Horticulture Innovation Australia

Final Report

DNA-marker development for flavour compounds in strawberry

Dr Eddie Pang
RMIT University

Project Number: BS12006

BS12006

This project has been funded by Horticulture Innovation Australia Limited using the strawberry industry levy and funds from the Australian Government.

ISBN 0 7341 3739 7

Published and distributed by:
Horticulture Innovation Australia Limited
Level 8, 1 Chifley Square
Sydney NSW 2000
Tel: (02) 8295 2300
Fax: (02) 8295 2399

Contents

Summary

Breeding strawberry (*Fragaria x ananassa*) with enhanced fruit flavour is one of the top priorities of the Australian Strawberry Breeding Program. Although several genes involved in the biosynthetic pathways of key volatile compounds have been identified, the development and application of molecular markers associated with this trait remains limited. This may be attributed to the complex genetic control and environmental influences, but also in the difficulty of assessing flavor in an objective and consistent manner. The current research aimed to firstly determine the attributes of Australian-grown strawberries that contribute to consumer liking. Concurrently, chemical analyses of the strawberries identified compounds contributing positively to consumer preferences. Finally, studies were performed to identify molecular markers closely linked to genes controlling the expression of these chemical compounds in strawberry.

To develop a platform for molecular marker identification, a DNA microarray was constructed using a broad subtraction approach between five strawberry genotypes and nine non-angiosperm species. Subtracted genomic library construction resulted in a microarray containing 287 features specific to the strawberry genome. Its efficacy was firstly evaluated by fingerprinting 15 strawberry genotypes, as well as correlating the aroma profiles and genetic data of three commercial cultivars and two breeding lines. Hierarchical clustering demonstrated that the strawberry-specific microarray was able to correctly cluster 15 strawberry genotypes based on known pedigree information. The aroma profiles and the microarray data of five selected genotypes produced highly similar hierarchical dendrograms. Three of the branch point markers, FLP1C6, FLP1E7 and FLP2E1 showed good correlation with ethyl hexanoate, methyl esters and linalool, respectively. These results indicate that the strawberry-specific microarray may be capable of detecting markers associated with aroma compound production.

For the identification of molecular markers associated with fruit flavour, the fruits from 50 F_1 progeny derived from 07-102-41 x Juliette cross were collected, and subjected to GC-MS analysis. The DNA from progeny exhibiting the extremes of a flavor attribute were pooled together, and used in hybridization experiments with the strawberry-specific microarray. Differential hybridisation patterns between the DNA pools were analysed using Discriminant Function Analysis (DFA), Fisher's ratio and Independent Samples *t*-Test to identify putative markers associated with fruit flavour. This approach successfully identified a set of the most predictive features that best discriminated between the extreme DNA pools. DNA sequence analyses showed that three DNA fragments (features), FLP1D7, FLP1A7 and FLP3E12 were linked to genes possibly involved in the biosynthesis of methyl butanoate, γ-dodecalactone and linalool.

The reliability of these potential markers, especially for methyl butanoate, is being assessed in an ongoing study. This robust platform should prove useful for future marker-trait association studies in strawberry. The application of marker technology for the strawberry program is now feasible, given the technological developments, and corresponding cost reduction in DNA analysis over the past five years. It is hoped that the DNA markers developed from this project, and our ongoing research will be used as a component of the strawberry breeder's toolbox for accelerating the production of new Australian strawberry varieties.

Keywords

DNA marker, Flavour, Strawberry, Marker-assisted Breeding, Sensory Panel, Gas Chromatography, Microarrays, Variety development

Introduction

The Strawberry Industry's Priorities 2009-2013 include a focus on consumer demands and preferences for high quality, flavoursome fruit. Strawberry aroma and flavour are thought to be the sum products of over 300 chemical compounds. The production of these compounds during the maturation phase of strawberry fruit has been found by a number of researchers, including our (RMIT) previous HAL project (BS06006), to be influenced by plant genotype, day/night temperature, light intensity and other factors such as relative humidity and soil moisture. In that project, we comprehensively analysed the flavour compounds of several Australian varieties (Adina, Juliette, Kiewa and Tallara) and Albion, a Californian variety over the growing season [1]. What we discovered was that the environment's influence varied with each flavour compound; some compounds are produced stably, while others fluctuated significantly over the season. Another finding was that for many compounds, the plant genotype exerted a greater influence on flavour production than the environment - good news for the plant breeder, as this factor is within his/her control [2].

