texture, sugar content, acidity and volatile aromatic compounds. Of these attributes, fruit sugars and volatile composition are the factors most closely associated with overall liking (Schwieterman *et al.* 2014).

Aroma compounds as key contributors to fruit flavor perception rely in a combination of taste, smell, appearance, and texture (Taylor *et al.* 2010). Thus, the volatile composition of strawberry fruits has been extensively studied and more than 350 constituents have been reported (Latrasse 1991). These compounds comprise esters, aldehydes, ketones, furanones, alcohols and terpenoids (Zabetakis and Holden 2015; Jetti *et al.* 2007; Isabelle *et al.* 2004). However, among these compounds only less than 20 compounds can be perceived by humans (Schieberle and Hofmann 1997; Ulrich *et al.* 1997; Ulrich *et al.* 2007). Aroma volatiles are represented in three main categories: fatty acid, amino acid and carbohydrate derivatives (Schwab *et al.* 2008). The degradation of fatty acids through the lipoxygenase pathway or through α -or β - oxidation is the main source of plant volatile compounds including straight-chain alcohols, aldehydes, ketones, acids, esters and lactones. Terpenoids, which are important plant volatiles are synthesized from acetyl-coA and pyruvate by prenyltransferases and terpene synthases. However, there are other sources of volatile compounds such as the degradation products of branched-chain and aromatic amino acids and the carbohydrate-derived compounds. Esters, which have been described as the majority compounds containing more than 130 different types in strawberry (Jetti *et al.* 2007), are the characteristic volatile compounds that define strawberry volatiles. Among them, ethyl butanoate, methyl butanoate, ethyl hexanoate, methyl hexanoate, hexyl acetate and (E)-2-hexenyl acetate have been reported as key aroma compounds for strawberry fruit, providing green and sweet fruity notes (Isabelle *et al.* 2004; Pérez *et al.* 2002; Schieberle and Hofmann 1997). (*Z*)-3-hexenal, (*E*)-2-hexenal and (*Z*)-3-hexen-1-ol which are described as 'green leaf volatile compounds' have been reported providing green leaf fresh aroma that decreases with ripeness (Ulrich *et al.* 1997; Schieberle and Hofmann 1997). Methyl 2-aminobenzoate is a noteworthy cultivar-specific compound conferring the typical "wild strawberry" aroma that is found in woodland strawberry (*F. vesca*) accessions, though very rarely present in commercial varieties. 2,5-dimethyl-4-hydroxy-3(2H)-furanone (furaneol) and 2,5-dimethyl-4-methoxy-3(2H)-furanone (mesifurane) are considered to be the two most important furanones in strawberry and give the characteristic caramel-like, sweet, floral, and fruity aroma (Jetti *et al.* 2007). The terpenoids linalool and nerolidol, constitute the other important group, providing pleasant citrus and flowery sweet notes in some cultivars of strawberry but not in diploid wild strawberries (Loughrin and Kasperbauer 2002). The compound γ-decalactone was

reported as a specific compound conferring peach-like flavor to strawberry fruit (Olbricht *et al.* 2008).

Researchers and breeders alike have struggled to identify genes and/or molecular markers linked to a specific taste attribute to speed up breeding for better-tasting strawberries. These efforts have resulted in the identification of several genes that contribute to the accumulation of specific volatile compounds. For instance, a fatty acid desaturase gene *FaFAD1* has been identified as a causative agent for γ-decalactone (fruity aroma) accumulation in garden strawberry (Chambers *et al.* 2014a). Accumulation of linalool and nerolidol, both of which confer a flowery sweet aroma, has been linked to a mutation in a nerolidol synthase gene (*FaNES1*) of garden strawberry (Aharoni *et al.* 2004). All the mentioned compounds are considered desirable in fresh strawberries. However, genes responsible for accumulation of fresh aroma, or green-leaf volatiles have not been identified in the garden strawberry.

Wild strawberries (*F. vesca*) have a more intense and fruity aroma than garden strawberry mainly due to their higher concentration in esters and terpenoids.The recent study by Urrutia *et al.* (2017) identified several QTL responsible for the accumulation of various volatile compounds in the NIL collection. One of the QTLs with the largest effect was identified at the distal end of linkage group 5 (LG5) for the accumulation of green leaf volatile (GLV) compounds. GLV compounds confer a characteristic green leaf odour, which is necessary for the fresh strawberry aroma (Ulrich *et al.* 1997; Schieberle and Hofmann 1997). GLVs are synthesized in chloroplasts (Chen *et al.* 2004) by a biochemical pathway that starts from linolenic acid and includes several sequential enzymatic steps (Figure 2-1). Linolenic acid is first peroxidazed by a lipoxygenase (LOX) (Chen *et al.* 2004), and then cleaved by a hydroperoxidelyase to produce (Z)-3-hexenal (Howe *et al.* 2000; Hatanaka *et al.* 1987). (Z)-3-hexenal may be further reduced to its alcohol and ester forms (Z)-3-hexenol and (Z)-3-hexenyl acetate, respectively (Yasuo *et al.* 2011; D'Auria *et al.* 2007; Bate NJ *et al.* 1998). However, (Z)-3-hexenal may also be isomerised into (E)-2-hexenal by a recently identified (Z)-3:(E)-2-hexenal isomerise (Kunishima *et al.* 2016). (E)-2-hexenal may be reduced to corresponding alcohol (E)-2-hexenol and further to the ester (E)-2-hexenyl acetate by the same enzymes as mentioned for (Z)-3-hexenal derivatives.

The GLV-associated QTL located at the end of LG5 of the diploid strawberry (Urrutia *et al.* 2017) was shown to reduce the accumulation of (E)-2-hexenal and (E)-2-hexenyl acetate and increase the accumulation of (Z)-3-hexenal and (Z)-3-hexenyl acetate. This suggested that an *F. bucharica* allele of a gene responsible for (Z)-3:(E)-2-hexenal conversion might be defective or differentially regulated in the NILs with an introgression at the end of LG5. Here, we describe the identification and functional characterization of *F. vesca* hexenalisomerase (*FvHI*) located within the GLV-QTL region at the end of LG5.

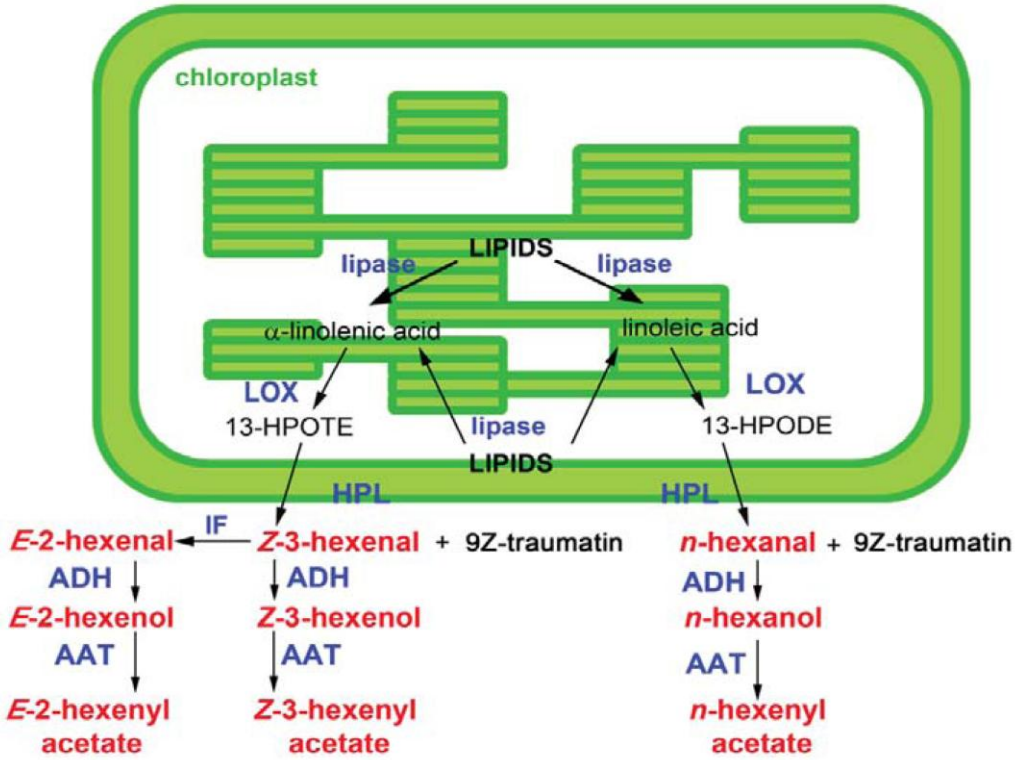

Figure 2-1. Green Leaf Volatile (GLV) biosynthesis. Lipase(s) release(s) α-linolenic and linoleic acid from galactolipids. 13-lipoxygenases (LOXs) catalyze the addition of oxygen to α-linolenic acid to form 13(S)-hydroperoxy 9*Z*,11*E*,15*Z*-octadecatrienoic acid (13-HPOTE), in Section 2 referred to as 13-hydroperoxide. 13-HPOTE is converted to *Z*-3-hexenal and 9*Z*-traumatin by 13-HPL (HPL). An isomerization factor, (3*Z*):(2*Z*)-enalisomerase (IF) is responsible for converting *Z*-3-hexenal into its isomer, *E*-2-hexenal. *Z*-3-hexenol and *E*-2-hexenyl acetate are converted to *Z*-3-hexenyl acetate and *E*-2-hexenyl acetate by alcohol acyltransferase(s) (AAT) (figure adopted from (Scala *et al.* 2013)).

Material and methods

Plant material

The aroma volatile compounds of diploid strawberry ripe fruits were studied using a near isogenic line (NIL) collection constructed in detail in Urrutia *et al.* (2015). The 11 NILs covering all the different regions of linkage group 5 (LG5) of diploid *Fragaria* and their recurrent parent 'Reine del Vallées' (RV) were used in all experiments (Figure2-2) and are described in detail in Table S2-1. Four NILs from LG7 were only used in verification of candidate genes for volatile compounds. All plant materials used in the experiments were propagated from seed.

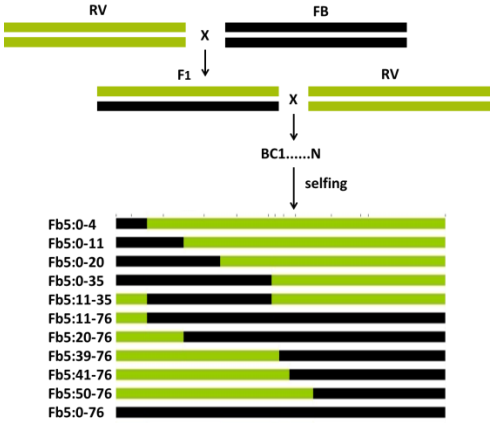

Figure 2-2. Graphical representations of the NILs used in the experiments.

Plant seeds were germinated in October 2017 as described in Urrutia *et al.* (2015). Six plants of each genotype were grown under greenhouse conditions at day temperature between 22 to 24°C and 17°C during night without artificial lighting, relative humidity 40-50% at the Centre for Research in Agricultural Genomics (CRAG) in Bellaterra (latitude:41° 29'N, longitude: 2° 06'E) until March 2018. In March, four individual plants of each genotype were transferred to a shaded greenhouse at Centre Torre Marimon in Caldes de Montbui (latitude: 41°36'N, longitude: 2°10'E at 203 m of altitude from sea level). The plants grew in natural conditions and photoperiod without receiving additional lighting or heating from March to September. The agronomical practices were the usual for strawberryfruit production.

Plants grown in controlled climate conditions were germinated as described in Urrutia *et al.* (2015) at the end of July 2018. Germinated seedlings were transferred onto soil in mid-August 2018 and cultivated in a growth chamber (16 hours of light emitted from LED lamps, temperature of 22°C and 70 % relative humidity) for 11 weeks. After this, the plants were moved to a greenhouse compartment supplemented with artificial lighting (16 hour daylength) emitted from high-pressure sodium lamps. The temperature in

the greenhouse compartment was 22°C and relative humidity 30-40%. The plants were fertirrigated in two-week intervals.

Sampling

For volatile analysis from shaded greenhouse plants, samples were harvested during July 2018. Fully ripe red berries from four individual plants were pooled together and analysed as independent biological replicates. Each biological replicate was mixed of 10 to 20 fruits (>10g) and immediately frozen on dry ice and stored at -75°C until volatile analysis.

Fruit samples used for RNA extraction from shaded greenhouse plantswerecollected during April to July 2018. Three stages offruits were harvested. The stage 1 samples correspond to green fruits, approximately 15 days after anthesis. When the fruit color was“turning” (i.e. seeds colored and light coloration on fruit skin), the stage 2 samples were collected. The stage 3 samples are determined by fully ripe fruits (Figure2-3). At least four biological replicates were harvested for every stage of each genotype. All samples were immediately frozen on dry ice and stored at -75° until RNA extraction.

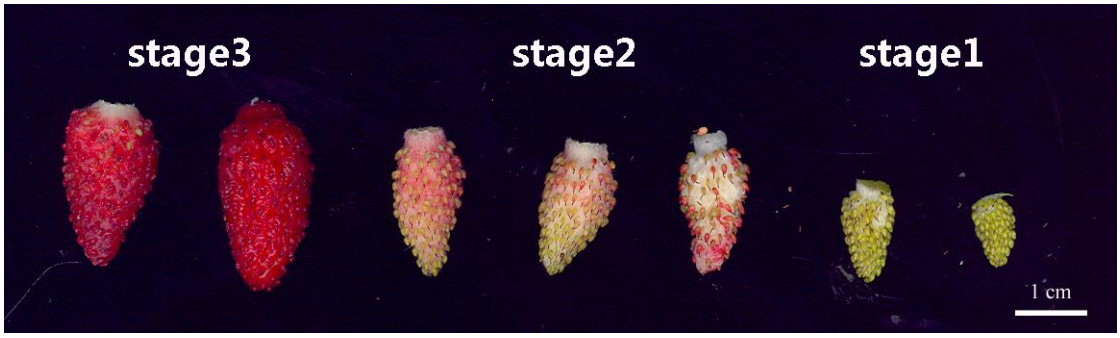

Figure 2-3. Three stages of wild strawberry fruit samples.

Plant tissue samples for studying tissue-specific expression patterns were collected from plants grown in a growth-chamber under long-day (LD) conditions for eleven weeks. Each genotype was represented by three individual plants. For leaf tissue, the youngest fully opened leaf was sampled. For root samples, two-centimetre root tips from actively growing roots were pooled together. Each plant had multiple branch crowns and for crown samples, three branch crowns from each plant were pooled together. Fruits were sampled at white stage (approximately 10days post-anthesis) and at fully ripe red stage. All samples were immediately frozen in liquid nitrogen and stored at -75°C until RNA extraction.

Transiently transformed berries used for RNA extraction and volatile compounds analyses were collected when the fruits were fully ripe. Each transiently transformed berry was cut in two halves and both halves were immediately frozen in liquid nitrogen and stored at -75°C until further processing. One of the halves

was used for RNA extraction and the other for volatile analysis.

Volatile compounds analysis

Volatile compounds were detected in a modification of the method described in Zorrilla-Fontanesi *et al.* (2012) at CRAG. Each biological replicate was analyzed as an independent sample. The samples were ground for 2 minutes at 30/sec to fine powder in liquid nitrogen using a Retsch MM400 ball mill (Retch GmbH, Germany). For each sample, 1g of powdered fruit was weighed in a 10ml screw cap headspace glass vial and 1ml of a saturated NaCl solution with 10 ppm of internal standard (3-hexanone) was added andthe mixture was homogenized gently. A Combi-PAL autosampler (CTC Analytics AG, Switzerland) was used for incubation, extraction and desorption. Vials were first heated at 50℃ during 10 min with agitation at 500 rpm, then SPME fibre (50/30µm DVB/CAR/PDMS; Supelco, PA, USA) were exposed 30 min more in the same conditions. The extracted volatiles were desorbed in the GC injection port at 250℃ for 5 min in splitless mode where volatile compounds were analysed on a 7890A gas chromatograph coupled with a 5975C mass spectrometer (Aligent Technology, CA, USA).A HP-5MS UI GC column (30 m, 0.250 mm, 0.25 µm) (J&W, CA, USA) and 1.2 ml min^{-1} constant helium flow was used for chromatographic separation. Oven condition was 40℃ for 2 mins, 5℃ min^{-1} ramp to 250℃, and then 5 min at 250℃. Compounds were identified with the same retention time and ion fragmentation as each commercial standard (Sigma-Aldrich, MO, USA). Areas were normalized by comparison with internal standard peaks. All profiles were analysed using Enhanced ChemStation software (Agilent Technologies, CA, USA) and with mass spectra libraries, NIST08 and NIST11. In this experiment, Z-3-hexenal and furaneol couldn't be identified properly.

The obtained value was expressed as the ratio between each compound and internal standard values. The data used for heatmap was normalized in relation to RV.

The samples used for identification Z-3-hexenal and E-2-hexenal from the transient transformation experiment and the samples from shaded greenhouse were detected as described in Urrutia *et al.*(2017) at Instituto de Biologia Molecular y Celular de Plantas (IBMCP, Valencia, Spain).

The data statistical analysis was done using the free source software R 3.5.1 (RCore Team 2018) with the Rstudio 0.92.501 interface (Rstudio, 2012). Heatmap was performed with *heatmap* function from the heatmap3 R package. To confirm these QTL and estimate their effect an interval mapping analysis was performed using MapQTL v.6 (Van Ooijen 2009). Stable regions that explained around 50% or more of the variability and showed LOD scores > 1.8 were considered major QTLs.

Identification of candidate genes for volatile compounds

A hundred volatile compounds were identified in the *F. vesca* strawberry fruits and many QTL explaining considerable percentages of their variability were mapped in Urrutia *et al.* (2017). Through whole transcriptome analysis of four selected NILs (Fb5:0-35, Fb5:50-76, Fb6:84-101 and Fb7:0-10), some candidate genes for volatile compounds were selected by combining expression data with the metabolic QTL mapped in Urrutia *et al.*(2015). Based on various metabolic pathways of volatile compounds, candidate genes were selected and their amino acid sequences were used for querying the *Fragaria vesca* V4.0 a1 genome database (Edger *et al.* 2018) by BLASTpin Genome Database for Rosaceae. Then RT-QPCR primers were designed based on the coding sequences of the identified volatile compound-related genes (Table S2-2).

Identification of *F. vesca* (Z)-3:(E)-2-hexenal isomerase genes

The amino acid sequence of cucumber (Z)-3:(E)-2-hexenal isomerase (accession number XP_004151504.1) was used for querying the *Fragaria vesca* genome V4 protein database (Jung *et al.* 2019) by BLASTp with default settings (blastp -max_target_seqs -evalue 0.001 -word_size 3 –gapopen 7 –gapextend 2 -culling_limit 0 -matrix PAM30). Matches with bit scores higher than 200, were reported.

For studying phylogenetic relationships of hexenalisomerase-like proteins, the amino acid sequences of HI- and HI-like proteins from various plant species (accession numbers in Table S2-3) were aligned by MAFFT (Katoh *et al.* 2005) using E-INS-I iterative refinement method with default settings and retaining gappy regions. The resulting amino acid alignment was used as input for MEGA6 (Tamura *et al.* 2013) software. Phylogenetic tree was constructed using maximum likelihood algorithm with default settings and 1000 bootstrap replications.

F. vesca and *C. annuum* amino acid sequences were aligned by MAFFT with default settings. The resulting amino acid alignments were used as input for the Color Align Properties tool of the DNA Sequence Manipulation Suite (Stothard 2000).

RNA extraction and qRT-PCR

The samples for RNA extraction were ground to fine powder in 2-ml Eppendorf tubes in liquid nitrogenusing 3-mm tungsten beads in a Retsch MM400 ball mill (Retch GmbH, Germany). Tissue samples (root, leaf and crown) were ground for 30 seconds at 30/sec once on each side, while fruit samples (all three stages) were ground for 2 minutes. The sample blocks were pre-cooled in liquid nitrogen and care was taken not to thaw the samples while grinding. RNA extraction protocol used was a modification of the CTAB method originally described by Monte and Somerville (2002) (Monte and Somerville 2002). Briefly, 800 µl pre-warmed (65°C) CTAB buffer consisting of 2% (w/v) CTAB, 2%

(w/v) PVP-40, 1 M NaCl, 100 mM Tris (pH 8.0), 1 mM EDTA (pH 8.0) and 0.5 g/l spermidine, supplemented with 2% (v/v) β-mercaptoethanol was mixed with the homogenized sample, extracted twice with chloroform:IAA (24:1) and precipitated overnight at +4°C in the presence of 1/3 volume of 8 M LiCl. Next day, samples were pelleted by centrifugation, dissolved in 500 µl SSTE buffer (1.0 M NaCl, 10 mM Tris-HCl [pH 8], 1 mM EDTA [pH 8] and 0.5% SDS), extracted once with chloroform:IAA and precipitated overnight at -20°C with absolute ethanol. Contaminating DNA was removed by DNAse I treatment following manufacturer's instructions (Thermo Scientific, USA). DNAse I was removed by chloform:IAA extraction, followed by overnight RNA precipitation at -20°C with 1/10 volume of 3 M NaCl and absolute ethanol. Prior to cDNA synthesis, all samples were diluted to approximate concentration of 100ng/µl.

cDNA was synthesised from 300 ng of total-RNA using PrimeScript reverse transcriptase according to manufacturer's instructions (Takara Bio Inc., Japan). The synthesised cDNA was diluted 6-fold prior to quantitative real-time PCR (RT-qPCR). RT-qPCR was run in a LightCycler 480 instrument (Roche, Mannheim, Germany) in a 10 µl total volume with final concentration of 1x SYBR Green I Master Mix (Roche), 0.45 µm primers and 3.5 µl of diluted cDNA. All samples were run with three technical replicates. The RT-qPCR temperature profile has been described in (Koskela *et al.* 2016). The primers used for RT-qPCR are described in Table S2-2. Relative expression levels were calculated using the $2^{-\Delta\Delta Ct}$ method (Pfaffl 2001) using *FvMSI1* as normalization gene. Log-transformed relative expression values were used for statistical tests.

Transient overexpression/silencing vector construction and transformation

Coding sequence of the gene *FvHI* (gene 5g29270) was amplified from cDNA synthesized from RV and NIL Fb5:41-76 fruit RNA samples (mature fruit) using 300 ng of total-RNA, which was diluted to a final volume of 50 µl. The PCR to amplify the coding sequence was carried out using Phusion Green High Fidelity DNA polymerase (Thermo Scientific, USA) and a gene-specific primer pair designed to target the gene 5g29270 (Table2-1). The sequence acquired from NIL Fb5:41-76 is named "Fb" for *F. bucharica*.

PCR reactions were carried out in a final volume of 50 µl containing: 5 µl cDNA, 5X Phusion Green HF buffer (Thermo Scientific, USA), 200 µM each dNTPs, 0.5 µM each primer and 0.5 µl Phusion DNA Polymerase (2 units/µl). The following touch-down PCR conditions were used: an initial denaturation step of 98℃ (30s), then 10 cycles of 98℃ (10s), 65-58℃ annealing temperature decreasing by 0.7℃ per cycle (30s), and 72℃ (45s), followed by 20 cycles of 98℃ (10s), 58℃ (30s) and 72℃ (45s), and a final elongation step of 72℃ (10min).

Table 2-1. Sequences of the primers used in the study.

Primer name	Primer sequence	Used for
C01 5g29270-F	ATGGCGGAAATGGATCTAACACC	Amplification of 3Z2E coding sequence
C02 5g29270-R	ACAGACTTCGATTGATGGTGCAGG	Amplification of 3Z2E coding sequence
attB1_5g29270-F	AAAAAGCAGGCTTCGAAGGAGATAGAACC ATGGCGGAAATGGATCTAACACC	Gene-specific Gateway primer for overexpression
attB2_5g29270-R	AGAAAGC TGGGTACAGACTTCGATTGATGGTGCAGG	Gene-specific Gateway primer for overexpression
attB1_5g29270-RNAi-F	AAAAAGCAGGCTCTGATGAGGTCACCAAAA ACC	Gene-specific Gateway primer for silencing
attB2_5g29270-RNAi-R	AGAAAGCTGGGTCCCAGCTTGGTTAAGAAA AGG	Gene-specific Gateway primer for silencing
attB1_adapter	GGGGACAAGTTTGTACAAAAAAGCAGGCT	Gateway adapter
attB2_adapter	GGGGACCACTTTGTACAAGAAAGCTGGGT	Gateway adapter

The PCR products were gel-purified using the agarose gel protocol of the High Pure PCR Product Purification kit (Roche, Switzerland). As the Phusion Green High Fidelity polymerase produces blunt ends, 3´-As were added in a reaction mixture containing 50 µl purified product, 10X NH_4 Reaction buffer (Bioline, UK), 3 mM $MgCl_2$, 100 µM dNTPs, 5 U BIOTAQ DNA polymerase in 60 µl final volume at 72℃ for 10 min.

The purified PCR products were ligated into pCR-TOPO-2.1 vector following manufacturer´s protocol (Invitrogen, USA). The ligations were transformed into competent DH5α cells by heat-shock transformation. Transformations were plated onto LB-agar plates containing kanamycin at 50 µg/ml. Positive colonies were identified by blue-white screening.

Positive colonies were cultured overnight (LB-kanamycin, 50 µg/ml), plasmids were extracted using GeneJET plasmid Miniprep kit (Thermo Fisher Scientific, USA) following manufacturer´s protocol and the plasmids were digested with Sac1 and Xba1 double enzyme digestion (Takara Bio Inc., Japan) to confirm that the plasmids contained inserts of the correct sizes. Positive colonies were sent to sequencing with the primers M13-forward and M13-reverse. Sequences were used to build a consensus sequence for RV and NIL Fb5:41-76 CDS of 5g29270 gene since the Fb5:41-76 contained the *F. bucharica* allele of gene 5g29270.

Based on the sequenceof thegene 5g29270 from RV and FB, Gateway-compatible gene-specific primers were designed (Table2-1) and the coding sequence of gene 5g29270 from RV and FB in TOPO-TA

vector (5g29270-RV/FB-TOPO) was amplified using Gateway-compatible gene-specific and adapter primers following the two-step attB adapter PCR according to the manufacturer's recommendations (Invitrogen (Life Techonogies), USA). The first template-specific PCR reactions were in a final volume of 25 µl containing: 60 ng DNA, 5X Phusion Green HF buffer (Thermo Scientific, USA), 400 µM each dNTPs, 0.5 µM each gene-specific-primer, and 0.25 µl Phusion DNA Polymerase. The following PCR conditions were used: an initial denaturation step of 98℃ (2min), then 10 cycles of 98℃ (15s), 60℃ annealing temperature (30s), and 72℃ (1min). The PCR products were used as template in second step adapter PCR. Adapter PCR reaction were in a final volume of 50 µl containing: 10 µl template, 5X Phusion Green HF buffer (Thermo Scientific, USA), 200 µM each dNTPs, 0.8 µM each attB adapter primer, and 0.5 µl Phusion DNA Polymerase. The following PCR conditions were used: an initial denaturation step of 98℃ (20s), then 5 cycles of 98℃ (15s), 45℃ annealing temperature (30s), and 72℃ (1min), followed by 30 cycles of 98℃ (15s), 55℃ (30s) and 72℃ (1min). In order to get enough PCR products, each PCR reaction was run in duplicate.

The PCR products were gel-purified using the agarose gel protocol of the High Pure PCR Product Purification kit (Roche, Switzerland). The purified PCR products were recombined into donor vector pDONR^{TM}221 via BP recombination reaction following the manufacturer's recommendations (Invitrogen, USA). The BP recombination reaction contained 50 femtomoles each of attB-PCR product and donor vector. The reactions were incubated at 25℃ overnight and the reactions were stopped by adding 1 µl of Proteinase K solution and incubating the reactions at 37℃ for 10 min. The BP reactions were transformed into competent DH5α cells by electroporation-shock transformation. Transformations were plated onto LB-agar plates containing kanamycin at 50 µg/ml.

Colonies were cultured overnight (LB-kanamycin, 50µg/ml), plasmids were extracted using GeneJET plasmid Miniprep kit (Thermo Fisher Scientific, USA) following manufacturer´s protocol and a colony-PCR was used with the primers M13-forward and c02 5g29270-reverse (the silencing entry-vector used M13-reverse) to confirm that the entry-vector contains the 5g29270 gene. Colonies with inserts of the correct size were sent to sequencing with the primers M13-forward and M13-reverse to confirm the insert sequence.

The positive plasmids were recombined into overexpression vector pK7WG2D (Karimi *et al.* 2002) or the silencing vector pK7GWIWG2D (ii) (Karimi *et al.* 2005) in a 10 µl LR recombination reaction containing 50 femtomoles each of the insert and destination vector according to manufacturer's recommendations (Invitrogen, USA). The reactions were incubated at 25℃ overnight and the reactions were stopped by

adding 1 µl of Proteinase K solution and incubating the reactions at 37℃ for 10 min. The LR reaction products were transformed into competent DH5α cells by electroporation-shock transformation. Transformations were plated onto LB-agar plates containing spectinomycin at 100 µg/ml.

Colonies were cultured overnight (LB-spectinomycin, 100 µg/ml), plasmids were extracted using GeneJET plasmid Miniprep kit (Thermo Fisher Scientific, USA) following manufacturer´s protocol and the plasmids were digested with BamHI enzyme digestion (Takara Bio Inc., Japan) to confirm that the plasmids contained inserts of the correct sizes. The positive colonies were transformed into Agrobacterium strain GV3101 cells by electroporation-shock transformation. Transformations were plated onto LB-agar plates containing spectinomycin at 100 µg/ml, rifampicin at 50 µg/ml and gentamicin at 25 µg/ml. Colonies were cultured overnight (LB-spectinomycin (100 µg/ml) – rifampicin (50 µg/ml) – gentamicin (25 µg/ml)) and stored at -80℃ for further use.

For transformation, the agrobacterium with 5g29270-overexpression construct was cultured in liquid LB (spectinomycin (100µg/ml) – rifampicin (50µg/ml) – gentamicin (25µg/ml)) at +28°C until the OD_{600} reached approximately 0.8. 2ml of the culture was pelleted by centrifugation at 5000 G for 10 minutes, after which the pellet was resuspended in 2ml of MS supplemented with 2% (w/v) of sucrose. Fruits at white or turning stage were injected with either the 5g29270-overexpression construct or with MS-sucrose solution (mock treatment) using a 1 ml sterile syringe.

Result

Volatile compounds analysis in LG5

A preliminary genetic analysis revealed that major QTLs related tonine key volatile compounds of wild strawberry aroma located in LG5 (Urrutia *et al.* 2017). Therefore, to further narrow down the genetic region affecting the accumulation of these key volatile compounds, a deep analysis was performed in the NILs of LG5.

Nineteen key volatile compounds were analyzed in ripe fruits of 11 NILs covering all the different regions of LG5 as well as the recurrent parent RV with two harvests in 2018. Three to five biological replicates were included for each line. An internal standard 3-hexanone was added in all samples in the analysis as reference. Finally, 17 compounds were identified by the comparison of both retention time and mass spectra with those of commercial standards run under the same conditions. However, Z-3-hexenal was not clearly separated and furaneol was not detected in our samples. The 17 identified compounds included one aldehyde ((E)-2-hexenal), one furan (mesifurane), one lactone (γ-decalactone), two terpenoids (linalool and nerolidol) and 12 esters (butyl acetate, butyl butanoate, (E)-2-hexenyl acetate, ethyl butanoate, ethyl hexanoate, hexyl acetate, methyl-2-aminobenzoate, methyl butanoate, methyl cinnamate, methyl hexanoate, myrtenyl acetate and (Z)-3-hexenyl acetate).

Data on each individual compound was transferred into a ratio (sample to internal standard). Mean and standard deviation (sd) values for all compounds for both RV and the 11 NILs were calculated independently. For each genotype, most compounds showed similar mean values in both harvests (Table S2-4).

It could be seen in Figure 2-4 that the NILs covering the central region (20-35cM) of LG5 (Fb5:0-35, Fb5:11-35, Fb5:11-76, Fb5:20-76 and Fb5:0-76) had lower accumulation of methyl 2-aminobenzoate and myrtenyl acetate in both harvests, indicating this region played a role as a negative effector in the accumulation of these two key compounds. Methyl butanoate was also lower accumulated in the NILs containing introgression of LG5:11-20cM than RV and the decreased methyl butanoate QTL was located in LG5:11-20cM. For the compound butyl butanoate and methyl hexanoate, the QTL was detected only in one harvest with low LOD value (Table2-2) and did not match to previous QTL regions, therefore we did not consider them as good QTL.

It was also easily observed that for the compound (Z)-3-hexenyl acetate, it had obviously higher accumulation in NILs with distal introgressions (50-76 cM) than RV and other NILs in both harvests. On

the contrary, (E)-2-hexenyl acetate had obviously lower accumulation in NILs with distal introgressions (50-76 cM) than RV and other NILs in both harvests.

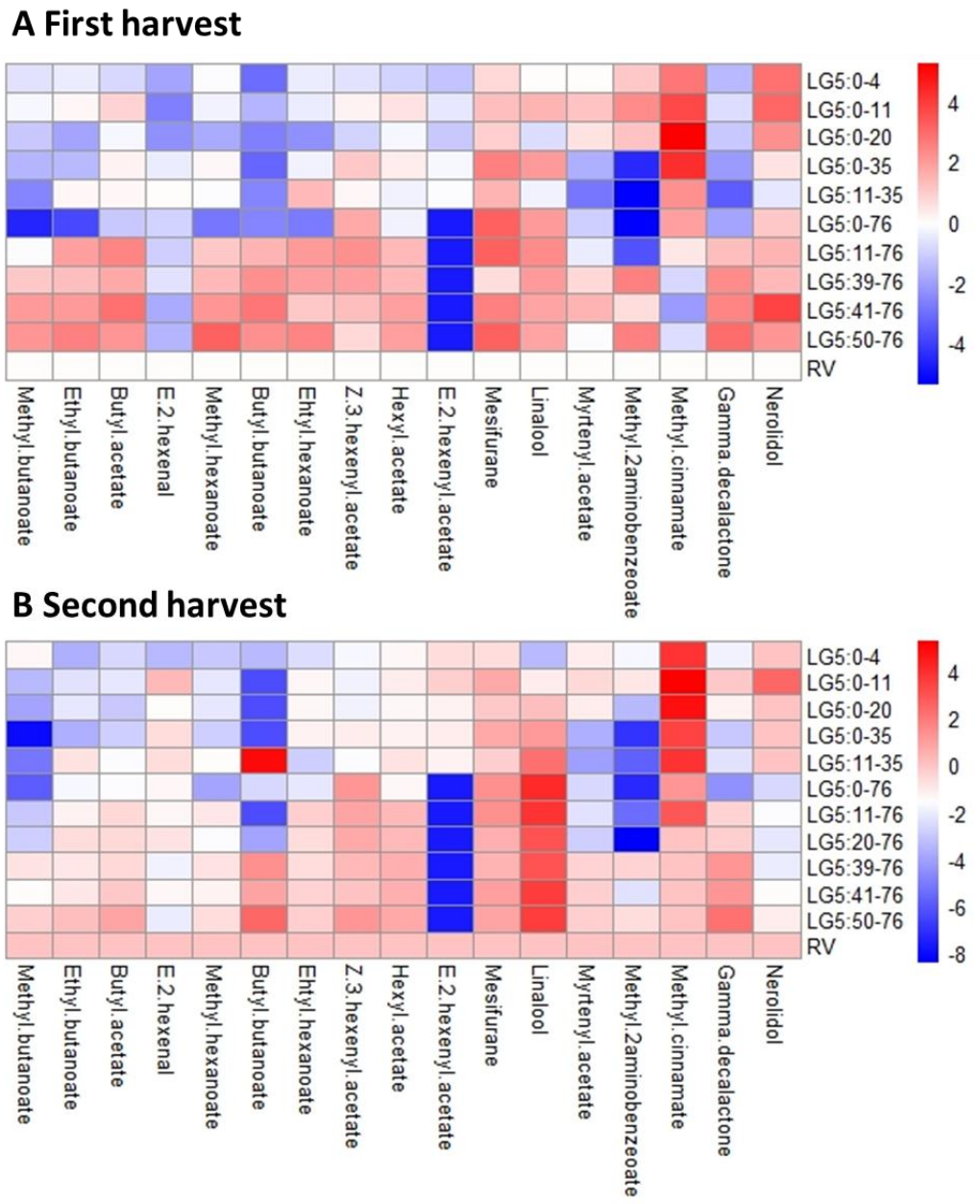

Figure 2-4. Heatmapof key volatile compounds levels detected. The normalized data of all studied volatile compoundsper genotype are shown in the heatmap on a blue (negative) to red (positive) scale.

Table 2-2. QTL for volatile compound detected in NILs with *F. bucharica* introgression in LG5.

compounds	direction	QTL(cM)	LOD	%explained	harvest
Butyl-butanoate_1	up	LG5:50-76	2.06	57.80%	1
Butyl-butanoate_2	-	-	-	-	2
E2-hexenylacetate_1	down	LG5:50-76	3.97	81%	1
E2-hexenylacetate_2	down	LG5:50-76	3.95	78%	2
M-2aminobenzeoate_1	down	LG5:20-35	1.82	53.20%	1
M-2aminobenzeoate_2	down	LG5:20-35	2.84	66.30%	2
Methyl-butanoate_1	down	LG5:11-20	1.81	53%	1
Methyl-butanoate_2	down	LG5:11-20	2.41	57.20%	2
Methyl-hexanoate_1	-	-	-	-	1
Methyl-hexanoate_2	down	LG5:0-4	2.35	59.20%	2
Myrtenyl-acetate_1	down	LG5:20-35	1.81	52.10%	1
Myrtenyl-acetate_2	down	LG5:20-35	4.99	85.20%	2
Z3-hexenylacetate_1	up	LG5:50-76	4.65	87.60%	1
Z3-hexenylacetate_2	up	LG5:50-76	3.27	71.50%	2

Green leaf volatile compounds in ripe strawberry fruits

The NILs carrying *F. bucharica* introgressions at the distal end of LG5 have been reported to accumulate higher amounts of (Z)-3-hexenal and (Z)-3-hexenyl acetate and lower amounts of (E)-2-hexenal and (E)-2-hexenyl acetate than the recurrent parent RV (Urrutia *et al.* 2017). Due to (Z)-3-hexenal could not be identified in previous GC-MS experiment, the samples used for detecting green leaf volatile compounds were analysed using equipment with higher resolution in IBMCP, Valencia. We measured the accumulation of these compounds in ripe strawberry fruits by GC-MS in all NILs carrying introgressions in LG5, as well as in the recurrent parent RV. Our results on accumulation of green leaf volatiles were well in line with the results reported by Urrutia *et al.* (2017), RV and the NILs with introgressions in the beginning or middle of LG5 accumulated lower amounts of (Z)-3-hexenal and (Z)-3-hexenyl acetate than the NILs harbouring an introgression at the end of LG5. Also the accumulation of (E)-2-hexenal and its derivative (E)-2-hexenyl acetate followed the expected pattern with RV and LG5:0-35 lines having higher levels of the compounds than the LG5:50-76 lines (Figure 2-5). Our findings confirmed the earlier reports (Urrutia *et al.* 2017) of the presence of a QTL affecting the accumulation of green leaf volatile compounds located within the introgression at the end of LG5.

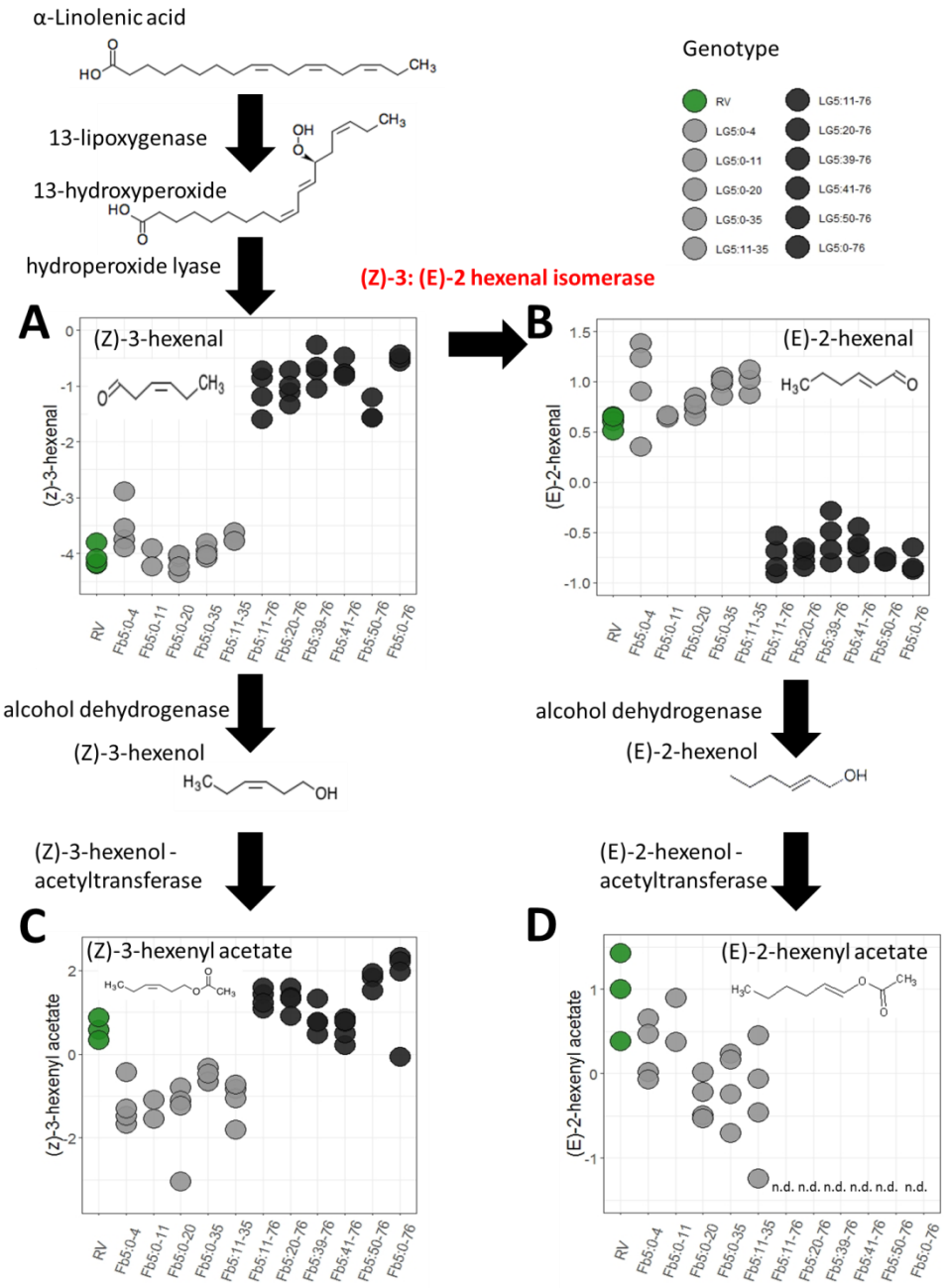

Figure 2-5. Accumulation of green leaf volatiles in RV, and LG5 introgression lines. Light gray and dark gray dots depict NILs with the RV background and F. bucharica introgression at the end of LG5 (50-76 cM), respectively. Accumulation of **A):** (Z)-3-hexenal, **B)**: (E)-2-hexenal, **C)**: (Z)-3-hexenyl acetate, **D)**: (E)-2-hexenyl acetate in NILs with introgressions covering different regions of the LG5.

Verification of candidate genes for volatile compounds

According to earlier results by Urrutia *et al.* (2017), QTL analysis revealed that NILs Fb5:0-35 and Fb7:0-10 harbor QTL linked to the accumulation of several volatile alcohols and esters, including five

'key compounds' for the strawberry aroma (methyl 2-aminobenzoate, myrtenyl acetate, methyl butanoate, butyl butanoate, methyl hexanoate), and Fb5:50-76 shows QTLs related with the accumulation of compounds derived from the lipoxygenase pathway associated with the green leaf volatile compounds ((*E*)-2-hexenal, (*Z*)-3-hexenal, (*E*)-2-hexenyl acetate and (*Z*)-3-hexenyl acetate). Therefore, three NILs (Fb5:0-35, Fb5:50-76and Fb7:0-10) were selected to perform a whole mRNA sequencing (RNAseq) analysis and evaluated differences in expression levels between NILs and the recurrent parental line RV (Urrutia *et al.* 2017). Based on the metabolic pathway of volatile compounds and candidate genes described in Urrutia *et al.*(2017), finally, we selected 18 genes as key candidate genes, five genes for LG5 from 0 to 35 cM, six for LG5 from 50 to 76 cM and seven for LG7 from 0 to 10 cM (Table 2-2) in order to be validated in these regions.

Five genes selected from 0 to 35 cM of LG5 were FvH4_5g06470.1, FvH4_5g06530.1, FvH4_5g02520.1, FvH4_5g06590.1 and FvH4_5g06920.1. The expression of these five genes was analyzed by RT-qPCR in fully ripe fruits from RV and all lines of LG5. Results of gene expression of NILs related to RV expression was analyzed using Dunnett's test (Table 2-3). The gene FvH4_5g06470 was not expressed in any NIL. Therefore, we discarded this gene as a candidate gene. Also, we did not consider the gene FvH_5g02520 as a candidate gene since the relative expression pattern of this gene did not show any correlation with the introgressions that NILs contain and also no significant difference was observed between the relative expression value of RV and any other NIL using Dunnett's test (Table2-3).

Table 2-3. Log-transformed fold-change values of candidate genes in LG5:0-35cM for volatile accumulation.

NIL	FvH4_5g06530	FvH4_5g02520	FvH4_5g06590	FvH4_5g06920
Fb5:0-4	-1.68±2.85	-0.21±0.79	-0.22±0.3	1.43±1.13
Fb5:0-11	-2.95±1.59**	0.72±0.67	-0.16±1.2	-0.34±0.59
Fb5:0-20	-2.28±2.33**	1.02±0.2	1.14±0.45	1.19±0.92
Fb5:0-35	-4.17±0.67**	0.46±0.2	3.44±1.42**	-
Fb5:0-76	-3.75±1.66**	0.6±0.65	2.21±0.47**	-
Fb5:11-35	-3.06±0.81**	1.17±0.61	1.82±0.49**	-
Fb5:11-76	-3.62±1.19**	-0.5±1.24	3.73±0.85**	-
Fb5:20-76	-1.81±1.67	-0.57±0.98	3.75±1.03**	-
Fb5:39-76	-0.24±1.16	-1.29±1.12	-0.3±0.85	-0.74±1
Fb5:41-76	0.43±0.46	-0.49±0	0.25±1.12	-1.09±0.55
Fb5:50-76	-2.94±1.78**	-0.18±0.95	2.83±2.33**	-2.74±0
RV	0±1.19	0.01±0.44	0±0.3	0±1.15

Note: "-" represents no expression in q-PCR data.Values are averages of 2-6 biological replicates ± standard deviation. Expression values that are significantly different from RV (Dunnett's test) are indicated by asterisks: **$p < 0.05$.