In the 2009/10 season we also began analysing the flavour compounds from the most promising breeding lines from the strawberry breeding program, including Dutch lines 07-102-41-FL, and 07-095-35-FL, and Japanese line 04-069-91-FL. The breeding lines mentioned above were subsequently crossed with Juliette, Adina, and several day-neutral lines in the hope of producing a new variety with the new, distinctive flavour characteristics during the summer of 2010/11. Several hundred hybrid plants from each cross were grown in the field and glasshouse, and the evaluation of these were performed in the 2011/12 for many characteristics, including flavour. The concern, however, was that during the selection process, the genes conferring the distinctive flavours may be lost. One possibility is to analyse the berries from each hybrid plant for the unique flavour compounds as we had been performing. However, this is impractical, as the chromatographic analysis (GCxGC-TOF), while producing accurate and extensive information, is labour intensive, slow, and unsuitable for screening thousands of individual plants. Further, all plants, including those producing fruit with undesirable flavour must be grown to maturity for harvesting, thereby increasing space and labour requirements.

Another important consideration is that the perceived flavour of the fruit is the result of complex stimulus involving response interactions between the food matrix and human sensory, perceptual and cognitive processes [3]. In taste, there are multi-quality interactions resulting from mixing of more than one compound (for example, bitter, salty, sour, sweet and umami). Many outcomes are possible, including perceptual enhancement and suppression, unmasking of a taste not initially observed, or possibly chemical synthesis of a new taste [3]. Therefore, accurate assessment of flavour attributes of strawberry varieties required using sensory evaluation techniques and a trained panel that is familiar with a range of strawberry flavour attributes [4]. We therefore proposed to characterise the aforementioned breeding lines, their progeny, and a number of commercially-grown varieties using descriptive analysis techniques and trained panel of judges. A trained panel should accurately evaluate and quantify the aroma, taste, aftertaste, texture, and appearance of the fruits.

DNA-based markers provide a more feasible solution for screening large numbers of plants rapidly at vegetative stage, and at a significantly lower cost. The final aim of this project was to develop DNA markers linked to the genes/alleles responsible for the production of

the main distinctive compounds identified by the current project BS06006 in the breeding lines 07-102-41-FL, 07-095-35-FL and 04-069-91-FL. These markers will be produced by correlating the genotype (genetic variation) and phenotype GC-MS analysis of desirable and off flavour compounds, and descriptive sensory analysis of aroma, taste, aftertaste, texture, and appearance of the fruits) of the aforementioned breeding lines and their progeny. Such a comprehensive analysis should allow us to produce DNA markers related to desirable/undesirable strawberry flavour. These markers, once produced, will be deployed in national breeding programs to track the introgression of these genes into a new Australian strawberry variety.

Methodology

Strawberry Varieties

A number of currently-cultivated varieties and advanced breeding lines were employed in this study for sensory analysis and for DNA fingerprinting and molecular marker development. They were:

Current Australian varieties: Juliette (short-day) and Melba (day-neutral)

Current American varieties: Albion, Camino Real, Portola, Palomar, San Andreas

Australian breeding lines: 07-102-41-FL (Dutch1), 07-095-35-FL (Dutch2), 04-069-91-FL (Japanese)

Australian advanced breeding lines: 27B, 80B, 165B, 168B, 190B, 202B

Older Varieties: Cambridge Rival (UK), Hokowase (JP), Chandler (USA), Adina (Aus), Lowanna (Aus), Alinta (Aus)

***Fragaria vesca*:** Fraises des Bois

Sampling Locations and Years

The project was conducted over the period of late 2012-2015, where strawberries from day-neutral varieties were sampled weekly from November to April for up to a maximum of 16 weeks (typically 12-16). Short-day varieties were sampled from between 6-8 weeks. The strawberries were sampled either from Mr Sam Violi's commercial farm in Coldstream, VIC. or from the Strawberry Breeding Farm, Wandin, VIC.

Strawberry Flavour Profiling using a Sensory Panel

Following screening of 58 potential panellists, 20 participants were trained in the descriptive language of strawberries. A lexicon was developed and defined to characterise the appearance, aroma, taste, texture and after taste of the strawberries (Table 1). Product references were agreed upon and supplied for each of the descriptive terms. Panellists were familiarised with each of these first individually, followed by the products spiked in strawberry samples. Panellists were trained to use 15cm line scales in the same way, through the use of spiked samples. Anchors for the ends of the scales were agreed upon. To ensure reproducibility and reliability of the results, panellists practiced these repeatedly. Following training consisting of 40 hours, 14 of the 20 panellists were chosen to continue on to the testing phase. Further training sessions were run regularly throughout the testing period, to ensure panellists were continuing to perform and to ensure reliable and repeatable results. Of the 14 panellists that completed year one, 12 went on to complete analyses in year two.

Table 1. Strawberry lexicon as devised by a trained panel

Attribute		Definition	Product reference
APPEARANCE			
Conical		Cone shape, the stem raised in the middle	
Heart shaped		The shape of a love heart, the stem dipping in the middle	
Square		Equal width to height	
Round		Circular, the stem raised in the middle	
Pointed tip		The tip of the strawberry pointed	
Rounded tip		The tip of the strawberry rounded	
Long		Elongated fruit	
Short		Short fruit	
Light red		Pale red	
Bright red		Fire truck red	
Dark red		Deep crimson or blood red	
Pink red		Pale red with pink tones	
Orange red		Pale red with orange tones	
AROMA			
Aniseed		Aromatics associated with liquorice and fennel	Liquorice
Apple		Aromatics associated with a combination of sweet and sour in both red and green apples and pear	Grated apple and pear
Banana		Characterised by a ripe banana	Ripe banana
Berry		The aroma associated with a combination of mixed berries (raspberries, blackberries, blueberries)	Combination of mixed frozen berries, thawed
Candy		An intensely artificial sweetness associated with lollies/candy	Crushed candy rolls
Caramel		The aromatics of cooked sugar with buttery and nutty attributes	Caramel sweets, freshly cooked caramel
Citrus		Characterised by sour notes present in lemon, lime and orange	Combination of cut lemon, lime and orange

Chemical		An artificial inedible pungent aroma characterised by bleach/ammonia	Bleach, ammonia
Earthy		Aromatic associated with dirt or soil	Dirt, fresh soil
Fermented		An over-ripe characteristic associated with fermented fruits	Mixture of over-ripe fruit
Floral		Sweet, fragrant aromatic associated with flowers, perfume and potpourri	Combination of fresh flowers
Fruity		Sweet, intense aromatic associated with a combination of mixed fruit; pineapple, melon, apple, grape	Combination of mixed, cut fruit
Grape		Aromatic associated with ripe green and red grapes	Combination of crushed red and green grapes
Green		An unripe aroma characterised by cut grass and unripe or green fruit	Cis-5-hexanal, cut grass
Honey		A sweet aroma associated with honey	Honey
Mango		Aromatic associated with a ripe, sweet mango	Ripe mango
Melon		Aromatic associated with rockmelon/cantaloupe and honey dew melon	Combination of ripe rockmelon/cantaloupe and honey dew melon
Musk		A sweet, floral, perfume candy aromatic associated with musk	Musk sticks
Off-odour		Aromatic associated with rotting fruit	Rotting fruit
Passionfruit		Sweet and sour aromatics associated with a ripe passionfruit	Ripe, cut pineapple
Peach		Sweet aromatics associated with ripe yellow peach	Cut yellow peach
Pineapple		Sweet and sour aromatics associated with a fresh, ripe pineapple	Cut fresh pineapple
Plum		A rich combination of sweet and sour notes associated with plum, characterised by wine	Fresh or tinned plum
Rose		A distinct floral aroma characterised by fresh rose	Rose oil, fresh roses
Tomato		Sour aromatics associated with a ripe tomato	Ripe tomato
Vanilla		A sweet aroma characterised by vanilla beans	Vanilla beans
Vinegar		A sharp, pungent and acidic aroma characterised by vinegar	White vinegar
Woody		Aromatics associated with bark, a cut tree stump, tree branch or wood	Tree branch, bark, fresh wood
TASTE			
Sweet		Taste associated with sugar	Sucrose solution