.

Table 2-4. Candidate genes for volatile compounds.

NIL	Genefor	Candidatefor	V2_number*	DEG direction**	V4_number of the best hit*
Fb5:0-35	terpene synthase	Volatiles (alpha-pinene)	gene09971	+	FvH4_5g06470.1
Fb5:0-35	terpene synthase	Volatiles (alpha-pinene)	gene09972	-	FvH4_5g06530.1
Fb5:0-35	GLABRA	Volatiles	gene32494	+	FvH4_5g02520.1
Fb5:0-35	putative monoterpene synthase	Volatiles	gene09988	+	FvH4_5g06590.1
Fb5:0-35	putative arginosuccinate synthase	Volatiles	gene09887	-	FvH4_5g06920.1
Fb5:50-76	LOX	Volatiles (fatty acid derivatives)	gene29184	-	FvH4_5g26250.1
Fb5:50-76	3Z-2E-enal isomerase	Volatiles (fatty acid derivatives)	gene10882		FvH4_5g29270.1
Fb5:50-76	TT2	Volatiles (nerol)	gene25060	-	FvH4_5g32460.1
Fb5:50-76	terpene synthase	Volatiles (myrtenol/nerol)	gene12094	-	FvH4_5g35700.1
Fb5:50-76	putative arginosuccinate synthase	Volatiles (nerol)	gene11807	+	FvH4_5g30030.1
Fb5:50-76	amino acid permease	Volatiles	gene26913	-	FvH4_5g33530.1
Fb7:0-10	LOX	Volatiles (fatty acid derivatives)	gene26949	-	FvH4_7g00440.1
Fb7:0-10	acyl-coA hydrolase	Volatiles (methyl benzoate/methyl 2-aminobenzoate)	gene05329	-	FvH4_7g09160.1
Fb7:0-10	acyl-coA hydrolase	Volatiles (methyl benzoate/methyl 2-aminobenzoate)	gene09200	-	FvH4_7g06750.1
Fb7:0-10	acyltransferase	Volatiles (esters)	gene19411	+	FvH4_7g11730.1
Fb7:0-10	acyltransferase	Volatiles (esters)	gene00424	-	FvH4_7g04560.1
Fb7:0-10	putative hydroxyisobutytyl-CoA hydrolase	Volatiles	augustus_masked-LG7-processed-gene-21.17-mRNA-1	+	FvH4_7g16010.1
Fb7:0-10	MYC2	Volatiles	gene09222	+	FvH4_7g07050.1

* V2 and V4 numbers refer to gene IDs from different versions of the *F. vesca* genome from GDR.

** Differentially expressed genes were identified by RNAseq experiment in Urrutia (2017)

According to the annotation of the *F. vesca* genome, FvH4_5g06530 is a putative terpene synthase gene (www.rosaceae.org/species/fragaria/fragaria_vesca). The relative expression of this gene in five NILs Fb5:0-11, Fb5:0-20, Fb5:0-35, Fb5:11-35 and Fb5:0-76 were significantly lower than in RV (Figure2-6). The expression analyses correlate with RNA-seq data from Urrutia *et al.*2017, which indicates that the gene FvH_5g06530 might be involved in down regulation of terpene synthesis in our lines compared with RV.

On the contrary, the relative expression of gene FvH4_5g06590, a putative monoterpene synthase gene in the five NILs Fb5:0-35, Fb5:11-35, Fb5:11-76, Fb5:20-76 and Fb5:0-76 were significantly higher than that in RV (Figure2-1) which agrees with the earlier result (Urrutia *et al.* 2017) that it was up regulated in Fb5:0-35. The gene FvH4_5g06920, a putative arginosuccinate synthase gene, was not expressed in five NILs Fb5:0-35, Fb5:11-35, Fb5:11-76, Fb5:20-76 and Fb5:0-76 while there was expression in the control RV and also in NILs with introgressions without 20-39cM. Therefore, *F. bucharica* introgression containing the gene FvH_5g06920 may abolish the expression of arginosuccinate synthase in lines containing introgression of 20-35cM.

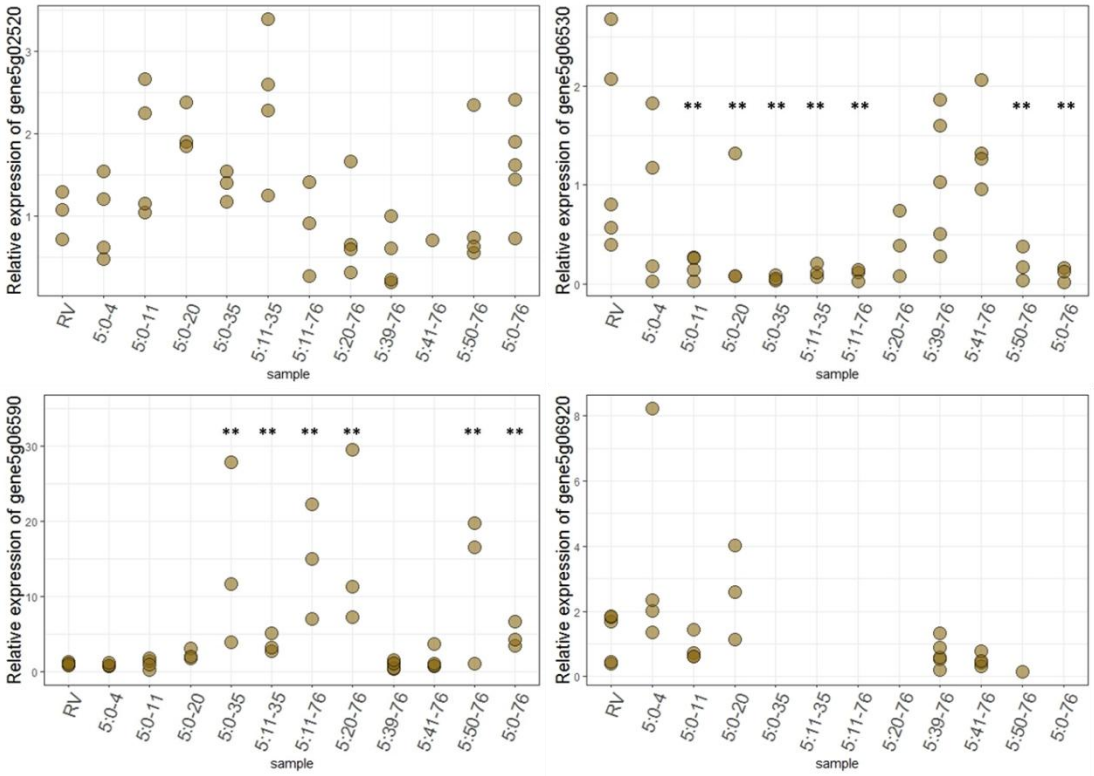

Figure 2-6. Relative expression of candidate genes from 0 to 35 cM in LG5 in different genotypes. Asterisks indicate statistically significant differences as compared to RV by Dunnett's test: ** - $p < 0.05$.

Six candidate genes selected from 50 to 76 cM of LG5 wereFvH4_5g26250.1, FvH4_5g29270.1, FvH4_5g32460.1, FvH4_5g35700.1, FvH4_5g30030.1 and FvH4_5g33530.1. The expression of these genes was detected the same way as described for that from 0 to 35 cM of LG5. The gene FvH4_5g26250 was not expressed in any line (data not show). Therefore, we discarded this gene as our candidate gene. Also, we did not consider the gene FvH4_5g32460 as a candidate gene since the relative expression pattern of this gene did not show any correlation with the introgressions that NILs contain and also no significant differences were observed between the relative expression value of RV and any other NIL using Dunnett's test (Table 2-4).

Table 2-5. Log-transformed fold-change values of candidate genes in LG5:50-76cM for volatile accumulation.

NIL	FvH4_5g29270	FvH4_5g32460	FvH4_5g35700	FvH4_5g30030	FvH4_5g33530
Fb5:0-4	-1.09±1.63	-4.69±7.49	1.74±1.5	0.14±0.36	0.02±0.22
Fb5:0-11	-1.57±0.96	-3.67±7.65	2.21±1.72	-1.02±1.21	0.46±0.83
Fb5:0-20	-1±1.27	-2.75±7.5	3.34±0.36	-0.28±0.28	0.72±0.24
Fb5:0-35	-0.48±1.24	-0.4±7.01	1.97±1.25	-1.48±1.59	0.18±0.82
Fb5:0-76	-9.4±0.46**	-2.1±7.79	-	1.77±1.63	-
Fb5:11-35	-0.7±0.85	-1.48±6.32	1.65±1.57	0.15±0.39	0.99±0.99
Fb5:11-76	-10.08±0.4**	3.44±1.47	-	0.7±0.82	-
Fb5:20-76	-11.28±1.58**	-3.1±8.32	-	1.95±1.29**	-
Fb5:39-76	-10.63±1.28**	-3.87±6.69	-	2.88±0.34**	-
Fb5:41-76	-11.29±1.27**	-1.63±7.63	-	2.77±0.92**	-
Fb5:50-76	-10.9±0.67**	-4.32±8	-	1.21±1.3	-
RV	0±0.59	0.24±0.48	0.01±1.87	0±0.48	0±0.55

Note: "-" represents no expression in q-PCR data. Values are averages of 2-6 biological replicates ± standard deviation. Expression values that are significantly different from RV (Dunnett's test) are indicated by asterisks: **$p < 0.05$.

The gene FvH4_5g30030 is a putative arginosuccinate synthase gene. The relative expression of this gene in three NILs Fb5:20-76, Fb5:39-76 and Fb5:41-76 was significantly higher than that in RV, however we did not consider this gene as a good candidate gene since the expression in NILs Fb5:50-76 and Fb5:0-76 were similar as RV. Gene FvH4_5g29270 is a putative 3Z-2E-enal isomerase gene. The relative expression levels of this gene were similar in RV and in NILs covering the LG5 until 35cM. However, the NILs harbouring an introgression at the end of LG5 showed extremely low levels of expression. This corroborated our hypothesis that gene FvH_5g29270 is the gene responsible for the low (Z)-3:(E)-2-hexenal conversion rate.

Gene FvH4_5g33530 is an amino acid permease gene and gene FvH_5g35700 is a putative terpene synthase gene. Both genes had similar relative expression patterns to the gene FvH4_5g29270, even

without any expression in the lines harbouring introgression at the end of LG5, indicating these genes could be considered as candidate genes for volatile pathway analysis (Figure 2-7).

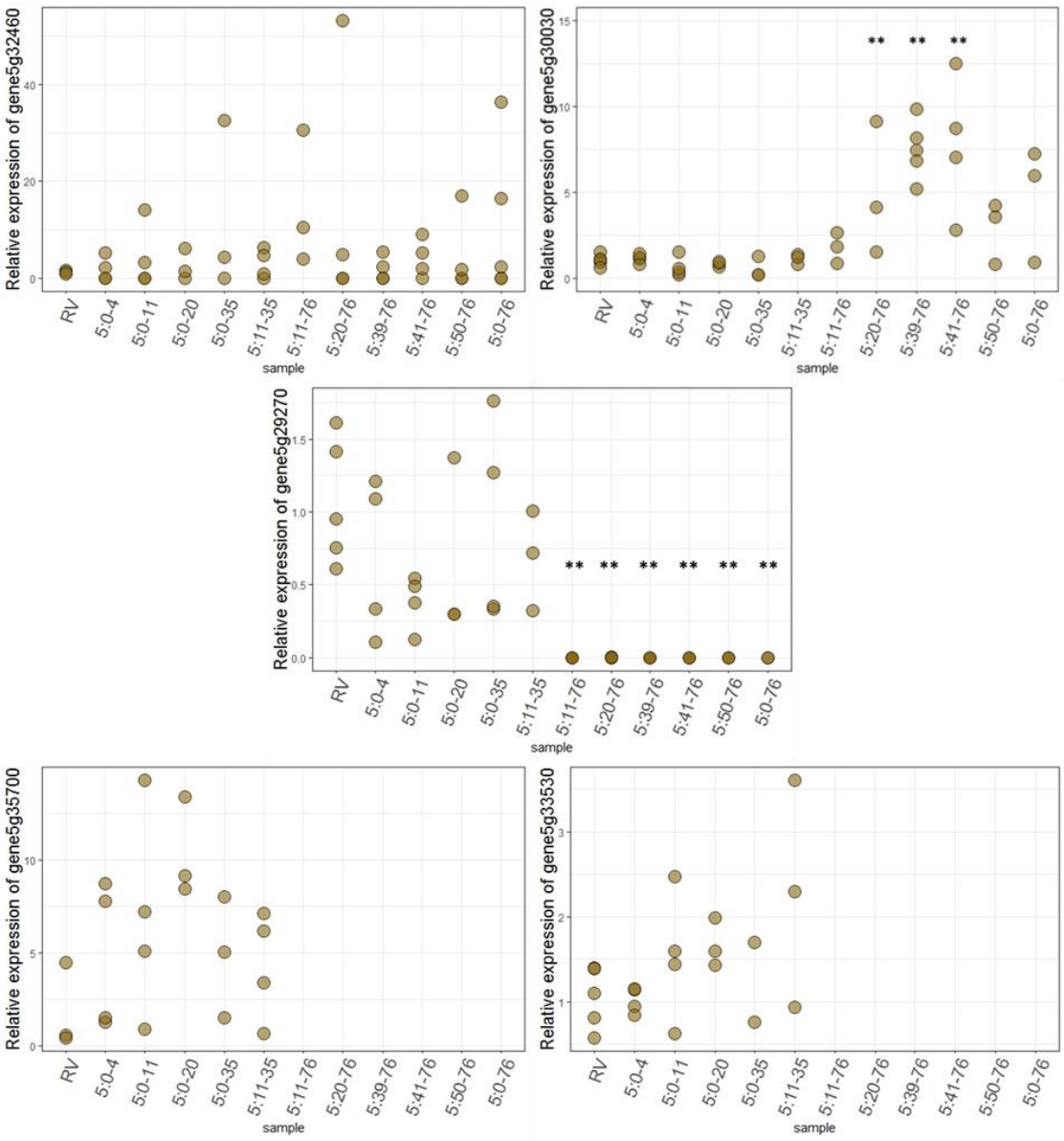

Figure 2-7. Relative expression of candidate genes from 50 to 76cM in LG5 in different genotypes. Asterisks indicate statistically significant differences as compared to RV by Dunnett's test: ** - $p < 0.05$.

Seven genes selected from 0 to 10 cM in LG7 were FvH4_7g00440.1, FvH4_7g09160.1, FvH4_7g06750.1, FvH4_7g11730.1, FvH4_7g04560.1 and FvH4_7g16010.1. The expression of these genes was analyzed in RV and three lines of LG7 (Fb7:0-10, Fb7:0-27 and Fb7:26-59) (Table 2-5). The gene FvH4_7g04560 and gene FvH4_7g16010 were not expressed in any line (data not shown). Therefore, we discarded these two genes as our candidate genes. And also, the gene FvH4_7g11730 did

not show any correlation with the introgressions that NILs contain and also no significant difference were observed between the relative expression values of RV and any other NIL using Dunnett's test. Thus, we did not consider the gene as a candidate gene for volatile compound accumulation. For the relative expression of genes FvH4_7g09160, FvH4_7g07050 and FvH4_7g06750, a significant difference as compared to RV was detected only in NIL Fb7:0-27, but in the NIL Fb7:0-10 their relative expression was similar to RV. We intend to check the positions of these genes in LG7 and compare them with the genotype markers to decide whether to consider them as good candidate genes for volatile compounds.

Table 2-6. Log-transformed fold-change values of candidate genes for volatile accumulation.

NIL	7g00440	7g09160	7g06750	7g11730	7g07050
Fb7:0-10	-2.83±1.49**	0.33±1.31	-1.18±1	-0.29±	0.97±0.97
Fb7:0-27	-3.4±1.33**	-8.77±0.61**	-9.21±0.68**	1.94±0.94	4.46±0.31**
Fb7:26-59	-0.11±1.67	-2.7±1.62	1.49±0.32	1.18±1.15	1.58±0.36
RV	0±0.32	0±0.7	0±0.55	0±0.9	0±1.78

Note: Values are averages of 3-5 biological replicates ± standard deviation. Expression values that are significantly different from RV (Dunnett's test) are indicated by asterisks: ** $p < 0.05$.

The gene FvH_7g00440 is a lipoxygenase (LOX) gene. The relative expression of this gene in NILs Fb7:0-10 and Fb7:0-27 was significantly lower than that in RV (Figure 2-8). The expression analysis correlates with RNA-seq data from Urrutia *et al.* 2015, which indicates the gene FvH_7g00440 might be involved in down regulation of LOX in our lines compared with RV. The LOX pathway is the most important pathway for synthesis of most fruit aroma volatiles like straight-chain aldehydes, alcohols, esters, lactones and ketones. According to the importance of this gene with the relation to aroma compounds synthesis and our expression analysis, we will consider it as one gene for functional validation in future studies.

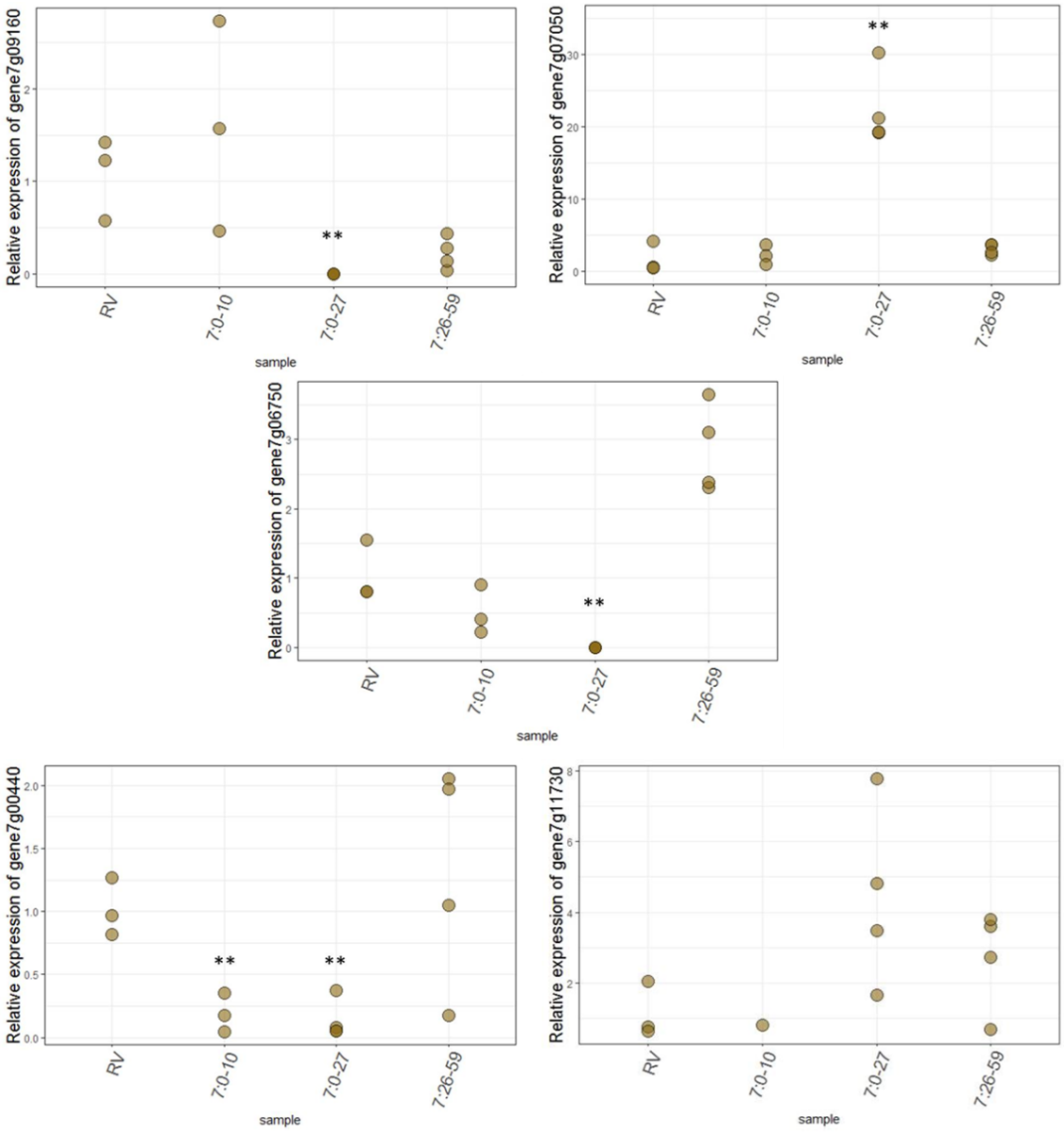

Figure 2-8. Relative expression of candidate genes from 0 to 10cM in LG7 in different genotypes. Asterisks indicate statistically significant differences as compared to RV by Dunnett's test: ** - $p < 0.05$.

F. vesca homologs of (Z)-3:(E)-2-hexenal isomerase

Based on our findings and earlier reports (Urrutia *et al.* 2017), the *F. bucharica* introgression within Fb5:50-76 harbours a QTL that affects the accumulation of (Z)-3-hexenal and (E)-2-hexenal and their respective derivatives (Z)-3-hexenyl acetate and (E)-2-hexenyl acetate. A key step in the biosynthesis pathway of these compounds was recently identified in bell pepper (*Capsicum annuum* L.), in which the isomerization of (Z)-3-hexenal into (E)-2-hexenal was shown to take place via the activity of a (Z)-3:(E)-2-hexenal isomerase (Kunishima *et al.* 2016). Similarly, in cucumber (*Cucumis sativus*) the conversion of (Z)-3-hexenal into (E)-2-hexenal is dependent on the activity of a cucumber (Z)-3:(E)-2-hexenal

isomerase (Spyropoulou *et al.* 2017). To identify *F. vesca* homologs of (Z)-3:(E)-2-hexenal isomerase (HI), we performed a BLAST query with cucumber HI amino acid sequence against *F. vesca* protein database. This resulted in the discovery of four proteins with bit scores higher than 200 (Table S2-5). The protein with the highest score was 5g29270, located in the 5th chromosome within the LG5:50-76 introgression.

Phylogenetic relationships of the four identified *Fragaria* HI and HI-like proteins were studied by comparing the *Fragaria* proteins to other cupin protein superfamily proteins (Table S2-3) from various plant species. All four *Fragaria* proteins were most closely associated with the HI and HI-like protein clade (Figure 2-9A). A more detailed analysis of the HI and HI-like clade showed that only the protein 5g29270 clustered together with HI proteins from other plant species (Figure 2-9B). These data suggested that only the protein 5g29270 is an actual *Fragaria* HI, while the other three *Fragaria* proteins are similar to HI-like proteins.

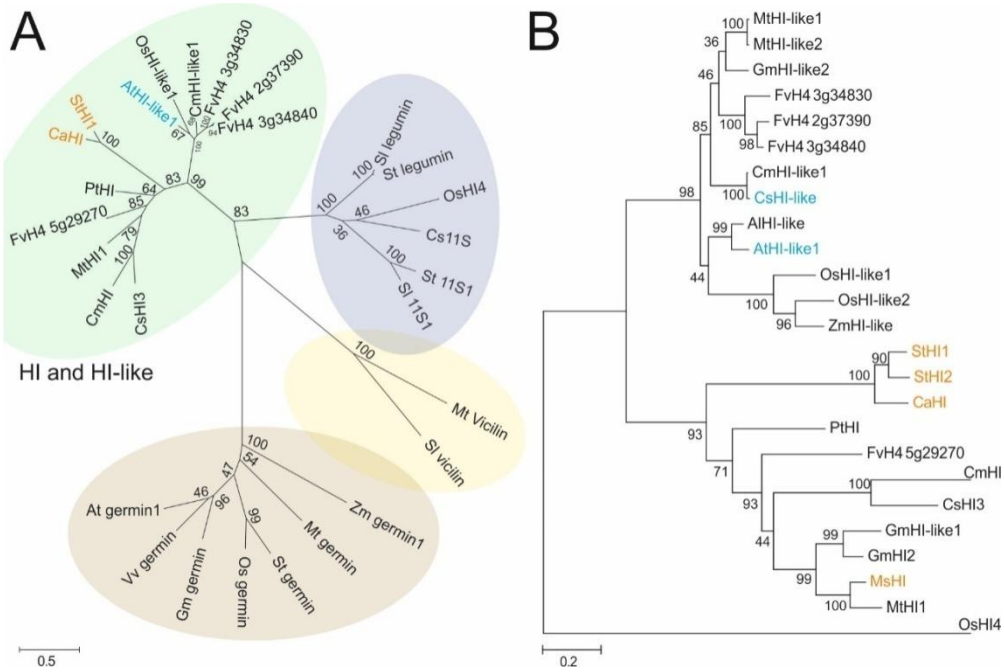

Figure 2-9. Phylogenetic relationships of Fragaria HI and HI-like proteins. A) Phylogenetic tree of cupin superfamily proteins from various plant species. Proteins belonging to HI and HI-like, 11S seed globulin, vicilin and germin clades are highlighted by green, blue, yellow and brown backgrounds, respectively. **B)** Phylogenetic tree of HI- and HI-like proteins. Proteins with and without demonstrated HI activity are shown in orange and blue fonts, respectively. The values next to branching points indicate the percentage of bootstrap support with 1000 replications.

Next, we investigated whether the identified FvHI proteins possess the amino acid residues essential for HI enzymatic activity (Kunishima *et al.* 2016). Alignment of the predicted *F. vesca* proteins against bell pepper HI showed that only the 5g29270 protein possesses the three amino acid residues essential for

hexenal isomerase activity (Figure 2-10A).

As the *F. vesca* protein 5g29270 is located within the *F. bucharica* introgression that affects the accumulation of (Z)-3-hexenal and (E)-2-hexenal, we decided to clone the gene encoding the protein from both parents of the NIL collection to see whether the protein itself is altered or non-functional in *F. bucharica*. Sequencing the coding sequence (Table S2-6) of the gene 5g29270 showed that the predicted proteins from RV and *F. bucharica* were virtually identical (Figure 2-10B). These data suggested that if the hexenalisomerase gene 5g29270 (referred to as *FvHI* from hereafter) is the causative agent behind the altered (Z)-3:(E)-2-hexenal conversion rate observed in near-isogenic lines harbouring an *F. bucharica* introgression, the difference probably occurs at transcriptional level.

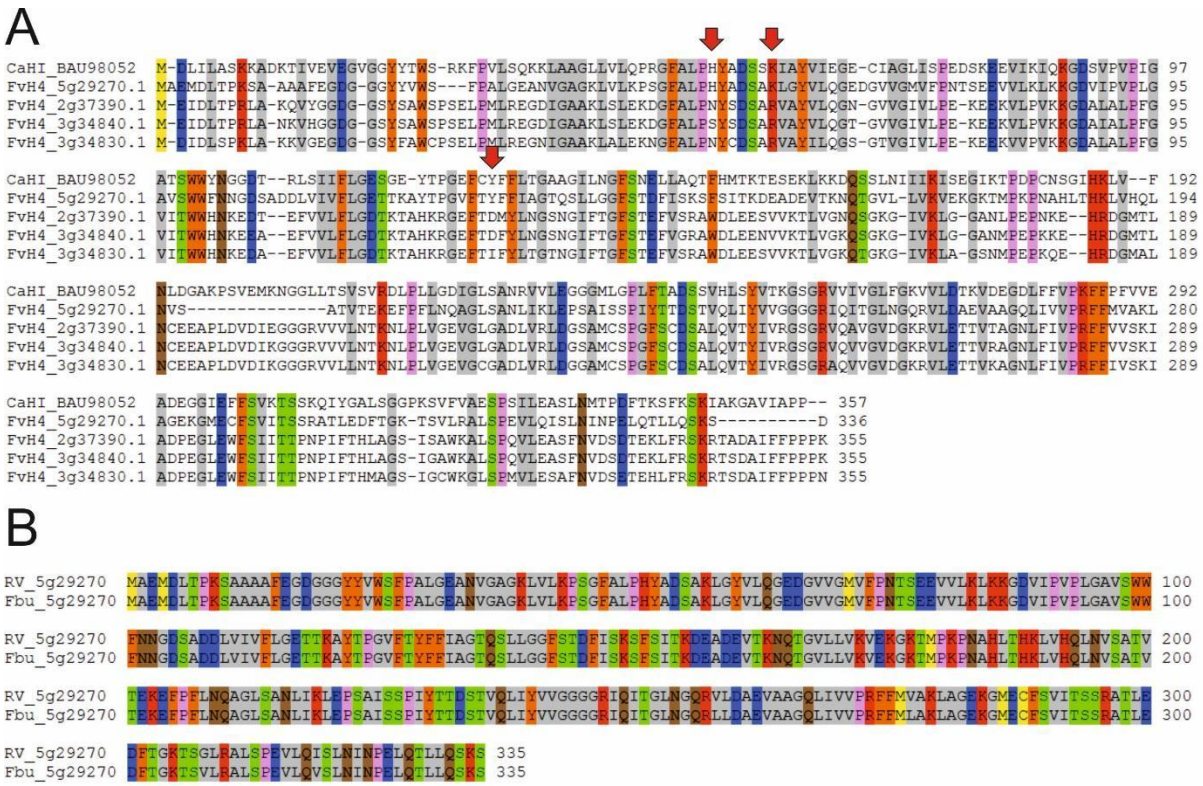

Figure 2-10. Amino acid alignments of HI and HI-like proteins from *F. vesca* and *F. bucharica*. A) Alignment of four F. vesca HI proteins with bell pepper HI (CaHI_BAU98052). The three functionally essential amino acids (H - K- Y) are highlighted by red arrows. Only the Fragaria protein 5g29270 possesses all three essential amino acids. **B)** Alignment of translated coding sequences from the recurrent parent RV and the donor parent *F. bucharica*.

Gene expression patterns of *FvHI* in RV and near-isogenic lines

Next, we investigated gene expression patterns of *FvHI* in the recurrent parent RV and in NILs with *F. bucharica* introgressions covering different regions of the LG5. We first analysed *FvHI* expression in

fully ripe fruits of field-grown plants. The mRNA levels of the gene *FvHI* were similar in RV and in NILs covering the LG5 until 35cM. However, the NILs harbouring an introgression at the end of LG5 showed extremely low levels of *FvHI* mRNA (Figure 2-11A). This corroborated our hypothesis that *FvHI* is the gene responsible of the low (Z)-3:(E)-2-hexenal conversion rate.

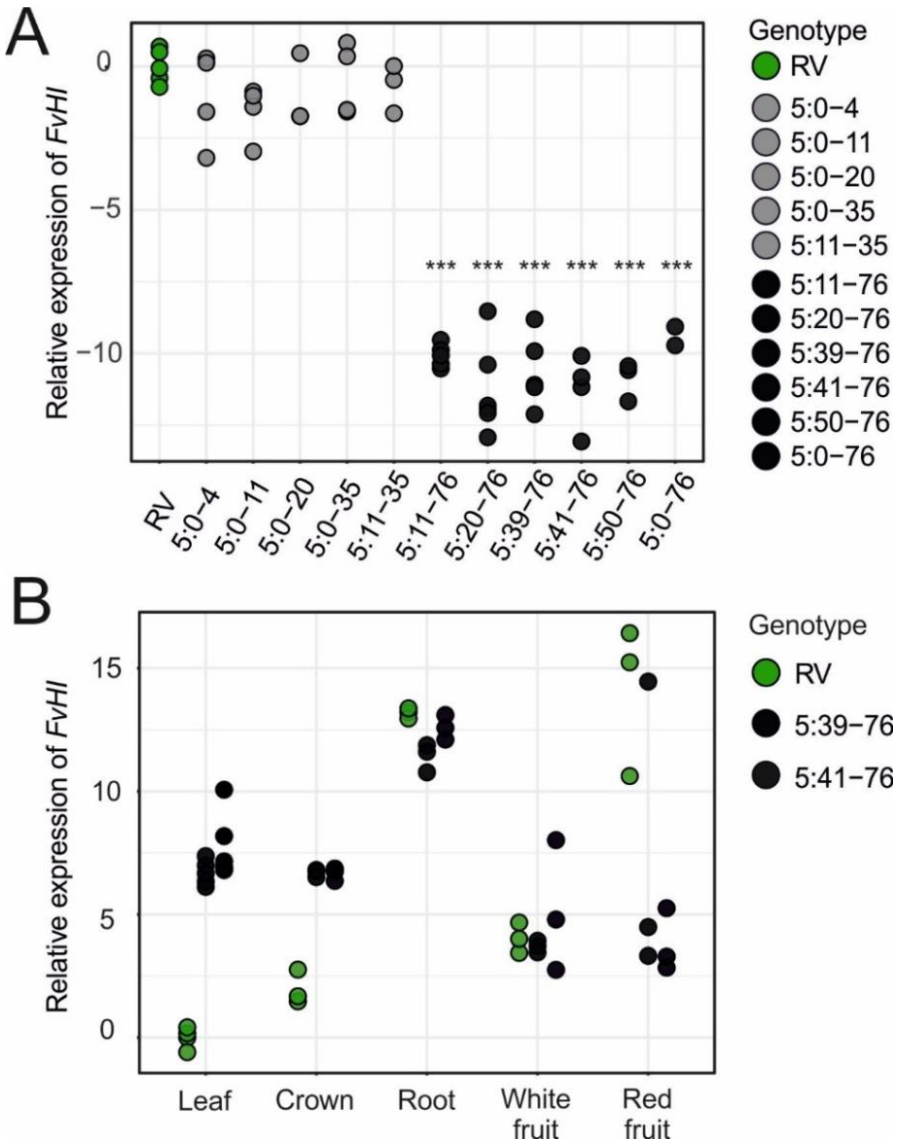

Figure 2-11. Relative expression of *FvHI* in different genotypes and plant tissues. A) Relative expression of *FvHI* in NILs with introgressions covering different regions of the LG5. Light gray and dark gray dots depict NILs with the RV and *F. bucharica* alleles of *FvHI*, respectively. **B)** Relative expression of FvHI in various plant tissues of RV and in two NILs with introgressions containing the *F. bucharica* allele of *FvHI*. Relative expression values have been normalized to *FvMSI1* and shown as log transformed fold change values. Asterisks indicate statistically significant differences as compared to RV by Dunnett's test: *** - $p < 0.001$. In A), n = (2 – 5); B) n = (3 – 5).

We also examined the tissue-specific expression patterns of *FvHI* in the recurrent parent RV and in two NILs with introgression at the end of LG5 (Figure 2-11B). The relative expression rate in fully ripe red fruits were in concordance with the results obtained from field-grown plants; *FvHI* mRNA was significantly more abundant in RV than in the two NILs. However, the level of *FvHI* expression in white fruits was comparable in the three genotypes. Interestingly, the NILs with introgressions at the end of LG5 had higher levels of *FvHI* expression in leaves and crowns than RV.

Function of (3Z):(2E)-enal isomerase in *F. vesca*

In order to verify the function of *FvHI* in isomerization of (Z)-3-hexenal into (E)-2-hexenal in *F. vesca*, transient expression was applied. As explained in material and methods, firstly, the *FvHI* of RV overexpression vector was constructed and then the *FvHI* gene of RV was over-expressed in the stage 2 fruits of RV, Fb5:39-76 and Fb5: 41-76 via agrobacterium mediated transfection. In total, 233 fruits were infected and 209 fruits (the number of fruits treated by mock/FvHI-OX: RV 56/58, Fb5:39-76 16/17, Fb5:41-76 24/20) were collected. Every fruit was cut into two halves, one half used for volatile compounds analyze and the other for extraction RNA.

The agro-infiltrated mature fruit of these three lines were collected for quantifying relative expression of *FvHI* as well as the content of (Z)-3-hexenal and (E)-2-hexenal. Regarding to the transcripts amount of *FvHI*, unexpectedly no significant difference was found in the over-expressed berries compared with mock infiltrated berries in any line (RV, Fb5:39-76 or Fb5: 41-76) (Figure 2-12). In each line, either mock infiltrated or over-expressed berries presented similar concentration for both (Z)-3-hexenal and (E)-2-hexenal (Table 2-6).

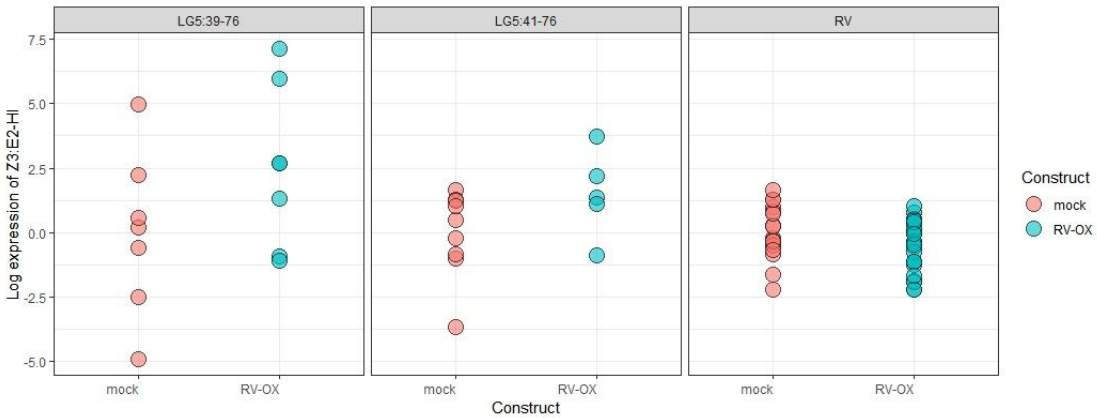

Figure 2-12. Relative expression of *FvHI* in mocks and agro-infiltrated fruits of different genotypes. Pink and blue dots depict mock and RV-overexpression (RV-OX) infiltrated, respectively. The goal was to compare the effect of the *FvHI* overexpression construct to the mock for each genotype separately. Relative expression values have been normalized to *FvMSI1* and shown as log transformed fold change values.

The agro-infiltrated mature fruits were used for detecting green leaf volatile compounds by GC-MS (In IBMCP, Valencia). The results are shown for each compound and expressed as average ratios between mock-transformed and FvHI-ox construct transformed fruit for each genotype (Table S2-7). The results showed that (E)-2-hexenal content of FvHI-overexpression fruits was nearly double of mock-treat fruits in both NILs Fb5:39-76 and Fb5:41-76, although the difference was not statistically significant. And also, the content of other three green leaf volatile compounds did not show significant differences between FvHI-overexpressed and mock-treated fruits.

Table 2-7. Green leaf volatile compounds in mocks and agro-infiltrated fruits of different genotypes.

	RV		Fb5:39-76		Fb5:41-76	
Compound	FvHI-ox/mock	p-value	FvHI-ox/mock	p-value	FvHI-ox/mock	p-value
(Z)-3-hexenal	0.7364	0.5022	0.6608	0.6085	0.9995	0.9992
(E)-2-hexenal	0.9429	0.7391	2.2476	0.1143	2.0540	0.1240
(Z)-3-hexenyl acetate	0.9582	0.9088	0.5183	0.1399	0.9035	0.8784
(E)-2-hexenyl acetate	0.9806	0.9505	0.8369	0.4446	1.5926	0.0508

Data expressed as the ratio between average values of *FvHI* overexpression construct and mock for each genotype, respectively. T-test result is shown.

Discussion

Strawberry is a popular fruit worldwide. However, during the domestication of the cultivated strawberries (*F. x ananassa*), it has lost apart of its flavor and fragrance. The diverse aroma patterns were required for the cultivars (Negri *et al.* 2015). As opposite to cultivated varieties, wild strawberries (*F. vesca*) are renowned for their intense flavor and fragrance due to the accumulation of higher levels and wider assortments of volatile compounds (Ulrich *et al.* 2007), therefore plant breeders regard wild strawberries as important donors of novel scented molecules. A volatile composition profile of NIL collection resulting from a cross between *F. vesca* and *F. bucharica* has been reported (Urrutia *et al.* 2017) and some QTLs for key volatile compounds in wild strawberry were mapped in linkage group 5 and 7. In this study we performed a deeper analysis on these key volatile compounds and some candidate genes for compounds were identified. The locations of two QTLs responsible for decreased accumulation of methyl 2-aminobenzoate and myrtenyl acetate were narrowed down to region LG5:20-35cM while they were mapped in LG5:11-35cM in the previous research (Urrutia *et al.* 2017). The QTL for decreased accumulation of methyl butanoate was narrowed down to region LG5:11-20cM while previously it was also mapped in LG5:11-35cM.

According to the previous reports, four QTLs for key volatile compounds associated to green leaf volatile (GLV) were located in LG5:50-76, including positive QTL for (*Z*)-3-hexenal and (*Z*)-3-hexenyl acetate and negative QTL for their respective trans-2 isomers (*E*)-2-hexenal and (*E*)-2-hexenyl acetate (Urrutia *et al.* 2017). Our results are consistent with the earlier report. Among these four green leaf volatile compounds, (*Z*)-3-hexenal is directly synthetized from linolenic acid (Howe *et al.* 2000; Hatanaka *et al.* 1987) and then it may be reduced to (Z)-3-hexenyl acetate (Yasuo *et al.* 2011) or isomerised into (E)-2-hexenal by (Z)-3:(E)-2-hexenal isomerase (Kunishima *et al.* 2016). The compound (*E*)-2-hexenyl acetate was produced by reduction of (E)-2-hexenal. In the region of LG5:50-76cM, a candidate gene (*FvHI*) putatively encoding a (Z)-3:(E)-2-hexenal isomerase (HI) was identified. The function of gene *HI* isomerising *Z*-3- to *E*-2-hexenal has been demonstrated in some plant species. In bell pepper fruits, an HI was identified and *Z*-3- to *E*-2-hexenal converting activity was confirmed by heterologous expression of a bell pepper HI and orthologous HIs of other plant species (Kunishima *et al.* 2016). Two HI genes were identified from cucumber and the function was verified through transient expressed HI homologs in *N. benthamiana* (Spyropoulou *et al.* 2017). A putative catalytic site (catalytic HKY) was verified as an important prerequisite for the enzymatic function of HIs. In this study, the identified FvHI protein possesses the amino acid residues essential for HI enzymatic activity and it was clustered together with HI proteins from other species. Therefore we considered *FvHI* is an actual HI protein encoding gene. First of all, to know whether differences existed in the gene coding sequence, we cloned *FvHI* gene from both

parental lines of the NIL collection and it was found that they were virtually identical. Then, the transcription levels of *FvHI*in RV and in lines with the *F.bucharica* allele was analyzed. In white fruit stage, all samples present similar expression level. However, in red fruit stage, *FvHI* was significantly more abundant in RV than in the NILs with *F. bucharica* allele of *FvHI*. Hence, the *FvHI* expression difference between RV and the NILs with *F. bucharica* allele is probably caused at transcriptional level and the expression of *FvHI* was up-regulated in RV while not in the NILs with *F. bucharica* allele of *FvHI* during fruit ripening.

To our knowledge, there has been no report related with the function of *FvHI* gene neither in *Rosaceae* nor in strawberry. While we identified the *FvHI* gene encoding the HI protein, the problem remains what is the main function of this gene for strawberry. When the cucumber HI homolog was transiently expressed in *N. benthamiana,* it caused a significant increase in (E)-2-hexenal and (E)-2-hexenyl acetate (Spyropoulou *et al.* 2017). We analyzed the (Z)-3 and (E)-2-hexenal content in ripen strawberry fruits from the NILs containing introgression from *F. bucharica* in LG5 and the recurrent parent RV, and found that (Z)-3-hexenal is higher in the NILs with *F. bucharica* allele of *HI* than in RV while (E)-2-hexenal was higher in RV. Combined with gene expression results, it indicated that the low expression level of *FvHI* mRNA is responsible of the low (Z)-3:(E)-2-hexenal conversion rate, which is similar to the protein's function in cucumber. To verify the function of HI in strawberry, the transient overexpression of *FvHI* in RV and two NILs with *F. bucharica* introgression in LG5:50-76cM was performed. However, no difference was detected either at gene expression level or in GLV content between overexpression samples and mock treated samples. The *Agrobacterium* mediated transient transformation has been successfully applied in *F. vesca* accessions Yellow Wonder (Hawkins *et al.* 2016; Luo *et al.* 2018) and *F. x ananassa* cultivated variety (Hoffmann *et al.* 2006). However, to our knowledge, no cases have been reported in the *F.vesca* accession 'Reine del Vallées'. In this study, we used an overexpression vector carrying a green fluorescent protein (GFP) marker and only faint GFP signals were detected in the treated fruits, suggesting that the transient transformation had very low efficiency. Many factors like temperature, the fruit stage for injection, the plant and fruits health condition have been reported affecting the transformation efficiency. Another reason may be thatwe did not inject enough berries so that there were not enough fruits for analyzing gene expression and GLV content. Further, we will try to improve the transient transformation efficiency by trying other strawberry genotypes, and by injecting more individuals to get enough positive transient transformants. The other approach is to verify the *FvHI* gene function using stable transformation, which is an under-going project for us. At the same time, more work can be done to study whether any differences exist within the introns, promoter regions or other transcription elements between RV and *F. bucharica* alleles of *FvHI*.

Besides that, we investigated *FvHI* mRNA expression level in different plant tissues. It is quite interesting that the NILs with introgressions at the end of LG5 had higher levels of *FvHI* expression in leaves and crowns than RV. It means that in leaves and crowns of the NILs with introgressions at the end of LG5 the conversion from (Z)-3- to (E)-2-GLVs is increased. (E)-2-hexenal is a commonly occurring compound in the volatile bouquet of stressed plants and plays an important role in transferring information to plants and insects either as a single molecule (James 2005; Kessler *et al.* 2006) or within a complex volatile mixture (Allmann *et al.* 2013). Increased (E)-2-hexenal can be beneficial for herbivore attacked plants by attracting natural enemies of the herbivores (Allmann and Baldwin 2010). We speculate that the increased accumulation of (E)-2-GLVs in NILs with introgressions at the end of LG5 is related toplant defence, which still needs to be verified.