Sour		Taste associated with acid	Citric acid solution
Bitter		Taste associated with toxins, eg. Caffeine or quinine	Caffeine solution
MOUTH FEEL			
Metallic		Feeling in the mouth associated with metal, tin or rare, bloody meat	Iron sulphate
Astringent		Feeling in the mouth characterised by drying associated with the presence of tannins	Tannic acid solution
TEXTURE			
Juicy		High presence of fluid	A juicy orange
Dry		Low presence of fluid	Apple
Soft		Soft to the touch	Mango
Firm		Hard to the touch	Apple
Gritty		Grainy, sandy texture, high presence of seeds	Grainy bread, chia, capeseed
Smooth		Low presence of seeds, slimy	Avocado, mango
Fibrous		Presence of long fibres, stringy	Unripe mango

Consumer hedonic ratings also were evaluated on three separate occasions for the available commercial varieties and elite lines. A total of two hundred consumers provided hedonic ratings at one of three sessions. This data was collected from two separate locations for a greater representation of consumer acceptance.

Strawberry Flavour Profiling by Gas Chromatography

Solid Phase Microextraction (SPME)

To assess the ability of the microarray in finding DNA markers related to strawberry flavour, phenotyping of the aroma profiles of the current varieties and breeding lines was performed. Volatile compounds were extracted from the fruit puree using Solid Phase Microextraction (SPME) method [5]. Five large berries from each variety/breeding line for each sampling time were thawed and homogenised with a hand blender. Approximately 1 g of puree was immediately dispensed into individual SPME vials with screw caps and stored at -80 °C. Prior to GC-MS analysis, the sample was thawed to room temperature for 20 min and pre-equilibrated at 60 °C in a heating block for 10 min. The volatile compounds were extracted using a 65 µm polydimethylxiloxane/divinylbenzene (PDMS/DVB)-coated fiber held in a SPME Holder 57330-U (Supelco, Bellafonte, PA) (Figure 1). This fiber was first conditioned at 250 °C for 30 min, and then exposed to the vial headspace for 30 min at 60 °C. After equilibrium, the fiber was removed from the sample and the analytes were thermally desorbed in a GC injector port at 250 °C for 3 min. Each sample was performed in duplicate for two sampling periods.