Bibliography

Aharoni A, Giri AP, Verstappen FW, Bertea CM, Sevenier R, Sun Z, Jongsma MA, Schwab W, Bouwmeester HJ (2004) Gain and loss of fruit flavor compounds produced by wild and cultivated strawberry species. The Plant Cell 16 (11):3110-3131

Allmann S, Baldwin IT (2010) Insects betray themselves in nature to predators by rapid isomerization of green leaf volatiles. Science 329 (5995):1075-1078

Allmann S, Späthe A, Bisch-Knaden S, Kallenbach M, Reinecke A, Sachse S, Baldwin IT, Hansson BS (2013) Feeding-induced rearrangement of green leaf volatiles reduces moth oviposition. Elife 2:e00421

Bate NJ, Riley JCM, JE T, SJ R (1998) Quantitative and qualitative difference in C6-volatile production from the lipoxygenase pathway in an alcohol dehydrogenase mutant of Arabidopsis thaliana. Physiol Plant 104:97–104

Chambers AH, Pillet J, Plotto A, Bai J, Whitaker VM, Folta KM (2014) Identification of a strawberry flavor gene candidate using an integrated genetic-genomic-analytical chemistry approach. Bmc Genomics 15 (1):217

Chen G, Hackett R, Walker D, Taylor A, Lin Z, Grierson D (2004) Identification of a specific isoform of tomato lipoxygenase (TomloxC) involved in the generation of fatty acid-derived flavor compounds. Plant physiology 136 (1):2641-2651.

D'Auria JC, Pichersky E, Schaub A, Hansel A, Gershenzon J (2007) Characterization of a BAHD acyltransferase responsible for producing the green leaf volatile (Z)-3-hexen-1-yl acetate in Arabidopsis thaliana. The Plant journal : for cell and molecular biology 49 (2):194-207.

Edger PP, VanBuren R, Colle M, Poorten TJ, Wai CM, Niederhuth CE, Alger EI, Ou S, Acharya CB, Wang J, Callow P, McKain MR, Shi J, Collier C, Xiong Z, Mower JP, Slovin JP, Hytonen T, Jiang N, Childs KL, Knapp SJ (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.

Hatanaka A, Kajiwara T, Sekiya J (1987) Biosynthetic pathway for C6-aldehydes formation from linolenic acid in green leaves.

Hawkins C, Caruana J, Schiksnis E, Liu Z (2016) Genome-scale DNA variant analysis and functional validation of a SNP underlying yellow fruit color in wild strawberry. Sci Rep 6:29017.

Hoffmann T, Kalinowski G, Schwab W (2006) RNAi‐induced silencing of gene expression in strawberry fruit (Fragaria× ananassa) by agroinfiltration: a rapid assay for gene function analysis. The Plant Journal 48 (5):818-826

Howe GA, Lee GI, Itoh A, ., Li L, ., Derocher AE (2000) Cytochrome P450-dependent metabolism of oxylipins in tomato. Cloning and expression of allene oxide synthase and fatty acid hydroperoxide lyase. Plant physiology 123 (2):711-724

Isabelle M, Michel J, Christophe A (2004) Changes in physicochemical characteristics and volatile constituents of strawberry (Cv. Cigaline) during maturation. Journal of Agricultural & Food Chemistry 52 (5):1248-1254

James DG (2005) Further field evaluation of synthetic herbivore-induced plan volatiles as attractants for beneficial insects. Journal of chemical ecology 31 (3):481-495

Jetti RR, Yang E, Kurnianta A, Finn C, Qian MC (2007) Quantification of selected aroma-active compounds in strawberries by headspace solid-phase microextraction gas chromatography and correlation with sensory descriptive analysis. Journal of food science 72 (7):S487-496.

Jung S, Lee T, Cheng CH, Buble K, Zheng P, Yu J, Humann J, Ficklin SP, Gasic K, Scott K, Frank M, Ru S, Hough H, Evans K, Peace C, Olmstead M, DeVetter LW, McFerson J, Coe M, Wegrzyn JL, Staton ME, Abbott AG, Main D (2019) 15 years of GDR: New data and functionality in the Genome Database for Rosaceae. Nucleic acids research 47 (D1):D1137-d1145.

Karimi M, De Meyer B, Hilson P (2005) Modular cloning in plant cells. Trends in plant science 10 (3):103-105

Karimi M, Inzé D, Depicker A (2002) GATEWAY™ vectors for Agrobacterium-mediated plant transformation. Trends in plant science 7 (5):193-195

Katoh K, Kuma K, Toh H, Miyata T (2005) MAFFT version 5: improvement in accuracy of multiple sequence alignment. Nucleic acids research 33 (2):511-518.

Kessler A, Halitschke R, Diezel C, Baldwin IT (2006) Priming of plant defense responses in nature by airborne signaling between Artemisia tridentata and Nicotiana attenuata. Oecologia 148 (2):280-292

Koskela EA, Sã¸Nsteby A, Flachowsky H, Heide OM, Hanke MV, Elomaa P, Hytã¶Nen T (2016) TERMINAL FLOWER1 is a breeding target for a novel everbearing trait and tailored flowering responses in cultivated strawberry (FragariaÂ Ã—Â ananassa Duch.). Plant Biotechnology Journal 14 (9):1852-1861

Kunishima M, Yamauchi Y, Mizutani M, Kuse M, Takikawa H, Sugimoto Y (2016) Identification of (Z)-3:(E)-2-hexenal isomerases essential to the production of the leaf aldehyde in plants. Journal of Biological Chemistry 291 (27):jbc.M116.726687

Latrasse A (1991) Fruits III, in Volatile Compounds in Foods and Beverages. New York, USA

Loughrin JH, Kasperbauer MJ (2002) Aroma of fresh strawberries is enhanced by ripening over red versus black mulch. J Agric Food Chem 50 (1):161-165.

Luo H, Dai C, Li Y, Feng J, Liu Z, Kang C (2018) Reduced Anthocyanins in Petioles codes for a GST anthocyanin transporter that is essential for the foliage and fruit coloration in strawberry. J Exp Bot 69 (10):2595-2608.

Monte D, Somerville S (2002) Pine tree method for isolation of plant RNA. In DNA Microarrays: A Molecular Cloning Manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor

Negri AS, Allegra D, Simoni L, Rusconi F, Tonelli C, Espen L, Galbiati M (2015) Comparative analysis of fruit aroma patterns in the domesticated wild strawberries "Profumata di Tortona"(F. moschata) and "Regina delle Valli"(F. vesca). Frontiers in plant science 6:56

Olbricht K, Grafe C, Weiss K, Ulrich D (2008) Inheritance of aroma compounds in a model population of Fragaria×ananassa Duch. Plant Breeding 127 (1):87-93

Pérez AG, Raquel O, Pilar L, Carlos S (2002) Biosynthesis of strawberry aroma compounds through amino acid metabolism. J Agric Food Chem 50 (14):4037-4042

Pfaffl MW (2001) A new mathematical model for relative quantification in real-time RT-PCR. Nucleic acids research 29 (9):e45.

Scala A, Allmann S, Mirabella R, Haring MA, Schuurink RC (2013) Green leaf volatiles: a plant's multifunctional weapon against herbivores and pathogens. International journal of molecular sciences 14 (9):17781-17811

Schieberle P, Hofmann T (1997) Evaluation of the Character Impact Odorants in Fresh Strawberry Juice by Quantitative Measurements and Sensory Studies on Model Mixtures. J Agric Food Chem 45: 227–232

Schwab W, Davidovich-Rikanati R, Lewinsohn E (2008) Biosynthesis of plant-derived flavor compounds. The Plant journal : for cell and molecular biology 54 (4):712-732.

Schwieterman ML, Colquhoun TA, Jaworski EA, Bartoshuk LM, Gilbert JL, Tieman DM, Odabasi AZ, Moskowitz HR, Folta KM, Klee HJ, Sims CA, Whitaker VM, Clark DG (2014) Strawberry flavor: diverse chemical compositions, a seasonal influence, and effects on sensory perception. PloS one 9 (2):e88446.

Spyropoulou EA, Dekker HL, Steemers L, van Maarseveen JH, de Koster CG, Haring MA, Schuurink RC, Allmann S (2017) Identification and Characterization of (3Z):(2E)-Hexenal Isomerases from Cucumber. Frontiers in Plant Science 8:1342-

Stothard P, . (2000) The sequence manipulation suite: JavaScript programs for analyzing and formatting protein and DNA sequences. Biotechniques 28 (6):1102, 1104

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Molecular biology and evolution 30 (12):2725-2729. Taylor AJ, Roberts DD, Taylor AJ, Roberts DD (2010) Flavor perception. Flavor Perception 41 (7):865-865

Ulrich D, Hoberg E, Rapp A, Kecke S (1997) Analysis of strawberry flavour – discrimination of aroma types by quantification of volatile compounds. Zeitschrift für Lebensmitteluntersuchung und -Forschung A 205 (3):218-223

Ulrich D, Komes D, Olbricht K, Hoberg E (2007) Diversity of aroma patterns in wild and cultivated *Fragaria* accessions. Genetic Resources and Crop Evolution 54 (6):1185

Urrutia M, Rambla JL, Alexiou KG, Granell A, Monfort A (2017) Genetic analysis of the wild strawberry (*Fragaria vesca*) volatile composition. Plant physiology and biochemistry : PPB 121:99-117.

Yasuo Y, Ayaka H, Ai T, Masaharu M, Yukihiro S (2011) NADPH-dependent reductases involved in the detoxification of reactive carbonyls in plants. Journal of Biological Chemistry 286 (9):6999-7009

Zabetakis I, Holden MA (2015) Strawberry Flavour: Analysis and Biosynthesis. Journal of the Science of Food & Agriculture 74 (4):421-434

Supplementary Material

Table S2-1. Near-isogenic lines used in the experiments.

NIL name	1st marker (reference)	1st marker position on LG5 (Mb*)	2nd marker (reference)	2nd marker position on LG5(Mb*)
Fb5:0-4**	CFV-3072	0.06	EMFvi108	0.05
Fb5:0-11	CFV-3072	0.06	CFV-3132	1.4
Fb5:0-20	CFV-3072	0.06	CEL2	2.6
Fb5:0-35	CFV-3072	0.06	FvH4095	5.7
Fb5:11-35	CFV-3132	1.4	FvH4095	5.7
Fb5:11-76	CFV-3132	1.4	EMFv024	23.0
Fb5:20-76	CEL2	2.6	EMFv024	23.0
Fb5:39-76	CFVCT024	8.3	EMFv024	23.0
Fb5:41-76	UDF009	8.6	EMFv024	23.0
Fb5:50-76	EMFn010	11.8	EMFv024	23.0
Fb5:0-76	CFV-3072	0.06	EMFv024	23.0
Fb7:0-10	CFV3096	5.2	EMFn201	12.3
Fb7:0-27	CFV3096	5.2	ARSFL099	19.6
Fb7:26-59	EMFvi008	19.2	ChFaM010	24.1

*megabase pairs of the *F. vesca* genome V4, pseudochromosome 5. Marker locations were determined by using the primer sequences as queries for a BLAST search against the *F. vesca* genome V4

** NIL names refer to the linkage group positions (in cMs) of the flanking markers.

Table S2-2. RT-qPCR primers for the volatile compounds related candidate genes

Primer name	Target gene	Direction	Fragariavesca v4.0 hit	Primer sequence 5'-	Length	Product size
q03_MSI1-R	MSI1(housekeeping gene)	R	FvH4_7g08380	ACACCATCAGTCTCCTGCCAAG	22	107
q04_MSI1-F		F		TCCCCACACCTTTGATTGCCA	21	
q19_5g02520-F	GLABRA3	F	FvH4_5g02520	TGTCTTTCGTCTTCAACATTGG	22	111
q20_5g02520-R		R		AGCGACTAAACACTTTACTATCTGC	25	
q32_3Z2E-F	3Z-2E-enal isomerase	F	FvH4_5G29270	GAGGGAGATGGTGGAGGATA	20	197
q33_3Z2E_R		R		ACCACCTCCTCCGATGTGT	19	
q34_5g26250-F	LOX	F	FvH4_5g26250	GCTCGGCAAAACCTTTCTTC	20	199
q35_5g26250_R		R		CTCCTCGTGGTGCTGATTTTC	21	
q36_5g06470-F	Terpene synthase	F	FvH4_5g06470	TCAGCATCTCAAGCTCAAACC	21	199
q37_5g06470-R		R		TCTGTGGAGAACTAAGTCGTAGC	23	
q38_5g06530-R	Terpene synthase	F	FvH4_5g06530	CAAACCCAAAATGGCAGTAGC	21	120
q39_5g06530-R		R		GGCTTCAATATCAGCAGTTTCC	22	
q40_5g32460-F	TT2 transcription factor	F	FvH4_5g32460	CACGGCGAAGGAAAGTGG	18	114
q41_5g32460-R		R		ATTGCCCCTCTTGATATTGG	20	
q42_5g35700-F	Terpene synthase	F	FvH4_5g35700	TCACTTTCTCTCTTACGCTACGG	23	116
q43_5g35700-R		R		TTTCTGAAGGCTCTTTCACTGG	22	
q54_5g06590-F	Putative monoterpene synthase	F	FvH4_5g06590	ACCAAATGAAGGGTGTCATCG	21	86
q55_5g06590-R		R		TAGCTCAGCCATAATCCAGACC	22	
q56_5g30030-F	putative arginosuccinate synthase	F	FvH4_5g30030	AAGAGTGAGTTACAGGGCAAGG	22	112
q57_5g30030-R		R		TTTCTGTCTCGCTGTCACTAGG	22	
q58_5g06920-F	putative arginosuccinate synthase	F	FvH4_5g06920	CAAAGCACATGCAGAAAGTAGG	22	104
q59_5g06920-R		R		GAACTCCGGCAATAAGATATGG	22	
q60_5g33530-F	amino acid permease	F	FvH4_5g33530	AACAGCAAGGACAACCTTAACC	22	128
q61_5g33530-R		R		CCAATGTGTCGGTAGAGTAAGC	22	
q62_7g00440-F	LOX	F	FvH4_7g00440	AGTGAAGGCAGTGGTTACGG	20	190
q63_7g00440-R		R		ATTCACTTTGTGCGCATACCC	21	
q64_7g09160-F	acyl-coA hydrolase	F	FvH4_7g09160	GTTTGCGATCGATTAGAGAAGG	22	99
q65_7g09160-R		R		AGCTGATTTTTCCTCGGATAGC	22	
q66_7g06750-F	acyl-coA hydrolase	F	FvH4_7g06750	ATAGCCTGTCGTATTGCTAGGG	22	117
q67_7g06750-R		R		GACGAGTTCCAATTCACTAGGC	22	
q68_7g11730-F	acyltransferase	F	FvH4_7g11730	ATCCACTCTCTTTCTTCACACC	22	113
q69_7g11730-R		R		GACCCCAACCAGGTATTCTAGG	22	
q70_7g07050-F	MYC2 TF	F	FvH4_7g07050	ACCAAGCACTTTGCTACTGAGC	22	114
q71_7g07050-R		R		CTCACTCACAGTCCTTTTGAGC	22	
q72_7g16010-F	putative hydroxyisobutytyl-CoA hydrolase	F	FvH4_7g16010	AAGTGGGAGCCTAGTAAGTTGG	22	105
q73_7g16010-R		R		TGCAGGGAGCTTAAATACTTCC	22	
q74_7g04560-F	acyltransferase	F	FvH4_7g04560	GCAAGCTCAGTAAAGGTGTTGG	22	99
q75_7g04560-R		R		CTGCAGTTGCTAGGAAAACTCC	22	

Table S2-3. NCBI accession numbers for proteins used for constructing phylogenetic trees.

Abbreviation	Plant species	NCBI/GDR accession
AtHI-like	*Arabidopsis thaliana*	XP_002892418.1
AtGermin1	*Arabidopsis thaliana*	NP_187070.1
AtHI-like1	*Arabidopsis thaliana*	NP_180436.1
AtHI-like2	*Arabidopsis thaliana*	NP_191481.1
CaHI	*Capsicum annuum*	LC146479
CmHI	*Cucumismelo*	XP_008456077.1
CmHI-like1	*Cucumismelo*	XP_008461502.1
Cs11S	*Cucumissativus*	XP_011651441.1
CsHI2	*Cucumissativus*	XP_004139714.2
CsHI3	*Cucumissativus*	XP_004151504.1
CsHI4	*Cucumis sativus*	XP_011651276.1
FvHI	*Fragariavesca*	FvH4_5g29270
FvHI-like1	*Fragariavesca*	FvH4_2g37390
FvHI-like2	*Fragariavesca*	FvH4_3g34840
FvHI-like3	*Fragariavesca*	FvH4_3g34830
GmGermin	*Glycine max*	NP_001241457.1
GmHI-like1	*Glycine max*	NP_001241132.1
GmHI-like2	*Glycine max*	XP_003538022.1
GmHI2	*Glycine max*	XP_003525010.1
MtGermin	*Medicagotruncatula*	XP_013470283.1
MtHI-like1	*Medicagotruncatula*	XP_003607149.1
MtHI-like2	*Medicagotruncatula*	XP_003605501.1
MtHI1	*Medicagotruncatula*	XP_003629975.1
MtHI2	*Medicagotruncatula*	XP_003629976.2
MtVicilin	*Medicagotruncatula*	XP_003624146.1
OsGermin	*Oryzasativa*	XP_015612310.1
OsHI-like1	*Oryzasativa*	XP_015639453.1
OsHI-like2	*Oryzasativa*	XP_015641111.1
OsHI4	*Oryzasativa*	XP_015632411.1
PtHI	*Populustrichocarpa*	XP_006370285.1
Sl11S1	*Solanumlycopersicum*	XP_004247523.1
SlLegumin	*Solanumlycopersicum*	XP_004234041.1
SlVicilin	*Solanumlycopersicum*	NP_001308118
St11S1	*Solanumtuberosum*	XP_006351693.1
StGermin	*Solanumtuberosum*	NP_001275369.1
StHI-like2	*Solanumtuberosum*	XP_006339780.1
StHI1	*Solanumtuberosum*	XP_006349431.1
StHI2	*Solanumtuberosum*	XP_006349432.1
StLegumin	*Solanumtuberosum*	XP_006356113.1
VvGermin	*Vitisvinifera*	NP_001267944.1
ZmGermin1	*Zea mays*	NP_001148746.1
ZmHI-like	*Zea mays*	NP_001105204.1

Table S2-4. Volatile compounds summary. Data are expressed as the ratio between samples and a reference. Mean ratios±standard deviation (sd) were calculated for each compound in the recurrent parental F. vesca (RV) and in the 11 NILscovering all LG5. **A)** First harvest on May in 2018. **B)** Second harvest on July in 2018.

A

Compound	Famlily	LG5:0-4	LG5:0-11	LG5:0-20	LG5:0-35	LG5:11-35	LG5:11-76	LG5:20-76	LG5:39-76	LG5:41-76	LG5:50-76	LG5:0-76	RV
E-2-hexenal	aldehyde	0.15±0.07	0.09±0.06	0.11±0.08	0.44±0.53	0.58±0.47	0.27±0.26	-	0.37±0.32	0.17±0.24	0.19±0.12	0.29±0.26	0.59±0.42
Butyl acetate	ester	2.31±2.05	7.79±6.63	3.66±4.14	4.76±2.45	4.62±2.57	21.93±2.43	-	14.07±5.41	31.64±9.13	18.16±18.17	1.81±2.93	4.26±1.7
Butyl butanoate	ester	0.08±0.07	0.22±0.24	0.1±0.1	0.07±0.06	0.11±0.08	1.86±0.53	-	2.98±2.52	4.25±2.54	2.92±2.36	0.11±0.06	0.65±0.34
E-2-hexenyl acetate	ester	2.87±1.18	4.75±2.88	3.06±1.56	5.96±2.49	6.41±2.82	0±0	-	0±0	0±0	0±0	0±0	7.25±1.55
Ehtylhexanoate	ester	4.53±4.79	4.29±2.93	1.16±0.67	4.78±0.97	15.6±9.29	25.09±8.56	-	22.29±17.08	12.41±3.98	33.24±31.89	0.86±0.92	6.1±2.56
Ethyl butanoate	ester	5.05±4.66	7.56±4.95	1.82±1.55	2.53±0.5	7.64±4.47	25.62±4.95	-	17.56±3.81	28.76±9.43	40.09±18.99	0.51±0.63	7.14±2.03
Hexyl acetate	ester	4.63±2.08	12.87±7.55	7.58±2.87	11.07±3.34	7.13±2.08	22.7±1.93	-	22.79±6.39	32.9±19.99	33.78±3.38	7.07±4.07	8.87±0.95
Methyl butanoate	ester	0.95±1.07	1.29±1.11	0.62±0.64	0.46±0.23	0.24±0.18	1.34±0.81	-	3.06±0.84	5.78±2.57	6.31±0.07	0.06±0.11	1.47±0.17
Methyl cinnamate	ester	0.13±0.13	0.25±0.18	0.74±0.58	0.36±0.07	0.09±0.08	0.03±0.01	-	0.01±0	0±0.01	0.01±0.01	0.07±0.02	0.02±0.02
Methyl hexanoate	ester	1.49±1.23	1.35±1.14	0.46±0.23	1.74±0.6	1.52±1.11	3.38±1.26	-	4.36±1.22	7.18±3.48	14.74±12.54	0.22±0.24	1.65±0.08
Methyl-2aminobenzeoate	ester	8.2±8.9	20.03±22.67	8.6±7.74	0.17±0.01	0.09±0.06	0.32±0.24	-	23.55±8.98	5.97±2.63	23.27±2.57	0.1±0.08	3.97±3.33
Myrtenyl acetate	ester	2.81±1.13	6.33±4.24	4.09±2.51	0.86±0.33	0.38±0.2	2.07±0.31	-	4.67±1.24	8.05±4.97	2.68±1.45	1.33±0.51	2.88±1.95
Z-3-hexenyl acetate	ester	1.55±0.87	2.82±0.46	1.21±0.69	4.86±0.38	2.54±0.8	10.95±5.01	-	8.99±3.42	5.83±2.05	3.96±2.58	7.94±10.13	2.41±1.14
Mesifurane	furan	0.12±0.11	0.17±0.25	0.14±0.17	0.42±0.13	0.2±0.11	0.63±0.3	-	0.11±0.07	0.41±0.16	0.65±0.41	0.66±0.24	0.07±0.09
Gamma-decalactone	lactone	0.23±0.19	0.4±0.35	0.28±0.37	0.14±0.11	0.06±0.04	1.52±1.16	-	3.24±1.04	3.5±1.94	5.11±2.01	0.17±0.09	0.66±0.12
Linalool	terpenoid	0.05±0.02	0.13±0.05	0.03±0.04	0.19±0.02	0.04±0.04	0.25±0.08	-	0.19±0.06	0.17±0.08	0.17±0.06	0.19±0.03	0.05±0.06
Nerolidol	terpenoid	0±0	0±0.01	0±0	0±0	0±0	0±0	-	0±0	0.01±0	0±0	0±0	0±0

B

Compound	Famlily	LG5:0-4	LG5:0-11	LG5:0-20	LG5:0-35	LG5:11-35	LG5:11-76	LG5:20-76	LG5:39-76	LG5:41-76	LG5:50-76	LG5:0-76	RV
E-2-hexenal	aldehyde	0.15±0.09	2.06±1.28	0.61±0.13	1.03±0.33	1.05±0.4	0.63±0.56	0.96±0.48	0.46±0.35	0.63±0.4	0.41±0.16	0.61±0.28	1.67±0.78
Butyl acetate	ester	0.86±0.53	1.13±0.07	0.64±0.38	0.75±0.17	1.65±0.55	3.51±0.69	3.8±1.08	3.67±2.18	4.93±1.81	9.5±2.42	1.74±0.44	5.48±3.78
Butyl butanoate	ester	0.01±0.01	0±0	0±0	0±0	2.43±1.76	0±0	0.01±0.01	0.21±0.37	0.15±0.18	0.43±0.25	0.01±0.03	0.08±0.06
E-2-hexenyl acetate	ester	1.23±0.3	1.58±0.4	0.82±0.15	0.94±0.26	0.87±0.4	0±0	0±0	0±0	0±0	0±0	0±0	2±0.69
Ehtylhexanoate	ester	2.2±1.45	4.71±0.32	4.51±0.66	4.9±1.14	1.62±3.23	9.48±2.12	7.4±3.36	7.22±3.48	8.7±3.73	9.48±1.82	2.58±1.73	12±8.15
Ethyl butanoate	ester	0.61±0.39	1.59±0.43	1.61±1.05	0.61±0.45	4.23±1.89	3.24±2.2	4.74±1.71	3.99±3.51	3.94±2.41	8.73±2.8	2.2±1.46	7.77±2.57
Hexyl acetate	ester	1.36±0.33	1.64±0.18	1.41±0.18	1.56±0.21	1.97±0.62	4.1±0.62	4.06±0.86	5.02±1.39	4.92±1.82	5.41±4.66	1.36±1.22	3.58±2.15
Methyl butanoate	ester	0.26±0.21	0.06±0.03	0.04±0.04	0±0	0.02±0.02	0.08±0.01	0.09±0.06	0.36±0.26	0.23±0.08	0.56±0.27	0.01±0.02	0.67±0.47
Methyl cinnamate	ester	0.01±0.01	0.04±0.03	0.03±0	0.01±0.01	0.01±0.02	0.01±0.01	0±0	0±0	0±0	0±0	0±0	0±0
Methyl hexanoate	ester	0.26±0.2	0.45±0.07	0.46±0.2	0.3±0.06	0.74±0.22	1.08±0.28	0.71±0.35	1.21±0.36	0.93±0.43	1.37±0.5	0.14±0.08	2.17±1.91
Methyl-2aminobenzeoate	ester	2.09±2.09	3.84±0.29	0.64±0.16	0.06±0.01	0.13±0.08	0.16±0.08	0.02±0.01	5.57±3.41	1.38±1.26	4.63±2.56	0.05±0.04	7.27±5.05
Myrtenyl acetate	ester	0.61±0.15	0.86±0.16	0.63±0.2	0.1±0.02	0.08±0.01	0.26±0.06	0.18±0.03	0.99±0.44	1.11±0.54	1.11±0.48	0.22±0.13	1.34±0.87
Z-3-hexenyl acetate	ester	0.46±0.2	0.41±0.09	0.4±0.2	0.7±0.08	0.48±0.14	2.55±0.4	2.51±0.46	1.81±0.43	1.57±0.28	3.43±0.5	3.63±1.83	1.53±0.28
Mesifurane	furan	0.03±0.03	0.07±0.03	0.04±0.02	0.07±0.04	0.04±0.04	0.11±0.13	0.06±0.07	0.06±0.05	0.08±0.08	0.07±0.06	0.11±0.08	0.04±0.05
Gamma-decalactone	lactone	0.03±0.03	0.1±0.02	0.05±0.01	0.01±0.02	0.02±0.02	0.08±0.07	0.09±0.09	0.27±0.19	0.25±0.14	0.51±0.24	0±0.01	0.11±0.15
Linalool	terpenoid	0±0	0±0	0.01±0	0.01±0	0.03±0.03	0.08±0	0.04±0.04	0.04±0.04	0.07±0.04	0.06±0.03	0.1±0.06	0.01±0.01
Nerolidol	terpenoid	0±0	0.01±0.01	0±0	0±0	0±0	0±0	0±0	0±0	0±0	0±0	0±0	0±0

Table S2-5. *F. vesca* proteins with the highest similarity to cucumber (Z)-3:(E)-2-hexenal isomerase.

BLAST hit	e-value	Location in *F. vesca* genome V4
FvH4_5g29270	2.123 e-97	chrFb5: 20294931..20297997
FvH4_2g37390	2.098 e-67	chr2: 27271351..27272846
FvH4_3g34840	3.436 3-67	chr3: 30118297..30121482
FvH4_3g34830	2.290 e-63	chr3: 30110113..30114706

Table S2-6. The coding sequence of FvH4_5g29270

>FvH4_5g29270_CDS

ATGGCGGAAATGGATCTAACACCAAAGTCAGCGGCAGCAGCGTTCGAGGGAGATGGTGGAGGATATTACGTATGGTCATTTCCG
GCGCTTGGCGAGGCCAACGTAGGTGCCGGAAAGCTTGTGCTGAAGCCTAGTGGCTTTGCTCTTCCTCACTATGCAGATTCTGCCAA
ACTGGGATATGTTCTTCAAGGCGAGGATGGAGTAGTTGGAATGGTATTCCCCAACACATCGGAGGAGGTGGTGTTGAAGCTTAAG
AAAGGAGACGTGATTCCGGTACCACTCGGAGCAGTCTCATGGTGGTTCAACAATGGCGACTCAGCCGATGACCTTGTCATCGTGT
TCTTGGGCGAAACAACAAAGGCTTACACTCCTGGTGTATTTACTTATTTCTTCATTGCAGGAACTCAAAGTCTTCTCGGAGGTTTCT
CTACTGACTTCATTAGCAAGTCATTCAGTATTACCAAAGATGAAGCTGATGAGGTCACCAAAAACCAGACAGGAGTCCTGCTAGT
TAAGGTAGAAAAGGGAAAGACCATGCCTAAGCCCAACGCCCACCTCACCCACAAACTTGTTCATCAACTCAATGTCAGTGCCACT
GTAACTGAGAAGGAGTTTCCTTTTCTTAACCAAGCTGGGTTAAGTGCCAACCTCATAAAACTTGAACCTTCTGCAATTTCCTCCCC
CATTTACACAACCGATTCTACGGTTCAATTGATCTATGTGGTTGGAGGAGGGGGTCGGATCCAAATCACGGGTCTTAATGGTCAGC
GTGTGTTGGATGCGGAAGTAGCTGCCGGTCAGTTGATCGTTGTGCCTAGGTTTTTCATGGTGGCGAAACTTGCCGGTGAAAAAGG
AATGGAATGTTTCTCTGTTATTACAAGTTCCCGGGCTACTCTGGAAGACTTTACTGGCAAGACATCAGTGTTGAGGGCATTATCAC
CTGAGGTGCTACAAATATCCCTCAATATAAACCCAGAATTGCAGACTCTTCTGCAGTCAAAGAGTGACTGAATCCTGCACCATCA
ATCGAAGTCTGT

Table S2-7. Summary of green leaf volatile compounds detected in Valencia.

	RV		Fb5:39-76		Fb5:41-76	
Compound	*FvHI*-ox	mock	*FvHI*-ox	mock	*FvHI*-ox	RV mock
(Z)-3-hexenal	675274.6667	916981.3	2518993	3811820	5050303	5052860
(E)-2-hexenal	22134801.67	23474574	11798330	5249297	9543522	4646213
(Z)-3-hexenyl acetate	1931563.333	2015721	5122034	9883199	6126429	6780770
(E)-2-hexenyl acetate	2037799.333	2078066	112588	134535.5	111849.5	70231.75

Data expressed as average values.

Chapter III

Genetic analysis of strawberry fruit colorand the function of *FvLhac4* in orange fruits

Introduction

The *Rosaceae* family is an economically important group of cultivated plants including most of the high commercial value fruits such as apple, pear, peach, strawberry and raspberry, as well as ornamental plants such as rose. In these fruits and flowers, color is a key quality trait since it directly affects the appearance and the consumer acceptability of the product. Diploid strawberry (*F.vesca*) has been developed as a model species for the cultivated strawberry (*F. x ananassa*) and *Rosaceae* family (Hawkins et al. 2016). Strawberry cultivars exhibit a continuous range in fruit color from white fruits of 'WeisseAnanas' or light orange of 'Madame Moutot', to blackish red of 'Rubina' (http://www.upov.int/en/publications/tg-rom/tg022/tg_22_10.pdf). Consumer's preferences for strawberry fruit color can change over timewith orange red fruits favored over dark red fruits nowadays in the majority of markets. Also, consumers from different locations prefer different color characteristics. For example, Finnish consumers tend to prefer darker-colored strawberries than Mediterranean consumers and Chinese and Japanese consumers prefer white fruits. Therefore, elucidating genes that control this trait is an important goal in strawberry breeding (Zorrilla-Fontanesi et al. 2011).

In strawberry fruit, both chlorophyll (Chl) and anthocyanin are important for fruit coloration (Kayesh et al. 2013). Chloroplasts play a vital photosynthetic role in green fruits and are degraded in the progress of fruit ripening and senescence accompanied with high amount of anthocyanin synthesis (Li et al. 2019a). The green color of unripe fruits is dueto accumulation of chlorophyll in fruit skin and flesh.

Diploid white strawberries are naturally present and co-exist with red varieties in the wild, including the sequenced *F. vesca* spp. *vesca* var. Hawaii-4 (Shulaev et al. 2011a), var. Yellow Wonder, var. White Solemacher, var. Pineapple Crush and var. White Soul (Urrutia and Monfort 2018). The fruit color of strawberry was long ago described to be controlled by one major gene locus(C) that was mapped in the end region of LG1, red being dominant to white fruit (Brown and Wareing 1965; Deng and Davis 2001). Later studies revealed that a natural mutation of a single nucleotide polymorphism producing a non-synonymous change in *FvMYB10* gene, which encodes a key transcription factor for anthocyanin biosynthesis, is responsible for changes in pigmentation between red and white diploid strawberry accessions (Hawkins et al. 2016). Recently, a *reduced anthocyanin in petioles* (*rap*) gene was identified to encode the RAP as the principal GST transporter of anthocyanins in the strawberry foliage and fruit, and it could be modified to alter the fruit color in strawberry (Luo et al. 2018).

Chlorophyll is contained in the photosystems in the thylakoid membranes in a set of proteins collectively designated as light-harvesting chlorophyll *a/b*-binding (LHC) polypeptides (Holtzegel 2016). In

higherplants, LHC proteins include a large gene family containing 10-12 members, constituting the peripheral light-harvesting antenna of the photosystem I (PSI) and photosystem II (PSII) (Figure 3-1). LHCAs are encoded by *lhca1* through *lhca6*, and LHCBs are encoded by *lhcb1* through *lhcb6* (Green and Durnford 1996; Jansson 1999; Dekker and Boekema 2005; Daum et al. 2010). All LHCs contain three membrane-spanning helices, and helix 1 and 3 are homologous. They may have originated via internal duplication and share the characteristic LHC motif (ExxxxRxAM), in which E (Glu) and R (Arg) areinvariant (Corbet et al. 2007). Each helix binds chlorophyll molecules (chlorophyll *a* and *b*) and some carotenoids within the thylakoid membrane. Thus far, the LHC superfamily has been extensively identified from several of algae and higher plants such *as Arabidopsis*, rice, cassava and coater bean (Klimmek et al. 2006; Mozzo et al. 2010; Jansson 1999; Zou and Yang 2019). Moreover, functional analysis has also been performed for a number of LHC superfamily genes (de Bianchi et al. 2011; Beck et al. 2017). Characterization of knockout lines for *Lhcb4* isoform of *Arabidopsis* suggested that Lhcb4 is unique among PSII antenna proteins and determinant for PSII macro-organization and photoprotection (de Bianchi et al. 2011). A *Brassica napus* accession was identified carrying the stay-green gene *NON-YELLOWING 1*(*NYE1*) deletion associated with increased chlorophyll content, and with upregulated expression of *Lhc* genes (Qian et al. 2016). The *delayed yellowing 1* (*DYE1*) gene encoding LHCA4 was cloned from a rice mutation *dye1-1* whose leaf yellowing was delayed in the field. In *dye1-1*, an amino acid substitution occurs at the location of a highly conserved amino acid residue involved in pigment binding which caused severely impaired structure of the PSI-LHCI super-complex, resulted in high accumulation of Lhcb1, higherchlorophyll content, and finally in leaves remaining green. By contrast, research on LHCAs in fruits and strawberry remains scarce.

To determine color composition has been described two methods. RGB system measures the strength of R (red), G (green) and B (blue) color in each pixel to reproduce other colors. The CIELab color space is able to approximate human visual perception. It is a spherical color space with the vertical axis representing lightness (+L*) to darkness (-L*). The chromaticity coordinates are a* and b* and their axis indicates color directions: +a* for the red direction, -a* for the green direction, +b* for the yellow direction and –b* for the blue direction. Hue and chroma are descriptors of color based on a* and b* values. Hue represents the basic color. It is an angular measurement in the quadrant between the a* and b* axes. Chroma is the saturation or vividness of color. It is measured radially from the center of each quadrant with the a* and b* axes (Strecker et al. 2010).

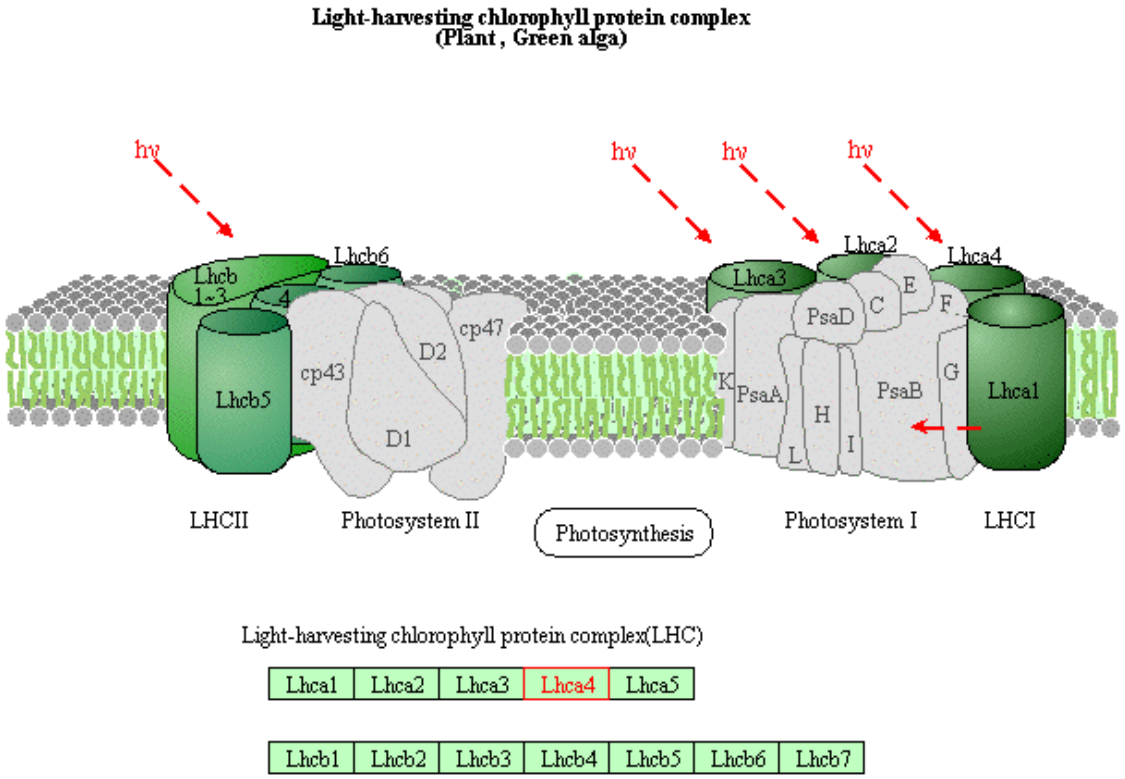

Figure 3-1.The pathway of photosynthesis-antenna proteins.(based on KEGG pathway, http://www.kegg.jp/dbget-bin/www_bget?map00196).

In this study, we observed ripe fruits of specific NILs presenting orange color while most of the NILshadred fruits. RGB color space and CIELab color space were applied to analyze the fruit color difference. The result of fruit color analysisshowed that the orange colored fruits were detected to have significantly higher values in the green parameter and lower values in a*. The QTL for orange color was mapped in LG5:35-39cM. In this region, a candidate gene was identified encoding the light-harvesting chlorophyll a/b binding protein and its genetic function was studied.

Material and methods

Plant material

The fruit color of diploid strawberry ripe fruits were studied using 49 lines of a near isogenic line (NIL) collection described in detail in Urrutia *et al* (2015), the recurrent parent 'Reine del Vallées' (RV) and the white fruited variety of *F.vesca* var. 'YellowWonder' (YW). Six plant replicates per genotype were grown under greenhouse conditions at day temperature between 22 to 24°C and 17°C during night without artificial lighting, relative humidity 40-50% at the Centre for Research in Agricultural Genomics (CRAG) in Bellaterra(latitude:41° 29'N, longitude: 2° 06'E) from October to March for three consecutive years (2016, 2017 and 2018).In March, four individual plants of each genotype were transferred to a shaded greenhouse at Centre Torre Marimon in Caldes de Montbui (latitude 41°36'N, longitude 2°10'E at 203 m of altitude from sea level).The plants grew in natural conditions and photoperiod without receiving additional lighting or heating from March to September. Each harvest year was considered an independent experiment. Three to five biological replicates of ripe fruits were harvested each year.

Sampling

For fruit color analysis from shaded greenhouse plants, samples were harvested from May to Julyeach year. Fully ripe red berries from individual plants were pooled together and analysed as independent biological replicates. Each biological replicate was mixed of at least 10 fruits kept in ice and processed immediately.

The fruit samples from six NILs (Fb5:0-35, Fb5:11-76, Fb5:20-76, Fb5:39-76, Fb5:41-76 and Fb5:0-76) and their recurrent parent RVwere used for RNA extraction at different fruit stages as described in chapter II.

Plant tissue samples for studying tissue-specific expression patterns were collected from the same plants which were used to collect fruits samplesin shaded greenhouse.For leaf tissue, the unopened leaf and fully opened leaf were sampled. For root samples, two-centimetre root tips from actively growing roots were pooled together. All samples were immediately frozen in liquid nitrogen and stored at -75°C until RNA extraction.

Fruit color analysis

For each biological replicate, six to ten fruit samples were randomly chosen from collected samples and whole fruits scanned one by one using a flatbed scanner (HP Scanjet 8200, USA). Then the fruits were

pooled together and crushed using a handheld blender W112 (Dynamix, Greece), and the puree was scanned in a transparent petri dish (Figure 3-1).

The pictures were analyzed employing Tomato Analyzer 3.0 (Strecker et al. 2010)designed to quantify the color parameters *R* (red), *G* (green), and *B* (blue) of the RGB color space(Strecker et al. 2010). The average RGB values were employed to calculate *L**, *a**, *b** of the CIELAB color space and hue and chroma color descriptors. The scanner color calibration was achieved employing Color checker Munsell Color X-write.

Statistical analyses were achieved using the JMP®8.0.1 statistical package (SAS Institute, NC, USA). The mean and ANOVA test were calculated for each trait. Means were compared with the recurrent parent by a Dunnett's test with $\alpha \leq 0.05$. QTL was determined when all significant lines shared the same introgression interval.

Identification of Light-harvesting complex (lhc) superfamily genes in *F.vesca*

An *Lhc* genes list of *Arabidopsis* has been reported by Umate (Umate 2010) and the protein sequences were acquired from TAIR10 (Lamesch *et al.* 2011) and used as queries to search for *F. vesca* LHC homologs in the *Fragaria vesca* genome V4 protein database (Jung et al. 2019)using BLASTP in GDR (https://www.rosaceae.org/species/fragaria/all). Matches with bit scores higher than 200 and lowest e-value were selected.

Protein sequence alignment and phylogenetic reconstruction

The amino acid sequences of proteins (accession numbers in Table S3-3) were aligned by MAFFT (Katoh et al. 2005) using E-INS-I iterative refinement method with default settings and retaining gappy regions. The resulting amino acid alignment was used as input for MEGA6 (Tamura *et al.* 2013) software. Phylogenetic tree was constructed using maximum likelihood algorithm with default settings and 1000 bootstrap replications.

Amino acid sequences of light-harvesting complex I chlorophyll a-b binding protein 4 (Lhca4) from *F. vesca* and various other specieswere aligned by MAFFT with default settings. The resulting amino acid alignments were used as input for the Color Align Properties tool of the DNA Sequence Manipulation Suite (Stothard 2000).

Measurement of chlorophyll content

Chlorophyll extraction used the modified method of (Hiscox and Israelstam 1979). One hundred milligrams of fruit tissue powder was placed in a vial containing 7 mL DMSO (Sigma, USA) and chlorophyll was extracted into the fluid without grinding at 65°C by incubating for two hours. The extract liquid was transferred to a graduated tube and made up to atotal volume of 10 ml with DMSO, assayed immediately.

A 3.0 mL sample of chlorophyll extract was transferred to a cuvette, and the OD values at 645 and 663 nm were read in a SpectraMax® M3 Microplate Reader (Molecular Devices LLC, USA)against a DMSO blank. Chlorophyll content was calculated following the equation used by Arnon (1949).

Chlorophyll a, chlorophyll b and total chlorophyll content (dissolved in DMSO) can be calculated based on absorbance values at 645 and 663 nm as follows(Arnon 1949):

Total chlorophyll (mg/l): C (total) = 20.2 * OD_{645} + 8.02 * OD_{663}

Chlorophyll a (mg/l): C_a = 12.7 * OD_{663}- 2.69 * OD_{645}

Chlorophyll b (mg/l): C_b= 22.9 * OD_{645} - 4.68 * OD_{663}

RNA extraction and qRT-PCR

Protocols for RNA extraction have been described in material and methods of chapter II (page 68). qRT-PCR method has been described also in chapter II (page 68).

Transient overexpression/silencing vector construction and transformation

All methods used in this progress are the same as described in chapter II with different gene-specific primers (Table3-1)

Table 3-1.Sequences of the primers used in the study.

primer name	Seq	used for
FC01_5g14770-F	ATGGCAACCGTCACGACTCA	Amplification of color gene coding sequence
FC02_5g14770-R	CTAGCCCCTTAGTGTTTGGAC	Amplification of color gene coding sequence
attB1-5g14770-F	AAAAAGCAGGCTTCGAAGGAGATAGAACCATG GCAACCGTCACGACTCA	Gene-specific Gateway primer for overexpression
attB2-5g14770-R	AGAAAGCTGGGTCTAGCCCCTTAGTGTTTGGAC	Gene-specific Gateway primer for overexpression
attB1_5g14770_RNAi_F	AAAAAGCAGGCTATCCACCCTGTTTGTGATCG	Gene-specific Gateway primer for silencing
attB2_5g14770_RNAi_R	AGAAAGCTGGGTACCTCATTTGGAGGCAAGC	Gene-specific Gateway primer for silencing
attB1_adapter	GGGGACAAGTTTGTACAAAAAAGCAGGCT	Gateway adapter
attB2_adapter	GGGGACCACTTTGTACAAGAAAGCTGGGT	Gateway adapter

Result

Fruit color analysis

Fruit color is an important trait for consumer perception, and can also be associated with anthocyanin and carotenoid content, which are related to the nutritional quality of the fruit. In this study, fruits from the entire NIL collection were analyzed for eight fruit color-related parameters (red, green, blue, ***L*, a*, b*,***chroma and hue) for three years. A main fruit color effect in NILs Fb5:0-76, Fb5:11-76 and Fb5:20-76 (Figure3-2A) was discovered by a scanner observation of fruits, with whole fruits and fruit puree showing an orange color. Color parameter analyses showed that in these NILs, significantly higher values for the red, green and hue values and significantly lower values for ***a****values could be observed (Figure3-2B). Scanner data values of three years were statistically analyzed (Table S3-1). These three NILs (Fb5:0-76, Fb5:11-76 and Fb5:20-76) define a major QTL affecting fruit color on chromosome 5 between 20 to 76cM. The minimum interval could be mapped between 35 -39 cM (Figure 3-3) since theline Fb5:39-76 shows a low level of green color. RV as recurrent parent and YW as external control line were also analyzed. YW presented significant difference with RV forall eight color parameters (Table S3-1). However, differences between yellow and orange lines are related to an increased value of the green parameter while they showed similarly high level of red parameter.