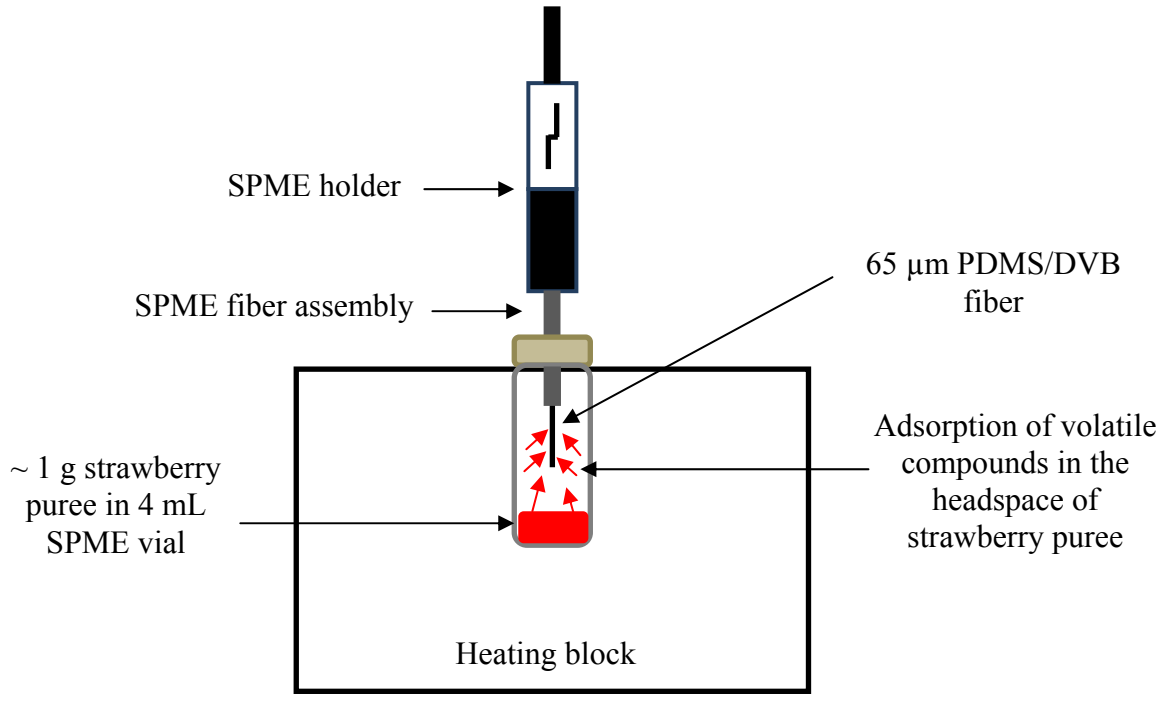


Figure 1. Extraction of volatile compounds from strawberry purees using SPME method.

Gas Chromatography-Mass Spectrometry (GC-MS)

The volatile compounds were analysed using the Agilent 6890 GC coupled with a 5973 MS detector through a heated transfer line at 280 °C. Compounds were separated using DB-5ms column with dimensions of 30 m x 0.25 mm I.D. x 0.25 µm film thickness. Helium was used as a carrier gas at a flow rate of 1.5 mL/min. 1.0 µL was injected using the splitless injection mode with a 2.5 min of solvent delay. The oven temperature was programmed initially at 40 °C for 1 min, then increased at a rate of 6 °C/min to 190 °C and kept constant at the same temperature for 26 min with a final isotherm at 190 °C for 4 min. The MS source temperature was 230 °C and the compounds were monitored over the mass range m/z 45 – 400.

GC-MS data analysis

A similarity search was carried out by comparing the retention times and quality of known compounds in Wiley and Adams mass spectra libraries. All chromatographic peaks found in two or more technical replicates of the same sample and with a quality greater than 80 were taken into account. The relative composition of volatile compounds in the headspace of the strawberry puree was quantitated based on area normalisation method with two assumptions: (1) detector response is the same for different compounds; and (2) compounds of the sample injected are completely detected and will produce peaks (Shimadzu, 2006). The calculation was done according to the equation below:

$C_i = A_i/A_t \times 100\ \%$
Where C_i = Content of a compound in the sample
A_i = Area of compound peak in the chromatogram
A_t = Total area of the peaks in the chromatogram

The mean relative composition and standard deviation of each compound was calculated from three technical replicates for the five strawberry genotypes. Volatile compounds detected by GC-MS and their relative compositions were categorised according to different chemical groups for each strawberry genotype.

Construction of a Strawberry-Specific Microarray

The details of the construction of the strawberry-specific microarray may be found in Appendix 1. A schematic summary of the process may be found in Fig. 2. Briefly, the DNA from several strawberry varieties were pooled together, and subtracted from the DNA of several non-flowering plant species (non-angiosperms). The subtracted DNA fragments was then cloned into *E. coli*, individually amplified by PCR and then gridded onto a microarray.

Fingerprinting of 15 Strawberry Genotypes

The procedure for conducting the DNA fingerprinting of the strawberry varieties may be found in Appendix 2. Briefly, DNA from the 15 varieties/lines were extracted, digested with restriction enzymes, fluorescently labelled, and individually hybridised onto the strawberry-specific microarray. A hierarchical dendrogram showing the genetic

relationships between the genotypes was generated using SPSS. This was compared with a hierarchical dendrogram generated using GC-MS volatile compound data for several common varieties.

Molecular-Marker Discovery

The procedure for the discovery of markers associated with flavor development is described at length in Appendix 3. Briefly, the fruits from 50 F_1 progeny derived from 07-102-41 x Juliette cross were collected, and subjected to GC-MS analysis. The DNA from progeny exhibiting the extremes of a flavour attribute were pooled together, and used in hybridization experiments with the strawberry-specific microarray. Differential hybridisation patterns between the DNA pools were analysed using Discriminant Function Analysis (DFA), Fisher's ratio and Independent Samples *t*-Test to identify putative markers associated with fruit flavour.

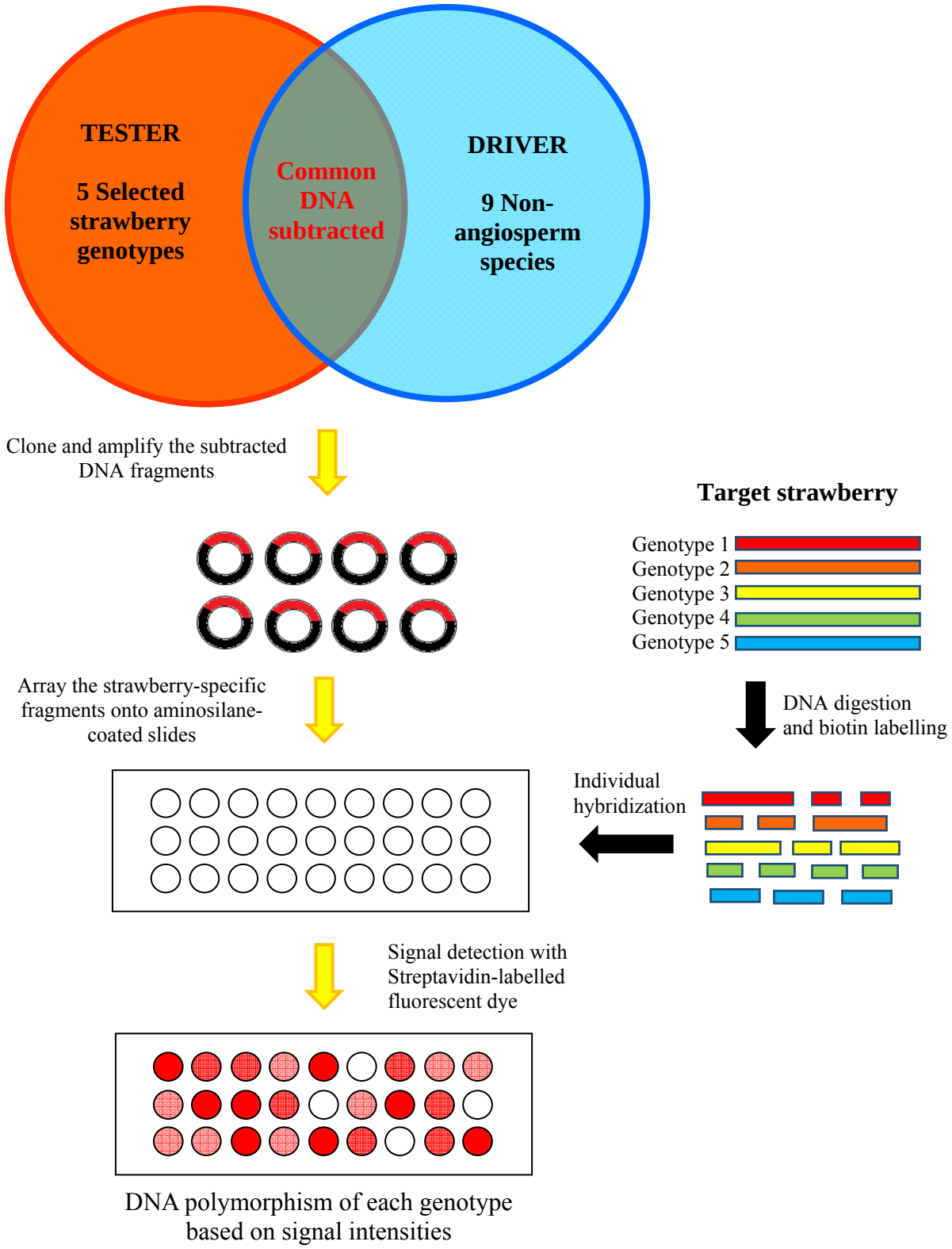


Figure 2. The workflow of strawberry-specific SDA construction and performance validation. Adapted and modified from [6]

Outputs

Procedure for training a panel for describing strawberry attributes

To conduct our sensory research, a procedure for training a panel and a lexicon for the description of different aroma and flavor attributes were developed by Ms Oliver, and her Deakin University supervisor A/Prof Russell Keast. A manuscript is currently being prepared for publication as part of Ms. Oliver's PhD thesis work. This protocol will be useful in future sensory studies, for example, on advanced breeding lines.