Identification of candidate genes for fruit orange color

Based on the results of fruit color analysis in the NIL collection, a QTL was mapped in 35-39cM in LG5. The physical positions of exotic introgressions in NIL collection were refined according to the strawberry reference genome version2.0(Urrutia Rosauro 2015). This QTL region corresponded to a physical distance of 2.99Mb (LG5: 5.953.948-8.945.335) and a reference sequence was acquired. Later, the *Fragaria vesca* V4.0 a1 genome database (Edger *et al.* 2018) was published and 416 genes were annotated in this region (Table S3-2). Based on the description of genes and our goal, the genes involved in the synthesis and degradation pathways for pigments related with fruit color such asanthocyanin, carotenoid and chlorophyll were selected. Finally, 18 genes were preliminary considered as candidate genes (Table 3-2) and RT-qPCR primers were designed (Table S3-3).

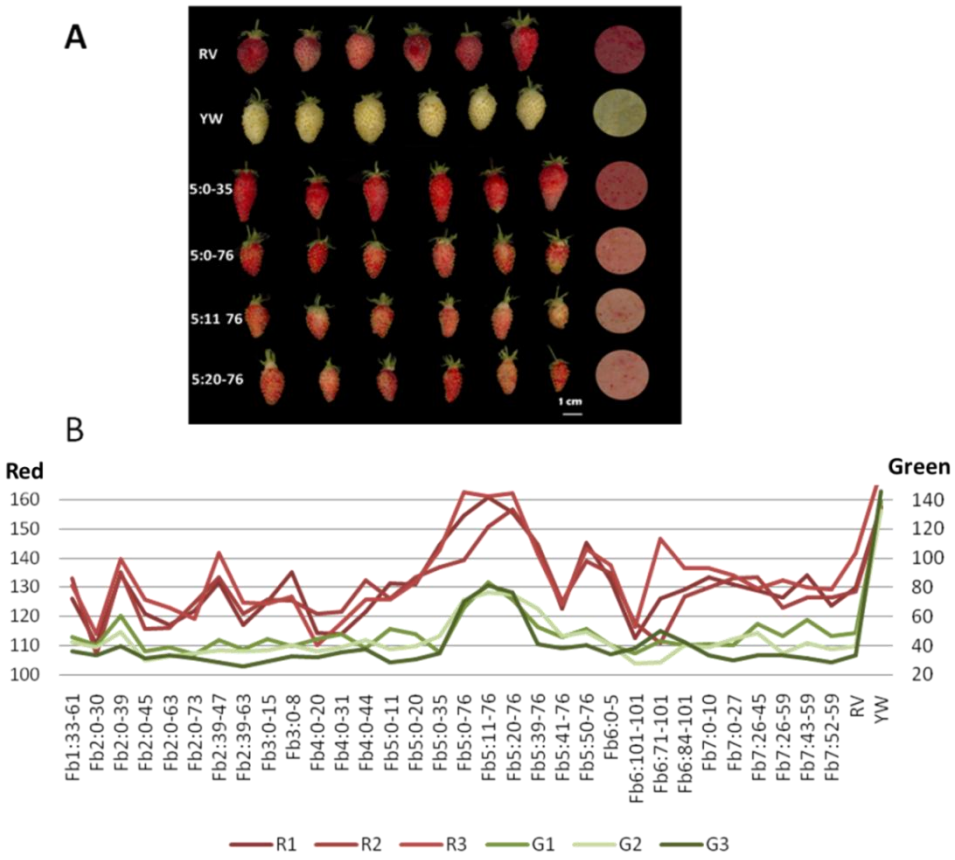

Figure 3-2.Strawberry fruit color analysis.A: Fruit color in whole and pureed fruit in RV, YW and NILs. **B**: Line chart of red and green factors of color in NIL collection for three years. R and G present red and green, respectively. 1, 2 and 3 mean in 2016, 2017 and 2018.

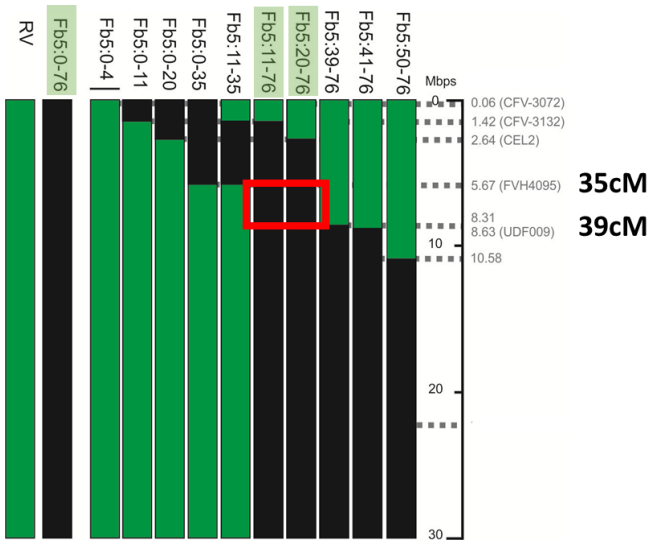

Figure 3-3.Graphical representation the QTL region for fruit orange color. The chromosomal location for the orange color QTL is marked by the red box.

Table 3-2.Candidate genes for fruit orange color presentwithinthe QTL region LG5:35-39.

V4_number	description	genome position
FvH4_5g10850.1	Transcription factor MYC/MYB N-terminal	Fvb5_v4.0.a1:6131847..6133470
FvH4_5g10880.1	Cytochrome P450	Fvb5_v4.0.a1:6163843..6166082
FvH4_5g11300.1	Transcription factor MYC/MYB N-terminal	Fvb5_v4.0.a1:6402371..6411521
FvH4_5g11430.1	SANT/Myb domain	Fvb5_v4.0.a1:6483578..6484431
FvH4_5g11930.1	SANT/Myb domain	Fvb5_v4.0.a1:6752057..6753723
FvH4_5g12340.1	Cytochrome cd1-nitrite reductase-like	Fvb5_v4.0.a1:6932229..6937606
FvH4_5g12400.1	Cytochrome P450	Fvb5_v4.0.a1:6972717..6986736
FvH4_5g12580.1	Cytochrome P450	Fvb5_v4.0.a1:7098823..7100337
FvH4_5g12670.1	SANT/Myb domain	Fvb5_v4.0.a1:7142582..7144426
FvH4_5g13590.1	Cytochrome P450	Fvb5_v4.0.a1:7695649..7698065
FvH4_5g13850.1	Cytochrome b245, heavy chain	Fvb5_v4.0.a1:7816138..7820214
FvH4_5g14010.1	Cytochrome P450	Fvb5_v4.0.a1:7931663..7935314
FvH4_5g14660.1	SANT/Myb domain	Fvb5_v4.0.a1:8312994..8323259
FvH4_5g14770.1	Chlorophyll A-B binding protein, plant	Fvb5_v4.0.a1:8390582..8392365
FvH4_5g14970.1	SANT/Myb domain	Fvb5_v4.0.a1:8477425..8479779
FvH4_5g15130.1	Cytochrome P450	Fvb5_v4.0.a1:8549105..8553891
FvH4_5g15200.1	SANT/Myb domain	Fvb5_v4.0.a1:8581206..8582486
FvH4_5g15580.1	Cytochrome P450	Fvb5_v4.0.a1:8784747..8786988

The expression of these genes was analyzed by RT-qPCR in fully ripe fruits from RV and five lines of LG5 (Fb5:0-35, Fb5:11-76, Fb5:20-76, Fb5:39-76 and Fb5:0-76). Eight candidate genes did not show any expression; for three of them (FvH4_5g11300, FvH4_5g14970 and FvH4_5g15200) we could not detect either genomic amplification or amplification in qPCR (data not shown). For five genes (FvH4_5g11930, FvH4_5g13850, FvH4_5g14010, FvH4_5g14660 and FvH4_5g15880) we were unable to detect expression in ripe fruits (data not shown). Therefore, we discarded these genes as our candidate genes.

The genes FvH4_5g10850, FvH4_5g10880, FvH4_5g12400 and FvH4_5g12580 also were not considered candidate genes because their relative expression patterns did not present correspondence with the introgression region where the fruit color QTL is located (Table 3-3; Figure 3-4 A). Moreover, the relative expression values did not show any significant differences between RV and other lines in LG5 by Dunnett's test (Table 3-3).

The genes FvH4_5g11430 and FvH4_5g12670 both encode MYB domain transcription factors and the relative expression of both of them in the line Fb5:20-76 showed a significant difference from RV. However, we did not consider these as candidate genes since they have similar relative expression in other lines containing the 35-39cM introgression and RV as shown in Figure3-3B. Similarly, the genes

FvH4_5g12340 and FvH4_5g15130 both encoding cytochrome superfamily proteins presented significant differences in relative expression in Fb5:11-76 compared to RV, but did not show any obvious differences in Fb5:20-76 and Fb5:0-76 (Figure 3-3C). The gene FvH4_5g13590 is a cytochrome P450 related gene. This gene showed a significant difference in the relative expression in Fb5:0-76 and Fb5:20-76 as compared to RV, but in Fb5:11-76 had similar value to RV. Therefore, it was not considered a good candidate gene (Figure 3-3D). The relative expression of gene FvH4_5g14770 which encodes a chlorophyll A-B binding protein in plants is significantly higher in Fb5:11-76, Fb5:20-76 and Fb5:0-76 than in RV (Figure 3-3E). The expression analysis correlated with the QTL for fruit orange color, making it an important candidate gene.

Table 3-3.Log-transformed fold-change values of candidate genes for orange fruit color.

Line	FvH4_5g10850	FvH4_5g10880	FvH4_5g11430	FvH4_5g12400	FvH4_5g12340
5:0-35	0.9±0.62	-0.49±0.74	1.8±0.71	1.46±0.81	-1.28±0.72
5:0-76	1.9±0.62	2±0.74	1.79±0.71	2.38±0.81	-1.86±0.72
5:11-76	1.93±0.76	0.99±0.74	1.26±0.71	1.16±0.66	2.66±0.72*
5:20-76	1.8±0.76	1.85±0.74	2.57±0.71*	1.2±0.66	-2.15±0.72
5:39-76	0.17±0.62	0.24±0.57	1.59±0.55	1.4±0.51	-0.11±0.56
RV	0±0.62	-0.13±0.57	-0.02±0.55	0.03±0.51	0±0.56

Line	FvH4_5g12580	FvH4_5g12670	FvH4_5g13590	FvH4_5g14770	FvH4_5g15130
5:0-35	1.68±0.85	2.65±0.95	2.56±0.67	1.85±0.84	-0.85±0.35
5:0-76	2.71±0.85	2.06±0.77	2.64±0.67*	3.66±0.84*	-0.45±0.35
5:11-76	3.01±0.85	2.11±0.77	2.05±0.67	3.9±0.84*	1.65±0.35*
5:20-76	3.02±1.04	3.17±0.77*	2.71±0.67*	3.85±0.84*	-0.1±0.35
5:39-76	3.88±1.47	0.34±0.6	1.05±0.52	1.79±0.65	0.13±0.27
RV	-0.23±0.73	0.19±0.67	0.23±0.52	0.03±0.65	0±0.27

Note: Values are averages of 2-5 biological replicates ± standard deviation. Expression values that are significantly different from RV (Dunnett's test) are indicated by asterisks: *$p < 0.05$.

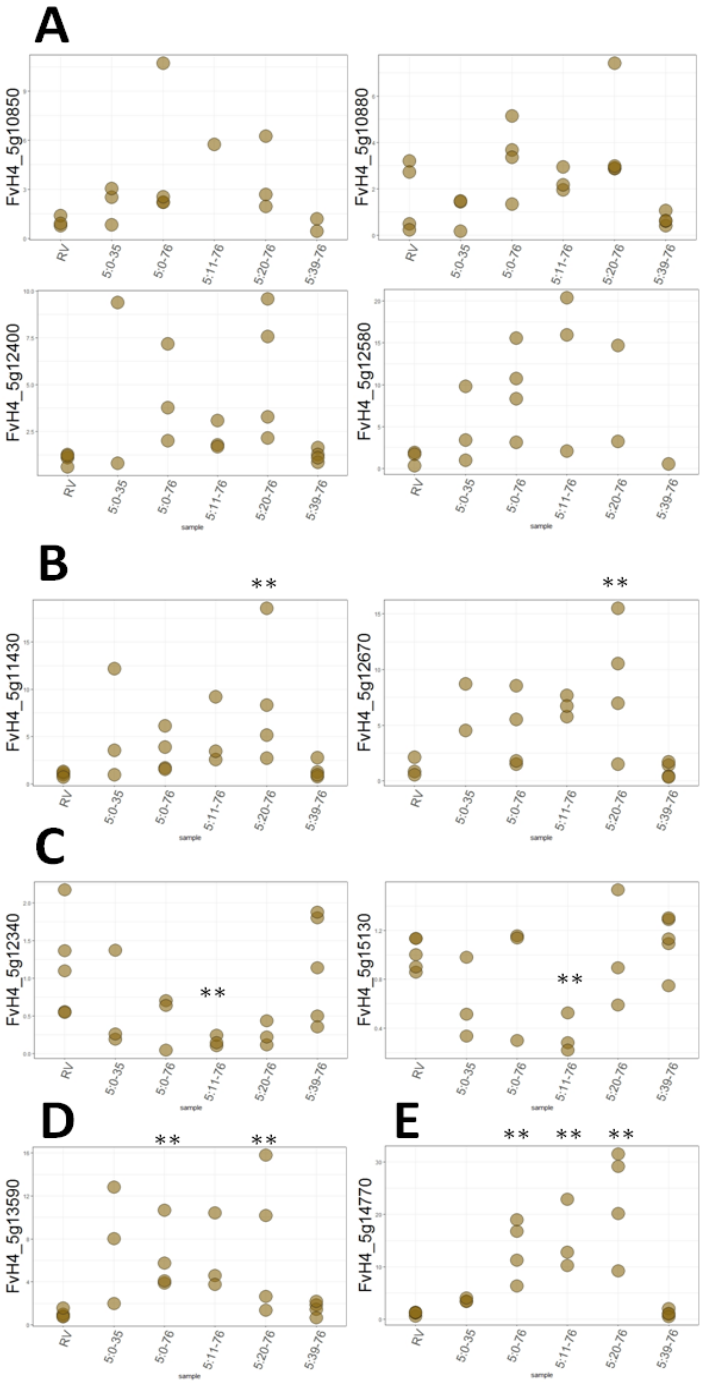

Figure 3-4.Relative expression of fruit orange color candidate genes in ripe fruits of different genotypes. Asterisks indicate statistically significant differences as compared to RV by Dunnett's test: ** - $p < 0.05$. **A)**FvH4_5g10850, FvH4_5g10880, FvH4_5g12400 and FvH4_5g12580did not show any significant differences in all samples. **B)** Relative expression of FvH4_5g11430 and FvH4_5g12670 in the line Fb5:20-76 showed significant difference from RV. **C)** FvH4_5g12340 and FvH4_5g15130 in the line Fb5:11-76 showed a significant difference from RV. **D)** FvH4_5g13590 showed a significant difference in the relative expression in Fb5:0-76 and Fb5:20-76 as compared to RV. **E)** FvH4_5g14770 is significantly higher in Fb5:11-76, Fb5:20-76 and Fb5:0-76 than in RV.

F.vesca homologs of *Lhca4*

In this study, an important candidate gene for orange fruit color FvH4_5g14770 was found and its function was annotated as encodinga chlorophyll A-B binding protein in plants. We acquired the gene DNA sequence from the *Fragaria vesca* V4.0 a1 genome database, and blasted the sequence in NCBI, to find one predicted gene (*FvLhca4*: XM_004299300, gene sequence shown in Table S3-4 with function encoding a light-harvesting complex I chlorophyll a-b binding protein P4 in chloroplasts (LOC101301492). Homologous genes have been reported in *Arabidopsis thaliana*(Zhang *et al.* 1991), rice(Yamatani et al. 2018)and *Chlamydomonasreinhardtii* (Koziol *et al.* 2007), and they belong to the light-harvesting complex (LHC) gene superfamily which contains 10-12 members in higher plants. The LHCa4 protein encoded by *Lhca4* gene is a light receptor that can bind with nine chlorophylls (Chls) and three potential His residues to extra Chls(Melkozernov and Blankenship 2003).

In order to assess the relationship between *FvLhca4* gene and *Lhc* gene family, phylogenetic relationships of the LHC proteins were studied by comparing the *Fragaria* proteins to other LHC superfamily proteins (Table S3-5) from various plant species.The FvLHCa4 belongs to a clade whose other members are all LHCa4 proteins (Figure 3-5) and other FvLHC proteins also cluster with their own respective subfamilies. Compared with other FvLHC proteins, FvLHCA4 showed a closer relationship with LHCa4 proteins from other plant species.

An alignment analysis of LHCa4 protein sequence of *Arabidopsis*, rice, diploid strawberry and soybean (Figure 3-6A) was performed to investigate whether the identified FvLHCa4 protein possesses the amino acidresidues essential for pigment binding. The alignment result showed that gene *Lhca4* is highly conserved in different plants and FvLHCa4 contained all nine essential amino acids for forming chlorophyll a and b ligands.

As the *F. vesca* protein 5g14770 is located within the *F. bucharica* introgression that affects fruit color, we decided to clone the gene encoding the protein from both parents of the NIL collection to see whether the protein itself is altered or non-functional in *F. bucharica*. Sequencing the coding sequence of the gene 5g14770 showed that the predicted proteins from RV and *F. bucharica* were virtually identical (Figure 3-6B), only a neutral amino acid change (N31T) was found between RV and *F.bucharica*. These data suggested that if the *Lhca4* gene 5g14770 is the causative agent behind the orange fruit color observed in near-isogenic lines harbouring an *F. bucharica* introgression, the difference probably occurs at transcriptional level.

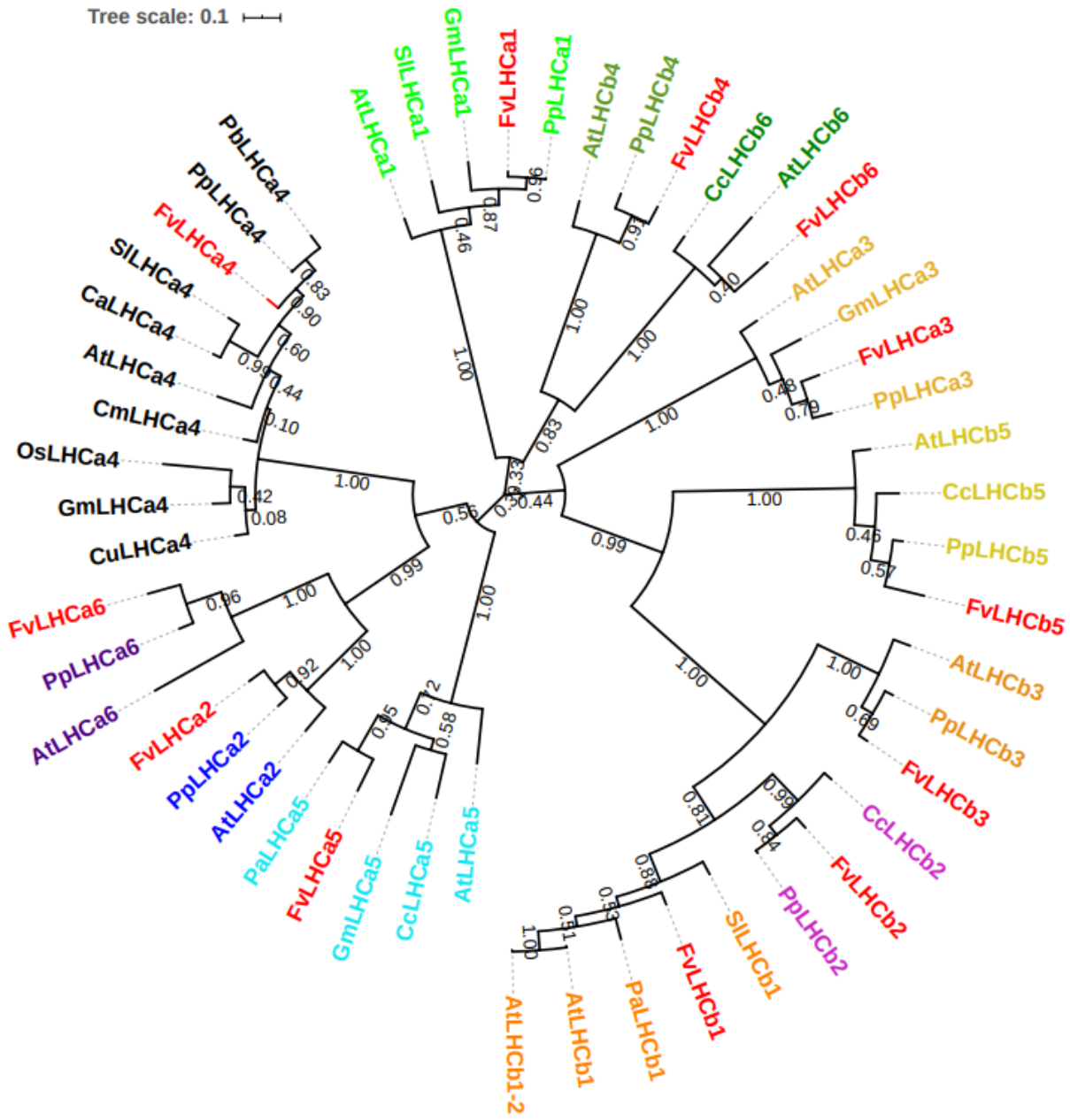

Figure 3-5.Phylogenetic analysis of FvLHC and other plant LHCs. Red labels mark F. vesca LHCs. Different colors present different LHC sub-families.The values next to branching points indicate the percentage of bootstrap support with 1000 replications. The phylogenetic tree was made in MEGA 6 and edited in the ITOL web version (https://itol.embl.de/).

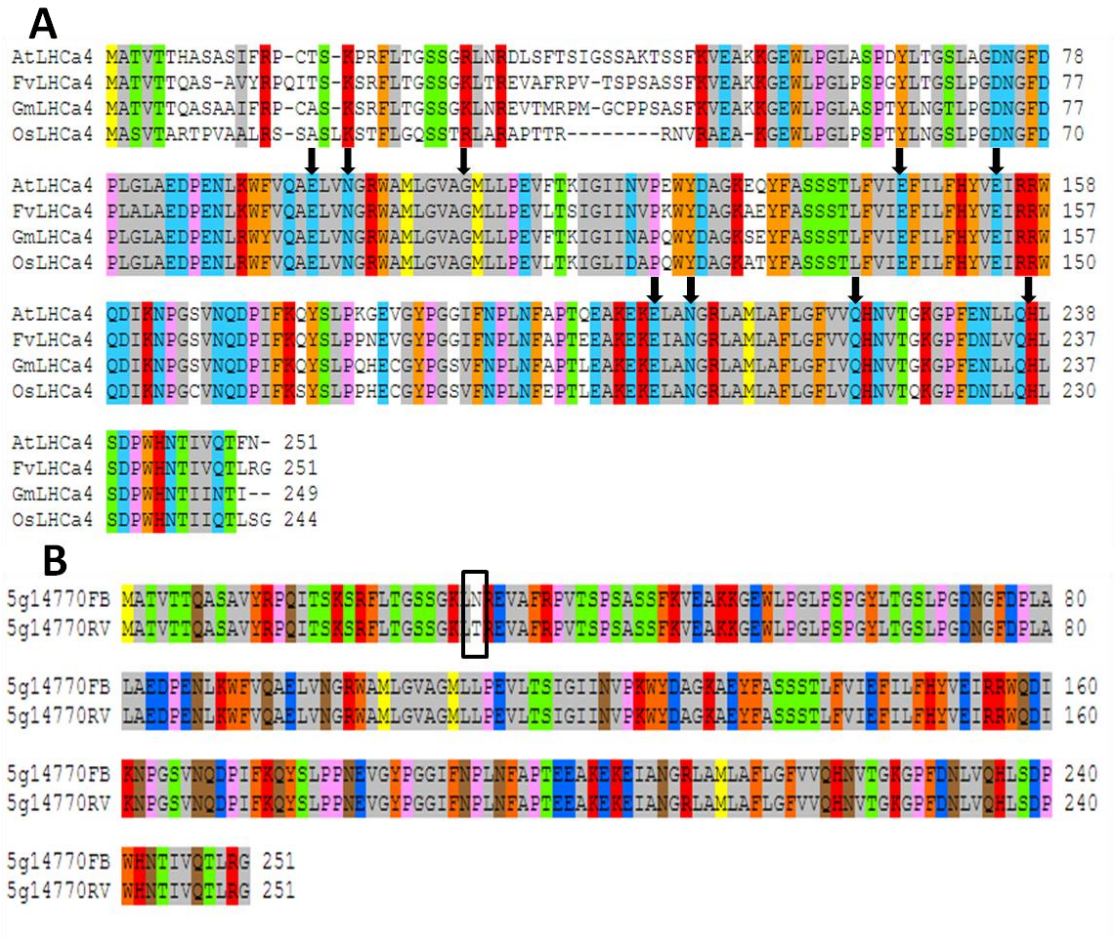

Figure 3-6.Amino acid alignments of LHCa4 proteins.A) Alignment of F. vesca LHCa4 protein with LHCa4 proteins from Arabidopsis (AtLHCa4), soybean (GmLHCa4) and rice (OsLHCa4). The nine essential pigment binding amino acids are highlighted by black arrows. **B)** Alignment of translated coding sequences from the recurrent parent RV and the donor parent F. bucharica. The amino acid different in RV versus F. bucharica is boxed.

Gene expression patterns of *FvLhca4* in RV and near-isogenic lines

To determine if differential expression exist, we investigated gene expression patterns of *FvLhca4* in the recurrent parent RV and in NILs with *F.bucharica* introgressions covering different regions of the LG5. We first analysed *FvLhca4*expression in different fruit stages of field-grown plants to detect changes in expression levels at fruit development stages. In the fruit stage 1 (green fruit), the mRNA levels of the gene *FvLhca4* were similar in RV and in NILs and then *FvLhca4*expression decreased in fruits at stage 2 (turning fruit) in all lines. In fruit stage 3, the NILs Fb5:0-35 and Fb5:39-76 presented similar mRNA level with RV, however, the NILs harbouring the introgression at 35-39cM of *F.bucharica* (Fb5:11-76, Fb5:20-76 and Fb5:0-76) showed high levels of *FvLhca4* mRNA (Figure 3-7). We also examined the tissue-specific expression patterns of *FvLhca4* in the recurrent parent RV and in NILs with different introgressions of LG5(Figure 3-7), the relative expression level of *FvLhca4* in root and young open-

leaves were comparable in all genotypes, but in unopened leaves it showed a bit lower expression level in Fb5:0-35 and Fb5:39-76.

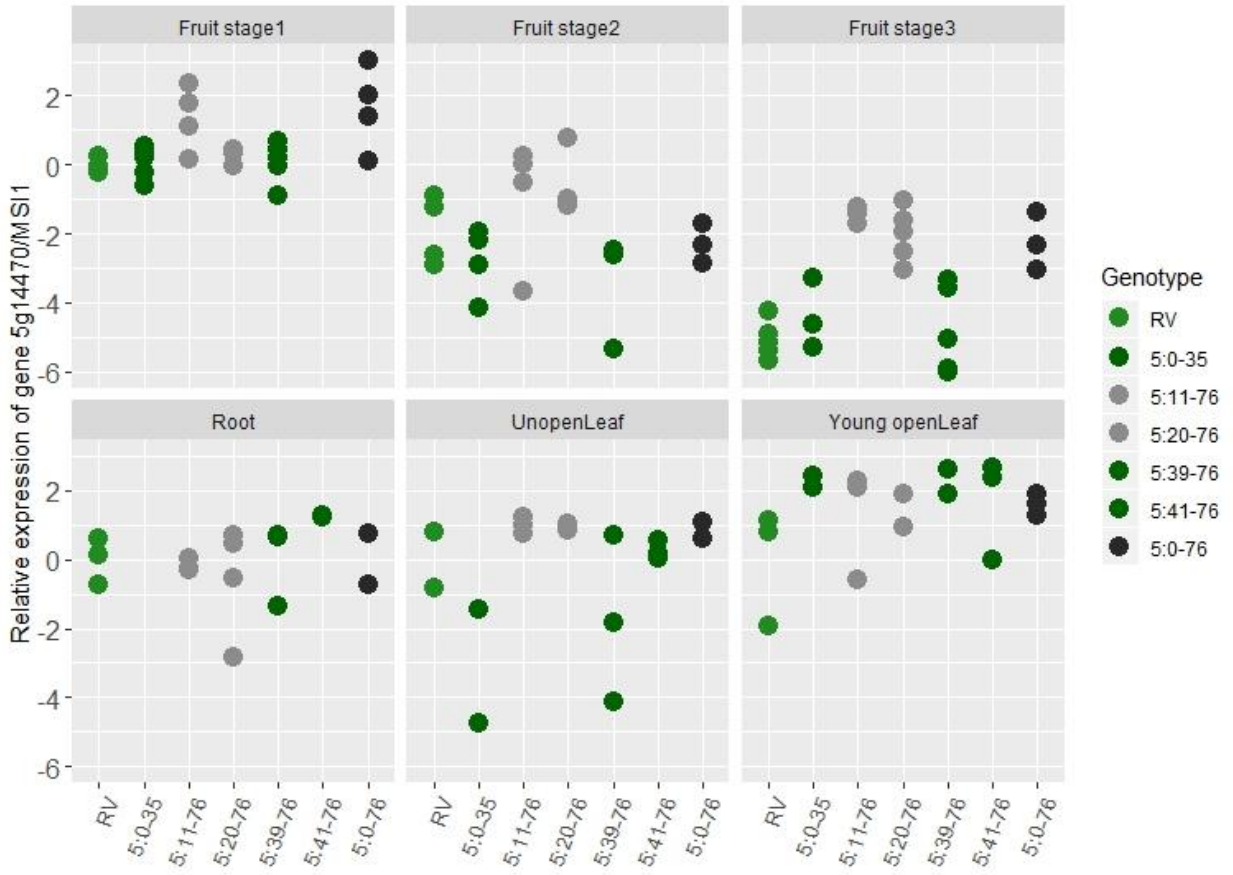

Figure 3-7.Relative expression of FvLhca4 gene in different genotypes and plant tissues. Relative expression values have been normalized to *FvMSI1* and shown as log transformed fold change values.

Chlorophyll content

In order to investigate relation between fruit orange color and chlorophyll, the content of Chl a, Chl b and total Chl were detected in the fruits of the recurrent parent RV and NILs with *F. bucharica* introgressions covering different regions of the LG5. Unexpected, the measured values are very low and did not show any obvious different between RV and NILs with orange fruit (Figure 3-8).

Transient transformation

In order to verify the function of FvLhca4 in *F. vesca*, we intended to overexpress and silence the *FvLhca4* gene in RV and NILs harboring the introgression of LG5:35-39 cM via transient expression. We constructed the *FvLhca4*-overexpression and silencing vectors and used them for *Agrobacterium*mediated transfection. 75 strawberry fruits of RV were transfected, half with the *FvLhca4*-overexpression construct and half with mock treatment. *Agrobacterium*carrying the silencing vector was transfected into 50 fruits

of NIL Fb5:20-76 and 34 fruits were mock treated. We have obtained the transient overexpression and silencing *FvLhca4* and mock treated fruits but the analysis is still on going (Figure 3-9). However, no obvious difference was observed in the fruit color either between mock and over expressed fruits or between mock and silenced fruits.

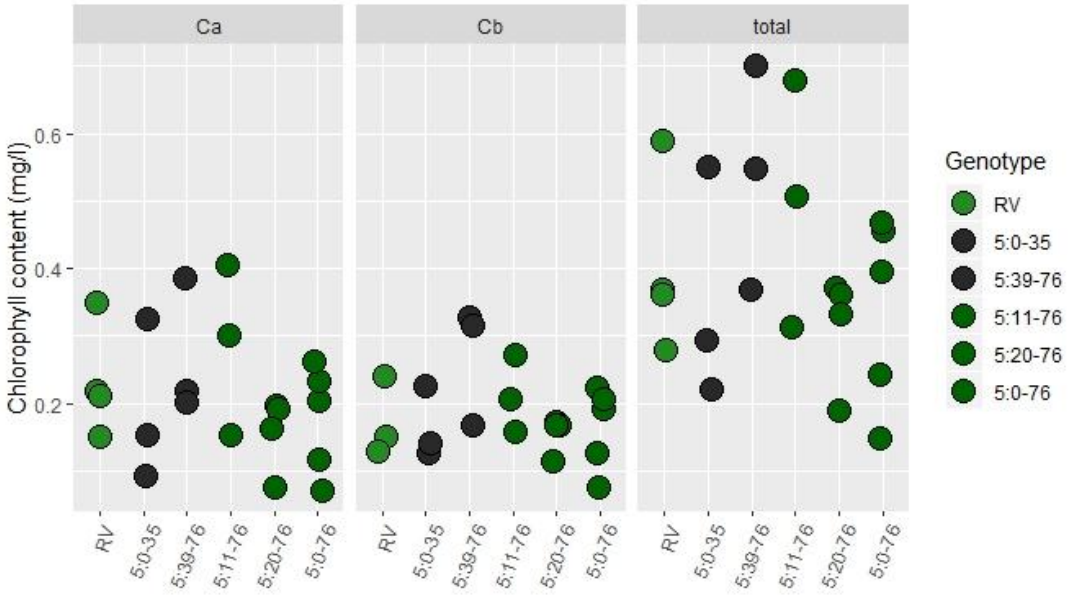

Figure 3-8.Chlorophyll content of fruits in different genotype. Ca: chlorophyll a; Cb: chlorophyll b; total: total chlorophyll.

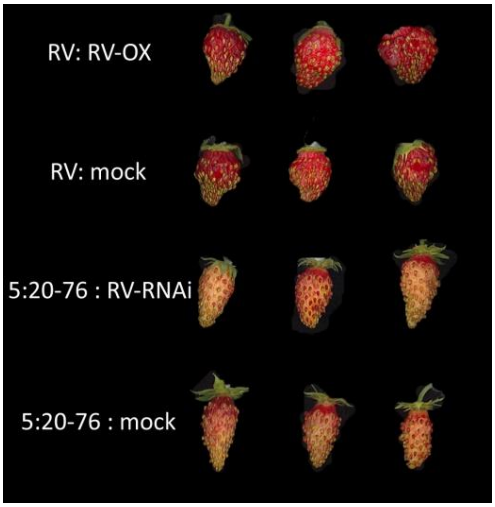

Figure 3-9.*FvLhca4* overexpression and silencing transient transformed fruit.RV and 5:20-76 are the plants used to transfected. RV-OX: *FvLhca4* gene overexpression treated. RV-RNAi: *FvLhca4* gene silencing treated.

Discussion

Fruit color is a primary attribute to the appearance and quality of strawberry. In general, color is important in attraction of dispersal agents (birds, animals, and primates), protection against ultraviolet damage, an indicator of ripeness, and contributes to polyphenolic content and their associated antioxidant properties (Selvaraj *et al.* 2016). As fruit ripens, color is one of many modifications that occur due to physiological and biochemical changes; including the increase in respiration rate, flesh softening, and formation of volatiles associated with development of flavor. During the ripening of strawberry, there is loss of chlorophyll pigment and anthocyanins are synthesized leading to development of red pigmentation(Pilkington *et al.* 2012).

In some diploid strawberry genotypes, such as Yellow Wonder (YW), white fruit color is naturally present. In this study, the fruits of three lines in our NIL collection, Fb5:11-76, Fb5:20-76 and Fb5:0-76 presented orange color. After analyzing eight color parameters of the yellow fruits from YW, red fruits of RV and NIL collection lines, a QTL for fruit orange color was mapped in LG5:35-39 cM. The fruits of the lines harboring introgressions including this region showed higher values in red and green parameters than RV and other NILs. The main difference between YW and lines with orange colored fruits was related to the increased green value.

To select candidate genes located in 35-39 cM in LG5, first of all, we considered the genes related to anthocyanin biosynthesis e.g., phenylalanine ammonia-lyase, chalcone synthase (CHS), flavanone 3β-hydroxylase (F3H), dihydroflavonol 4-reductase (DFR), anthocyanin synthase (ANS), and flavonoid 3-O-glucosyltransferase (FGT) (Carbone et al. 2009) and some transcription factors such as MYB (Hawkins et al. 2016) and MYC (Cultrone et al. 2010; Hawkins et al. 2016). Other genes related to the synthesis of pigments in fruits such ascarotenoids and chlorophyll were included. In total, 18 genes were selected and analyzed through detection of their mRNA expression level. Among these genes, the gene FvH_5g14770 encoding a chlorophyll A-B binding protein LHCA4 in plants was considered as a good candidate gene since it consistently had higher accumulation of mRNA in all the lines harboring introgression 35-39 cM than in RV. The *Lhca4* gene has been cloned in *Arabidopsis* (Jansson 1999), rice (Umate 2010), poplar (Klimmek et al. 2006) and some other plants (Zou et al. 2013; Zou and Yang 2019). LHCA4 protein encoded by *Lhca4* gene has been verified to bindto nine chlorophylls (Chls) and three potential *His* residues to extra Chls(Melkozernov and Blankenship 2003). It was associated with LHCA1 in the LHCI-730 (showing a fluorescence emission maximum at 730 nm) complex serving as antenna to capture light energy and deliver it to the photosystem reaction centers (Corbet et al. 2007). In our study, the identified FvLHCA4 proteins possesses the amino acid residues essential for binding nine Chls and it

was clustered together with LHCA4 proteins from other species. Therefore, *FvLhca4* is an actual LHCA4 protein encoding gene. To investigate whether differences existed in the gene coding sequence between parental lines of the NIL collection, we cloned *FvLhca4* gene from theparental lines and found that they had a three nucleotide difference. When the DNA sequences were *in vitro* translated to protein, only one non-conservative amino acid change was detected but with little effect to the function of the protein. The expression of *FvLhca4* was down-regulated in RV while not in the NILs with *F.bucharica* allele of *FvLhca4* during fruit ripening suggesting that *FvLhca4* expression is differentially regulated in the two parents of the NIL collection.

To our knowledge, there has been no report related with the function of *FvLhca4* gene neither in Rosaceae nor in strawberry. Previous research suggested that absence of LHCA4 protein in transgenic *Arabidopsis* causes chlorophyll content reduction and a drastic decrease in the longest-wavelength fluorescence (Zhang et al. 1997). While we identified the *FvLhca4* gene encoding the LHCA4 protein, the problem remains what is the main function of this gene in strawberry and whether the orange fruit color isrelated to this gene. In order to know whether orange fruit color is related to chlorophyll content, a preliminary experiment was performed to detect the content of Chls in ripen fruits. However, a relatively low Chls content was detected mainly caused by too smallsample size as well as abundant degradation of Chlsin the progress of fruit ripening. Therefore more samples should be used to measure Chl content in ripen fruits in the following experiment. At the same time, to verify the function of *FvLhca4* in strawberry, a transient overexpression experiment of *FvLhca4* in RV and two NILs with *F.bucharica*introgression in LG5:50-76cM has been done, and the result is progressing.

In the previous study, a QTL for short floral stem in *F. vesca* NIL collection was also mapped in LG5:29-39cM(Urrutia *et al.* 2015). Actually, the plants producing orange colored fruits presented short floral stemsin our study (data not shown). When the short floral stemmed inflorescences produced fruits, the fruits were covered by leaves and received less light than fruits borne onnormal floral stems. The chlorophyll synthesis and degradation are both regulated by light (Hörtensteiner 2006; Yin *et al.* 2016; Waters and Langdale 2009). Therefore, we speculated that the fruit orange color may also be related to the short floral stem since the fruits borne on short floral stemsreceived less light which may delay the degradation of chlorophyll and LHCA4 protein. On the other hand, accumulating evidence suggests that there is functional interaction between LHCI and LHCII (Wientjes *et al.* 2013; Grieco *et al.* 2015). In a recent research, a mutation of the *delayed yellowing 1* (*DYE1*) gene encoding LHCA4 protein was observed in rice and the amino acid substitution caused impaired structure of the PSI-LHCI super-complex and high accumulation of Lhcb1, more chlorophyll content, and finally the leaves remaining green (Yamatani et al. 2018). In pear fruits, *Lhca4* was downregulated in PE-bagged fruit but upregulated

in non-woven fabric-bagged fruit which suggested that *Lhca4* is related to the fruit photosynthesis and lenticels since it effect light-harvesting ability of fruits (Wang et al. 2017c). During peach fruit ripening, the transcription factor PpGLK1 affected chloroplast development by regulating its downstream targets including LHCA I family gene, leading to chlorophyll accumulation and finally affecting fruit color(Chen et al. 2018).Hence, if *FvLhca4* gene was responsible for the fruit orange color, it will be a complicated progress. We intend to measure the Chls content first to verify that the fruit orange color is caused by increased Chls content and then LHCA4 and other LHC protein content in ripen fruits will be measured to see if there are interactions between the proteins.

Bibliography

Arnon DI (1949) Copper enzymes in isolated chloroplasts. Polyphenoloxidase in Beta vulgaris. Plant physiology 24 (1):1

Beck J, Lohscheider JN, Albert S, Andersson U, Mendgen KW, Rojas-Stütz MC, Adamska I, Funck D (2017) Small one-helix proteins are essential for photosynthesis in *Arabidopsis*. Frontiers in plant science 8:7

Brown T, Wareing P (1965) The genetical control of the everbearing habit and three other characters in varieties of Fragaria vesca. Euphytica 14 (1):97-112

Carbone F, Preuss A, De Vos RC, D'AMICO E, Perrotta G, Bovy AG, Martens S, Rosati C (2009) Developmental, genetic and environmental factors affect the expression of flavonoid genes, enzymes and metabolites in strawberry fruits. Plant, Cell & Environment 32 (8):1117-1131

Chen M, Liu X, Jiang S, Wen B, Yang C, Xiao W, Fu X, Li D, Chen X, Gao D (2018) Transcriptomic and functional analyses reveal that ppglk1 regulates chloroplast development in peach (*Prunus persica*). Frontiers in Plant Science 9:34

Corbet D, Schweikardt T, Paulsen H, Schmid VH (2007) Amino acids in the second transmembrane helix of the Lhca4 subunit are important for formation of stable heterodimeric light-harvesting complex LHCI-730. Journal of molecular biology 370 (1):170-182

Cultrone A, Cotroneo PS, Recupero GR (2010) Cloning and molecular characterization of R2R3-MYB and bHLH-MYC transcription factors from Citrus sinensis. Tree genetics & genomes 6 (1):101-112

Daum B, Nicastro D, Austin J, McIntosh JR, Kühlbrandt W (2010) Arrangement of photosystem II and ATP synthase in chloroplast membranes of spinach and pea. The Plant Cell 22 (4):1299-1312

de Bianchi S, Betterle N, Kouril R, Cazzaniga S, Boekema E, Bassi R, Dall'Osto L (2011) *Arabidopsis* Mutants Deleted in the Light-Harvesting Protein Lhcb4 Have a Disrupted Photosystem II Macrostructure and Are Defective in Photoprotection. The Plant Cell 23 (7):2659-2679.

Dekker JP, Boekema EJ (2005) Supramolecular organization of thylakoid membrane proteins in green plants. Biochimica et Biophysica Acta (BBA)-Bioenergetics 1706 (1-2):12-39

Deng C, Davis T (2001) Molecular identification of the yellow fruit color (c) locus in diploid strawberry: a candidate gene approach. Theoretical and Applied Genetics 103 (2-3):316-322

Edger PP, VanBuren R, Colle M, Poorten TJ, Wai CM, Niederhuth CE, Alger EI, Ou S, Acharya CB, Wang J, Callow P, McKain MR, Shi J, Collier C, Xiong Z, Mower JP, Slovin JP, Hytonen T, Jiang N, Childs KL, Knapp SJ (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.

Green BR, Durnford DG (1996) The chlorophyll-carotenoid proteins of oxygenic photosynthesis. Annual review of plant biology 47 (1):685-714

Grieco M, Suorsa M, Jajoo A, Tikkanen M, Aro E-M (2015) Light-harvesting II antenna trimers connect energetically the entire photosynthetic machinery—including both photosystems II and I. Biochimica et Biophysica Acta (BBA)-Bioenergetics 1847 (6-7):607-619

Hawkins C, Caruana J, Schiksnis E, Liu Z (2016) Genome-scale DNA variant analysis and functional validation of a SNP underlying yellow fruit color in wild strawberry. Sci Rep 6:29017.

Hiscox J, Israelstam G (1979) A method for the extraction of chlorophyll from leaf tissue without maceration. Canadian journal of botany 57 (12):1332-1334

Holtzegel U (2016) The Lhc family of Arabidopsis thaliana. Endocytobiosis and Cell Research 27 (2):71-89

Hörtensteiner S (2006) Chlorophyll degradation during senescence. Annu Rev Plant Biol 57:55-77

Jansson S (1999) A guide to the Lhc genes and their relatives in Arabidopsis. Trends in plant science 4 (6):236-240

Jung S, Lee T, Cheng CH, Buble K, Zheng P, Yu J, Humann J, Ficklin SP, Gasic K, Scott K, Frank M, Ru S, Hough H, Evans K, Peace C, Olmstead M, DeVetter LW, McFerson J, Coe M, Wegrzyn JL, Staton ME, Abbott AG, Main D (2019) 15 years of GDR: New data and functionality in the Genome Database for Rosaceae. Nucleic acids research 47 (D1):D1137-d1145.

Katoh K, Kuma K, Toh H, Miyata T (2005) MAFFT version 5: improvement in accuracy of multiple sequence alignment. Nucleic acids research 33 (2):511-518.

Kayesh E, Shangguan L, Korir NK, Sun X, Bilkish N, Zhang Y, Han J, Song C, Cheng Z-M, Fang J (2013) Fruit skin color and the role of anthocyanin. Acta physiologiae plantarum 35 (10):2879-2890

Klimmek F, Sjödin A, Noutsos C, Leister D, Jansson S (2006) Abundantly and rarely expressed Lhc protein genes exhibit distinct regulation patterns in plants. Plant physiology 140 (3):793-804

Koziol AG, Borza T, Ishida K-I, Keeling P, Lee RW, Durnford DG (2007) Tracing the evolution of the light-harvesting antennae in chlorophyll a/b-containing organisms. Plant physiology 143 (4):1802-1816

Lamesch P, Berardini TZ, Li D, Swarbreck D, Wilks C, Sasidharan R, Muller R, Dreher K, Alexander DL, Garcia-Hernandez M (2011) The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. Nucleic acids research 40 (D1):D1202-D1210

Li D, Zhang X, Li L, Aghdam MS, Wei X, Liu J, Xu Y, Luo Z (2019) Elevated CO2 delayed the chlorophyll degradation and anthocyanin accumulation in postharvest strawberry fruit. Food chemistry 285:163-170

Luo H, Dai C, Li Y, Feng J, Liu Z, Kang C (2018) Reduced Anthocyanins in Petioles codes for a GST anthocyanin transporter that is essential for the foliage and fruit coloration in strawberry. J Exp Bot 69 (10):2595-2608.

Melkozernov AN, Blankenship RE (2003) Structural modeling of the Lhca4 subunit of LHCI-730 peripheral antenna in photosystem I based on similarity with LHCII. Journal of Biological Chemistry 278 (45):44542-44551

Mozzo M, Mantelli M, Passarini F, Caffarri S, Croce R, Bassi R (2010) Functional analysis of Photosystem I light-harvesting complexes (Lhca) gene products of Chlamydomonas reinhardtii. Biochimica et Biophysica Acta (BBA)-Bioenergetics 1797 (2):212-221

Pilkington SM, Montefiori M, Jameson PE, Allan AC (2012) The control of chlorophyll levels in maturing kiwifruit. Planta 236 (5):1615-1628.

Qian L, Voss-Fels K, Cui Y, Jan HU, Samans B, Obermeier C, Qian W, Snowdon RJ (2016) Deletion of a stay-green gene associates with adaptive selection in Brassica napus. Molecular plant 9 (12):1559-1569

Selvaraj D, Sherif S, Dek MSP, Paliyath G, El-Sharkawy I, Subramanian J (2016) Identification and Characterization of Genes Involved in the Fruit Color Development of European Plum. Journal of the American Society for Horticultural Science 141 (5):467-474

Shulaev V, Sargent DJ, Crowhurst RN, Mockler TC, Folkerts O, Delcher AL, Jaiswal P, Mockaitis K, Liston A, Mane SP (2011) The genome of woodland strawberry (*Fragaria vesca*). Nature genetics 43 (2):109

Stothard P, . (2000) The sequence manipulation suite: JavaScript programs for analyzing and formatting protein and DNA sequences. Biotechniques 28 (6):1102, 1104

Strecker J, Rodríguez G, Njanji I, Thomas J, Jack A, Darrigues A, Hall J, Dujmovic N, Gray S, van der Knaap E (2010) Tomato analyzer color test user manual version 3.

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Molecular biology and evolution 30 (12):2725-2729.

Umate P (2010) Genome-wide analysis of the family of light-harvesting chlorophyll a/b-binding proteins in *Arabidopsis* and rice. Plant signaling & behavior 5 (12):1537-1542

Urrutia M, Bonet J, Arus P, 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. TAG Theoretical and applied genetics Theoretische und angewandte Genetik 128 (7):1261-1275.

Urrutia M, Monfort A (2018) Fruit Quality and the Use of Near-Isogenic Lines for Functional Characterization in Fragaria vesca. In: The Genomes of Rosaceous Berries and Their Wild Relatives. Springer, pp 49-62

Urrutia Rosauro M (2015) Fragaria vesca NIL collection: development and genetic characterization of agronomical, nutritional and organoleptic traits.

Wang Y, Zhang X, Wang R, Bai Y, Liu C, Yuan Y, Yang Y, Yang S (2017) Differential gene expression analysis of 'Chili'(Pyrus bretschneideri) fruit pericarp with two types of bagging treatments. Horticulture research 4:17005

Waters MT, Langdale JA (2009) The making of a chloroplast. The EMBO journal 28 (19):2861-2873

Wientjes E, van Amerongen H, Croce R (2013) LHCII is an antenna of both photosystems after long-term acclimation. Biochimica et Biophysica Acta (BBA)-Bioenergetics 1827 (3):420-426

Yamatani H, Kohzuma K, Nakano M, Takami T, Kato Y, Hayashi Y, Monden Y, Okumoto Y, Abe T, Kumamaru T (2018) Impairment of Lhca4, a subunit of LHCI, causes high accumulation of chlorophyll and the stay-green phenotype in rice. Journal of experimental botany 69 (5):1027-1035

Yin Xr, Xie Xl, Xia Xj, Yu Jq, Ferguson IB, Giovannoni JJ, Chen Ks (2016) Involvement of an ethylene response factor in chlorophyll degradation during citrus fruit degreening. The Plant Journal 86 (5):403-412

Zhang H, Goodman HM, Jansson S (1997) Antisense inhibition of the photosystem I antenna protein Lhca4 in *Arabidopsis thaliana*. Plant physiology 115 (4):1525-1531

Zhang H, Hanley S, Goodman HM (1991) Isolation, Characterization, and Chromosomal Location of a New *cab* Gene from *Arabidopsis thaliana.* Plant physiology 96 (4):1387-1388.

Zorrilla-Fontanesi Y, Cabeza A, Domínguez P, Medina JJ, Valpuesta V, Denoyes-Rothan B, Sánchez-Sevilla JF, Amaya I (2011) Quantitative trait loci and underlying candidate genes controlling agronomical and fruit quality traits in octoploid strawberry (*Fragaria× ananassa*). Theoretical and applied genetics 123 (5):755-778

Zou Z, Huang Q, An F (2013) Genome-wide Identification, Classification and Expression Analysis of Lhc Supergene Family in Castor Bean (Ricinus communis L.). Agricultural Biotechnology (2164-4993) 2 (6)

Zou Z, Yang J (2019) Genomics analysis of the light-harvesting chlorophyll a/b-binding (Lhc) superfamily in cassava (Manihot esculenta Crantz). Gene 702:171-181

Supplementary Material

Table S3-1. Dunnett's test results of eight color parameters in RV, YW and NILs in the years 2016, 2017 and 2018.Bold p-values representa significant p-value(p<0.05).

2016	red		green		blue		*L**		*a**		*b**		hue		chroma	
Line	mean	p-Value	mean	p-Value	mean	p-Value	mean	p-Value	mean	p-Value	mean	p-Value	mean	p-Value	mean	p-Value
Fb1:33-61	126.09	0.9962	45.67	1	37.53	0.3726	27.81	0.9987	28.08	1	20.67	0.9687	36.21	0.9154	35.02	1
Fb2:0-30	110.50	**<.0001**	40.65	0.3497	40.95	1	24.15	**0.0002**	25.73	0.1822	14.35	**<.0001**	28.96	**0.0005**	29.52	**0.0005**
Fb2:0-39	134.34	0.9862	60.74	0.0941	49.52	0.1643	32.06	0.2007	25.58	0.274	19.61	1	37.97	0.1413	32.48	0.8485
Fb2:0-45	120.80	0.1259	36.33	**0.0358**	38.74	0.7545	25.36	**0.0313**	30.13	0.9398	17.25	0.2151	29.71	**0.0147**	34.76	1
Fb2:0-63	117.00	**0.0024**	39.04	0.1772	39.91	0.9817	25.15	**0.0124**	28.13	1	16.28	**0.0071**	30.03	**0.0232**	32.55	0.7875
Fb2:0-73	122.04	0.2842	33.85	**0.0043**	37.03	0.2407	25.27	**0.024**	31.06	0.3583	18.12	0.8971	30.13	**0.04**	35.99	0.9955
Fb2:39-47	131.60	1	43.59	0.987	38.29	0.7227	28.43	1	30.54	0.8145	21.19	0.7055	34.79	1	37.23	0.5449
Fb2:39-63	117.02	**0.0055**	37.90	0.1282	37.36	0.3543	24.85	**0.0098**	28.37	1	17.26	0.2548	31.04	0.2608	33.31	0.9996
Fb3:0-15	125.23	0.9271	44.64	0.997	39.86	0.976	27.60	0.9746	28.30	1	19.27	1	34.41	1	34.35	1
Fb3:0-8	135.07	0.8769	39.57	0.2665	35.41	**0.0421**	28.58	1	32.24	**0.027**	22.98	**0.0023**	35.52	0.9997	39.82	**0.0007**
Fb4:0-20	114.19	**0.0004**	44.47	0.9988	44.53	1	25.64	0.0999	25.60	0.2799	14.27	**<.0001**	29.10	**0.0062**	29.38	**0.0025**
Fb4:0-31	114.16	**<.0001**	47.98	1	46.47	0.8805	26.26	0.205	24.39	**0.0079**	13.97	**<.0001**	29.56	**0.0068**	28.26	**<.0001**
Fb4:0-44	121.69	0.2503	38.04	0.1306	38.11	0.5653	25.75	0.0932	29.71	0.999	18.05	0.8662	31.24	0.336	34.82	1
Fb4:0-78	128.15	1	47.51	1	43.61	1	28.53	1	28.48	1	18.50	0.9984	33.00	0.9998	34.07	1
Fb5:0-11	131.51	1	51.77	1	47.11	0.7452	29.91	1	28.20	1	18.32	0.9826	32.89	0.9992	33.66	1
Fb5:0-20	131.06	1	48.24	1	42.02	1	29.18	1	28.98	1	20.05	1	34.88	1	35.41	1
Fb5:0-35	144.68	**0.0042**	35.18	0.0578	32.68	**0.007**	29.68	1	36.45	**<.0001**	25.78	**<.0001**	35.27	1	44.67	**<.0001**
Fb5:0-76	154.59	**<.0001**	65.89	**0.0011**	42.97	1	36.35	**<.0001**	29.17	1	28.09	**<.0001**	44.11	**<.0001**	40.59	**0.0001**
Fb5:11-76	160.71	**<.0001**	83.33	**<.0001**	54.57	**0.0001**	40.57	**<.0001**	24.97	0.0737	27.24	**<.0001**	47.92	**<.0001**	37.12	0.5356
Fb5:20-76	155.59	**<.0001**	71.84	**<.0001**	47.52	0.5574	37.60	**<.0001**	27.45	0.9999	27.29	**<.0001**	45.18	**<.0001**	38.88	**0.0118**
Fb5:39-76	144.67	**0.0004**	53.01	0.999	37.51	0.3574	32.28	0.0874	30.89	0.4612	26.09	**<.0001**	40.24	**0.0002**	40.55	**<.0001**
Fb5:41-76	122.67	0.39	45.72	1	40.31	0.9983	27.23	0.8455	27.34	0.9995	18.58	0.9995	33.89	1	33.19	0.9969
Fb5:50-76	145.19	**0.0002**	51.71	1	39.38	0.9258	32.16	0.1145	31.61	0.1318	25.15	**<.0001**	38.78	**0.0156**	40.52	**<.0001**
Fb6:0-5	132.44	1	40.31	0.4202	37.71	0.4387	28.02	1	31.98	0.0655	21.12	0.6703	33.44	1	38.36	0.0596
Fb6:101-101	112.45	**0.0003**	34.52	**0.0449**	36.87	0.4255	23.40	**0.0008**	28.41	1	15.77	**0.012**	28.83	**0.0119**	32.55	0.9514
Fb6:71-101	126.03	0.9981	43.04	0.9537	39.05	0.9098	27.24	0.9201	29.34	1	19.38	1	33.33	1	35.26	1
Fb6:84-101	129.25	1	40.05	0.3756	34.84	**0.0263**	27.44	0.9572	30.86	0.5087	21.78	0.1872	35.24	1	37.87	0.1602
Fb7:0-10	133.46	0.9988	40.87	0.4997	39.81	0.9807	28.37	1	32.12	**0.0403**	20.47	0.9967	32.53	0.9723	38.14	0.0847
Fb7:0-27	131.06	1	40.14	0.3537	36.91	0.2076	27.84	0.9989	31.47	0.1651	21.24	0.5197	34.04	1	38.03	0.0997
Fb7:0-59	137.18	0.3226	54.46	0.9059	43.09	1	31.42	0.3576	28.27	1	22.20	**0.031**	38.42	**0.0216**	36.29	0.9215
Fb7:26-45	128.89	1	54.76	0.9209	47.65	0.5647	29.94	1	26.28	0.6241	17.91	0.7588	34.17	1	32.03	0.5156
Fb7:26-59	126.60	0.9995	46.50	1	40.11	0.9942	27.96	0.9998	28.28	1	19.57	1	34.66	1	34.53	1
Fb7:43-59	134.19	0.9926	57.91	0.4222	47.31	0.7559	31.48	0.5115	26.51	0.8444	20.04	1	37.25	0.4273	33.51	1
Fb7:52-59	123.77	0.6852	46.34	1	41.17	1	27.54	0.9789	27.45	1	18.44	0.9964	33.80	1	33.27	0.9992
RV	129.92	1	48.89	1	42.88	1	29.03	1	28.48	1	19.47	1	34.26	1	34.57	1
YW	157.66	**<.0001**	138.67	**<.0001**	71.58	**<.0001**	52.80	**<.0001**	1.16	**<.0001**	32.15	**<.0001**	88.00	**<.0001**	32.20	0.6275

2017	red		green		blue		*L**		*a**		*b**		Hue		chroma	
LINE	mean	p-Value	mean	p-Value	mean	p-Value	mean	p-Value	mean	p-Value	mean	p-Value	mean	p-Value	mean	p-Value
Fb1:0-61	134.08	0.5697	44.31	0.9803	29.91	0.7814	29.11	0.7147	30.76	1	25.84	**0.0015**	40.21	0.1125	40.46	0.5329
Fb1:33-61	133.06	0.8063	42.30	1	33.12	1	28.66	0.9597	31.17	1	23.79	0.8385	37.12	1	39.44	0.9997
Fb1:50-61	143.01	**<.0001**	51.34	**0.0132**	39.12	**0.0038**	31.84	**<.0001**	31.36	1	24.64	0.1603	38.12	0.9734	39.98	0.9034
Fb2:0-30	107.12	**<.0001**	38.85	1	38.14	**0.0371**	23.23	**0.0002**	25.53	**<.0001**	14.46	**<.0001**	29.38	**0.0001**	29.47	**<.0001**
Fb2:0-39	135.17	0.234	49.44	0.0709	39.85	**0.0005**	30.13	0.0572	29.82	0.9987	22.23	1	37.02	1	37.44	0.9988
Fb2:0-45	115.78	**<.0001**	29.97	0.0866	33.36	1	23.34	**0.0004**	30.90	1	17.47	**<.0001**	29.20	**<.0001**	35.61	0.0695
Fb2:0-63	116.03	**0.0001**	32.99	0.5745	34.91	0.9942	23.88	**0.0045**	30.01	1	17.29	**<.0001**	29.69	**0.0004**	34.80	**0.0042**
Fb2:0-73	124.72	0.951	35.00	0.9678	35.65	0.8392	25.93	0.8917	31.65	1	19.50	**0.0099**	31.52	0.0466	37.35	0.9974
Fb2:39-47	133.37	0.7781	36.75	1	32.06	1	27.77	1	33.28	0.4176	23.56	0.9849	35.48	1	40.94	0.1941
Fb2:39-63	120.97	0.0873	36.24	0.9996	36.30	0.4968	25.43	0.4492	30.16	1	18.49	**<.0001**	31.69	0.0557	35.49	**0.0416**
Fb3:0-15	125.85	0.9989	37.04	1	32.01	1	26.35	0.9955	31.21	1	21.74	1	34.94	1	38.22	1
Fb3:0-8	125.59	0.9947	40.21	1	34.00	1	26.88	1	30.02	0.9999	21.25	0.8909	35.30	1	37.01	0.8544
Fb4:0-20	120.99	0.0444	36.47	0.9995	34.45	0.9996	25.41	0.305	30.19	1	19.33	**0.001**	32.38	0.1149	36.00	0.101
Fb4:0-31	121.73	0.1792	39.24	1	35.41	0.9102	26.07	0.9511	29.31	0.8933	19.67	0.0164	33.82	0.8969	35.38	**0.0287**
Fb4:0-44	132.50	0.9819	43.91	0.9978	36.41	0.5507	28.63	0.9934	30.99	1	22.31	1	35.79	1	38.30	1
Fb4:0-78	126.75	1	45.48	0.8436	38.52	**0.0214**	27.82	1	28.97	0.689	20.17	0.1277	34.82	1	35.34	**0.033**
Fb4:58-78	125.28	0.9836	28.27	**0.0092**	27.35	**0.015**	24.84	0.0794	33.89	0.0841	22.46	1	33.37	0.6031	40.72	0.2605
Fb5:0-11	126.24	0.9999	37.31	1	29.42	0.4018	26.48	0.9993	31.01	1	23.17	1	36.83	1	38.85	1
Fb5:0-20	133.58	0.6591	39.88	1	29.74	0.6279	28.37	0.9994	32.01	0.9998	25.20	0.0164	38.46	0.8442	40.93	0.1598
Fb5:0-35	136.74	0.0435	46.81	0.4265	31.61	1	29.97	0.0862	30.70	1	26.03	**0.0002**	40.49	0.0456	40.44	0.4787
Fb5:0-76	139.38	0.0882	71.94	**<.0001**	44.99	**<.0001**	35.27	**<.0001**	22.20	**<.0001**	25.22	0.2806	49.44	**<.0001**	34.29	**0.0384**
Fb5:11-76	150.76	**<.0001**	76.60	**<.0001**	44.44	**<.0001**	37.67	**<.0001**	24.31	**<.0001**	28.37	**<.0001**	49.93	**<.0001**	37.64	1
Fb5:20-76	156.80	**<.0001**	74.95	**<.0001**	50.24	**<.0001**	38.22	**<.0001**	27.28	0.1534	26.49	**0.0059**	44.56	**0.0003**	38.42	1
Fb5:39-76	144.48	**<.0001**	65.31	**<.0001**	40.60	**0.0001**	34.54	**<.0001**	26.56	**0.0009**	26.72	**<.0001**	46.00	**<.0001**	38.27	1
Fb5:41-76	125.21	0.9804	46.57	0.4831	35.80	0.6965	27.81	1	27.75	0.037	21.35	0.9609	37.45	1	35.37	**0.0191**
Fb5:50-76	138.86	0.0406	49.38	0.2941	33.66	1	30.75	0.0609	30.56	1	26.10	**0.0076**	40.95	0.1283	40.32	0.8947
Fb6:0-5	134.77	0.3121	40.20	1	34.68	0.9982	28.59	0.9845	32.60	0.886	23.18	1	35.66	1	40.12	0.7786
Fb6:101-101	117.81	**0.0137**	28.08	**0.0646**	32.68	1	23.57	**0.0096**	31.96	1	18.18	**0.0002**	29.49	**0.003**	36.81	0.9344
Fb6:71-101	110.87	**<.0001**	28.30	**0.0213**	30.86	0.9977	22.23	**<.0001**	29.84	0.9996	17.43	**<.0001**	30.35	**0.0034**	34.65	**0.0026**
Fb6:84-101	126.94	1	40.66	1	32.01	1	27.40	1	29.74	0.996	22.86	1	38.08	0.9716	37.76	1
Fb7:0-10	129.90	1	39.22	1	30.44	0.9342	27.58	1	31.27	1	23.91	0.7431	37.58	0.9997	39.55	0.9981
Fb7:0-27	132.97	0.9045	44.16	0.9899	33.49	1	29.00	0.83	30.56	1	24.04	0.7022	38.23	0.9651	38.98	1
Fb7:0-59	128.30	1	50.94	0.0586	33.49	1	29.35	0.6449	26.57	**0.0043**	24.12	0.739	42.94	**0.0007**	36.37	0.5671
Fb7:26-45	133.29	0.7918	49.00	0.1139	35.27	0.9432	29.63	0.2581	29.29	0.8752	23.80	0.8766	39.28	0.4223	37.98	1
Fb7:26-59	122.99	0.6354	34.57	0.9664	31.44	1	25.51	0.6846	31.13	1	21.01	0.8721	34.00	0.9857	37.70	1
Fb7:43-59	126.58	1	41.55	1	31.85	1	27.34	1	29.55	0.9708	22.78	1	37.93	0.9898	37.86	1
Fb7:52-59	126.53	1	37.58	1	30.57	0.9684	26.65	1	30.92	1	22.71	1	36.25	1	38.79	1
RV	128.68	1	39.77	1	32.84	1	27.41	1	30.97	1	22.48	1	36.12	1	38.49	1
YW	157.11	**<.0001**	132.42	**<.0001**	61.60	**<.0001**	51.09	**<.0001**	3.01	**<.0001**	34.17	**<.0001**	84.96	**<.0001**	34.36	**<.0001**

2018	red		green		blue		*L**		*a**		*b**		Hue		chroma	
LINE	mean	p-Value	mean	p-Value	mean	p-Value	mean	p-Value	mean	p-Value	mean	p-Value	mean	p-Value	mean	p-Value
Fb1:0-61	136.80	0.9122	40.67	0.4556	38.88	**0.0005**	29.83	0.9562	35.41	0.6681	22.23	**0.035**	32.04	0.4048	41.86	0.0686
Fb1:33-61	130.61	**0.0206**	35.98	1	37.77	**0.0033**	27.96	1	35.36	0.6273	20.59	**0.0004**	30.07	**0.0123**	40.95	**0.0088**
Fb1:50-61	131.80	0.0604	31.59	1	35.11	0.0956	27.52	1	37.02	1	21.53	**0.0062**	30.12	**0.0139**	42.84	0.3428
Fb2:0-30	113.94	**<.0001**	33.68	1	29.89	1	23.19	**<.0001**	29.63	**<.0001**	18.83	**<.0001**	31.19	0.1279	35.38	**<.0001**
Fb2:0-39	139.53	1	39.97	0.6477	30.21	1	28.83	1	34.85	0.3164	25.66	1	35.99	1	43.43	0.7041
Fb2:0-45	125.66	**0.0001**	31.46	1	28.80	1	25.08	0.0537	33.59	**0.0184**	21.89	**0.0236**	32.22	0.5541	40.30	**0.0028**
Fb2:0-63	122.94	**<.0001**	33.58	1	34.58	0.1997	25.05	**0.0491**	32.16	**0.0001**	19.06	**<.0001**	30.40	**0.0372**	37.45	**<.0001**
Fb2:0-73	119.12	**<.0001**	31.53	1	34.51	0.1963	23.91	**0.001**	31.98	**<.0001**	17.71	**<.0001**	28.75	**0.0006**	36.61	**<.0001**
Fb2:39-47	141.93	1	28.87	0.996	29.21	1	27.92	1	38.66	0.9943	25.68	1	33.53	0.9984	46.43	1
Fb2:39-63	124.88	**<.0001**	25.62	0.5043	30.32	1	24.11	**0.0027**	35.31	0.6497	20.44	**0.0005**	29.99	**0.016**	40.82	**0.0102**
Fb3:0-15	124.37	**<.0001**	29.05	0.9968	32.67	0.678	24.43	**0.0067**	34.22	0.0877	19.47	**<.0001**	29.47	**0.0039**	39.40	**0.0002**
Fb3:0-8	126.86	**0.0001**	32.38	1	34.10	0.1963	25.46	0.0771	33.71	**0.0111**	19.89	**<.0001**	30.49	**0.0206**	39.19	**<.0001**
Fb4:0-20	110.30	**<.0001**	32.25	1	34.00	0.4835	22.53	**<.0001**	28.62	**<.0001**	15.96	**<.0001**	28.18	**0.0012**	32.88	**<.0001**
Fb4:0-31	117.90	**<.0001**	35.37	1	37.56	**0.0062**	24.33	**0.0047**	30.41	**<.0001**	16.46	**<.0001**	27.76	**<.0001**	34.65	**<.0001**
Fb4:0-44	125.72	**<.0001**	37.57	0.99	35.61	0.0615	26.10	0.4004	31.74	**<.0001**	19.79	**<.0001**	31.84	0.3288	37.46	**<.0001**
Fb4:58-78	124.36	**0.0004**	34.88	1	28.23	1	25.54	0.3253	31.75	**0.0004**	22.38	0.1762	33.81	1	39.02	**0.001**
Fb5:0-11	125.66	**0.0001**	28.20	0.9683	33.28	0.5066	24.65	**0.016**	34.85	0.3376	19.47	**<.0001**	29.16	**0.0022**	39.93	**0.001**
Fb5:0-20	131.65	0.0655	30.76	1	34.40	0.2151	26.18	0.4718	35.68	0.8856	20.78	**0.0011**	30.18	**0.0207**	41.31	**0.0272**
Fb5:0-35	142.74	1	34.58	1	35.70	0.0623	28.94	1	37.33	1	23.51	0.3891	32.20	0.5252	44.15	0.9806
Fb5:0-76	162.59	**<.0001**	69.99	**<.0001**	48.31	**<.0001**	38.18	**<.0001**	31.08	**<.0001**	27.72	0.9999	42.09	**<.0001**	41.76	0.0618
Fb5:11-76	161.25	**<.0001**	81.55	**<.0001**	54.18	**<.0001**	40.29	**<.0001**	26.22	**<.0001**	27.09	1	46.55	**<.0001**	37.98	**<.0001**
Fb5:20-76	162.20	**<.0001**	76.42	**<.0001**	50.28	**<.0001**	39.25	**<.0001**	28.59	**<.0001**	27.90	0.9988	44.75	**<.0001**	40.14	**0.001**
Fb5:39-76	141.49	1	41.31	0.3694	35.44	0.0818	29.51	0.9991	35.01	0.4155	24.14	0.7529	34.48	1	42.63	0.2886
Fb5:41-76	123.94	**<.0001**	38.38	0.9478	38.98	**0.0009**	25.87	0.3	31.31	**<.0001**	17.70	**<.0001**	29.34	**0.0034**	36.08	**<.0001**
Fb5:50-76	143.18	1	39.99	0.6433	36.07	**0.0414**	29.68	0.991	35.78	0.9308	24.11	0.7361	33.98	1	43.20	0.5696
Fb6:0-5	137.71	0.9935	34.02	1	35.51	0.0758	27.77	1	36.41	1	22.14	**0.036**	31.19	0.1388	42.65	0.2949
Fb6:101-101	116.35	**<.0001**	38.50	0.9437	39.49	**0.0005**	24.69	**0.021**	28.77	**<.0001**	15.86	**<.0001**	28.74	**0.0009**	32.92	**<.0001**
Fb6:71-101	146.48	0.943	50.03	**0.0002**	39.77	**0.0002**	32.10	**0.0178**	33.05	**0.0034**	24.95	0.9969	37.39	0.9124	41.53	**0.0488**
Fb6:84-101	136.68	0.9596	42.59	0.2793	35.77	0.1079	28.98	1	32.97	**0.0068**	23.26	0.3945	35.25	1	40.39	**0.009**
Fb7:0-10	136.46	0.8765	33.42	1	30.80	0.998	27.41	0.9999	35.93	0.9761	23.85	0.5808	33.08	0.944	43.34	0.6542
Fb7:0-27	134.17	0.359	30.02	1	28.54	1	26.62	0.8016	36.12	0.9961	24.11	0.7331	33.25	0.9776	43.62	0.8086
Fb7:26-45	129.72	**0.0112**	33.21	1	33.32	0.4734	26.10	0.419	34.36	0.1191	20.92	**0.0016**	30.72	0.0614	40.49	**0.0038**
Fb7:26-59	132.37	0.114	33.37	1	31.24	0.984	26.64	0.8163	34.86	0.3242	22.69	0.1083	32.59	0.7394	41.90	0.0883
Fb7:43-59	130.04	**0.0154**	31.39	1	32.22	0.8144	25.90	0.2996	35.01	0.4149	21.31	**0.0048**	30.88	0.0823	41.29	**0.0259**
Fb7:52-59	129.11	**0.0059**	28.58	0.9851	29.35	1	25.27	0.0769	35.54	0.7962	21.98	**0.0251**	30.90	0.0845	42.11	0.1297
RV	141.66	1	33.17	1	28.01	1	28.39	1	37.35	1	26.49	1	35.17	1	45.96	1
YW	169.99	**<.0001**	146.07	**<.0001**	75.46	**<.0001**	55.87	**<.0001**	2.17	**<.0001**	34.04	**<.0001**	86.40	**<.0001**	34.12	**<.0001**

Table S3-2. 416 genes annotation in 35-39cM in LG5 from Fragaria vesca V4.0 a1 genome database.(Edger et al. 2018)

Query	Match	Description
FvH4_5g10460.1	IPR000911	Ribosomal protein L11/L12
FvH4_5g10460.1	IPR006519	Ribosomal protein L11, bacterial-type
FvH4_5g10460.1	IPR020783	Ribosomal protein L11, C-terminal
FvH4_5g10460.1	IPR020784	Ribosomal protein L11, N-terminal
FvH4_5g10460.1	IPR020785	Ribosomal protein L11, conserved site
FvH4_5g10460.1	IPR036769	Ribosomal protein L11, C-terminal domain superfamily
FvH4_5g10460.1	IPR036796	Ribosomal protein L11/L12, N-terminal domain superfamily
FvH4_5g10470.1	IPR036852	Peptidase S8/S53 domain superfamily
FvH4_5g10480.1	IPR000209	Peptidase S8/S53 domain
FvH4_5g10480.1	IPR010259	Peptidase S8 propeptide/proteinase inhibitor I9
FvH4_5g10480.1	IPR015500	Peptidase S8, subtilisin-related
FvH4_5g10480.1	IPR023828	Peptidase S8, subtilisin, Ser-active site
FvH4_5g10480.1	IPR036852	Peptidase S8/S53 domain superfamily
FvH4_5g10480.1	IPR037045	Peptidase S8 propeptide/proteinase inhibitor I9 superfamily
FvH4_5g10490.1	IPR000209	Peptidase S8/S53 domain
FvH4_5g10490.1	IPR002259	Equilibrative nucleoside transporter
FvH4_5g10490.1	IPR010259	Peptidase S8 propeptide/proteinase inhibitor I9
FvH4_5g10490.1	IPR015500	Peptidase S8, subtilisin-related
FvH4_5g10490.1	IPR023828	Peptidase S8, subtilisin, Ser-active site
FvH4_5g10490.1	IPR036259	MFS transporter superfamily
FvH4_5g10490.1	IPR036852	Peptidase S8/S53 domain superfamily
FvH4_5g10490.1	IPR037045	Peptidase S8 propeptide/proteinase inhibitor I9 superfamily
FvH4_5g10510.1	IPR002659	Glycosyl transferase, family 31
FvH4_5g10510.1	IPR025298	Domain of unknown function DUF4094
FvH4_5g10520.1	IPR002913	START domain
FvH4_5g10520.1	IPR023393	START-like domain superfamily
FvH4_5g10580.1	IPR001296	Glycosyl transferase, family 1
FvH4_5g10580.1	IPR011835	Bacterial/plant glycogen synthase
FvH4_5g10580.1	IPR013534	Starch synthase, catalytic domain
FvH4_5g10590.1	IPR000711	ATPase, OSCP/delta subunit
FvH4_5g10590.1	IPR026015	F1F0 ATP synthase OSCP/delta subunit, N-terminal domain superfamily
FvH4_5g10600.1	IPR000490	Glycoside hydrolase family 17
FvH4_5g10600.1	IPR017853	Glycoside hydrolase superfamily
FvH4_5g10610.1	IPR001394	Peptidase C19, ubiquitin carboxyl-terminal hydrolase
FvH4_5g10610.1	IPR006615	Peptidase C19, ubiquitin-specific peptidase, DUSP domain
FvH4_5g10610.1	IPR018200	Ubiquitin specific protease, conserved site
FvH4_5g10610.1	IPR028134	Ubiquitin carboxyl-terminal hydrolase USP4
FvH4_5g10610.1	IPR028889	Ubiquitin specific protease domain
FvH4_5g10610.1	IPR035927	DUSP-like superfamily
FvH4_5g10620.1	IPR001965	Zinc finger, PHD-type
FvH4_5g10620.1	IPR011011	Zinc finger, FYVE/PHD-type
FvH4_5g10620.1	IPR013083	Zinc finger, RING/FYVE/PHD-type
FvH4_5g10620.1	IPR019786	Zinc finger, PHD-type, conserved site
FvH4_5g10630.1	IPR000504	RNA recognition motif domain
FvH4_5g10630.1	IPR009818	Ataxin-2, C-terminal
FvH4_5g10630.1	IPR035979	RNA-binding domain superfamily
FvH4_5g10640.1	IPR002328	Alcohol dehydrogenase, zinc-type, conserved site
FvH4_5g10640.1	IPR011032	GroES-like superfamily
FvH4_5g10640.1	IPR013149	Alcohol dehydrogenase, C-terminal
FvH4_5g10640.1	IPR013154	Alcohol dehydrogenase, N-terminal
FvH4_5g10640.1	IPR036291	NAD(P)-binding domain superfamily
FvH4_5g10650.1	IPR002328	Alcohol dehydrogenase, zinc-type, conserved site
FvH4_5g10650.1	IPR011032	GroES-like superfamily
FvH4_5g10650.1	IPR013154	Alcohol dehydrogenase, N-terminal
FvH4_5g10650.1	IPR036291	NAD(P)-binding domain superfamily

FvH4_5g10670.1	IPR005061	Vacuolar protein sorting-associated protein Ist1
FvH4_5g10680.1	IPR025322	Protein of unknown function DUF4228, plant
FvH4_5g10690.1	IPR003441	NAC domain
FvH4_5g10690.1	IPR036093	NAC domain superfamily
FvH4_5g10700.1	IPR002109	Glutaredoxin
FvH4_5g10700.1	IPR036249	Thioredoxin-like superfamily
FvH4_5g10710.1	IPR000048	IQ motif, EF-hand binding site
FvH4_5g10710.1	IPR025064	Domain of unknown function DUF4005
FvH4_5g10710.1	IPR027417	P-loop containing nucleoside triphosphate hydrolase
FvH4_5g10720.1	IPR000719	Protein kinase domain
FvH4_5g10720.1	IPR008271	Serine/threonine-protein kinase, active site
FvH4_5g10720.1	IPR011009	Protein kinase-like domain superfamily
FvH4_5g10730.1	IPR000157	Toll/interleukin-1 receptor homology (TIR) domain
FvH4_5g10730.1	IPR035897	Toll/interleukin-1 receptor homology (TIR) domain superfamily
FvH4_5g10740.1	IPR000157	Toll/interleukin-1 receptor homology (TIR) domain
FvH4_5g10740.1	IPR035897	Toll/interleukin-1 receptor homology (TIR) domain superfamily
FvH4_5g10750.1	IPR014977	WRC domain
FvH4_5g10750.1	IPR014978	Glutamine-Leucine-Glutamine, QLQ
FvH4_5g10750.1	IPR031137	Growth-regulating factor
FvH4_5g10780.1	IPR009646	Root cap
FvH4_5g10800.1	IPR009637	Lung seven transmembrane receptor-like
FvH4_5g10810.1	IPR001701	Glycoside hydrolase family 9
FvH4_5g10810.1	IPR008928	Six-hairpin glycosidase-like
FvH4_5g10810.1	IPR018221	Glycoside hydrolase family 9, His active site
FvH4_5g10810.1	IPR033126	Glycosyl hydrolases family 9, Asp/Glu active sites
FvH4_5g10820.1	IPR025114	Domain of unknown function DUF4033
FvH4_5g10830.1	IPR001087	GDSL lipase/esterase
FvH4_5g10830.1	IPR036514	SGNH hydrolase superfamily
FvH4_5g10840.1	IPR000073	Alpha/beta hydrolase fold-1
FvH4_5g10840.1	IPR029058	Alpha/Beta hydrolase fold
FvH4_5g10850.1	IPR011598	Myc-type, basic helix-loop-helix (bHLH) domain
FvH4_5g10850.1	IPR025610	Transcription factor MYC/MYB N-terminal
FvH4_5g10850.1	IPR036638	Helix-loop-helix DNA-binding domain superfamily
FvH4_5g10860.1	IPR009880	Glyoxal oxidase, N-terminal
FvH4_5g10860.1	IPR011043	Galactose oxidase/kelch, beta-propeller
FvH4_5g10860.1	IPR013783	Immunoglobulin-like fold
FvH4_5g10860.1	IPR014756	Immunoglobulin E-set
FvH4_5g10860.1	IPR015202	Galactose oxidase-like, Early set domain
FvH4_5g10860.1	IPR037293	Galactose oxidase, beta-propeller
FvH4_5g10880.1	IPR001128	Cytochrome P450
FvH4_5g10880.1	IPR002401	Cytochrome P450, E-class, group I
FvH4_5g10880.1	IPR017972	Cytochrome P450, conserved site
FvH4_5g10880.1	IPR036396	Cytochrome P450 superfamily
FvH4_5g10890.1	IPR000277	Cys/Met metabolism, pyridoxal phosphate-dependent enzyme
FvH4_5g10890.1	IPR015421	Pyridoxal phosphate-dependent transferase, major region, subdomain 1
FvH4_5g10890.1	IPR015422	Pyridoxal phosphate-dependent transferase, subdomain 2
FvH4_5g10890.1	IPR015424	Pyridoxal phosphate-dependent transferase
FvH4_5g10900.1	IPR000073	Alpha/beta hydrolase fold-1
FvH4_5g10900.1	IPR000952	AB hydrolase 4, conserved site
FvH4_5g10900.1	IPR012020	AB hydrolase 4 family
FvH4_5g10900.1	IPR029058	Alpha/Beta hydrolase fold
FvH4_5g10920.1	IPR007493	Protein of unknown function DUF538
FvH4_5g10920.1	IPR036758	At5g01610-like superfamily
FvH4_5g10940.1	IPR000195	Rab-GTPase-TBC domain
FvH4_5g10940.1	IPR035969	Rab-GTPase-TBC domain superfamily
FvH4_5g10960.1	IPR011990	Tetratricopeptide-like helical domain superfamily
FvH4_5g10960.1	IPR011993	PH-like domain superfamily
FvH4_5g10970.1	IPR004813	Oligopeptide transporter, OPT superfamily
FvH4_5g10990.1	IPR000719	Protein kinase domain

FvH4_5g10990.1	IPR011009	Protein kinase-like domain superfamily
FvH4_5g10990.1	IPR017441	Protein kinase, ATP binding site
FvH4_5g11020.1	IPR003226	Metal-dependent protein hydrolase
FvH4_5g11030.1	IPR008509	Molybdate-anion transporter
FvH4_5g11030.1	IPR036259	MFS transporter superfamily
FvH4_5g11040.1	IPR003245	Phytocyanin domain
FvH4_5g11040.1	IPR008972	Cupredoxin
FvH4_5g11050.1	IPR007757	MT-A70-like
FvH4_5g11050.1	IPR029063	S-adenosyl-L-methionine-dependent methyltransferase
FvH4_5g11060.1	IPR004648	Tetrapeptide transporter, OPT1/isp4
FvH4_5g11060.1	IPR004813	Oligopeptide transporter, OPT superfamily
FvH4_5g11070.1	IPR004910	Yippee/Mis18/Cereblon
FvH4_5g11070.1	IPR034751	Yippee domain
FvH4_5g11080.1	IPR025124	Domain of unknown function DUF4050
FvH4_5g11090.1	IPR008808	Powdery mildew resistance protein, RPW8 domain
FvH4_5g11100.1	IPR001810	F-box domain
FvH4_5g11100.1	IPR008808	Powdery mildew resistance protein, RPW8 domain
FvH4_5g11100.1	IPR025886	Phloem protein 2-like
FvH4_5g11100.1	IPR036047	F-box-like domain superfamily
FvH4_5g11120.1	IPR008808	Powdery mildew resistance protein, RPW8 domain
FvH4_5g11130.1	IPR002182	NB-ARC
FvH4_5g11130.1	IPR027417	P-loop containing nucleoside triphosphate hydrolase
FvH4_5g11170.1	IPR025886	Phloem protein 2-like
FvH4_5g11180.1	IPR008808	Powdery mildew resistance protein, RPW8 domain
FvH4_5g11200.1	IPR008808	Powdery mildew resistance protein, RPW8 domain
FvH4_5g11210.1	IPR001810	F-box domain
FvH4_5g11210.1	IPR008808	Powdery mildew resistance protein, RPW8 domain
FvH4_5g11210.1	IPR025886	Phloem protein 2-like
FvH4_5g11210.1	IPR036047	F-box-like domain superfamily
FvH4_5g11220.1	IPR008808	Powdery mildew resistance protein, RPW8 domain
FvH4_5g11230.1	IPR008808	Powdery mildew resistance protein, RPW8 domain
FvH4_5g11250.1	IPR008808	Powdery mildew resistance protein, RPW8 domain
FvH4_5g11250.1	IPR012416	CALMODULIN-BINDING PROTEIN60
FvH4_5g11270.1	IPR001012	UBX domain
FvH4_5g11270.1	IPR006577	UAS
FvH4_5g11270.1	IPR029071	Ubiquitin-like domain superfamily
FvH4_5g11270.1	IPR036249	Thioredoxin-like superfamily
FvH4_5g11280.1	IPR001841	Zinc finger, RING-type
FvH4_5g11280.1	IPR013083	Zinc finger, RING/FYVE/PHD-type
FvH4_5g11280.1	IPR024766	Zinc finger, RING-H2-type
FvH4_5g11290.1	IPR010399	Tify domain
FvH4_5g11290.1	IPR018467	CO/COL/TOC1, conserved site
FvH4_5g11300.1	IPR011598	Myc-type, basic helix-loop-helix (bHLH) domain
FvH4_5g11300.1	IPR020966	Aluminum-activated malate transporter
FvH4_5g11300.1	IPR025610	Transcription factor MYC/MYB N-terminal
FvH4_5g11300.1	IPR036638	Helix-loop-helix DNA-binding domain superfamily
FvH4_5g11340.1	IPR001554	Glycoside hydrolase, family 14
FvH4_5g11340.1	IPR017853	Glycoside hydrolase superfamily
FvH4_5g11350.1	IPR008948	L-Aspartase-like
FvH4_5g11350.1	IPR013539	Adenylosuccinate lyase PurB, C-terminal
FvH4_5g11350.1	IPR022761	Fumarate lyase, N-terminal
FvH4_5g11350.1	IPR024083	Fumarase/histidase, N-terminal
FvH4_5g11380.1	IPR000210	BTB/POZ domain
FvH4_5g11380.1	IPR011333	SKP1/BTB/POZ domain superfamily
FvH4_5g11390.1	IPR007203	ORMDL family
FvH4_5g11410.1	IPR003851	Zinc finger, Dof-type
FvH4_5g11420.1	IPR001680	WD40 repeat
FvH4_5g11420.1	IPR015943	WD40/YVTN repeat-like-containing domain superfamily
FvH4_5g11420.1	IPR017986	WD40-repeat-containing domain

FvH4_5g11420.1	IPR020472	G-protein beta WD-40 repeat
FvH4_5g11420.1	IPR036322	WD40-repeat-containing domain superfamily
FvH4_5g11430.1	IPR001005	SANT/Myb domain
FvH4_5g11430.1	IPR009057	Homeobox-like domain superfamily
FvH4_5g11430.1	IPR017930	Myb domain
FvH4_5g11440.1	IPR015324	Ribosomal protein Rsm22-like
FvH4_5g11440.1	IPR029063	S-adenosyl-L-methionine-dependent methyltransferase
FvH4_5g11450.1	IPR002347	Short-chain dehydrogenase/reductase SDR
FvH4_5g11450.1	IPR036291	NAD(P)-binding domain superfamily
FvH4_5g11460.1	IPR001661	Glycoside hydrolase, family 37
FvH4_5g11460.1	IPR008928	Six-hairpin glycosidase-like
FvH4_5g11480.1	IPR003439	ABC transporter-like
FvH4_5g11480.1	IPR003593	AAA+ ATPase domain
FvH4_5g11480.1	IPR017871	ABC transporter, conserved site
FvH4_5g11480.1	IPR027417	P-loop containing nucleoside triphosphate hydrolase
FvH4_5g11480.1	IPR032781	ABC-transporter extension domain
FvH4_5g11490.1	IPR003010	Carbon-nitrogen hydrolase
FvH4_5g11490.1	IPR003694	NAD(+) synthetase
FvH4_5g11490.1	IPR014445	Glutamine-dependent NAD(+) synthetase
FvH4_5g11490.1	IPR014729	Rossmann-like alpha/beta/alpha sandwich fold
FvH4_5g11490.1	IPR022310	NAD/GMP synthase
FvH4_5g11490.1	IPR036526	Carbon-nitrogen hydrolase superfamily
FvH4_5g11500.1	IPR013126	Heat shock protein 70 family
FvH4_5g11500.1	IPR018181	Heat shock protein 70, conserved site
FvH4_5g11500.1	IPR029047	Heat shock protein 70kD, peptide-binding domain superfamily
FvH4_5g11500.1	IPR029048	Heat shock protein 70kD, C-terminal domain superfamily
FvH4_5g11510.1	IPR001494	Importin-beta, N-terminal domain
FvH4_5g11510.1	IPR011989	Armadillo-like helical
FvH4_5g11510.1	IPR016024	Armadillo-type fold
FvH4_5g11510.1	IPR027140	Importin subunit beta-1, plants
FvH4_5g11520.1	IPR001554	Glycoside hydrolase, family 14
FvH4_5g11520.1	IPR017853	Glycoside hydrolase superfamily
FvH4_5g11550.1	IPR011990	Tetratricopeptide-like helical domain superfamily
FvH4_5g11550.1	IPR013026	Tetratricopeptide repeat-containing domain
FvH4_5g11550.1	IPR019734	Tetratricopeptide repeat
FvH4_5g11550.1	IPR036388	Winged helix-like DNA-binding domain superfamily
FvH4_5g11560.1	IPR004316	SWEET sugar transporter
FvH4_5g11570.1	IPR000270	PB1 domain
FvH4_5g11570.1	IPR003035	RWP-RK domain
FvH4_5g11570.1	IPR029016	GAF-like domain superfamily
FvH4_5g11610.1	IPR001141	Ribosomal protein L27e
FvH4_5g11610.1	IPR008991	Translation protein SH3-like domain superfamily
FvH4_5g11610.1	IPR014722	Ribosomal protein L2, domain 2
FvH4_5g11630.1	IPR018838	Domain of unknown function DUF2439
FvH4_5g11640.1	IPR028457	ABI family
FvH4_5g11650.1	IPR006746	26S proteasome non-ATPase regulatory subunit Rpn12
FvH4_5g11650.1	IPR033464	CSN8/PSMD8/EIF3K
FvH4_5g11650.1	IPR036388	Winged helix-like DNA-binding domain superfamily
FvH4_5g11660.1	IPR000529	Ribosomal protein S6
FvH4_5g11660.1	IPR014717	Translation elongation factor EF1B/ribosomal protein S6
FvH4_5g11660.1	IPR020814	Ribosomal protein S6, plastid/chloroplast
FvH4_5g11660.1	IPR035980	Ribosomal protein S6 superfamily
FvH4_5g11670.1	IPR001841	Zinc finger, RING-type
FvH4_5g11670.1	IPR013083	Zinc finger, RING/FYVE/PHD-type
FvH4_5g11680.1	IPR005150	Cellulose synthase
FvH4_5g11680.1	IPR029044	Nucleotide-diphospho-sugar transferases
FvH4_5g11690.1	IPR000270	PB1 domain
FvH4_5g11690.1	IPR003340	B3 DNA binding domain
FvH4_5g11690.1	IPR010525	Auxin response factor