Development of GC-MS methodologies for volatile compounds analysis in strawberries

A GCxGC TOF-MS method for the detection of strawberry volatiles was already developed in the course of our previous project, BS06006 [1,2]. While this method is superior in terms of the total number of compounds detected, often a simpler method, such as GC-MS is sufficient for the assessment of volatiles for comparing different genotypes of strawberry for flavor intensity. Ms Thish DeSilva was able to develop a SPME GC-MS system for the detection of key volatile compounds in strawberry fruit. The information obtained from this method may be used in correlation analyses to establish which volatile compound contributes most to consumer "liking".

Strawberry-specific microarray and subtraction DNA products

A strawberry-specific microarray was produced by randomly gridding several hundred cloned DNA products after subtractive hybridization. This microarray may be used for a variety of purposes, for example, DNA fingerprinting of unknown varieties, producing DNA fingerprints for a new variety for PVR purposes, or be used to assess germplasm diversity in the strawberry breeding program.

The remaining unused subtracted DNA, which is currently stored at -80^{o}C is an untapped resource. If funding is available, we could sequence all the fragments in a next-generation sequencer, and perhaps discover strawberry-specific gene fragments related to flowering, and fruit production. This research could lead to a deeper understanding of the mechanisms of short-day versus day-neutral plants, and of the ethylene insensitivity of strawberry fruit, which currently prevents the implementation of artificial ripening, as in e.g. bananas.

DNA marker for Methyl Butanoate

In the course of this project we were able to identify a potential DNA marker for an important ester compound, methyl butanoate. We are currently conducting studies to further validate the reliability of the marker. A manuscript is currently being prepared for publication. An Honours project commencing in 2016 will attempt to discover a DNA marker for another important volatile compound in segregating progeny from a different cross to the one employed in this study.

Capacity Building

Three PhD students were trained in this project on different aspects of strawberry research-sensory, chemical, and genetics/breeding. It is hoped that further funding may be obtained to retain them within the horticultural industries.

Future Publications

Although we have only produced one publication thus far, all three PhD students are currently preparing manuscripts for publication. It is anticipated that within the next year, we will have at least six publications from this project.

Outcomes

RESULTS

Strawberry Flavour Profiling using a Sensory Panel

Strawberries were evaluated weekly in a single two hour session. In the first season twelve weeks of evaluations were conducted over the period from November 2013 until April 2014. In the second season twelve weeks of evaluations were completed over the November 2014 to April 2015 period. Due to the effects of fatigue, not all strawberry varieties were evaluated during each session. The appearance, aroma, taste, texture and aftertaste were evaluated in each session for each of the strawberry varieties. Availability of strawberries determined which strawberries were evaluated at each time point. In the first year, Camino Real, Palomar and Portola varieties were evaluated until the end of December. Albion, San Andreas and Melba were evaluated through until April. In year two, the short-day elite lines were evaluated from November until the middle of January. The day-neutral elite lines were evaluated to the beginning of April.

For the first year of samples, weekly flavour profiles were compiled for each of the varieties tested. Overall flavour profiles were then compiled for each of these varieties. A few flavour profiles are included below as examples (Fig. 3). It is not possible to show all the profiles in this report. At the time of writing of this report, statistical analyses of both years' data is ongoing.

The hedonic ratings for the 2014/15 season were collated to determine the most liked strawberry. It may be observed that in this season, the Australian variety was the least liked by consumers, and that the Australia elite breeding line "27B" was the most-liked. It is interesting to note however, that we noticed a deterioration in the quality of Melba fruits picked during the 2014/15 season compared to those picked in 2013/15. The reasons for this are currently being investigated. This information is currently being combined with the flavour profiles and statistically analysed to determine which flavours are most liked to inform breeders. This analysis is ongoing.