FvH4_5g11690.1	IPR015300	DNA-binding pseudobarrel domain superfamily
FvH4_5g11690.1	IPR033389	AUX/IAA domain
FvH4_5g11700.1	IPR014030	Beta-ketoacyl synthase, N-terminal
FvH4_5g11700.1	IPR014031	Beta-ketoacyl synthase, C-terminal
FvH4_5g11700.1	IPR016039	Thiolase-like
FvH4_5g11700.1	IPR017568	3-oxoacyl-[acyl-carrier-protein] synthase 2
FvH4_5g11700.1	IPR018201	Beta-ketoacyl synthase, active site
FvH4_5g11700.1	IPR020841	Polyketide synthase, beta-ketoacyl synthase domain
FvH4_5g11710.1	IPR001752	Kinesin motor domain
FvH4_5g11710.1	IPR019821	Kinesin motor domain, conserved site
FvH4_5g11710.1	IPR021881	NPK1-activating kinesin-like protein, C-terminal
FvH4_5g11710.1	IPR027417	P-loop containing nucleoside triphosphate hydrolase
FvH4_5g11710.1	IPR027640	Kinesin-like protein
FvH4_5g11710.1	IPR036961	Kinesin motor domain superfamily
FvH4_5g11720.1	IPR008507	Protein of unknown function DUF789
FvH4_5g11730.1	IPR001701	Glycoside hydrolase family 9
FvH4_5g11730.1	IPR008928	Six-hairpin glycosidase-like
FvH4_5g11730.1	IPR018221	Glycoside hydrolase family 9, His active site
FvH4_5g11740.1	IPR025287	Wall-associated receptor kinase, galacturonan-binding domain
FvH4_5g11740.1	IPR032872	Wall-associated receptor kinase, C-terminal
FvH4_5g11760.1	IPR000719	Protein kinase domain
FvH4_5g11760.1	IPR008271	Serine/threonine-protein kinase, active site
FvH4_5g11760.1	IPR011009	Protein kinase-like domain superfamily
FvH4_5g11760.1	IPR017441	Protein kinase, ATP binding site
FvH4_5g11760.1	IPR032872	Wall-associated receptor kinase, C-terminal
FvH4_5g11770.1	IPR003822	Paired amphipathic helix
FvH4_5g11770.1	IPR013194	Histone deacetylase interacting domain
FvH4_5g11770.1	IPR036600	Paired amphipathic helix superfamily
FvH4_5g11780.1	IPR000719	Protein kinase domain
FvH4_5g11780.1	IPR000742	EGF-like domain
FvH4_5g11780.1	IPR001245	Serine-threonine/tyrosine-protein kinase, catalytic domain
FvH4_5g11780.1	IPR008271	Serine/threonine-protein kinase, active site
FvH4_5g11780.1	IPR011009	Protein kinase-like domain superfamily
FvH4_5g11780.1	IPR013032	EGF-like, conserved site
FvH4_5g11780.1	IPR017441	Protein kinase, ATP binding site
FvH4_5g11780.1	IPR025287	Wall-associated receptor kinase, galacturonan-binding domain
FvH4_5g11780.1	IPR032872	Wall-associated receptor kinase, C-terminal
FvH4_5g11790.1	IPR032872	Wall-associated receptor kinase, C-terminal
FvH4_5g11800.1	IPR012416	CALMODULIN-BINDING PROTEIN60
FvH4_5g11810.1	IPR003406	Glycosyl transferase, family 14
FvH4_5g11820.1	IPR003406	Glycosyl transferase, family 14
FvH4_5g11830.1	IPR001810	F-box domain
FvH4_5g11830.1	IPR006527	F-box associated domain, type 1
FvH4_5g11830.1	IPR015915	Kelch-type beta propeller
FvH4_5g11830.1	IPR017451	F-box associated interaction domain
FvH4_5g11830.1	IPR036047	F-box-like domain superfamily
FvH4_5g11840.1	IPR001965	Zinc finger, PHD-type
FvH4_5g11840.1	IPR011011	Zinc finger, FYVE/PHD-type
FvH4_5g11840.1	IPR013083	Zinc finger, RING/FYVE/PHD-type
FvH4_5g11840.1	IPR019786	Zinc finger, PHD-type, conserved site
FvH4_5g11850.1	IPR002553	Clathrin/coatomer adaptor, adaptin-like, N-terminal
FvH4_5g11850.1	IPR011989	Armadillo-like helical
FvH4_5g11850.1	IPR016024	Armadillo-type fold
FvH4_5g11850.1	IPR017105	Adaptor protein complex AP-3, delta subunit
FvH4_5g11860.1	IPR001841	Zinc finger, RING-type
FvH4_5g11860.1	IPR008913	Zinc finger, CHY-type
FvH4_5g11860.1	IPR012312	Haemerythrin-like
FvH4_5g11860.1	IPR013083	Zinc finger, RING/FYVE/PHD-type
FvH4_5g11860.1	IPR017921	Zinc finger, CTCHY-type

FvH4_5g11860.1	IPR037274	Zinc finger, CHY-type superfamily
FvH4_5g11860.1	IPR037275	Zinc finger, CTCHY-type superfamily
FvH4_5g11870.1	IPR012442	Protein of unknown function DUF1645, plant
FvH4_5g11880.1	IPR000555	JAB1/MPN/MOV34 metalloenzyme domain
FvH4_5g11880.1	IPR015063	USP8 dimerisation domain
FvH4_5g11880.1	IPR037518	MPN domain
FvH4_5g11890.1	IPR000719	Protein kinase domain
FvH4_5g11890.1	IPR001245	Serine-threonine/tyrosine-protein kinase, catalytic domain
FvH4_5g11890.1	IPR008271	Serine/threonine-protein kinase, active site
FvH4_5g11890.1	IPR011009	Protein kinase-like domain superfamily
FvH4_5g11890.1	IPR024171	S-receptor-like serine/threonine-protein kinase
FvH4_5g11900.1	IPR004158	Protein of unknown function DUF247, plant
FvH4_5g11910.1	IPR004158	Protein of unknown function DUF247, plant
FvH4_5g11920.1	IPR000554	Ribosomal protein S7e
FvH4_5g11930.1	IPR001005	SANT/Myb domain
FvH4_5g11930.1	IPR001345	Phosphoglycerate/bisphosphoglycerate mutase, active site
FvH4_5g11930.1	IPR009057	Homeobox-like domain superfamily
FvH4_5g11930.1	IPR017930	Myb domain
FvH4_5g11940.1	IPR000620	EamA domain
FvH4_5g11950.1	IPR011598	Myc-type, basic helix-loop-helix (bHLH) domain
FvH4_5g11950.1	IPR036638	Helix-loop-helix DNA-binding domain superfamily
FvH4_5g11960.1	IPR006927	Protein of unknown function DUF639
FvH4_5g11980.1	IPR008889	VQ
FvH4_5g11990.1	IPR002075	Nuclear transport factor 2
FvH4_5g11990.1	IPR018222	Nuclear transport factor 2, eukaryote
FvH4_5g11990.1	IPR032710	NTF2-like domain superfamily
FvH4_5g12010.1	IPR001841	Zinc finger, RING-type
FvH4_5g12010.1	IPR013083	Zinc finger, RING/FYVE/PHD-type
FvH4_5g12010.1	IPR020966	Aluminum-activated malate transporter
FvH4_5g12010.1	IPR024766	Zinc finger, RING-H2-type
FvH4_5g12040.1	IPR000008	C2 domain
FvH4_5g12040.1	IPR035892	C2 domain superfamily
FvH4_5g12060.1	IPR002421	5'-3' exonuclease, N-terminal
FvH4_5g12060.1	IPR008918	Helix-hairpin-helix motif, class 2
FvH4_5g12060.1	IPR020045	DNA polymerase I-like, H3TH domain
FvH4_5g12060.1	IPR020046	5'-3' exonuclease, alpha-helical arch, N-terminal
FvH4_5g12060.1	IPR029060	PIN-like domain superfamily
FvH4_5g12060.1	IPR036279	5'-3' exonuclease, C-terminal domain superfamily
FvH4_5g12070.1	IPR001890	RNA-binding, CRM domain
FvH4_5g12070.1	IPR035920	YhbY-like superfamily
FvH4_5g12080.1	IPR001890	RNA-binding, CRM domain
FvH4_5g12080.1	IPR035920	YhbY-like superfamily
FvH4_5g12090.1	IPR003441	NAC domain
FvH4_5g12090.1	IPR036093	NAC domain superfamily
FvH4_5g12100.1	IPR008195	Ribosomal protein L34Ae
FvH4_5g12100.1	IPR012870	Protein of unknown function DUF1666
FvH4_5g12130.1	IPR001715	Calponin homology domain
FvH4_5g12130.1	IPR001752	Kinesin motor domain
FvH4_5g12130.1	IPR027417	P-loop containing nucleoside triphosphate hydrolase
FvH4_5g12130.1	IPR027640	Kinesin-like protein
FvH4_5g12130.1	IPR036872	Calponin-like domain superfamily
FvH4_5g12130.1	IPR036961	Kinesin motor domain superfamily
FvH4_5g12150.1	IPR000315	B-box-type zinc finger
FvH4_5g12150.1	IPR010402	CCT domain
FvH4_5g12160.1	IPR001611	Leucine-rich repeat
FvH4_5g12160.1	IPR006553	Leucine-rich repeat, cysteine-containing subtype
FvH4_5g12160.1	IPR032675	Leucine-rich repeat domain superfamily
FvH4_5g12170.1	IPR006553	Leucine-rich repeat, cysteine-containing subtype
FvH4_5g12170.1	IPR032675	Leucine-rich repeat domain superfamily

FvH4_5g12180.1	IPR006553	Leucine-rich repeat, cysteine-containing subtype
FvH4_5g12180.1	IPR032675	Leucine-rich repeat domain superfamily
FvH4_5g12190.1	IPR006553	Leucine-rich repeat, cysteine-containing subtype
FvH4_5g12190.1	IPR032675	Leucine-rich repeat domain superfamily
FvH4_5g12200.1	IPR006553	Leucine-rich repeat, cysteine-containing subtype
FvH4_5g12200.1	IPR032675	Leucine-rich repeat domain superfamily
FvH4_5g12210.1	IPR001611	Leucine-rich repeat
FvH4_5g12210.1	IPR006553	Leucine-rich repeat, cysteine-containing subtype
FvH4_5g12210.1	IPR032675	Leucine-rich repeat domain superfamily
FvH4_5g12220.1	IPR018392	LysM domain
FvH4_5g12220.1	IPR036779	LysM domain superfamily
FvH4_5g12230.1	IPR005135	Endonuclease/exonuclease/phosphatase
FvH4_5g12230.1	IPR036691	Endonuclease/exonuclease/phosphatase superfamily
FvH4_5g12240.1	IPR001611	Leucine-rich repeat
FvH4_5g12240.1	IPR006553	Leucine-rich repeat, cysteine-containing subtype
FvH4_5g12240.1	IPR032675	Leucine-rich repeat domain superfamily
FvH4_5g12250.1	IPR001471	AP2/ERF domain
FvH4_5g12250.1	IPR016177	DNA-binding domain superfamily
FvH4_5g12250.1	IPR036955	AP2/ERF domain superfamily
FvH4_5g12270.1	IPR034577	Protein NIM1-INTERACTING 2
FvH4_5g12310.1	IPR016140	Bifunctional inhibitor/plant lipid transfer protein/seed storage helical domain
FvH4_5g12310.1	IPR036312	Bifunctional inhibitor/plant lipid transfer protein/seed storage helical domain superfamily
FvH4_5g12320.1	IPR036163	Heavy metal-associated domain superfamily
FvH4_5g12330.1	IPR000357	HEAT repeat
FvH4_5g12330.1	IPR011989	Armadillo-like helical
FvH4_5g12330.1	IPR016024	Armadillo-type fold
FvH4_5g12330.1	IPR020966	Aluminum-activated malate transporter
FvH4_5g12330.1	IPR021133	HEAT, type 2
FvH4_5g12340.1	IPR001680	WD40 repeat
FvH4_5g12340.1	IPR006692	Coatomer, WD associated region
FvH4_5g12340.1	IPR010714	Coatomer, alpha subunit, C-terminal
FvH4_5g12340.1	IPR011048	Cytochrome cd1-nitrite reductase-like, haem d1 domain superfamily
FvH4_5g12340.1	IPR011990	Tetratricopeptide-like helical domain superfamily
FvH4_5g12340.1	IPR015943	WD40/YVTN repeat-like-containing domain superfamily
FvH4_5g12340.1	IPR016391	Coatomer alpha subunit
FvH4_5g12340.1	IPR017986	WD40-repeat-containing domain
FvH4_5g12340.1	IPR019775	WD40 repeat, conserved site
FvH4_5g12340.1	IPR020472	G-protein beta WD-40 repeat
FvH4_5g12340.1	IPR036322	WD40-repeat-containing domain superfamily
FvH4_5g12350.1	IPR001128	Cytochrome P450
FvH4_5g12350.1	IPR002401	Cytochrome P450, E-class, group I
FvH4_5g12350.1	IPR017972	Cytochrome P450, conserved site
FvH4_5g12350.1	IPR036396	Cytochrome P450 superfamily
FvH4_5g12360.1	IPR001194	cDENN domain
FvH4_5g12360.1	IPR037516	Tripartite DENN domain
FvH4_5g12370.1	IPR010714	Coatomer, alpha subunit, C-terminal
FvH4_5g12380.1	IPR001128	Cytochrome P450
FvH4_5g12380.1	IPR002401	Cytochrome P450, E-class, group I
FvH4_5g12380.1	IPR017972	Cytochrome P450, conserved site
FvH4_5g12380.1	IPR036396	Cytochrome P450 superfamily
FvH4_5g12400.1	IPR001128	Cytochrome P450
FvH4_5g12400.1	IPR002401	Cytochrome P450, E-class, group I
FvH4_5g12400.1	IPR017972	Cytochrome P450, conserved site
FvH4_5g12400.1	IPR036396	Cytochrome P450 superfamily
FvH4_5g12420.1	IPR006692	Coatomer, WD associated region
FvH4_5g12420.1	IPR011044	Quinoprotein amine dehydrogenase, beta chain-like
FvH4_5g12420.1	IPR015943	WD40/YVTN repeat-like-containing domain superfamily
FvH4_5g12430.1	IPR001128	Cytochrome P450
FvH4_5g12430.1	IPR002401	Cytochrome P450, E-class, group I

FvH4_5g12430.1	IPR017972	Cytochrome P450, conserved site
FvH4_5g12430.1	IPR036396	Cytochrome P450 superfamily
FvH4_5g12440.1	IPR001128	Cytochrome P450
FvH4_5g12440.1	IPR002401	Cytochrome P450, E-class, group I
FvH4_5g12440.1	IPR036396	Cytochrome P450 superfamily
FvH4_5g12450.1	IPR001194	cDENN domain
FvH4_5g12450.1	IPR001680	WD40 repeat
FvH4_5g12450.1	IPR005112	dDENN domain
FvH4_5g12450.1	IPR005113	uDENN domain
FvH4_5g12450.1	IPR015943	WD40/YVTN repeat-like-containing domain superfamily
FvH4_5g12450.1	IPR017986	WD40-repeat-containing domain
FvH4_5g12450.1	IPR019775	WD40 repeat, conserved site
FvH4_5g12450.1	IPR020472	G-protein beta WD-40 repeat
FvH4_5g12450.1	IPR036322	WD40-repeat-containing domain superfamily
FvH4_5g12450.1	IPR037516	Tripartite DENN domain
FvH4_5g12460.1	IPR006461	PLAC8 motif-containing protein
FvH4_5g12490.1	IPR011205	Uncharacterised conserved protein UCP015417, vWA
FvH4_5g12490.1	IPR024553	Domain of unknown function DUF2828
FvH4_5g12490.1	IPR024796	T4 endonuclease V
FvH4_5g12490.1	IPR036465	von Willebrand factor A-like domain superfamily
FvH4_5g12500.1	IPR001328	Peptidyl-tRNA hydrolase
FvH4_5g12500.1	IPR018171	Peptidyl-tRNA hydrolase, conserved site
FvH4_5g12500.1	IPR036416	Peptidyl-tRNA hydrolase superfamily
FvH4_5g12510.1	IPR003591	Leucine-rich repeat, typical subtype
FvH4_5g12510.1	IPR013210	Leucine-rich repeat-containing N-terminal, plant-type
FvH4_5g12510.1	IPR032675	Leucine-rich repeat domain superfamily
FvH4_5g12520.1	IPR001461	Aspartic peptidase A1 family
FvH4_5g12520.1	IPR001969	Aspartic peptidase, active site
FvH4_5g12520.1	IPR021109	Aspartic peptidase domain superfamily
FvH4_5g12520.1	IPR032799	Xylanase inhibitor, C-terminal
FvH4_5g12520.1	IPR032861	Xylanase inhibitor, N-terminal
FvH4_5g12520.1	IPR033121	Peptidase family A1 domain
FvH4_5g12520.1	IPR033148	ASPG1
FvH4_5g12530.1	IPR002885	Pentatricopeptide repeat
FvH4_5g12530.1	IPR011990	Tetratricopeptide-like helical domain superfamily
FvH4_5g12540.1	IPR003769	Adaptor protein ClpS, core
FvH4_5g12540.1	IPR014719	Ribosomal protein L7/L12, C-terminal/adaptor protein ClpS-like
FvH4_5g12550.1	IPR001128	Cytochrome P450
FvH4_5g12550.1	IPR002401	Cytochrome P450, E-class, group I
FvH4_5g12550.1	IPR017972	Cytochrome P450, conserved site
FvH4_5g12550.1	IPR036396	Cytochrome P450 superfamily
FvH4_5g12570.1	IPR001128	Cytochrome P450
FvH4_5g12570.1	IPR002401	Cytochrome P450, E-class, group I
FvH4_5g12570.1	IPR017972	Cytochrome P450, conserved site
FvH4_5g12570.1	IPR036396	Cytochrome P450 superfamily
FvH4_5g12580.1	IPR001128	Cytochrome P450
FvH4_5g12580.1	IPR002401	Cytochrome P450, E-class, group I
FvH4_5g12580.1	IPR017972	Cytochrome P450, conserved site
FvH4_5g12580.1	IPR036396	Cytochrome P450 superfamily
FvH4_5g12590.1	IPR005135	Endonuclease/exonuclease/phosphatase
FvH4_5g12590.1	IPR036691	Endonuclease/exonuclease/phosphatase superfamily
FvH4_5g12600.1	IPR006693	Partial AB-hydrolase lipase domain
FvH4_5g12600.1	IPR029058	Alpha/Beta hydrolase fold
FvH4_5g12610.1	IPR021419	Mediator complex, subunit Med25, von Willebrand factor type A
FvH4_5g12610.1	IPR036465	von Willebrand factor A-like domain superfamily
FvH4_5g12620.1	IPR007034	Ribosome biogenesis protein BMS1/TSR1, C-terminal
FvH4_5g12620.1	IPR012948	AARP2CN
FvH4_5g12620.1	IPR030387	Bms1/Tsr1-type G domain
FvH4_5g12660.1	IPR000286	Histone deacetylase superfamily

FvH4_5g12660.1	IPR001876	Zinc finger, RanBP2-type
FvH4_5g12660.1	IPR023696	Ureohydrolase domain superfamily
FvH4_5g12660.1	IPR023801	Histone deacetylase domain
FvH4_5g12660.1	IPR037138	Histone deacetylase domain superfamily
FvH4_5g12670.1	IPR001005	SANT/Myb domain
FvH4_5g12670.1	IPR006447	Myb domain, plants
FvH4_5g12670.1	IPR009057	Homeobox-like domain superfamily
FvH4_5g12670.1	IPR017884	SANT domain
FvH4_5g12670.1	IPR017930	Myb domain
FvH4_5g12680.1	IPR024224	DENND6
FvH4_5g12680.1	IPR037516	Tripartite DENN domain
FvH4_5g12690.1	IPR001841	Zinc finger, RING-type
FvH4_5g12690.1	IPR019396	Transmembrane Fragile-X-F-associated protein
FvH4_5g12700.1	IPR000432	DNA mismatch repair protein MutS, C-terminal
FvH4_5g12700.1	IPR007695	DNA mismatch repair protein MutS-like, N-terminal
FvH4_5g12700.1	IPR007696	DNA mismatch repair protein MutS, core
FvH4_5g12700.1	IPR007860	DNA mismatch repair protein MutS, connector domain
FvH4_5g12700.1	IPR007861	DNA mismatch repair protein MutS, clamp
FvH4_5g12700.1	IPR011184	DNA mismatch repair Msh2-type
FvH4_5g12700.1	IPR016151	DNA mismatch repair protein MutS, N-terminal
FvH4_5g12700.1	IPR027417	P-loop containing nucleoside triphosphate hydrolase
FvH4_5g12700.1	IPR032642	DNA mismatch repair protein Msh2
FvH4_5g12700.1	IPR036187	DNA mismatch repair protein MutS, core domain superfamily
FvH4_5g12700.1	IPR036678	MutS, connector domain superfamily
FvH4_5g12710.1	IPR005333	Transcription factor, TCP
FvH4_5g12710.1	IPR015943	WD40/YVTN repeat-like-containing domain superfamily
FvH4_5g12710.1	IPR017887	Transcription factor TCP subgroup
FvH4_5g12710.1	IPR017888	CYC/TB1, R domain
FvH4_5g12730.1	IPR011989	Armadillo-like helical
FvH4_5g12730.1	IPR014782	Peptidase M1, membrane alanine aminopeptidase, N-terminal
FvH4_5g12730.1	IPR016024	Armadillo-type fold
FvH4_5g12740.1	IPR002853	Transcription initiation factor IIE subunit alpha, N-terminal
FvH4_5g12740.1	IPR017919	Transcription factor TFE/TFIIEalpha HTH domain
FvH4_5g12740.1	IPR024550	TFIIEalpha/SarR/Rpc3 HTH domain
FvH4_5g12740.1	IPR036388	Winged helix-like DNA-binding domain superfamily
FvH4_5g12740.1	IPR036390	Winged helix DNA-binding domain superfamily
FvH4_5g12750.1	IPR002853	Transcription initiation factor IIE subunit alpha, N-terminal
FvH4_5g12750.1	IPR017919	Transcription factor TFE/TFIIEalpha HTH domain
FvH4_5g12750.1	IPR024550	TFIIEalpha/SarR/Rpc3 HTH domain
FvH4_5g12750.1	IPR029037	YfgJ-like
FvH4_5g12750.1	IPR036388	Winged helix-like DNA-binding domain superfamily
FvH4_5g12750.1	IPR036390	Winged helix DNA-binding domain superfamily
FvH4_5g12770.1	IPR000136	Oleosin
FvH4_5g12780.1	IPR013057	Amino acid transporter, transmembrane domain
FvH4_5g12790.1	IPR000031	PurE domain
FvH4_5g12790.1	IPR003135	ATP-grasp fold, ATP-dependent carboxylate-amine ligase-type
FvH4_5g12790.1	IPR005875	Phosphoribosylaminoimidazole carboxylase, ATPase subunit
FvH4_5g12790.1	IPR011054	Rudiment single hybrid motif
FvH4_5g12790.1	IPR011761	ATP-grasp fold
FvH4_5g12790.1	IPR013815	ATP-grasp fold, subdomain 1
FvH4_5g12790.1	IPR016185	Pre-ATP-grasp domain superfamily
FvH4_5g12790.1	IPR016301	Phosphoribosylaminoimidazole carboxylase, fungi/plant
FvH4_5g12790.1	IPR033747	Class I PurE
FvH4_5g12790.1	IPR035893	PurE domain superfamily
FvH4_5g12800.1	IPR000649	Initiation factor 2B-related
FvH4_5g12800.1	IPR037171	NagB/RpiA transferase-like
FvH4_5g12810.1	IPR006016	UspA
FvH4_5g12810.1	IPR014729	Rossmann-like alpha/beta/alpha sandwich fold
FvH4_5g12820.1	IPR000424	Primosome PriB/single-strand DNA-binding

FvH4_5g12820.1	IPR012340	Nucleic acid-binding, OB-fold
FvH4_5g12830.1	IPR012935	Zinc finger, C3HC-like
FvH4_5g12840.1	IPR002130	Cyclophilin-type peptidyl-prolyl cis-trans isomerase domain
FvH4_5g12840.1	IPR024936	Cyclophilin-type peptidyl-prolyl cis-trans isomerase
FvH4_5g12840.1	IPR029000	Cyclophilin-like domain superfamily
FvH4_5g12850.1	IPR032675	Leucine-rich repeat domain superfamily
FvH4_5g12860.1	IPR003245	Phytocyanin domain
FvH4_5g12860.1	IPR008972	Cupredoxin
FvH4_5g12870.1	IPR000629	ATP-dependent RNA helicase DEAD-box, conserved site
FvH4_5g12870.1	IPR001650	Helicase, C-terminal
FvH4_5g12870.1	IPR011545	DEAD/DEAH box helicase domain
FvH4_5g12870.1	IPR014001	Helicase superfamily 1/2, ATP-binding domain
FvH4_5g12870.1	IPR014014	RNA helicase, DEAD-box type, Q motif
FvH4_5g12870.1	IPR025313	Domain of unknown function DUF4217
FvH4_5g12870.1	IPR027417	P-loop containing nucleoside triphosphate hydrolase
FvH4_5g12880.1	IPR001752	Kinesin motor domain
FvH4_5g12880.1	IPR019821	Kinesin motor domain, conserved site
FvH4_5g12880.1	IPR027417	P-loop containing nucleoside triphosphate hydrolase
FvH4_5g12880.1	IPR027640	Kinesin-like protein
FvH4_5g12880.1	IPR036961	Kinesin motor domain superfamily
FvH4_5g12900.1	IPR003958	Transcription factor CBF/NF-Y/archaeal histone domain
FvH4_5g12900.1	IPR006460	Protein MIZU-KUSSEI 1-like, plant
FvH4_5g12900.1	IPR009072	Histone-fold
FvH4_5g12920.1	IPR000061	SWAP/Surp
FvH4_5g12920.1	IPR001757	P-type ATPase
FvH4_5g12920.1	IPR006539	P-type ATPase, subfamily IV
FvH4_5g12920.1	IPR008250	P-type ATPase, A domain superfamily
FvH4_5g12920.1	IPR011666	G patch domain-containing protein, N-terminal
FvH4_5g12920.1	IPR018303	P-type ATPase, phosphorylation site
FvH4_5g12920.1	IPR023214	HAD superfamily
FvH4_5g12920.1	IPR023298	P-type ATPase, transmembrane domain superfamily
FvH4_5g12920.1	IPR023299	P-type ATPase, cytoplasmic domain N
FvH4_5g12920.1	IPR032630	P-type ATPase, C-terminal
FvH4_5g12920.1	IPR032631	P-type ATPase, N-terminal
FvH4_5g12920.1	IPR035967	SWAP/Surp superfamily
FvH4_5g12920.1	IPR036412	HAD-like superfamily
FvH4_5g12930.1	IPR002048	EF-hand domain
FvH4_5g12930.1	IPR011992	EF-hand domain pair
FvH4_5g12930.1	IPR018247	EF-Hand 1, calcium-binding site
FvH4_5g12940.1	IPR001471	AP2/ERF domain
FvH4_5g12940.1	IPR016177	DNA-binding domain superfamily
FvH4_5g12940.1	IPR036955	AP2/ERF domain superfamily
FvH4_5g12950.1	IPR001763	Rhodanese-like domain
FvH4_5g12950.1	IPR036873	Rhodanese-like domain superfamily
FvH4_5g12960.1	IPR000873	AMP-dependent synthetase/ligase
FvH4_5g12960.1	IPR020845	AMP-binding, conserved site
FvH4_5g12960.1	IPR021788	Protein CHAPERONE-LIKE PROTEIN OF POR1-like
FvH4_5g12970.1	IPR018499	Tetraspanin/Peripherin
FvH4_5g12980.1	IPR006597	Sel1-like repeat
FvH4_5g12990.1	IPR000941	Enolase
FvH4_5g12990.1	IPR020810	Enolase, C-terminal TIM barrel domain
FvH4_5g12990.1	IPR020811	Enolase, N-terminal
FvH4_5g12990.1	IPR029017	Enolase-like, N-terminal
FvH4_5g12990.1	IPR036849	Enolase-like, C-terminal
FvH4_5g13000.1	IPR008685	Centromere protein Mis12
FvH4_5g13010.1	IPR011051	RmlC-like cupin domain superfamily
FvH4_5g13010.1	IPR012864	Cysteine oxygenase/2-aminoethanethiol dioxygenase
FvH4_5g13010.1	IPR014710	RmlC-like jelly roll fold
FvH4_5g13020.1	IPR004531	Phenylalanyl-tRNA synthetase, class IIc, beta subunit, archae/euk cytosolic

FvH4_5g13020.1	IPR005146	B3/B4 tRNA-binding domain
FvH4_5g13020.1	IPR005147	tRNA synthetase, B5-domain
FvH4_5g13020.1	IPR009061	Putative DNA-binding domain superfamily
FvH4_5g13020.1	IPR020825	Phenylalanyl-tRNA synthetase, B3/B4
FvH4_5g13030.1	IPR002415	H/ACA ribonucleoprotein complex, subunit Nhp2, eukaryote
FvH4_5g13030.1	IPR004038	Ribosomal protein L7Ae/L30e/S12e/Gadd45
FvH4_5g13030.1	IPR018492	Ribosomal protein L7Ae/L8/Nhp2 family
FvH4_5g13030.1	IPR029064	50S ribosomal protein L30e-like
FvH4_5g13040.1	IPR006083	Phosphoribulokinase/uridine kinase
FvH4_5g13040.1	IPR023577	CYTH domain
FvH4_5g13040.1	IPR027417	P-loop containing nucleoside triphosphate hydrolase
FvH4_5g13040.1	IPR033469	CYTH-like domain superfamily
FvH4_5g13050.1	IPR007466	Peptidyl-arginine deiminase, Porphyromonas-type
FvH4_5g13050.1	IPR017754	Agmatine deiminase
FvH4_5g13060.1	IPR000719	Protein kinase domain
FvH4_5g13060.1	IPR008271	Serine/threonine-protein kinase, active site
FvH4_5g13060.1	IPR011009	Protein kinase-like domain superfamily
FvH4_5g13070.1	IPR006634	TRAM/LAG1/CLN8 homology domain
FvH4_5g13070.1	IPR016439	Sphingosine N-acyltransferase Lag1/Lac1-like
FvH4_5g13110.1	IPR002142	Peptidase S49
FvH4_5g13110.1	IPR004634	Peptidase S49, protease IV
FvH4_5g13110.1	IPR004635	Peptidase S49, SppA
FvH4_5g13110.1	IPR029045	ClpP/crotonase-like domain superfamily
FvH4_5g13120.1	IPR002142	Peptidase S49
FvH4_5g13120.1	IPR029045	ClpP/crotonase-like domain superfamily
FvH4_5g13180.1	IPR004252	Probable transposase, Ptta/En/Spm, plant
FvH4_5g13190.1	IPR021319	Protein of unknown function DUF2921
FvH4_5g13190.1	IPR036249	Thioredoxin-like superfamily
FvH4_5g13210.1	IPR001461	Aspartic peptidase A1 family
FvH4_5g13210.1	IPR001969	Aspartic peptidase, active site
FvH4_5g13210.1	IPR021109	Aspartic peptidase domain superfamily
FvH4_5g13210.1	IPR032861	Xylanase inhibitor, N-terminal
FvH4_5g13210.1	IPR033121	Peptidase family A1 domain
FvH4_5g13240.1	IPR002885	Pentatricopeptide repeat
FvH4_5g13240.1	IPR011990	Tetratricopeptide-like helical domain superfamily
FvH4_5g13240.1	IPR032867	DYW domain
FvH4_5g13260.1	IPR004827	Basic-leucine zipper domain
FvH4_5g13260.1	IPR008917	Transcription factor, Skn-1-like, DNA-binding domain superfamily
FvH4_5g13270.1	IPR001841	Zinc finger, RING-type
FvH4_5g13270.1	IPR013083	Zinc finger, RING/FYVE/PHD-type
FvH4_5g13280.1	IPR001680	WD40 repeat
FvH4_5g13280.1	IPR015943	WD40/YVTN repeat-like-containing domain superfamily
FvH4_5g13280.1	IPR017986	WD40-repeat-containing domain
FvH4_5g13280.1	IPR019775	WD40 repeat, conserved site
FvH4_5g13280.1	IPR022100	Minichromosome loss protein Mcl1, middle region
FvH4_5g13280.1	IPR036322	WD40-repeat-containing domain superfamily
FvH4_5g13300.1	IPR002213	UDP-glucuronosyl/UDP-glucosyltransferase
FvH4_5g13310.1	IPR001841	Zinc finger, RING-type
FvH4_5g13310.1	IPR001965	Zinc finger, PHD-type
FvH4_5g13310.1	IPR011011	Zinc finger, FYVE/PHD-type
FvH4_5g13310.1	IPR013083	Zinc finger, RING/FYVE/PHD-type
FvH4_5g13310.1	IPR017907	Zinc finger, RING-type, conserved site
FvH4_5g13310.1	IPR019787	Zinc finger, PHD-finger
FvH4_5g13320.1	IPR021715	Pre-mRNA splicing Prp18-interacting factor
FvH4_5g13330.1	IPR013187	F-box associated domain, type 3
FvH4_5g13330.1	IPR017451	F-box associated interaction domain
FvH4_5g13340.1	IPR000403	Phosphatidylinositol 3-/4-kinase, catalytic domain
FvH4_5g13350.1	IPR011598	Myc-type, basic helix-loop-helix (bHLH) domain
FvH4_5g13350.1	IPR036638	Helix-loop-helix DNA-binding domain superfamily

FvH4_5g13360.1	IPR019378	GDP-fucose protein O-fucosyltransferase
FvH4_5g13360.1	IPR024709	Putative O-fucosyltransferase, plant
FvH4_5g13370.1	IPR005607	BSD domain
FvH4_5g13370.1	IPR035925	BSD domain superfamily
FvH4_5g13390.1	IPR003851	Zinc finger, Dof-type
FvH4_5g13410.1	IPR000626	Ubiquitin domain
FvH4_5g13410.1	IPR001650	Helicase, C-terminal
FvH4_5g13410.1	IPR011545	DEAD/DEAH box helicase domain
FvH4_5g13410.1	IPR014001	Helicase superfamily 1/2, ATP-binding domain
FvH4_5g13410.1	IPR018973	DEAD/DEAH-box helicase, putative
FvH4_5g13410.1	IPR027417	P-loop containing nucleoside triphosphate hydrolase
FvH4_5g13410.1	IPR029071	Ubiquitin-like domain superfamily
FvH4_5g13420.1	IPR000246	Peptidase T2, asparaginase 2
FvH4_5g13420.1	IPR029055	Nucleophile aminohydrolases, N-terminal
FvH4_5g13430.1	IPR000863	Sulfotransferase domain
FvH4_5g13430.1	IPR027417	P-loop containing nucleoside triphosphate hydrolase
FvH4_5g13440.1	IPR000863	Sulfotransferase domain
FvH4_5g13440.1	IPR027417	P-loop containing nucleoside triphosphate hydrolase
FvH4_5g13450.1	IPR002495	Glycosyl transferase, family 8
FvH4_5g13450.1	IPR029044	Nucleotide-diphospho-sugar transferases
FvH4_5g13450.1	IPR029993	Plant galacturonosyltransferase GAUT
FvH4_5g13460.1	IPR001320	Ionotropic glutamate receptor
FvH4_5g13460.1	IPR001638	Solute-binding protein family 3/N-terminal domain of MltF
FvH4_5g13460.1	IPR001828	Receptor, ligand binding region
FvH4_5g13460.1	IPR017103	Ionotropic glutamate receptor, plant
FvH4_5g13460.1	IPR028082	Periplasmic binding protein-like I
FvH4_5g13470.1	IPR002067	Mitochondrial carrier protein
FvH4_5g13470.1	IPR002113	Adenine nucleotide translocator 1
FvH4_5g13470.1	IPR018108	Mitochondrial substrate/solute carrier
FvH4_5g13470.1	IPR023395	Mitochondrial carrier domain superfamily
FvH4_5g13480.1	IPR006016	UspA
FvH4_5g13480.1	IPR014729	Rossmann-like alpha/beta/alpha sandwich fold
FvH4_5g13490.1	IPR006016	UspA
FvH4_5g13490.1	IPR014729	Rossmann-like alpha/beta/alpha sandwich fold
FvH4_5g13500.1	IPR002100	Transcription factor, MADS-box
FvH4_5g13500.1	IPR002487	Transcription factor, K-box
FvH4_5g13500.1	IPR036879	Transcription factor, MADS-box superfamily
FvH4_5g13510.1	IPR002100	Transcription factor, MADS-box
FvH4_5g13510.1	IPR002487	Transcription factor, K-box
FvH4_5g13510.1	IPR036879	Transcription factor, MADS-box superfamily
FvH4_5g13520.1	IPR004333	SBP domain
FvH4_5g13520.1	IPR036893	SBP domain superfamily
FvH4_5g13530.1	IPR004901	Reversibly glycosylated polypeptide
FvH4_5g13530.1	IPR029044	Nucleotide-diphospho-sugar transferases
FvH4_5g13530.1	IPR037595	Reversibly glycosylated polypeptide family
FvH4_5g13540.1	IPR006918	COBRA, plant
FvH4_5g13550.1	IPR006918	COBRA, plant
FvH4_5g13560.1	IPR006918	COBRA, plant
FvH4_5g13570.1	IPR001680	WD40 repeat
FvH4_5g13570.1	IPR015943	WD40/YVTN repeat-like-containing domain superfamily
FvH4_5g13570.1	IPR017986	WD40-repeat-containing domain
FvH4_5g13570.1	IPR019775	WD40 repeat, conserved site
FvH4_5g13570.1	IPR020472	G-protein beta WD-40 repeat
FvH4_5g13570.1	IPR036322	WD40-repeat-containing domain superfamily
FvH4_5g13580.1	IPR002885	Pentatricopeptide repeat
FvH4_5g13590.1	IPR001128	Cytochrome P450
FvH4_5g13590.1	IPR002401	Cytochrome P450, E-class, group I
FvH4_5g13590.1	IPR017972	Cytochrome P450, conserved site
FvH4_5g13590.1	IPR036396	Cytochrome P450 superfamily

FvH4_5g13610.1	IPR022764	Peptidase S54, rhomboid domain
FvH4_5g13610.1	IPR035952	Rhomboid-like superfamily
FvH4_5g13620.1	IPR022764	Peptidase S54, rhomboid domain
FvH4_5g13620.1	IPR035952	Rhomboid-like superfamily
FvH4_5g13630.1	IPR003690	Transcription termination factor, mitochondrial/chloroplastic
FvH4_5g13640.1	IPR022764	Peptidase S54, rhomboid domain
FvH4_5g13640.1	IPR035952	Rhomboid-like superfamily
FvH4_5g13650.1	IPR011043	Galactose oxidase/kelch, beta-propeller
FvH4_5g13650.1	IPR015915	Kelch-type beta propeller
FvH4_5g13670.1	IPR000571	Zinc finger, CCCH-type
FvH4_5g13670.1	IPR004274	FCP1 homology domain
FvH4_5g13670.1	IPR023214	HAD superfamily
FvH4_5g13670.1	IPR036412	HAD-like superfamily
FvH4_5g13670.1	IPR036855	Zinc finger, CCCH-type superfamily
FvH4_5g13690.1	IPR025486	Domain of unknown function DUF4378
FvH4_5g13690.1	IPR032795	DUF3741-associated sequence motif
FvH4_5g13690.1	IPR033334	Protein LONGIFOLIA 1/2
FvH4_5g13710.1	IPR005333	Transcription factor, TCP
FvH4_5g13710.1	IPR017887	Transcription factor TCP subgroup
FvH4_5g13730.1	IPR000504	RNA recognition motif domain
FvH4_5g13730.1	IPR002075	Nuclear transport factor 2
FvH4_5g13730.1	IPR018222	Nuclear transport factor 2, eukaryote
FvH4_5g13730.1	IPR032710	NTF2-like domain superfamily
FvH4_5g13730.1	IPR035979	RNA-binding domain superfamily
FvH4_5g13740.1	IPR009500	Protein of unknown function DUF1118
FvH4_5g13760.1	IPR031307	Ninja family
FvH4_5g13760.1	IPR032308	Jas TPL-binding domain
FvH4_5g13770.1	IPR017946	PLC-like phosphodiesterase, TIM beta/alpha-barrel domain superfamily
FvH4_5g13770.1	IPR030395	Glycerophosphodiester phosphodiesterase domain
FvH4_5g13790.1	IPR004140	Exocyst complex component Exo70
FvH4_5g13790.1	IPR016159	Cullin repeat-like-containing domain superfamily
FvH4_5g13800.1	IPR001108	Peptidase A22A, presenilin
FvH4_5g13800.1	IPR006639	Presenilin/signal peptide peptidase
FvH4_5g13810.1	IPR000504	RNA recognition motif domain
FvH4_5g13810.1	IPR035979	RNA-binding domain superfamily
FvH4_5g13820.1	IPR000504	RNA recognition motif domain
FvH4_5g13820.1	IPR035979	RNA-binding domain superfamily
FvH4_5g13840.1	IPR023190	Phosphoserine phosphatase, domain 2
FvH4_5g13840.1	IPR023214	HAD superfamily
FvH4_5g13840.1	IPR036412	HAD-like superfamily
FvH4_5g13850.1	IPR000778	Cytochrome b245, heavy chain
FvH4_5g13850.1	IPR011992	EF-hand domain pair
FvH4_5g13850.1	IPR013112	FAD-binding 8
FvH4_5g13850.1	IPR013121	Ferric reductase, NAD binding domain
FvH4_5g13850.1	IPR013130	Ferric reductase transmembrane component-like domain
FvH4_5g13850.1	IPR013623	NADPH oxidase Respiratory burst
FvH4_5g13850.1	IPR017927	Ferredoxin reductase-type FAD-binding domain
FvH4_5g13850.1	IPR017938	Riboflavin synthase-like beta-barrel
FvH4_5g13850.1	IPR018247	EF-Hand 1, calcium-binding site
FvH4_5g13860.1	IPR000131	ATP synthase, F1 complex, gamma subunit
FvH4_5g13860.1	IPR023632	ATP synthase, F1 complex, gamma subunit conserved site
FvH4_5g13860.1	IPR035968	ATP synthase, F1 complex, gamma subunit superfamily
FvH4_5g13870.1	IPR002067	Mitochondrial carrier protein
FvH4_5g13870.1	IPR018108	Mitochondrial substrate/solute carrier
FvH4_5g13870.1	IPR023395	Mitochondrial carrier domain superfamily
FvH4_5g13880.1	IPR001623	DnaJ domain
FvH4_5g13880.1	IPR013087	Zinc finger C2H2-type
FvH4_5g13880.1	IPR022755	Zinc finger, double-stranded RNA binding
FvH4_5g13880.1	IPR036236	Zinc finger C2H2 superfamily

FvH4_5g13880.1	IPR036869	DnaJ domain superfamily
FvH4_5g13890.1	IPR000719	Protein kinase domain
FvH4_5g13890.1	IPR000985	Legume lectin, alpha chain, conserved site
FvH4_5g13890.1	IPR001220	Legume lectin domain
FvH4_5g13890.1	IPR008271	Serine/threonine-protein kinase, active site
FvH4_5g13890.1	IPR011009	Protein kinase-like domain superfamily
FvH4_5g13890.1	IPR013320	Concanavalin A-like lectin/glucanase domain superfamily
FvH4_5g13890.1	IPR017441	Protein kinase, ATP binding site
FvH4_5g13910.1	IPR000286	Histone deacetylase superfamily
FvH4_5g13910.1	IPR023696	Ureohydrolase domain superfamily
FvH4_5g13910.1	IPR023801	Histone deacetylase domain
FvH4_5g13910.1	IPR037138	Histone deacetylase domain superfamily
FvH4_5g13920.1	IPR001623	DnaJ domain
FvH4_5g13920.1	IPR003604	Matrin/U1-C-like, C2H2-type zinc finger
FvH4_5g13920.1	IPR013087	Zinc finger C2H2-type
FvH4_5g13920.1	IPR018253	DnaJ domain, conserved site
FvH4_5g13920.1	IPR022755	Zinc finger, double-stranded RNA binding
FvH4_5g13920.1	IPR036236	Zinc finger C2H2 superfamily
FvH4_5g13920.1	IPR036869	DnaJ domain superfamily
FvH4_5g13930.1	IPR011011	Zinc finger, FYVE/PHD-type
FvH4_5g13930.1	IPR013083	Zinc finger, RING/FYVE/PHD-type
FvH4_5g13940.1	IPR011011	Zinc finger, FYVE/PHD-type
FvH4_5g13940.1	IPR013083	Zinc finger, RING/FYVE/PHD-type
FvH4_5g13960.1	IPR012946	X8 domain
FvH4_5g13990.1	IPR001650	Helicase, C-terminal
FvH4_5g13990.1	IPR003593	AAA+ ATPase domain
FvH4_5g13990.1	IPR004179	Sec63 domain
FvH4_5g13990.1	IPR011545	DEAD/DEAH box helicase domain
FvH4_5g13990.1	IPR014001	Helicase superfamily 1/2, ATP-binding domain
FvH4_5g13990.1	IPR014756	Immunoglobulin E-set
FvH4_5g13990.1	IPR027417	P-loop containing nucleoside triphosphate hydrolase
FvH4_5g13990.1	IPR035892	C2 domain superfamily
FvH4_5g13990.1	IPR036390	Winged helix DNA-binding domain superfamily
FvH4_5g14000.1	IPR018781	Transmembrane protein adipocyte-associated 1
FvH4_5g14010.1	IPR001128	Cytochrome P450
FvH4_5g14010.1	IPR002401	Cytochrome P450, E-class, group I
FvH4_5g14010.1	IPR017972	Cytochrome P450, conserved site
FvH4_5g14010.1	IPR036396	Cytochrome P450 superfamily
FvH4_5g14020.1	IPR001270	ClpA/B family
FvH4_5g14020.1	IPR003593	AAA+ ATPase domain
FvH4_5g14020.1	IPR003959	ATPase, AAA-type, core
FvH4_5g14020.1	IPR004176	Clp, N-terminal
FvH4_5g14020.1	IPR018368	ClpA/B, conserved site 1
FvH4_5g14020.1	IPR019489	Clp ATPase, C-terminal
FvH4_5g14020.1	IPR027417	P-loop containing nucleoside triphosphate hydrolase
FvH4_5g14020.1	IPR028299	ClpA/B, conserved site 2
FvH4_5g14020.1	IPR036628	Clp, N-terminal domain superfamily
FvH4_5g14030.1	IPR007149	Leo1-like protein
FvH4_5g14060.1	IPR002213	UDP-glucuronosyl/UDP-glucosyltransferase
FvH4_5g14080.1	IPR003441	NAC domain
FvH4_5g14080.1	IPR036093	NAC domain superfamily
FvH4_5g14110.1	IPR029055	Nucleophile aminohydrolases, N-terminal
FvH4_5g14140.1	IPR000073	Alpha/beta hydrolase fold-1
FvH4_5g14140.1	IPR000639	Epoxide hydrolase-like
FvH4_5g14140.1	IPR029058	Alpha/Beta hydrolase fold
FvH4_5g14160.1	IPR002213	UDP-glucuronosyl/UDP-glucosyltransferase
FvH4_5g14170.1	IPR003441	NAC domain
FvH4_5g14170.1	IPR036093	NAC domain superfamily
FvH4_5g14180.1	IPR000719	Protein kinase domain

FvH4_5g14180.1	IPR008271	Serine/threonine-protein kinase, active site
FvH4_5g14180.1	IPR011009	Protein kinase-like domain superfamily
FvH4_5g14200.1	IPR001266	Ribosomal protein S19e
FvH4_5g14200.1	IPR018277	Ribosomal protein S19e, conserved site
FvH4_5g14200.1	IPR036390	Winged helix DNA-binding domain superfamily
FvH4_5g14210.1	IPR000756	Diacylglycerol kinase, accessory domain
FvH4_5g14210.1	IPR001206	Diacylglycerol kinase, catalytic domain
FvH4_5g14210.1	IPR002219	Protein kinase C-like, phorbol ester/diacylglycerol-binding domain
FvH4_5g14210.1	IPR016064	NAD kinase/diacylglycerol kinase-like domain superfamily
FvH4_5g14210.1	IPR017438	Inorganic polyphosphate/ATP-NAD kinase, N-terminal
FvH4_5g14230.1	IPR001732	UDP-glucose/GDP-mannose dehydrogenase, N-terminal
FvH4_5g14230.1	IPR008927	6-phosphogluconate dehydrogenase-like, C-terminal domain superfamily
FvH4_5g14230.1	IPR014026	UDP-glucose/GDP-mannose dehydrogenase, dimerisation
FvH4_5g14230.1	IPR014027	UDP-glucose/GDP-mannose dehydrogenase, C-terminal
FvH4_5g14230.1	IPR017476	UDP-glucose/GDP-mannose dehydrogenase
FvH4_5g14230.1	IPR021157	Cytochrome c1, transmembrane anchor, C-terminal
FvH4_5g14230.1	IPR028356	UDP-glucose 6-dehydrogenase, eukaryotic type
FvH4_5g14230.1	IPR036220	UDP-glucose/GDP-mannose dehydrogenase, C-terminal domain superfamily
FvH4_5g14230.1	IPR036291	NAD(P)-binding domain superfamily
FvH4_5g14240.1	IPR001650	Helicase, C-terminal
FvH4_5g14240.1	IPR002464	DNA/RNA helicase, ATP-dependent, DEAH-box type, conserved site
FvH4_5g14240.1	IPR007502	Helicase-associated domain
FvH4_5g14240.1	IPR011709	Domain of unknown function DUF1605
FvH4_5g14240.1	IPR014001	Helicase superfamily 1/2, ATP-binding domain
FvH4_5g14240.1	IPR027417	P-loop containing nucleoside triphosphate hydrolase
FvH4_5g14250.1	IPR001611	Leucine-rich repeat
FvH4_5g14250.1	IPR003591	Leucine-rich repeat, typical subtype
FvH4_5g14250.1	IPR032675	Leucine-rich repeat domain superfamily
FvH4_5g14270.1	IPR013087	Zinc finger C2H2-type
FvH4_5g14270.1	IPR036236	Zinc finger C2H2 superfamily
FvH4_5g14280.1	IPR006566	FBD domain
FvH4_5g14290.1	IPR003340	B3 DNA binding domain
FvH4_5g14290.1	IPR015300	DNA-binding pseudobarrel domain superfamily
FvH4_5g14300.1	IPR008979	Galactose-binding-like domain superfamily
FvH4_5g14300.1	IPR010325	Rhamnogalacturonate lyase
FvH4_5g14300.1	IPR011013	Galactose mutarotase-like domain superfamily
FvH4_5g14300.1	IPR013784	Carbohydrate-binding-like fold
FvH4_5g14300.1	IPR014718	Glycoside hydrolase-type carbohydrate-binding
FvH4_5g14300.1	IPR029411	Rhamnogalacturonan lyase, domain III
FvH4_5g14300.1	IPR029413	Rhamnogalacturonan lyase, domain II
FvH4_5g14310.1	IPR000727	Target SNARE coiled-coil homology domain
FvH4_5g14320.1	IPR005199	Glycoside hydrolase, family 79
FvH4_5g14320.1	IPR017853	Glycoside hydrolase superfamily
FvH4_5g14340.1	IPR000330	SNF2-related, N-terminal domain
FvH4_5g14340.1	IPR001650	Helicase, C-terminal
FvH4_5g14340.1	IPR002711	HNH endonuclease
FvH4_5g14340.1	IPR014001	Helicase superfamily 1/2, ATP-binding domain
FvH4_5g14340.1	IPR027417	P-loop containing nucleoside triphosphate hydrolase
FvH4_5g14340.1	IPR032446	S phase cyclin A-associated protein in the endoplasmic reticulum, N-terminal
FvH4_5g14350.1	IPR005516	Remorin, C-terminal
FvH4_5g14360.1	IPR011598	Myc-type, basic helix-loop-helix (bHLH) domain
FvH4_5g14360.1	IPR036638	Helix-loop-helix DNA-binding domain superfamily
FvH4_5g14370.1	IPR012417	Calmodulin-binding domain, plant
FvH4_5g14380.1	IPR008581	Protein of unknown function DUF863, plant
FvH4_5g14390.1	IPR000960	Flavin monooxygenase FMO
FvH4_5g14390.1	IPR020946	Flavin monooxygenase-like
FvH4_5g14390.1	IPR036188	FAD/NAD(P)-binding domain superfamily
FvH4_5g14400.1	IPR009039	EAR
FvH4_5g14400.1	IPR025984	dCTP pyrophosphatase 1

FvH4_5g14410.1	IPR000719	Protein kinase domain
FvH4_5g14410.1	IPR001611	Leucine-rich repeat
FvH4_5g14410.1	IPR003591	Leucine-rich repeat, typical subtype
FvH4_5g14410.1	IPR008271	Serine/threonine-protein kinase, active site
FvH4_5g14410.1	IPR011009	Protein kinase-like domain superfamily
FvH4_5g14410.1	IPR013210	Leucine-rich repeat-containing N-terminal, plant-type
FvH4_5g14410.1	IPR017441	Protein kinase, ATP binding site
FvH4_5g14410.1	IPR032675	Leucine-rich repeat domain superfamily
FvH4_5g14420.1	IPR001841	Zinc finger, RING-type
FvH4_5g14420.1	IPR013083	Zinc finger, RING/FYVE/PHD-type
FvH4_5g14430.1	IPR008545	WEB family
FvH4_5g14440.1	IPR007493	Protein of unknown function DUF538
FvH4_5g14440.1	IPR036758	At5g01610-like superfamily
FvH4_5g14450.1	IPR027075	Cleavage and polyadenylation specificity factor subunit 2
FvH4_5g14470.1	IPR000719	Protein kinase domain
FvH4_5g14470.1	IPR008271	Serine/threonine-protein kinase, active site
FvH4_5g14470.1	IPR011009	Protein kinase-like domain superfamily
FvH4_5g14470.1	IPR017441	Protein kinase, ATP binding site
FvH4_5g14490.1	IPR014020	Tensin phosphatase, C2 domain
FvH4_5g14490.1	IPR015425	Formin, FH2 domain
FvH4_5g14490.1	IPR029021	Protein-tyrosine phosphatase-like
FvH4_5g14490.1	IPR029023	Tensin-type phosphatase domain
FvH4_5g14490.1	IPR035892	C2 domain superfamily
FvH4_5g14510.1	IPR011990	Tetratricopeptide-like helical domain superfamily
FvH4_5g14510.1	IPR015947	PUA-like superfamily
FvH4_5g14520.1	IPR000719	Protein kinase domain
FvH4_5g14520.1	IPR001245	Serine-threonine/tyrosine-protein kinase, catalytic domain
FvH4_5g14520.1	IPR008271	Serine/threonine-protein kinase, active site
FvH4_5g14520.1	IPR011009	Protein kinase-like domain superfamily
FvH4_5g14520.1	IPR017441	Protein kinase, ATP binding site
FvH4_5g14520.1	IPR024788	Malectin-like carbohydrate-binding domain
FvH4_5g14540.1	IPR008630	Glycosyltransferase 34
FvH4_5g14550.1	IPR007705	Vesicle transport v-SNARE, N-terminal
FvH4_5g14550.1	IPR010989	SNARE
FvH4_5g14550.1	IPR027027	GOSR2/Membrin/Bos1
FvH4_5g14560.1	IPR001789	Signal transduction response regulator, receiver domain
FvH4_5g14560.1	IPR010402	CCT domain
FvH4_5g14560.1	IPR011006	CheY-like superfamily
FvH4_5g14580.1	IPR001789	Signal transduction response regulator, receiver domain
FvH4_5g14580.1	IPR011006	CheY-like superfamily
FvH4_5g14590.1	IPR012337	Ribonuclease H-like superfamily
FvH4_5g14590.1	IPR013520	Exonuclease, RNase T/DNA polymerase III
FvH4_5g14590.1	IPR036397	Ribonuclease H superfamily
FvH4_5g14600.1	IPR002885	Pentatricopeptide repeat
FvH4_5g14600.1	IPR011990	Tetratricopeptide-like helical domain superfamily
FvH4_5g14610.1	IPR002885	Pentatricopeptide repeat
FvH4_5g14610.1	IPR011990	Tetratricopeptide-like helical domain superfamily
FvH4_5g14620.1	IPR001841	Zinc finger, RING-type
FvH4_5g14620.1	IPR013083	Zinc finger, RING/FYVE/PHD-type
FvH4_5g14630.1	IPR000056	Ribulose-phosphate 3-epimerase-like
FvH4_5g14630.1	IPR011060	Ribulose-phosphate binding barrel
FvH4_5g14630.1	IPR013785	Aldolase-type TIM barrel
FvH4_5g14640.1	IPR004938	Xyloglucan fucosyltransferase
FvH4_5g14650.1	IPR001623	DnaJ domain
FvH4_5g14650.1	IPR018253	DnaJ domain, conserved site
FvH4_5g14650.1	IPR036249	Thioredoxin-like superfamily
FvH4_5g14650.1	IPR036869	DnaJ domain superfamily
FvH4_5g14660.1	IPR001005	SANT/Myb domain
FvH4_5g14660.1	IPR009057	Homeobox-like domain superfamily

FvH4_5g14660.1	IPR017930	Myb domain
FvH4_5g14670.1	IPR003441	NAC domain
FvH4_5g14670.1	IPR036093	NAC domain superfamily
FvH4_5g14680.1	IPR009305	Protein of unknown function DUF962
FvH4_5g14690.1	IPR034570	Photosynthetic NDH subunit of subcomplex B 4, chloroplastic
FvH4_5g14700.1	IPR013766	Thioredoxin domain
FvH4_5g14700.1	IPR036249	Thioredoxin-like superfamily
FvH4_5g14710.1	IPR000070	Pectinesterase, catalytic
FvH4_5g14710.1	IPR006501	Pectinesterase inhibitor domain
FvH4_5g14710.1	IPR011050	Pectin lyase fold/virulence factor
FvH4_5g14710.1	IPR012334	Pectin lyase fold
FvH4_5g14710.1	IPR035513	Invertase/pectin methylesterase inhibitor domain superfamily
FvH4_5g14750.1	IPR006873	Protein of unknown function DUF620
FvH4_5g14760.1	IPR002963	Expansin
FvH4_5g14760.1	IPR007112	Expansin/pollen allergen, DPBB domain
FvH4_5g14760.1	IPR007117	Expansin, cellulose-binding-like domain
FvH4_5g14760.1	IPR007118	Expansin/Lol pI
FvH4_5g14760.1	IPR009009	RlpA-like protein, double-psi beta-barrel domain
FvH4_5g14760.1	IPR036749	Expansin, cellulose-binding-like domain superfamily
FvH4_5g14760.1	IPR036908	RlpA-like domain superfamily
FvH4_5g14770.1	IPR001344	Chlorophyll A-B binding protein, plant
FvH4_5g14770.1	IPR022796	Chlorophyll A-B binding protein
FvH4_5g14770.1	IPR023329	Chlorophyll a/b binding domain superfamily
FvH4_5g14780.1	IPR002100	Transcription factor, MADS-box
FvH4_5g14780.1	IPR036879	Transcription factor, MADS-box superfamily
FvH4_5g14790.1	IPR004143	Biotinyl protein ligase (BPL) and lipoyl protein ligase (LPL), catalytic domain
FvH4_5g14800.1	IPR001932	PPM-type phosphatase domain
FvH4_5g14800.1	IPR036457	PPM-type phosphatase domain superfamily
FvH4_5g14810.1	IPR002048	EF-hand domain
FvH4_5g14810.1	IPR011992	EF-hand domain pair
FvH4_5g14810.1	IPR018247	EF-Hand 1, calcium-binding site
FvH4_5g14820.1	IPR002711	HNH endonuclease
FvH4_5g14820.1	IPR003615	HNH nuclease
FvH4_5g14830.1	IPR003851	Zinc finger, Dof-type
FvH4_5g14850.1	IPR000242	PTP type protein phosphatase
FvH4_5g14850.1	IPR000387	Tyrosine specific protein phosphatases domain
FvH4_5g14850.1	IPR003595	Protein-tyrosine phosphatase, catalytic
FvH4_5g14850.1	IPR016130	Protein-tyrosine phosphatase, active site
FvH4_5g14850.1	IPR029021	Protein-tyrosine phosphatase-like
FvH4_5g14860.1	IPR001087	GDSL lipase/esterase
FvH4_5g14860.1	IPR036514	SGNH hydrolase superfamily
FvH4_5g14870.1	IPR002218	tRNA uridine 5-carboxymethylaminomethyl modification enzyme MnmG-related
FvH4_5g14870.1	IPR010253	Geranylgeranyl reductase, plant/prokaryotic
FvH4_5g14870.1	IPR011774	Geranylgeranyl reductase, plant/cyanobacteria
FvH4_5g14870.1	IPR011777	Geranylgeranyl reductase family
FvH4_5g14870.1	IPR036188	FAD/NAD(P)-binding domain superfamily
FvH4_5g14890.1	IPR001236	Lactate/malate dehydrogenase, N-terminal
FvH4_5g14890.1	IPR010097	Malate dehydrogenase, type 1
FvH4_5g14890.1	IPR015955	Lactate dehydrogenase/glycoside hydrolase, family 4, C-terminal
FvH4_5g14890.1	IPR022383	Lactate/malate dehydrogenase, C-terminal
FvH4_5g14890.1	IPR036291	NAD(P)-binding domain superfamily
FvH4_5g14900.1	IPR002885	Pentatricopeptide repeat
FvH4_5g14900.1	IPR011990	Tetratricopeptide-like helical domain superfamily
FvH4_5g14900.1	IPR032867	DYW domain
FvH4_5g14910.1	IPR003035	RWP-RK domain
FvH4_5g14930.1	IPR011016	Zinc finger, RING-CH-type
FvH4_5g14930.1	IPR013083	Zinc finger, RING/FYVE/PHD-type
FvH4_5g14930.1	IPR022143	Protein of unknown function DUF3675
FvH4_5g14930.1	IPR033275	E3 ubiquitin-protein ligase MARCH-like

FvH4_5g14940.1	IPR004263	Exostosin-like
FvH4_5g14960.1	IPR006456	ZF-HD homeobox protein, Cys/His-rich dimerisation domain
FvH4_5g14970.1	IPR001005	SANT/Myb domain
FvH4_5g14970.1	IPR009057	Homeobox-like domain superfamily
FvH4_5g14970.1	IPR017930	Myb domain
FvH4_5g14980.1	IPR002885	Pentatricopeptide repeat
FvH4_5g14980.1	IPR011990	Tetratricopeptide-like helical domain superfamily
FvH4_5g14980.1	IPR032867	DYW domain
FvH4_5g14990.1	IPR004045	Glutathione S-transferase, N-terminal
FvH4_5g14990.1	IPR010987	Glutathione S-transferase, C-terminal-like
FvH4_5g14990.1	IPR036249	Thioredoxin-like superfamily
FvH4_5g14990.1	IPR036282	Glutathione S-transferase, C-terminal domain superfamily
FvH4_5g15000.1	IPR004045	Glutathione S-transferase, N-terminal
FvH4_5g15000.1	IPR010987	Glutathione S-transferase, C-terminal-like
FvH4_5g15000.1	IPR036249	Thioredoxin-like superfamily
FvH4_5g15000.1	IPR036282	Glutathione S-transferase, C-terminal domain superfamily
FvH4_5g15010.1	IPR004045	Glutathione S-transferase, N-terminal
FvH4_5g15010.1	IPR010987	Glutathione S-transferase, C-terminal-like
FvH4_5g15010.1	IPR036249	Thioredoxin-like superfamily
FvH4_5g15010.1	IPR036282	Glutathione S-transferase, C-terminal domain superfamily
FvH4_5g15020.1	IPR004045	Glutathione S-transferase, N-terminal
FvH4_5g15020.1	IPR010987	Glutathione S-transferase, C-terminal-like
FvH4_5g15020.1	IPR036249	Thioredoxin-like superfamily
FvH4_5g15020.1	IPR036282	Glutathione S-transferase, C-terminal domain superfamily
FvH4_5g15030.1	IPR002885	Pentatricopeptide repeat
FvH4_5g15030.1	IPR011990	Tetratricopeptide-like helical domain superfamily
FvH4_5g15040.1	IPR002885	Pentatricopeptide repeat
FvH4_5g15040.1	IPR011990	Tetratricopeptide-like helical domain superfamily
FvH4_5g15050.1	IPR009349	Zinc finger, C2HC5-type
FvH4_5g15060.1	IPR002885	Pentatricopeptide repeat
FvH4_5g15060.1	IPR011990	Tetratricopeptide-like helical domain superfamily
FvH4_5g15070.1	IPR000157	Toll/interleukin-1 receptor homology (TIR) domain
FvH4_5g15070.1	IPR035897	Toll/interleukin-1 receptor homology (TIR) domain superfamily
FvH4_5g15080.1	IPR004159	Putative S-adenosyl-L-methionine-dependent methyltransferase
FvH4_5g15080.1	IPR029063	S-adenosyl-L-methionine-dependent methyltransferase
FvH4_5g15100.1	IPR004045	Glutathione S-transferase, N-terminal
FvH4_5g15100.1	IPR010987	Glutathione S-transferase, C-terminal-like
FvH4_5g15100.1	IPR036249	Thioredoxin-like superfamily
FvH4_5g15100.1	IPR036282	Glutathione S-transferase, C-terminal domain superfamily
FvH4_5g15110.1	IPR004045	Glutathione S-transferase, N-terminal
FvH4_5g15110.1	IPR010987	Glutathione S-transferase, C-terminal-like
FvH4_5g15110.1	IPR036249	Thioredoxin-like superfamily
FvH4_5g15110.1	IPR036282	Glutathione S-transferase, C-terminal domain superfamily
FvH4_5g15120.1	IPR004045	Glutathione S-transferase, N-terminal
FvH4_5g15120.1	IPR010987	Glutathione S-transferase, C-terminal-like
FvH4_5g15120.1	IPR036249	Thioredoxin-like superfamily
FvH4_5g15120.1	IPR036282	Glutathione S-transferase, C-terminal domain superfamily
FvH4_5g15130.1	IPR001128	Cytochrome P450
FvH4_5g15130.1	IPR002401	Cytochrome P450, E-class, group I
FvH4_5g15130.1	IPR017972	Cytochrome P450, conserved site
FvH4_5g15130.1	IPR036396	Cytochrome P450 superfamily
FvH4_5g15140.1	IPR001128	Cytochrome P450
FvH4_5g15140.1	IPR002401	Cytochrome P450, E-class, group I
FvH4_5g15140.1	IPR017972	Cytochrome P450, conserved site
FvH4_5g15140.1	IPR036396	Cytochrome P450 superfamily
FvH4_5g15150.1	IPR005333	Transcription factor, TCP
FvH4_5g15150.1	IPR017887	Transcription factor TCP subgroup
FvH4_5g15170.1	IPR011598	Myc-type, basic helix-loop-helix (bHLH) domain
FvH4_5g15170.1	IPR036638	Helix-loop-helix DNA-binding domain superfamily

FvH4_5g15180.1	IPR004345	TB2/DP1/HVA22-related protein
FvH4_5g15190.1	IPR001715	Calponin homology domain
FvH4_5g15190.1	IPR004953	EB1, C-terminal
FvH4_5g15190.1	IPR027328	Microtubule-associated protein RP/EB
FvH4_5g15190.1	IPR036133	EB1, C-terminal domain superfamily
FvH4_5g15190.1	IPR036872	Calponin-like domain superfamily
FvH4_5g15200.1	IPR001005	SANT/Myb domain
FvH4_5g15200.1	IPR009057	Homeobox-like domain superfamily
FvH4_5g15200.1	IPR017930	Myb domain
FvH4_5g15210.1	IPR006501	Pectinesterase inhibitor domain
FvH4_5g15210.1	IPR035513	Invertase/pectin methylesterase inhibitor domain superfamily
FvH4_5g15220.1	IPR036638	Helix-loop-helix DNA-binding domain superfamily
FvH4_5g15230.1	IPR007528	RINT-1/Tip20
FvH4_5g15240.1	IPR011598	Myc-type, basic helix-loop-helix (bHLH) domain
FvH4_5g15240.1	IPR036638	Helix-loop-helix DNA-binding domain superfamily
FvH4_5g15250.1	IPR000979	Phosphodiesterase MJ0936/Vps29
FvH4_5g15250.1	IPR024654	Calcineurin-like phosphoesterase domain, lpxH type
FvH4_5g15250.1	IPR028661	Vacuolar protein sorting-associated protein 29
FvH4_5g15250.1	IPR029052	Metallo-dependent phosphatase-like
FvH4_5g15260.1	IPR001841	Zinc finger, RING-type
FvH4_5g15260.1	IPR013083	Zinc finger, RING/FYVE/PHD-type
FvH4_5g15270.1	IPR003439	ABC transporter-like
FvH4_5g15270.1	IPR003593	AAA+ ATPase domain
FvH4_5g15270.1	IPR008183	Aldose 1-/Glucose-6-phosphate 1-epimerase
FvH4_5g15270.1	IPR011013	Galactose mutarotase-like domain superfamily
FvH4_5g15270.1	IPR014718	Glycoside hydrolase-type carbohydrate-binding
FvH4_5g15270.1	IPR017871	ABC transporter, conserved site
FvH4_5g15270.1	IPR026082	ABC transporter A
FvH4_5g15270.1	IPR027417	P-loop containing nucleoside triphosphate hydrolase
FvH4_5g15280.1	IPR003439	ABC transporter-like
FvH4_5g15280.1	IPR003593	AAA+ ATPase domain
FvH4_5g15280.1	IPR017871	ABC transporter, conserved site
FvH4_5g15280.1	IPR026082	ABC transporter A
FvH4_5g15280.1	IPR027417	P-loop containing nucleoside triphosphate hydrolase
FvH4_5g15290.1	IPR000048	IQ motif, EF-hand binding site
FvH4_5g15290.1	IPR025064	Domain of unknown function DUF4005
FvH4_5g15320.1	IPR005174	Domain unknown function DUF295
FvH4_5g15320.1	IPR015915	Kelch-type beta propeller
FvH4_5g15330.1	IPR000608	Ubiquitin-conjugating enzyme E2
FvH4_5g15330.1	IPR016135	Ubiquitin-conjugating enzyme/RWD-like
FvH4_5g15330.1	IPR023313	Ubiquitin-conjugating enzyme, active site
FvH4_5g15340.1	IPR003657	WRKY domain
FvH4_5g15340.1	IPR036576	WRKY domain superfamily
FvH4_5g15350.1	IPR015590	Aldehyde dehydrogenase domain
FvH4_5g15350.1	IPR016160	Aldehyde dehydrogenase, cysteine active site
FvH4_5g15350.1	IPR016161	Aldehyde/histidinol dehydrogenase
FvH4_5g15350.1	IPR016162	Aldehyde dehydrogenase, N-terminal
FvH4_5g15360.1	IPR002125	Cytidine and deoxycytidylate deaminase domain
FvH4_5g15360.1	IPR016192	APOBEC/CMP deaminase, zinc-binding
FvH4_5g15360.1	IPR016193	Cytidine deaminase-like
FvH4_5g15380.1	IPR001279	Metallo-beta-lactamase
FvH4_5g15380.1	IPR036866	Metallo-hydrolase/oxidoreductase superfamily
FvH4_5g15390.1	IPR011989	Armadillo-like helical
FvH4_5g15390.1	IPR013918	Nucleotide exchange factor Fes1
FvH4_5g15390.1	IPR016024	Armadillo-type fold
FvH4_5g15390.1	IPR034085	TOG domain
FvH4_5g15410.1	IPR027417	P-loop containing nucleoside triphosphate hydrolase
FvH4_5g15420.1	IPR004561	Isochorismate synthase
FvH4_5g15420.1	IPR005801	ADC synthase

FvH4_5g15420.1	IPR015890	Chorismate-utilising enzyme, C-terminal
FvH4_5g15430.1	IPR002022	Pectate lyase
FvH4_5g15430.1	IPR007524	Pectate lyase, N-terminal
FvH4_5g15430.1	IPR011050	Pectin lyase fold/virulence factor
FvH4_5g15430.1	IPR012334	Pectin lyase fold
FvH4_5g15430.1	IPR018082	AmbAllergen
FvH4_5g15440.1	IPR002068	Alpha crystallin/Hsp20 domain
FvH4_5g15440.1	IPR008978	HSP20-like chaperone
FvH4_5g15460.1	IPR000008	C2 domain
FvH4_5g15460.1	IPR013583	Phosphoribosyltransferase C-terminal
FvH4_5g15460.1	IPR035892	C2 domain superfamily
FvH4_5g15480.1	IPR027246	Eukaryotic porin/Tom40
FvH4_5g15490.1	IPR001107	Band 7 domain
FvH4_5g15490.1	IPR036013	Band 7/SPFH domain superfamily
FvH4_5g15500.1	IPR009500	Protein of unknown function DUF1118
FvH4_5g15510.1	IPR000109	Proton-dependent oligopeptide transporter family
FvH4_5g15510.1	IPR036259	MFS transporter superfamily
FvH4_5g15520.1	IPR001356	Homeobox domain
FvH4_5g15520.1	IPR003106	Leucine zipper, homeobox-associated
FvH4_5g15520.1	IPR009057	Homeobox-like domain superfamily
FvH4_5g15520.1	IPR017970	Homeobox, conserved site
FvH4_5g15530.1	IPR007065	HPP
FvH4_5g15540.1	IPR002347	Short-chain dehydrogenase/reductase SDR
FvH4_5g15540.1	IPR020904	Short-chain dehydrogenase/reductase, conserved site
FvH4_5g15540.1	IPR036291	NAD(P)-binding domain superfamily
FvH4_5g15550.1	IPR002347	Short-chain dehydrogenase/reductase SDR
FvH4_5g15550.1	IPR036291	NAD(P)-binding domain superfamily
FvH4_5g15560.1	IPR002347	Short-chain dehydrogenase/reductase SDR
FvH4_5g15560.1	IPR020904	Short-chain dehydrogenase/reductase, conserved site
FvH4_5g15560.1	IPR036291	NAD(P)-binding domain superfamily
FvH4_5g15570.1	IPR000719	Protein kinase domain
FvH4_5g15570.1	IPR008271	Serine/threonine-protein kinase, active site
FvH4_5g15570.1	IPR011009	Protein kinase-like domain superfamily
FvH4_5g15570.1	IPR013210	Leucine-rich repeat-containing N-terminal, plant-type
FvH4_5g15570.1	IPR017441	Protein kinase, ATP binding site
FvH4_5g15570.1	IPR032675	Leucine-rich repeat domain superfamily
FvH4_5g15580.1	IPR001128	Cytochrome P450
FvH4_5g15580.1	IPR002401	Cytochrome P450, E-class, group I
FvH4_5g15580.1	IPR017972	Cytochrome P450, conserved site
FvH4_5g15580.1	IPR036396	Cytochrome P450 superfamily
FvH4_5g15590.1	IPR001128	Cytochrome P450
FvH4_5g15590.1	IPR002401	Cytochrome P450, E-class, group I
FvH4_5g15590.1	IPR017972	Cytochrome P450, conserved site
FvH4_5g15590.1	IPR036396	Cytochrome P450 superfamily
FvH4_5g15600.1	IPR001128	Cytochrome P450
FvH4_5g15600.1	IPR002401	Cytochrome P450, E-class, group I
FvH4_5g15600.1	IPR017972	Cytochrome P450, conserved site
FvH4_5g15600.1	IPR036396	Cytochrome P450 superfamily
FvH4_5g15610.1	IPR000109	Proton-dependent oligopeptide transporter family
FvH4_5g15610.1	IPR036259	MFS transporter superfamily
FvH4_5g15620.1	IPR000109	Proton-dependent oligopeptide transporter family
FvH4_5g15620.1	IPR036259	MFS transporter superfamily
FvH4_5g15640.1	IPR036396	Cytochrome P450 superfamily
FvH4_5g15650.1	IPR001128	Cytochrome P450
FvH4_5g15650.1	IPR002401	Cytochrome P450, E-class, group I
FvH4_5g15650.1	IPR017972	Cytochrome P450, conserved site
FvH4_5g15650.1	IPR036396	Cytochrome P450 superfamily
FvH4_5g15660.1	IPR000109	Proton-dependent oligopeptide transporter family
FvH4_5g15660.1	IPR036259	MFS transporter superfamily

FvH4_5g15670.1	IPR000109	Proton-dependent oligopeptide transporter family
FvH4_5g15670.1	IPR036259	MFS transporter superfamily
FvH4_5g15700.1	IPR001757	P-type ATPase
FvH4_5g15700.1	IPR004014	Cation-transporting P-type ATPase, N-terminal
FvH4_5g15700.1	IPR006534	P-type ATPase, subfamily IIIA
FvH4_5g15700.1	IPR008250	P-type ATPase, A domain superfamily
FvH4_5g15700.1	IPR018303	P-type ATPase, phosphorylation site
FvH4_5g15700.1	IPR023298	P-type ATPase, transmembrane domain superfamily
FvH4_5g15700.1	IPR023299	P-type ATPase, cytoplasmic domain N
FvH4_5g15700.1	IPR036412	HAD-like superfamily
FvH4_5g15710.1	IPR001623	DnaJ domain
FvH4_5g15710.1	IPR002939	Chaperone DnaJ, C-terminal
FvH4_5g15710.1	IPR008971	HSP40/DnaJ peptide-binding
FvH4_5g15710.1	IPR018253	DnaJ domain, conserved site
FvH4_5g15710.1	IPR036869	DnaJ domain superfamily
FvH4_5g15720.1	IPR000719	Protein kinase domain
FvH4_5g15720.1	IPR002048	EF-hand domain
FvH4_5g15720.1	IPR008271	Serine/threonine-protein kinase, active site
FvH4_5g15720.1	IPR011009	Protein kinase-like domain superfamily
FvH4_5g15720.1	IPR011992	EF-hand domain pair
FvH4_5g15720.1	IPR017441	Protein kinase, ATP binding site
FvH4_5g15720.1	IPR018247	EF-Hand 1, calcium-binding site
FvH4_5g15730.1	IPR002625	Smr domain
FvH4_5g15730.1	IPR002885	Pentatricopeptide repeat
FvH4_5g15730.1	IPR036063	Smr domain superfamily
FvH4_5g15740.1	IPR001394	Peptidase C19, ubiquitin carboxyl-terminal hydrolase
FvH4_5g15740.1	IPR006865	Domain of unknown function DUF629
FvH4_5g15740.1	IPR006866	Domain of unknown function DUF627, N-terminal
FvH4_5g15740.1	IPR011990	Tetratricopeptide-like helical domain superfamily
FvH4_5g15740.1	IPR013087	Zinc finger C2H2-type
FvH4_5g15740.1	IPR028889	Ubiquitin specific protease domain
FvH4_5g15750.1	IPR012476	GLE1-like
FvH4_5g15770.1	IPR004883	Lateral organ boundaries, LOB
FvH4_5g15780.1	IPR001841	Zinc finger, RING-type
FvH4_5g15780.1	IPR008913	Zinc finger, CHY-type
FvH4_5g15780.1	IPR012312	Haemerythrin-like
FvH4_5g15780.1	IPR013083	Zinc finger, RING/FYVE/PHD-type
FvH4_5g15780.1	IPR017921	Zinc finger, CTCHY-type
FvH4_5g15780.1	IPR037274	Zinc finger, CHY-type superfamily
FvH4_5g15780.1	IPR037275	Zinc finger, CTCHY-type superfamily
FvH4_5g15790.1	IPR010658	Nodulin-like
FvH4_5g15790.1	IPR020846	Major facilitator superfamily domain
FvH4_5g15790.1	IPR036259	MFS transporter superfamily

Table S3-3. RT-qPCR primers for the fruit orange color related candidate genes

Primer name	Target gene	Direction	Fragaria vesca v4.0 hit	Primer sequence 5'-	Length
q76_5g10850_F	Transcription factor MYC/MYB N-terminal	F	FvH4_5g10850	GTATTCCTACGCCCGATGG	19
q77_5g10850_R		R		CATAGTGTTTTGCAGCTGAGG	21
q78_5g10880_F:	Cytochrome P450	F	FvH4_5g10880:	CCGACCACGACATGATAGC	19
q79_5g10880_R:		R		TTGATTGGATATCGGGATGG	20
q80_5g11300_F:	Transcription factor MYC/MYB N-terminal	F	FvH4_5g11300:	CACAGGTCCACATGCTTGG	19
q81_5g11300_R:		R		GATCAAAACGCCAGAAAACC	20
q82_5g11430_F:	SANT/Myb domain	F	FvH4_5g11430:	TTGTTCAAAGGAGGGTTTGC	20
q83_5g11430_R:		R		CTGCAACTCTTCCCACATCG	21
q84_5g11930_F:	SANT/Myb domain	F	FvH4_5g11930:	CCCCATTCTTTTGTTTCTGG	20
q85_5g11930_R:		R		TTTCCACGCTTAACATCTGG	20
q86_5g12340_F:	Cytochrome cd1-nitrite reductase-like	F	FvH4_5g12340:	TGTCAATTGGGCTTCATTCC	20
q87_5g12340_R:		R		TGTGTCCTCTCAACGTGTCC	20
q88_5g12400_F:	Cytochrome P450	F	FvH4_5g12400:	GTCAAAAGCCAAACACTTGC	20
q89_5g12400_R:		R		CTCCGCTGTCAGTTTGAGC	19
q90_5g12580_F:	Cytochrome P450	F	FvH4_5g12580:	CTGTTTTGAGTCCACCGG	18
q91_5g12580_R:		R		CTTTTATCGTCTAGGGCAAG	20
q92_5g12670_F:	SANT/Myb domain	F	FvH4_5g12670:	GGCTCATTAGTGTAAGCTCTCC	22
q93_5g12670_R:		R		CACTATGGAAGTGAAGGGATGG	22
q94_5g13590_F:	Cytochrome P450	F	FvH4_5g13590:	TGCTGCTTTGTGAGATCTGG	20
q95_5g13590_R:		R		AAACCGAAGTGGGAAAAAGC	20
q96_5g13850_F:	Cytochrome b245, heavy chain	F	FvH4_5g13850:	TGTAAAGGACCGGTTTGACC	20
q97_5g13850_R:		R		AGCTGATTAGCGAACTCTTGC	21
q98_5g14010_F:	Cytochrome P450	F	FvH4_5g14010:	GTGTGGCCAGTTTGATACCC	20
q99_5g14010_R:		R		CCAGGTTCACTACCTTTGACC	21
q100_5g14660_F:	SANT/Myb domain	F	FvH4_5g14660:	GTTTGCAGGAAGACCAGACC	20
q101_5g14660_R:		R		TTGGTAACTGTGCAGCTATGG	21
q102_5g14770_F:	Chlorophyll A-B binding protein, plant	F	FvH4_5g14770:	CCTGTGACATCTCCTTCTGC	20
q103_5g14770_R:		R		GGATCAAACCCATTGTCACC	20
q104_5g14970_F:	SANT/Myb domain	F	FvH4_5g14970:	AACACTCCCCAAGTGAATCG	20
q105_5g14970_R:		R		GAAGAAGCAAACGTCCAAGG	20
q106_5g15130_F:	Cytochrome P450	F	FvH4_5g15130:	CAAGACCCAAGACGAGAACC	20
q107_5g15130_R:		R		GGACCTTATCACTGGCTTCC	20
q108_5g15200_F:	SANT/Myb domain	F	FvH4_5g15200:	TGTGATTCTGTACGTGGTAGCC	22
q109_5g15200_R:		R		GAGTCCAAGGACCTCTGTGC	20
q110_5g15580_F:	Cytochrome P450	F	FvH4_5g15580:	GTTCAACCTCGCATCATCG	19
q111_5g15580_R:		R		AAACCTCCCAAAACCAAAGC	20

Table S3-4. *FvLhca4* gene sequence in NCBI

>XM_004299300.2 PREDICTED: Fragaria vesca subsp. vesca chlorophyll a-b binding protein P4, chloroplastic (LOC101301492), mRNA
CCACCACAACACTGAGCACTGTCATTGACTCATTGACATGATATAATACAAAGCCATGAACACAAACACCATATTGTACA
TTTTGCTTTAGCCAAACCACCTTCCCTCTACCTGAAAATTCCATTTTCTTGACCCAAATTCAGGTAGAATAGTATGGCAA
CCGTCACGACTCAGGCGTCCGCGGTCTACCGGCCACAAATCACCTCGAAATCGAGGTTCCTTACCGGCTCTTCCGGCAAG
TTAACTAGGGAAGTTGCTTTTAGGCCTGTGACATCTCCTTCTGCTAGTTCTTTTAAAGTTGAGGCCAAGAAGGGCGAGTG
GTTACCTGGCTTGCCATCACCAGGTTACCTAACTGGCAGTCTTCCTGGTGACAATGGGTTTGATCCTTTGGCACTTGCTG
AGGACCCAGAGAACCTGAAATGGTTTGTCCAAGCTGAGCTTGTGAACGGTCGTTGGGCTATGTTGGGTGTAGCTGGGATG
CTATTGCCTGAAGTCTTGACAAGCATTGGGATCATTAACGTTCCAAAATGGTACGATGCTGGGAAAGCTGAGTACTTTGC
TTCTTCATCCACCCTGTTTGTGATCGAGTTCATCTTGTTCCACTATGTCGAGATCAGACGGTGGCAAGACATCAAGAACC
CAGGAAGTGTGAACCAAGACCCCATCTTTAAGCAATACAGCTTGCCTCCAAATGAGGTAGGTTACCCTGGTGGCATCTTC
AACCCTCTCAACTTTGCACCTACTGAGGAGGCCAAGGAGAAGGAAATCGCCAACGGGAGACTGGCAATGTTGGCATTCCT
GGGATTTGTGGTTCAACACAACGTGACTGGCAAAGGACCCTTTGACAACCTAGTGCAGCATCTTTCTGACCCATGGCACA
ACACCATTGTCCAAACACTAAGGGGCTAGATCCATCAAAAGCAAGCAAGAAATGGAGAAAACCTGTTTTCATTTGTTTTG
TAACAACGATTTAATATGGCTTTTGAGTGATGAAATGGGAGTGAGCATCAATGTAGACAAGGTTTTACCAGTGTACAGGAACCAT
CATGGCAATATGGCATTATATTATTATTACCATTCTCATTACAATTCACTAA

Table S3-5. NCBI accession numbers for proteins used for constructing phylogenetic trees.

Abbreviation	Plant species	NCBI/GDR accession number
AtLHCa1	*Arabidopsis thaliana*	NP_191049.1
AtLHCa2	*Arabidopsis thaliana*	Q9SYW8
AtLHCa3	*Arabidopsis thaliana*	AAA18206.1
AtLHCa4	*Arabidopsis thaliana*	NP_190331.3
AtLHCa5	*Arabidopsis thaliana*	AAD28768.1
AtLHCa6	*Arabidopsis thaliana*	AAA57542.1
AtLHCb1	*Arabidopsis thaliana*	CAA27540.1
AtLHCb1-2	*Arabidopsis thaliana*	CAA27542.1
AtLHCb2	*Arabidopsis thaliana*	AAD28769.1
AtLHCb3	*Arabidopsis thaliana*	NP_200238.1
AtLHCb4	*Arabidopsis thaliana*	CAA50712.1
AtLHCb5	*Arabidopsis thaliana*	AAD28776.1
AtLHCb6	*Arabidopsis thaliana*	AAD28777.1
CaLHCa4	*Capsicum annuum*	PHT88346.1
CcLHCa4	*Citrus clementina*	XP_006449643.1
CcLHCa5	*Citrus clementina*	XP_006434778.1
CcLHCb2	*Citrus clementina*	XP_006447758.1
CcLHCb5	*Citrus clementina*	XP_006441835.1
CcLHCb6	*Citrus clementina*	XP_006423921.1
CsLHCa6	*Citrus sinensis*	XP_006474704.1
CuLHCa4	*Citrus unshiu*	GAY55275.1
CmLHCa4	*Cucurbita moschata*	XP_022921659.1
FvLHCa1	*Fragaria vesca*	XP_004307395.1
FvLHCa2	*Fragaria vesca*	XP_004310079.1
FvLHCa3	*Fragaria vesca*	XP_004303913.1
FvLHCa4	*Fragaria vesca*	FvH4_5g14770
FvLHCa5	*Fragaria vesca*	XP_004291112.2
FvLHCa6	*Fragaria vesca*	XP_004288997.1
FvLHCb1	*Fragaria vesca*	XP_004293579.1
FvLHCb2	*Fragaria vesca*	XP_004293460.1
FvLHCb3	*Fragaria vesca*	XP_004294489.1
FvLHCb4	*Fragaria vesca*	XP_004302636.1
FvLHCb5	*Fragaria vesca*	XP_004300672.1
FvLHCb6	*Fragaria vesca*	XP_004300703.1
GmLHCa1	*Glycine max*	NP_001235996.2
GmLHCa3	*Glycine max*	NP_001347182.1
GmLHCa4	*Glycine max*	XP_003527087.3
GmLHCa5	*Glycine max*	XP_003528272.2
OsLHCa4	*Oryza sativa Japonica Group*	XP_015649730.1
PaLHCa5	*Prunus avium*	XP_021820218.1

PaLHCb1	*Prunus avium*	XP_021811685.1
Abbreviation	Plant species	NCBI/GDR accession number
PpLHCa1	*Prunus persica*	XP_007201239.1
PpLHCa2	*Prunus persica*	XP_007222715.1
PpLHCa3	*Prunus persica*	XP_020416135.1
PpLHCa4	*Prunus persica*	XP_007209497.1
PpLHCa6	*Prunus persica*	XP_007202451.1
PpLHCb2	*Prunus persica*	XP_007215825.1
PpLHCb3	*Prunus persica*	XP_007211811.1
PpLHCb4	*Prunus persica*	XP_020421586.1
PpLHCb5	*Prunus persica*	XP_007201856.1
PbLHCa4	*Pyrus x bretschneideri*	NP_001306741.1
SlLHCa1	*Solanum lycopersicum*	XP_004239673.1
SlLHCa4	*Solanum lycopersicum*	NP_001316908.1
SlLHCb1	*Solanum lycopersicum*	XP_010317413.1

General discussion

The diploid strawberry *F. vesca* serves as an ideal model plant for cultivated strawberry as well as the *Rosaceae* family. *F. vesca* is the most widely distributed diploid *Fragaria* species naturally and considered to be one of the progenitors of the cultivated strawberry (Liston et al. 2014). In recent research, *F. vesca* subgenome has been confirmed to be the dominant genome over the other three subgenomes of cultivated strawberry, since it retained more ancestral genes and a greater number of tandemly duplicated genes, and has replaced large portions of the submissive subgenomes via homoeologous exchanges (Edger et al. 2019). And also, strawberry flavor, color, and aroma were found being largely controlled by a single dominant *F. vesca* subgenome (Edger et al. 2019). Due to these features, many genetic tools have been released in the recent years including dense genetic maps (Sargent et al. 2006; Sargent et al. 2008; Sargent et al. 2011) and IStraw90 Axiom® SNP array (Nagano et al. 2017). With the genome assembly (Shulaev et al. 2011) and updates of genome annotation of *F. vesca* (Darwish et al. 2015; Tennessen et al. 2014; Edger et al. 2018), the research on improving strawberry fruit quality traits can be largely promoted.

In order to understand the genetic control of fruit quality traits in strawberry, a diploid strawberry NIL collection has been constructed using *F. vesca* cv. Reine des Vallées as recurrent parent and *F. bucharica* as introgression donor parent (Urrutia et al. 2015). The collection contains 39 homozygous NILs covering 88.5% of the *F. bucharica* genome and was genotyped throughout all steps of development using mainly microsatellites (SSRs) markers. In this study, 10 new NILs were developed to supplement previous NIL collection. Until now, the 94.8% of genome has been covered by introgressions from *F. bucharica*, only in the middle of linkage group six (39-71 cM) was gapped since the plants harboring heterozygous introgressions covering this region had difficulties to produce homozygous offspring and did not produce any recombination in this region. Thereafter, we would try to expend the offspring population to select desired NILs.

The NIL collection, as a powerful tool for QTL analysis, has been used to map QTLs related to agronomical, nutritional and organoleptic traits (Urrutia et al. 2015; Urrutia et al. 2016; Urrutia et al. 2017). Differentially expressed genes in ripe fruits from NILs Fb5:0-35, Fb5:50-76, Fb6:84-101 and Fb7:0-10 with respect to *F. vesca* were detected and candidate genes for polyphenolic QTL and volatile QTL were highlighted (Urrutia Rosauro 2015). Here, we focused mainly on flavor and color traits in ripe strawberry fruits.

In a previous study, the QTLs located in LG1 were preliminarily mapped since only two NILs harboring introgressions in LG1 were used for analyzing the QTLs for agronomical traits, phenolic compounds and volatile compounds. For instance, the QTLs located in 26-61cM in LG1 for phenolic compounds:

flavonol, flavan-3-ol, flavanone and ellagic acid (Urrutia et al. 2016) and for key volatile compounds like butyl acetate and ethyl hexanoate (Urrutia et al. 2017). Four new NILs harboring introgressions in LG1 were acquired in this study, and they can be directly used for narrowing down the previously mapped QTLs in LG1. Now, there are eight NILs harboring introgression in LG1 available, they can provide more precisely mapped QTL and candidate genes. Phenotypic data, by GC-MS and LC-MS, is needed for these new lines covering unknown region.

Four new NILs with exotic introgressions in LG5 were added to the NIL collection also. They were produced for the purpose of narrowing down the QTLs mapped in LG5:0-35cM for polyphenol and aroma volatile compounds. Actually, at least for the QTLs for key volatile compounds in this region, we achieved this goal. The QTL for methyl-2-aminobenzoate and myrtenyl-acetate were narrowed down from 11-35cM to 20-35cM and the QTL for methyl-butanoate which mapped to11-35cM before was narrowed down to 11-20cM. The improvement of NILs in this linkage group greatly helped us in candidate gene selection and also made it easier for fine mapping. Methyl-2-aminobenzoateis a major component in the pleasant aroma of the diploid strawberry *F. vesca* (Ulrich et al. 1997) but absent in most commercial varieties, imparting a key grape note to fruit aroma. In order to enhance its presence in modern fruit varieties, it is of interest to identify the genes required for its synthesis because of phenotyping challenges. The gene *FanAAMT* (*Anthranilic Acid Methyl Transferase*) in cultivated strawberry which was matched to gene04119 (*F. vesca* genome version 1.1) in diploid strawberry was identified and it was verified to modulate the accumulation of methyl-2-aminobenzoate but the genes responsible for basal methyl-2-aminobenzoate production were undiscovered (Pillet et al. 2017). The biosynthesis of methyl-2-aminobenzoate involves an alcohol acyltransferase that catalyzes the formation of methyl anthranilate from anthraniloyl-coenzyme A (CoA) and methanol (Wang and Luca 2005). Therefore, based on the previous analysis result on differential expression between LG5:0-35cM and RV (Urrutia et al. 2017), the genes encoding putative acyl-coA hydrolases and acyltransferases that were differentially expressed in LG5 between 20 and 35cM (only 2.7Mb) could be considered good candidate genes for methyl-2-aminobenzoate accumulation.

In this study, a new QTL for fruit acidity was mapped in LG4:20-31cM. In cultivated strawberry a QTL for pH was also reported in LG4 CII (Verma et al. 2017) and HG4 (Lerceteau-Köhler et al. 2012; Zorrilla-Fontanesi et al. 2012) but precise region was unknown. Thus, our result could be a good starting point to investigate candidate genes for fruit acidity in a small region (1.5Mb) and provide research base for improvement of strawberry taste.

The genome sequence of *Fragaria vesca* was a scientific milestone in strawberry research (Shulaev et al. 2011). The sequencing and annotation provides access to the genes, their putative functions, and their genomic locations, as well as a reference that can be used to assay nucleotide variation and gene expression across individuals and conditions (Bevan and Uauy 2013). It provides us with a solid starting point to investigate the genetic elements and functions which affect the fruit quality and flavor that entices the consumer's senses. Depending on the differential expression gene analysis result described in Urrutia *et al.* 2017, 16 genes related with volatile compounds content were selected and verified using qPCR. Finally, seven genes were considered as good candidate genes for various volatile compounds. Among them, two genes are putative terpene synthase genes. Terpenoids are of great importance for the characteristic flavor and aroma for most soft fruits (Maarse 2017). The terpenoid profile of cultivated strawberry species is dominated by the monoterpene linalool and the sesquiterpene nerolidol which add flowery sweet notes to strawberry aroma, whereas the fruits of wild strawberry species give off mainly olefinic monoterpenes α-pinene and myrtenyl acetate, which are not found in the cultivated species (Aharoni et al. 2004). The gene FvH4_5g06530 was downregulated in the lines harboring introgression in LG5:4-20cM. Combined QTLs for content of different terpenoids have been reported previously (Urrutia et al. 2017), only a QTL for α-pinene was located in LG5:0-11cM, therefore, we speculated the gene FvH4_5g06530 is an α-pinene synthase gene. Similarly, the gene FvH4_5g35700 was considered as candidate for myrtenol or nerol synthase gene. In *F. vesca*, the *FvPINS* gene located in LG6 encoding the enzyme for catalyzing the biosynthesis of multiple monoterpenes (major product is α-pinene) has been cloned whereas it is nonfunctional in cultivated strawberry (Aharoni et al. 2004). Therefore, these two genes may provide a good starting point for further studies on this trait.

The gene FvH4_5g29270 was identified as a (Z)-3:(E)-2-hexenal isomerasegene (*FvHI*). (E)-2- and (Z)-3- hexenal are related to green notes and fresh green odors characteristically associated with unripe fruits that is not desired in commercial strawberry. The major QTL regulating the accumulation of (E)-2- and (Z)-3- hexenal and hexenyl acetates which were derived from the lipoxygenase pathway were mapped in the region LG5:50-76cM in *F. vesca* (Urrutia et al. 2017). In this study, we identified the gene *FvHI* encoding the key enzyme catalyzing the (Z)-3-hexenal isomerization to (E)-2-hexenal. The lower expression of *FvHI* caused the increased accumulation of (Z)-3-hexenal and decrease of (E)-2-hexenal. Our results might provide a new point for improvement of aroma compounds in cultivated strawberry. And also, because of the blend of inherent green, fresh and fruity aromas, green leaf volatiles (GLVs) are widely applied in the food and beverage industry (Fukushige and Hildebrand 2005), our research might lead to advances in application of GLVs in food and aroma industry. On the other hand, many studies mention that GLVs are often related with plant defense. GLVs can act as signals not only within the plant,

inducing a series of internal defense responses, but also between plants, priming the volatile-receiver plant to respond more effectively to subsequent attacks (Heil and Bueno 2007; Jung et al. 2009; Wang et al. 2015). Moreover, (*E*)-2-hexenal and (*Z*)-3-hexenal were able to induce the expression of defense related genes (Gomi et al. 2003; Wei and Kang 2011). It has been verified that the lower levels of (*Z*)-3-hexenal enhanced resistance to the white-backed plant hopper infestation in rice (Wang et al. 2015). Thus, to investigate the effect of GLVs to plants defense in strawberry would be interesting.

Fruit color of strawberry is an important trait concerning fruit quality, since the appearance is one of the main determinants for consumer's choice in the market. In this study, an orange fruit color QTL was mapped in LG5:35-39cM and via analyzing the mRNA accumulation level, the gene FvH4_5g14770 was considered as a candidate gene for this trait. FvH4_5g14770 encodes LHCA4 protein which is a light receptor that can bind with chlorophylls. The green color of vegetables and most unripe fruits mainly is due to the amount of chlorophylls. Dramatic Chl degradation and anthocyanin accumulation occurred in strawberry fruit at the white stage of maturity (Martinez et al. 1994). If the chlorophyll catabolism was delayed or inhibited, it would cause the fruit to stay green, as has been described in a number of species, including *Festuca*, pea, soybean, rice, pepper, sorghum and *Arabidopsis* (Bachmann et al. 1994; Cha et al. 2002; Luquez and Guiamét 2002; Armstead et al. 2007; Ren et al. 2007; Shimoda et al. 2016; Manasa and Deshpande 2017). In tomato, a green flesh mutation was reported and suggested that the LHCII polypeptides were protected from senescent degradation and persevered chlorophyll and carotenoids in the thylakoids, which caused the tomato flesh to stay green. In this study, we speculated the LHCA4 protein accumulation and delayed Chl degradation affected strawberry color. In strawberry fruits, a verified delay of Chl degradation could efficiently delay the ripening in strawberry fruit under storage condition and thus prolong the storage time and maintain the organoleptic features of the fruits.

In brief, this study presented genetic studies of strawberry quality traits, especially flavor related traits using a NIL collection. The stable QTLs for acidity and ^{o}Brix were detected and provide a good opportunity for further research to increase taste quality in fruit. The investigation of the gene functions of *FvHI* and *FvLhca4* revealed their probable relations with accumulation of (E)-2- and (Z)-3- hexenal, and chlorophyll determined fruit color in strawberry, respectively. All these results provide a fundamental basis for strawberry aroma and color breeding.

Conclusion

Conclusion

1. An improved strawberry NIL collection has been developed by generating ten new NILs that complement the previously developed *F. vesca* NIL collection. Finally, the NIL collection consisted of 49 NILs covering 94.8% of the genetic background of *F. vesca* with homozygous introgressions of *F. bucharica* and with an average bin resolution of 11.6 cM covering 2.1% of the genome.

2. Acidity as a taste quality trait of ripe strawberry fruits were measured by pH and citric acid test in three harvests of all NILs. A major and stable QTL for increasing pH and decreasing citric acid content was mapped in LG4:20-31cM. Other major QTLs decreasing pH in LG1:50-61cM and LG2:39-45cM should be related to the content of other acids.

3. Soluble Solid Content measured as ºBrix has been also related to taste, and a major and stable QTL for increasing ºBrix of ripe strawberry fruits was mapped in LG1:50-61cM in NIL collection.

4. After a GC-MS analysis comparing the recurrent parent *F. vesca* and all NILs harboring overlapping introgressions in LG5, seventeen volatile compounds were identified and quantified. Five QTL for key volatile compounds accumulation were mapped, including a major QTL for methyl 2-aminobenzoate, responsible for the "wild strawberry" aroma, now mapped in LG5:20-35 reducing the QTL size by 9 cM and a major QTL for green leaf volatile compounds in ripe strawberry fruits has been mapped in LG5:50-76.

5. According to the earlier differential expression analysis results between three NILs (Fb5:0-35, Fb5:50-76and Fb7:0-10) and their recurrent parental line RV, five genes in LG5:0-35cM, six genes in LG5:50-76cM and seven genes in LG7:0-10cM were selected as candidate genes for volatile compounds. The transcription level analysis of these candidate genes in ripe fruits show that three genes (FvH4_5g06530, FvH4_5g06590 and FvH4_5g06920) in LG5:0-35cM, three genes (FvH4_5g29270, FvH4_5g33530 and FvH4_5g35700) in LG5:50-76cM and one gene FvH4_7g00440 in LG7:0-10cM were considered as good candidate genes.

6. The gene FvH4_5g29270 encoding a 3Z-2E-enal isomerase was selected as candidate gene for green leaf volatile compounds and functionally verified via transient overexpression of the *F. vesca* allele of FvH4_5g29270 in *F. vesca* and in NILs harboring *F. bucharica* introgression in LG5:50-76.

7. Differences in fruit color of ripe strawberries were observed within the NIL collection, and a major and stable QTL for orange fruit color was mapped in LG5:35-39cM by analyzing the NIL collection, the recurrent parent *F. vesca* "Reine des Valles" and the yellow strawberry *F. vesca* "Yellow Wonder". Inside this region, 18 genes were considered as candidate genes.

8. The analysis of transcription level of 18 candidate genes selected to be related to changes in fruit color were done and the gene FvH4_5g14770 encoding light-harvesting complex chlorophyll A-B binding protein was selected as good candidate gene.

Bibliography

Aharoni A, Giri AP, Verstappen FW, Bertea CM, Sevenier R, Sun Z, Jongsma MA, Schwab W, Bouwmeester HJ (2004) Gain and loss of fruit flavor compounds produced by wild and cultivated strawberry species. The Plant Cell 16 (11):3110-3131

Akhatou I, Gonzalez-Dominguez R, Fernandez-Recamales A (2016) Investigation of the effect of genotype and agronomic conditions on metabolomic profiles of selected strawberry cultivars with different sensitivity to environmental stress. Plant physiology and biochemistry : PPB 101:14-22.

Allan AC, Hellens RP, Laing WA (2008) MYB transcription factors that colour our fruit. Trends Plant Sci 13 (3):99-102.

Armstead I, Donnison I, Aubry S, Harper J, Hörtensteiner S, James C, Mani J, Moffet M, Ougham H, Roberts L (2007) Cross-species identification of Mendel's I locus. science 315 (5808):73-73

Bachmann A, FERNÁNDEZ‐LÓPEZ J, Ginsburg S, Thomas H, Bouwkamp JC, Solomos T, Matile P (1994) Stay‐green genotypes of Phaseolus vulgaris L.: chloroplast proteins and chlorophyll catabolites during foliar senescence. New Phytologist 126 (4):593-600

Bassil N, Gunn M, Folta K, Lewers K (2006) Microsatellite markers for Fragaria from 'Strawberry Festival'expressed sequence tags. Molecular Ecology Notes 6 (2):473-476

Bassil NV, Davis TM, Zhang H, Ficklin S, Mittmann M, Webster T, Mahoney L, Wood D, Alperin ES, Rosyara UR (2015) Development and preliminary evaluation of a 90 K Axiom® SNP array for the allo-octoploid cultivated strawberry Fragaria× ananassa. BMC genomics 16 (1):155

Besseau S, Hoffmann L, Geoffroy P, Lapierre C, Pollet B, Legrand M (2007) Flavonoid accumulation in Arabidopsis repressed in lignin synthesis affects auxin transport and plant growth. The Plant Cell 19 (1):148-162

Bevan MW, Uauy C (2013) Genomics reveals new landscapes for crop improvement. Genome biology 14 (6):206

Bogs J, Ebadi A, McDavid D, Robinson SP (2006) Identification of the flavonoid hydroxylases from grapevine and their regulation during fruit development. Plant physiology 140 (1):279-291.

Botha AM, van Eck L, Burger NF, Swanevelder ZH (2014) Near-isogenic lines of Triticum aestivum with distinct modes of resistance exhibit dissimilar transcriptional regulation during Diuraphis noxia feeding. Biology open 3 (11):1116-1126.

Buti M, Sargent DJ, Mhelembe KG, Delfino P, Tobutt KR, Velasco R (2016) Genotyping-by-sequencing in an orphan plant species Physocarpus opulifolius helps identify the evolutionary origins of the genus Prunus. BMC research notes 9:268.

Cha K-W, Lee Y-J, Koh H-J, Lee B-M, Nam Y-W, Paek N-C (2002) Isolation, characterization, and mapping of the stay green mutant in rice. Theoretical and Applied Genetics 104 (4):526-532

Chai YM, Jia HF, Li CL, Dong QH, Shen YY (2011) FaPYR1 is involved in strawberry fruit ripening. J Exp Bot 62 (14):5079-5089.

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

Cheng J, Wen S, Xiao S, Lu B, Ma M, Bie Z (2018a) Overexpression of the tonoplast sugar transporter CmTST2 in melon fruit increases sugar accumulation. J Exp Bot 69 (3):511-523.

Cheng MN, Huang ZJ, Hua QZ, Shan W, Kuang JF, Lu WJ, Qin YH, Chen JY (2017) The WRKY transcription factor HpWRKY44 regulates CytP450-like1 expression in red pitaya fruit (Hylocereus polyrhizus). Horticulture Research 4:17039.

Cheng R, Cheng Y, Lu J, Chen J, Wang Y, Zhang S, Zhang H (2018b) The gene *PbTMT4* from pear (*Pyrus bretschneideri*) mediates vacuolar sugar transport and strongly affects sugar accumulation in fruit. Physiol Plant 164 (3):307-319.

Cruz-Rus E, Sesmero R, Angel-Perez JA, Sanchez-Sevilla JF, Ulrich D, 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 : new strategies in plant improvement 37 (10):131.

Darwish O, Shahan R, Liu Z, Slovin JP, Alkharouf NW (2015) Re-annotation of the woodland strawberry (*Fragaria vesca*) genome. BMC genomics 16 (1):29

Davik J, Sargent DJ, Brurberg MB, Lien S, Kent M, Alsheikh M (2015) A ddRAD Based Linkage Map of the Cultivated Strawberry, Fragaria xananassa. PloS one 10 (9):e0137746.

Davis TM, Denoyes-Rothan B, Lerceteau-Köhler E (2007) Strawberry. In: Kole C (ed) Fruits and Nuts. Springer Berlin Heidelberg, Berlin, Heidelberg, pp 189-205.

Desnoues E, Gibon Y, Baldazzi V, Signoret V, Genard M, Quilot-Turion B (2014) Profiling sugar metabolism during fruit development in a peach progeny with different fructose-to-glucose ratios. BMC Plant Biol 14:336.

Ding X, Li X, Xiong L (2011) Evaluation of near-isogenic lines for drought resistance QTL and fine mapping of a locus affecting flag leaf width, spikelet number, and root volume in rice. TAG Theoretical and applied genetics Theoretische und angewandte Genetik 123 (5):815-826.

Edger PP, Poorten TJ, VanBuren R, Hardigan MA, Colle M, McKain MR, Smith RD, Teresi SJ, Nelson ADL, Wai CM, Alger EI, Bird KA, Yocca AE, Pumplin N, Ou S, Ben-Zvi G, Brodt A, Baruch K, Swale T, Shiue L, Acharya CB, Cole GS, Mower JP, Childs KL, Jiang N, Lyons E, Freeling M, Puzey JR, Knapp SJ (2019) Origin and evolution of the octoploid strawberry genome. Nature genetics 51 (3):541-547.

Edger PP, VanBuren R, Colle M, Poorten TJ, Wai CM, Niederhuth CE, Alger EI, Ou S, Acharya CB, Wang J, Callow P, McKain MR, Shi J, Collier C, Xiong Z, Mower JP, Slovin JP, Hytonen T, Jiang N, Childs KL, Knapp SJ (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.

Eduardo I, Arus P, Monforte AJ (2005) Development of a genomic library of near isogenic lines (NILs) in melon (Cucumis melo L.) from the exotic accession PI161375. TAG Theoretical and applied genetics Theoretische und angewandte Genetik 112 (1):139-148.

Etienne A, Genard M, Lobit P, Mbeguie AMD, Bugaud C (2013) What controls fleshy fruit acidity? A review of malate and citrate accumulation in fruit cells. J Exp Bot 64 (6):1451-1469.

Fukushige H, Hildebrand DF (2005) Watermelon (Citrullus lanatus) hydroperoxide lyase greatly increases C6 aldehyde formation in transgenic leaves. Journal of agricultural and food chemistry 53 (6):2046-2051

Giampieri F, Alvarez-Suarez JM, Battino M (2014) Strawberry and human health: effects beyond antioxidant activity. J Agric Food Chem 62 (18):3867-3876.

Giampieri F, Forbes-Hernandez TY, Gasparrini M, Alvarez-Suarez JM, Afrin S, Bompadre S, Quiles JL, Mezzetti B, Battino M (2015) Strawberry as a health promoter: an evidence based review. Food & function 6 (5):1386-1398.

Gomi K, Yamasaki Y, Yamamoto H, Akimitsu K (2003) Characterization of a hydroperoxide lyase gene and effect of C6-volatiles on expression of genes of the oxylipin metabolism in Citrus. Journal of plant physiology 160 (10):1219-1231

Goodstal FJ, Kohler GR, Randall LB, Bloom AJ, Clair DAS (2005) A major QTL introgressed from wild Lycopersicon hirsutum confers chilling tolerance to cultivated tomato (Lycopersiconesculentum). Theoretical and Applied Genetics 111 (5):898-905

Govan C, Simpson D, Johnson A, Tobutt K, Sargent D (2008) A reliable multiplexed microsatellite set for genotyping Fragaria and its use in a survey of 60 F.× ananassa cultivars. Molecular Breeding 22 (4):649-661

Guo J, Wang S, Yu X, Dong R, Li Y, Mei X, Shen Y (2018) Polyamines Regulate Strawberry Fruit Ripening by Abscisic Acid, Auxin, and Ethylene. Plant physiology 177 (1):339-351.

Hadonou AM, Sargent DJ, Wilson F, James CM, Simpson DW (2004) Development of microsatellite markers in Fragaria, their use in genetic diversity analysis, and their potential for genetic linkage mapping. Genome 47 (3):429-438.

Heil M, Bueno JCS (2007) Within-plant signaling by volatiles leads to induction and priming of an indirect plant defense in nature. Proceedings of the National Academy of Sciences 104 (13):5467-

5472

Hirakawa H, Shirasawa K, Kosugi S, Tashiro K, Nakayama S, Yamada M, Kohara M, Watanabe A, Kishida Y, Fujishiro T (2014) Dissection of the octoploid strawberry genome by deep sequencing of the genomes of Fragaria species. DNA research 21 (2):169-181

Hollender CA, Geretz AC, Slovin JP, Liu Z (2012) Flower and early fruit development in a diploid strawberry, Fragaria vesca. Planta 235 (6):1123-1139.

Hossain MR, Kim HT, Shanmugam A, Nath UK, Goswami G, Song JY, Park JI, Nou IS (2018) Expression Profiling of Regulatory and Biosynthetic Genes in Contrastingly Anthocyanin Rich Strawberry (*Fragaria x ananassa*) Cultivars Reveals Key Genetic Determinants of Fruit Color. International journal of molecular sciences 19 (3).

Hou BZ, Xu C, Shen YY (2018) A leu-rich repeat receptor-like protein kinase, FaRIPK1, interacts with the ABA receptor, FaABAR, to regulate fruit ripening in strawberry. J Exp Bot 69 (7):1569-1582.

Hu DG, Sun CH, Ma QJ, You CX, Cheng L, Hao YJ (2016) MdMYB1 Regulates Anthocyanin and Malate Accumulation by Directly Facilitating Their Transport into Vacuoles in Apples. Plant physiology 170 (3):1315-1330.

Hummer KE, Hancock J (2009) Strawberry genomics: botanical history, cultivation, traditional breeding, and new technologies. In: Genetics and genomics of Rosaceae. Springer, pp 413-435

Hummer KE, Nathewet P, Yanagi T (2009) Decaploidy in Fragaria iturupensis (Rosaceae). American journal of botany 96 (3):713-716.

Illa E, Sargent DJ, Girona EL, Bushakra J, Cestaro A, Crowhurst R, Pindo M, Cabrera A, Van der Knaap E, Iezzoni A (2011) Comparative analysis of rosaceous genomes and the reconstruction of a putative ancestral genome for the family. BMC Evolutionary biology 11 (1):9

Isobe SN, Hirakawa H, Sato S, Maeda F, Ishikawa M, Mori T, Yamamoto Y, Shirasawa K, Kimura M, Fukami M, Hashizume F, Tsuji T, Sasamoto S, Kato M, Nanri K, Tsuruoka H, Minami C, Takahashi C, Wada T, Ono A, Kawashima K, Nakazaki N, Kishida Y, Kohara M, Nakayama S, Yamada M, Fujishiro T, Watanabe A, 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 : an international journal for rapid publication of reports on genes and genomes 20 (1):79-92.

James C, Wilson F, Hadonou A, Tobutt K (2003) Isolation and characterization of polymorphic microsatellites in diploid strawberry (Fragaria vesca L.) for mapping, diversity studies and clone identification. Molecular Ecology Notes 3 (2):171-173

Jia D, Shen F, Wang Y, Wu T, Xu X, Zhang X, Han Z (2018) Apple fruit acidity is genetically diversified by natural variations in three hierarchical epistatic genes: MdSAUR37, MdPP2CH and

MdALMTII. The Plant journal : for cell and molecular biology 95 (3):427-443.

Jia H, Wang Y, Sun M, Li B, Han Y, Zhao Y, Li X, Ding N, Li C, Ji W, Jia W (2013) Sucrose functions as a signal involved in the regulation of strawberry fruit development and ripening. The New phytologist 198 (2):453-465.

Jin W, Wang H, Li M, Wang J, Yang Y, Zhang X, Yan G, Zhang H, Liu J, Zhang K (2016) The R2R3 MYB transcription factor PavMYB10.1 involves in anthocyanin biosynthesis and determines fruit skin colour in sweet cherry (Prunus avium L.). Plant Biotechnol J 14 (11):2120-2133.

Jo C, Kim S (2019) Transposition of a non-autonomous DNA transposon in the gene coding for a bHLH transcription factor results in a white bulb color of onions (Allium cepa L.). TAG Theoretical and applied genetics Theoretische und angewandte Genetik.

Jung HW, Tschaplinski TJ, Wang L, Glazebrook J, Greenberg JT (2009) Priming in systemic plant immunity. Science 324 (5923):89-91

Koes R, Verweij W, Quattrocchio F (2005) Flavonoids: a colorful model for the regulation and evolution of biochemical pathways. Trends Plant Sci 10 (5):236-242.

Lerceteau-Kohler E, Moing A, Guerin G, Renaud C, Petit A, Rothan C, Denoyes B (2012) Genetic dissection of fruit quality traits in the octoploid cultivated strawberry highlights the role of homoeo-QTL in their control. TAG Theoretical and applied genetics Theoretische und angewandte Genetik 124 (6):1059-1077.

Lerceteau-Köhler E, Moing A, Guérin G, Renaud C, Petit A, Rothan C, 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 (6):1059-1077

Li R, Liu HP, Hong CL, Dai ZX, Liu JW, Zhou J, Hu CQ, Weng HX (2017a) Iodide and iodate effects on the growth and fruit quality of strawberry. J Sci Food Agric 97 (1):230-235.

Li SJ, Yin XR, Wang WL, Liu XF, Zhang B, Chen KS (2017b) Citrus CitNAC62 cooperates with CitWRKY1 to participate in citric acid degradation via up-regulation of CitAco3. J Exp Bot 68 (13):3419-3426.

Li SJ, Yin XR, Xie XL, Allan AC, Ge H, Shen SL, Chen KS (2016) The Citrus transcription factor, CitERF13, regulates citric acid accumulation via a protein-protein interaction with the vacuolar proton pump, CitVHA-c4. Sci Rep 6:20151.

Li Y, Pi M, Gao Q, Liu Z, Kang C (2019) Updated annotation of the wild strawberry Fragaria vesca V4 genome. Hortic Res 6:61.

Lin-Wang K, McGhie TK, Wang M, Liu Y, Warren B, Storey R, Espley RV, Allan AC (2014) Engineering the anthocyanin regulatory complex of strawberry (Fragaria vesca). Front Plant Sci 5:651.

Lin Q, Li S, Dong W, Feng C, Yin X, Xu C, Sun C, Chen K (2015) Involvement of CitCHX and CitDIC

in developmental-related and postharvest-hot-air driven citrate degradation in citrus fruits. PloS one 10 (3):e0119410.

Liston A, Cronn R, Ashman TL (2014) Fragaria: a genus with deep historical roots and ripe for evolutionary and ecological insights. American journal of botany 101 (10):1686-1699.

Liu B, Poulsen EG, Davis TM (2016) Insight into octoploid strawberry (Fragaria) subgenome composition revealed by GISH analysis of pentaploid hybrids. Genome 59 (2):79-86.

Liu X, Hu XM, Jin LF, Shi CY, Liu YZ, Peng SA (2014) Identification and transcript analysis of two glutamate decarboxylase genes, CsGAD1 and CsGAD2, reveal the strong relationship between CsGAD1 and citrate utilization in citrus fruit. Molecular biology reports 41 (9):6253-6262.

Luquez VM, Guiamét JJ (2002) The stay green mutations d1 and d2 increase water stress susceptibility in soybeans. Journal of experimental botany 53 (373):1421-1428

Ma B, Liao L, Fang T, Peng Q, Ogutu C, Zhou H, Ma F, Han Y (2019) A Ma10 gene encoding P-type ATPase is involved in fruit organic acid accumulation in apple. Plant Biotechnol J 17 (3):674-686.

Maarse H (2017) Volatile compounds in foods and beverages. Routledge,

Maeda H, Yamaguchi T, Omoteno M, Takarada T, Fujita K, Murata K, Iyama Y, Kojima Y, Morikawa M, Ozaki H, Mukaino N, Kidani Y, Ebitani T (2014) Genetic dissection of black grain rice by the development of a near isogenic line. Breed Sci 64 (2):134-141.

Manasa K, Deshpande S (2017) Utilizing genomic resources for understanding the stay-green QTLs interactions in Sorghum.

Martinez G, Chaves A, Anon M (1994) Effect of gibberellic acid on ripening of strawberry fruits (Fragaria annanassa Duch.). Journal of Plant Growth Regulation 13 (2):87

Mikulic-Petkovsek M, Schmitzer V, Slatnar A, Stampar F, Veberic R (2012) Composition of sugars, organic acids, and total phenolics in 25 wild or cultivated berry species. Journal of food science 77 (10):C1064-1070.

Molina-Hidalgo FJ, Medina-Puche L, Canete-Gomez C, Franco-Zorrilla JM, Lopez-Vidriero I, Solano R, Caballero JL, Rodriguez-Franco A, Blanco-Portales R, Munoz-Blanco J, Moyano E (2017) The fruit-specific transcription factor FaDOF2 regulates the production of eugenol in ripe fruit receptacles. J Exp Bot 68 (16):4529-4543.

Monfort A, Vilanova S, Davis T, 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

Moreno E, Obando JM, Dos-Santos N, Fernández-Trujillo JP, Monforte AJ, Garcia-Mas J (2008) Candidate genes and QTLs for fruit ripening and softening in melon. Theoretical and Applied Genetics 116 (4):589-602

Nagano S, Shirasawa K, Hirakawa H, Maeda F, Ishikawa M, Isobe SN (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.

Oren E, Tzuri G, Vexler L, Dafna A, Meir A, Faigenboim A, Kenigswald M, Portnoy V, Schaffer AA, Levi A, Buckler ES, Katzir N, Burger J, Tadmor Y, Gur A (2019) The multi-allelic APRR2 gene is associated with fruit pigment accumulation in melon and watermelon. J Exp Bot 70 (15):3781-3794.

Palmieri L, Masuero D, Martinatti P, Baratto G, Martens S, Vrhovsek U (2017) Genotype-by-environment effect on bioactive compounds in strawberry (Fragaria x ananassa Duch.). J Sci Food Agric 97 (12):4180-4189.

Pillet J, Chambers AH, Barbey C, Bao Z, Plotto A, Bai J, Schwieterman M, Johnson T, Harrison B, Whitaker VM (2017) Identification of a methyltransferase catalyzing the final step of methyl anthranilate synthesis in cultivated strawberry. BMC plant biology 17 (1):147

Pombo MA, Martinez GA, Civello PM (2011) Cloning of FaPAL6 gene from strawberry fruit and characterization of its expression and enzymatic activity in two cultivars with different anthocyanin accumulation. Plant science : an international journal of experimental plant biology 181 (2):111-118.

Prat L, Espinoza MI, Agosin E, Silva H (2014) Identification of volatile compounds associated with the aroma of white strawberries (Fragaria chiloensis). J Sci Food Agric 94 (4):752-759.

Qiao L, Cao M, Zheng J, Zhao Y, Zheng ZL (2017) Gene coexpression network analysis of fruit transcriptomes uncovers a possible mechanistically distinct class of sugar/acid ratio-associated genes in sweet orange. BMC Plant Biol 17 (1):186.

Ren G, An K, Liao Y, Zhou X, Cao Y, Zhao H, Ge X, Kuai B (2007) Identification of a novel chloroplast protein AtNYE1 regulating chlorophyll degradation during leaf senescence in Arabidopsis. Plant physiology 144 (3):1429-1441

Rousseau-Gueutin M, Lerceteau-Köhler E, Barrot L, Sargent DJ, Monfort A, Simpson D, Arus P, Guérin G, Denoyes-Rothan B (2008) Comparative genetic mapping between octoploid and diploid Fragaria species reveals a high level of colinearity between their genomes and the essentially disomic behavior of the cultivated octoploid strawberry. Genetics 179 (4):2045-2060

Sanchez-Sevilla JF, Cruz-Rus E, Valpuesta V, Botella MA, 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.

Sargent D, Hadonou A, Simpson D (2003) Development and characterization of polymorphic microsatellite markers from Fragaria viridis, a wild diploid strawberry. Molecular Ecology Notes

3 (4):550-552

Sargent DJ, Cipriani G, Vilanova S, Gil-Ariza D, Arus P, Simpson DW, Tobutt KR, Monfort A (2008) The development of a bin mapping population and the selective mapping of 103 markers in the diploid Fragaria reference map. Genome 51 (2):120-127.

Sargent DJ, Clarke J, Simpson DW, Tobutt KR, Arus P, Monfort A, Vilanova S, Denoyes-Rothan B, Rousseau M, Folta KM, Bassil NV, Battey NH (2006) An enhanced microsatellite map of diploid Fragaria. TAG Theoretical and applied genetics Theoretische und angewandte Genetik 112 (7):1349-1359.

Sargent DJ, Davis TM, Tobutt KR, Wilkinson MJ, Battey NH, Simpson DW (2004) A genetic linkage map of microsatellite, gene-specific and morphological markers in diploid Fragaria. TAG Theoretical and applied genetics Theoretische und angewandte Genetik 109 (7):1385-1391.

Sargent DJ, Kuchta P, Girona EL, Zhang H, Davis TM, Celton J-M, Marchese A, Korbin M, Folta KM, Shulaev V (2011) Simple sequence repeat marker development and mapping targeted to previously unmapped regions of the strawberry genome sequence. The Plant Genome 4 (3):165-177

Schaart JG, Dubos C, Romero De La Fuente I, van Houwelingen AM, de Vos RC, Jonker HH, Xu W, Routaboul JM, Lepiniec L, Bovy AG (2013) Identification and characterization of MYB-bHLH-WD40 regulatory complexes controlling proanthocyanidin biosynthesis in strawberry (Fragaria x ananassa) fruits. The New phytologist 197 (2):454-467.

Sheng L, Shen D, Luo Y, Sun X, Wang J, Luo T, Zeng Y, Xu J, Deng X, Cheng Y (2017) Exogenous gamma-aminobutyric acid treatment affects citrate and amino acid accumulation to improve fruit quality and storage performance of postharvest citrus fruit. Food Chem 216:138-145. doi:10.1016/j.foodchem.2016.08.024

Shi CY, Hussain SB, Yang H, Bai YX, Khan MA, Liu YZ (2019) CsPH8, a P-type proton pump gene, plays a key role in the diversity of citric acid accumulation in citrus fruits. Plant science : an international journal of experimental plant biology 289:110288.

Shimoda Y, Ito H, Tanaka A (2016) Arabidopsis STAY-GREEN, Mendel's green cotyledon gene, encodes magnesium-dechelatase. The Plant Cell 28 (9):2147-2160

Shlizerman L, Marsh K, Blumwald E, Sadka A (2007) Iron-shortage-induced increase in citric acid content and reduction of cytosolic aconitase activity in Citrus fruit vesicles and calli. Physiol Plant 131 (1):72-79.

Shulaev V, Sargent DJ, Crowhurst RN, Mockler TC, Folkerts O, Delcher AL, Jaiswal P, Mockaitis K, Liston A, Mane SP (2011a) The genome of woodland strawberry (Fragaria vesca). Nature genetics 43 (2):109

Shulaev V, Sargent DJ, Crowhurst RN, Mockler TC, Folkerts O, Delcher AL, Jaiswal P, Mockaitis K, Liston A, Mane SP, Burns P, Davis TM, Slovin JP, Bassil N, Hellens RP, Evans C, Harkins T, Kodira C, Desany B, Crasta OR, Jensen RV, Allan AC, Michael TP, Setubal JC, Celton JM, Rees DJ, Williams KP, Holt SH, Ruiz Rojas JJ, Chatterjee M, Liu B, Silva H, Meisel L, Adato A, Filichkin SA, Troggio M, Viola R, Ashman TL, Wang H, Dharmawardhana P, Elser J, Raja R, Priest HD, Bryant DW, Jr., Fox SE, Givan SA, Wilhelm LJ, Naithani S, Christoffels A, Salama DY, Carter J, Lopez Girona E, Zdepski A, Wang W, Kerstetter RA, Schwab W, Korban SS, Davik J, Monfort A, Denoyes-Rothan B, Arus P, Mittler R, Flinn B, Aharoni A, Bennetzen JL, Salzberg SL, Dickerman AW, Velasco R, Borodovsky M, Veilleux RE, Folta KM (2011b) The genome of woodland strawberry (*Fragaria vesca*). Nature genetics 43 (2):109-116.

Slovin JP, Schmitt K, Folta KM (2009) An inbred line of the diploid strawberry Fragaria vesca f. semperflorens for genomic and molecular genetic studies in the Rosaceae. Plant methods 5 (1):15

Song C, Hong X, Zhao S, Liu J, Schulenburg K, Huang FC, Franz-Oberdorf K, Schwab W (2016) Glucosylation of 4-Hydroxy-2,5-Dimethyl-3(2H)-Furanone, the Key Strawberry Flavor Compound in Strawberry Fruit. Plant physiology 171 (1):139-151.

Sun CH, Zhang QY, Sun MH, Hu DG (2016) MdSOS2L1 forms a complex with MdMYB1 to control vacuolar pH by transcriptionally regulating MdVHA-B1 in apples. Plant Signal Behav 11 (3):e1146846.

Tanksley S, Grandillo S, Fulton T, Zamir D, Eshed Y, Petiard V, Lopez J, Beck-Bunn T (1996) Advanced backcross QTL analysis in a cross between an elite processing line of tomato and its wild relative L. pimpinellifolium. Theoretical and applied genetics 92 (2):213-224

Tennessen JA, Govindarajulu R, Ashman T-L, 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

Tohge T, Nishiyama Y, Hirai MY, Yano M, Nakajima Ji, Awazuhara M, Inoue E, Takahashi H, Goodenowe DB, Kitayama M (2005) Functional genomics by integrated analysis of metabolome and transcriptome of Arabidopsis plants over‐expressing an MYB transcription factor. The Plant Journal 42 (2):218-235

Ulrich D, Hoberg E, Rapp A, Kecke S (1997) Analysis of strawberry flavour – discrimination of aroma types by quantification of volatile compounds. Zeitschrift für Lebensmitteluntersuchung und -Forschung A 205 (3):218-223

Ulrich D, Kecke S, Olbricht K (2018) What Do We Know about the Chemistry of Strawberry Aroma? J Agric Food Chem 66 (13):3291-3301.

Urrutia M, Bonet J, Arus P, 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. TAG Theoretical and applied genetics Theoretische und angewandte Genetik 128 (7):1261-1275.

Urrutia M, Rambla JL, Alexiou KG, Granell A, Monfort A (2017) Genetic analysis of the wild strawberry (*Fragaria vesca*) volatile composition. Plant physiology and biochemistry : PPB 121:99-117.

Urrutia M, Schwab W, Hoffmann T, Monfort A (2016) Genetic dissection of the (poly)phenol profile of diploid strawberry (Fragaria vesca) fruits using a NIL collection. Plant science : an international journal of experimental plant biology 242:151-168.

Urrutia Rosauro M (2015) *Fragaria vesca* NIL collection: development and genetic characterization of agronomical, nutritional and organoleptic traits.

Vallarino JG, de Abreu ELF, Soria C, Tong H, Pott DM, Willmitzer L, Fernie AR, Nikoloski Z, Osorio S (2018) Genetic diversity of strawberry germplasm using metabolomic biomarkers. Sci Rep 8 (1):14386.

Vegas J, Garcia-Mas J, Monforte AJ (2013) Interaction between QTLs induces an advance in ethylene biosynthesis during melon fruit ripening. Theoretical and applied genetics 126 (6):1531-1544

Verma S, Zurn JD, Salinas N, Mathey MM, Denoyes B, Hancock JF, Finn CE, Bassil NV, Whitaker VM (2017) Clarifying sub-genomic positions of QTLs for flowering habit and fruit quality in U.S. strawberry (*Fragaria x ananassa*) breeding populations using pedigree-based QTL analysis. Hortic Res 4:17062.

Vladimir S, Sargent DJ, Crowhurst RN, Mockler TC, Otto F, Delcher AL, Pankaj J, Keithanne M, Aaron L, Mane SP (2011) The genome of woodland strawberry (*Fragaria vesca*). Nature genetics 43 (2):109-116

Wang B, Zhou G, Xin Z, Ji R, Lou Y (2015) (Z)-3-Hexenal, One of the Green Leaf Volatiles, Increases Susceptibility of Rice to the White-Backed Planthopper Sogatella furcifera. Plant Molecular Biology Reporter 33 (3):377-387.

Wang H, Zhang H, Yang Y, Li M, Zhang Y, Liu J, Dong J, Li J, Butelli E, Xue Z, Wang A, Wang G, Martin C, Jin W (2019) The control of red color by a family of MYB transcription factors in octoploid strawberry (*Fragaria x ananassa*) fruits. Plant Biotechnol J.

Wang J, Luca VD (2005) The biosynthesis and regulation of biosynthesis of Concord grape fruit esters, including 'foxy'methylanthranilate. The Plant Journal 44 (4):606-619

Wang L, Zhang XL, Wang L, Tian Y, Jia N, Chen S, Shi NB, Huang X, Zhou C, Yu Y, Zhang ZQ, Pang XQ (2017a) Regulation of ethylene-responsive SlWRKYs involved in color change during tomato fruit ripening. Sci Rep 7 (1):16674.

Wang Y, Zhang J, Cui W, Guan C, Mao W, Zhang Z (2017b) Improvement in Fruit Quality by Overexpressing miR399a in Woodland Strawberry. J Agric Food Chem 65 (34):7361-7370.

Weber N, Zupanc V, Jakopic J, Veberic R, Mikulic-Petkovsek M, Stampar F (2017) Influence of deficit irrigation on strawberry (Fragaria x ananassa Duch.) fruit quality. J Sci Food Agric 97 (3):849-857.

Wei J, Kang L (2011) Roles of (Z)-3-hexenol in plant-insect interactions. Plant signaling & behavior 6 (3):369-371

Wei L, Mao W, Jia M, Xing S, Ali U, Zhao Y, Chen Y, Cao M, Dai Z, Zhang K, Dou Z, Jia W, Li B (2018) FaMYB44.2, a transcriptional repressor, negatively regulates sucrose accumulation in strawberry receptacles through interplay with FaMYB10. J Exp Bot 69 (20):4805-4820.

Witkowicz J, Urbanczyk-Wochniak E, Przybecki Z (2003) AFLP marker polymorphism in cucumber (Cucumis sativus L.) near isogenic lines differing in sex expression. Cellular & molecular biology letters 8 (2):375-381

Xie XB, Li S, Zhang RF, Zhao J, Chen YC, Zhao Q, Yao YX, You CX, Zhang XS, Hao YJ (2012) The bHLH transcription factor MdbHLH3 promotes anthocyanin accumulation and fruit colouration in response to low temperature in apples. Plant Cell Environ 35 (11):1884-1897.

Xie Z, Fan J, Charles MT, Charlebois D, Khanizadeh S, Rolland D, Roussel D, Zhang Z (2016) Preharvest ultraviolet-C irradiation: Influence on physicochemical parameters associated with strawberry fruit quality. Plant physiology and biochemistry : PPB 108:337-343.

Xu H, Cao Y, Xu Y, Ma P, Ma F, Song L, Li L, An D (2017) Marker-Assisted Development and Evaluation of Near-Isogenic Lines for Broad-Spectrum Powdery Mildew Resistance Gene Pm2b Introgressed into Different Genetic Backgrounds of Wheat. Front Plant Sci 8:1322.

Xu W, Peng H, Yang T, Whitaker B, Huang L, Sun J, Chen P (2014) Effect of calcium on strawberry fruit flavonoid pathway gene expression and anthocyanin accumulation. Plant physiology and biochemistry : PPB 82:289-298.

Yan JW, Ban ZJ, Lu HY, Li D, Poverenov E, Luo ZS, Li L (2018) The aroma volatile repertoire in strawberry fruit: a review. J Sci Food Agric 98 (12):4395-4402.

Yuan Y, Mei L, Wu M, Wei W, Shan W, Gong Z, Zhang Q, Yang F, Yan F, Zhang Q, Luo Y, Xu X, Zhang W, Miao M, Lu W, Li Z, Deng W (2018) SlARF10, an auxin response factor, is involved in chlorophyll and sugar accumulation during tomato fruit development. J Exp Bot 69 (22):5507-5518.

Zhang C, Bian Y, Hou S, Li X (2018) Sugar transport played a more important role than sugar biosynthesis in fruit sugar accumulation during Chinese jujube domestication. Planta 248 (5):1187-1199.

Zhang Q, Ma B, Li H, Chang Y, Han Y, Li J, Wei G, Zhao S, Khan MA, Zhou Y, Gu C, Zhang X, Han Z, Korban SS, Li S, Han Y (2012) Identification, characterization, and utilization of genome-wide

simple sequence repeats to identify a QTL for acidity in apple. BMC Genomics 13:537.

Zhang Y, Li W, Dou Y, Zhang J, Jiang G, Miao L, Han G, Liu Y, Li H, Zhang Z (2015) Transcript Quantification by RNA-Seq Reveals Differentially Expressed Genes in the Red and Yellow Fruits of Fragaria vesca. PloS one 10 (12):e0144356.

Zhao F, Li G, Hu P, Zhao X, Li L, Wei W, Feng J, Zhou H (2018) Identification of basic/helix-loop-helix transcription factors reveals candidate genes involved in anthocyanin biosynthesis from the strawberry white-flesh mutant. Sci Rep 8 (1):2721.

Zhou R, Zhu Z, Kong X, Huo N, Tian Q, Li P, Jin C, Dong Y, Jia J (2005) Development of wheat near-isogenic lines for powdery mildew resistance. TAG Theoretical and applied genetics Theoretische und angewandte Genetik 110 (4):640-648.

Zhou Y, He W, Zheng W, Tan Q, Xie Z, Zheng C, Hu C (2018) Fruit sugar and organic acid were significantly related to fruit Mg of six citrus cultivars. Food Chem 259:278-285.

Zorrilla-Fontanesi Y, Rambla JL, Cabeza A, Medina JJ, Sanchez-Sevilla JF, Valpuesta V, Botella MA, Granell A, Amaya I (2012) Genetic analysis of strawberry fruit aroma and identification of O-methyltransferase FaOMT as the locus controlling natural variation in mesifurane content. Plant physiology 159 (2):851-870.