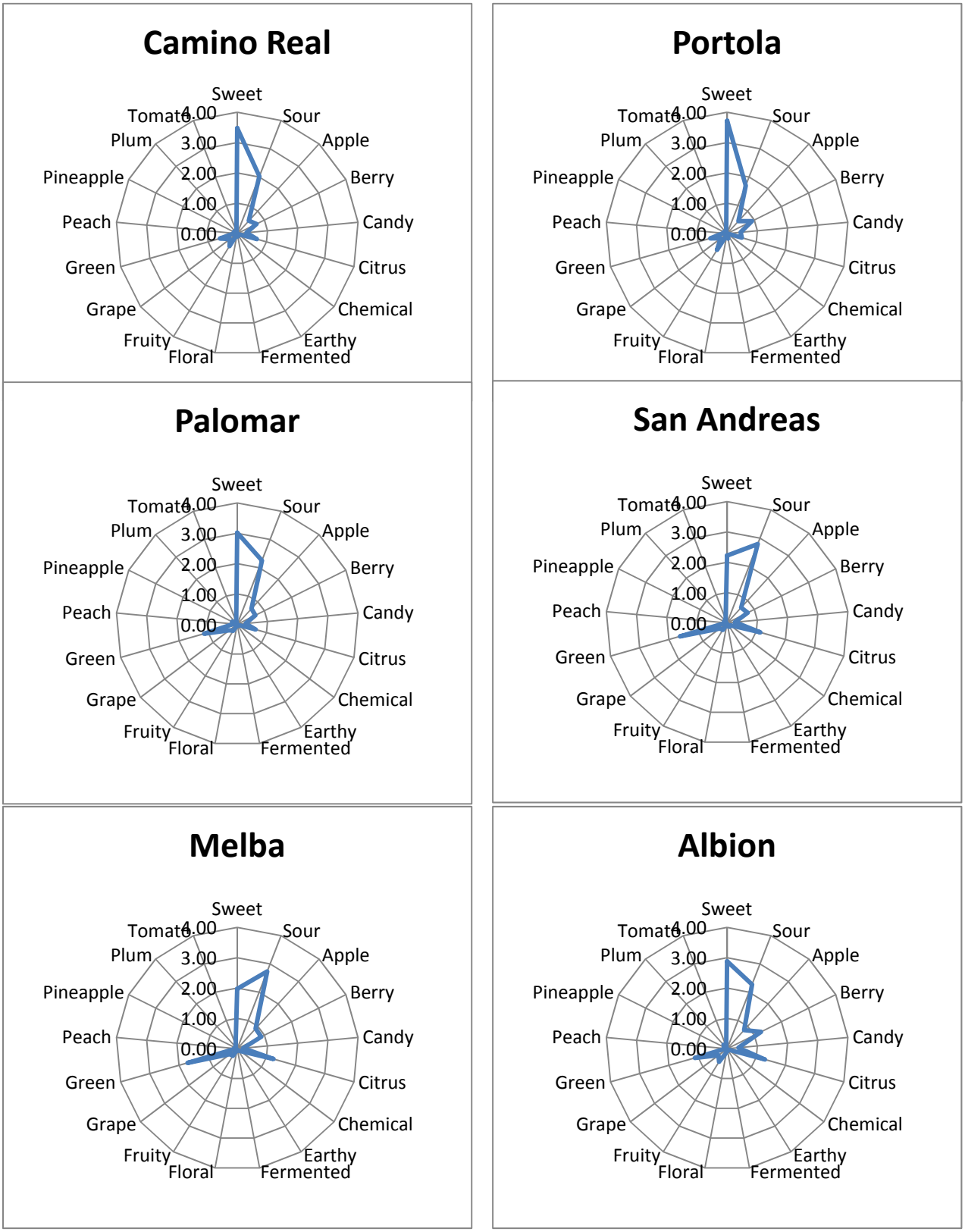


Figure 3. Flavour profiles of commercial varieties sampled during the 2013/14 season

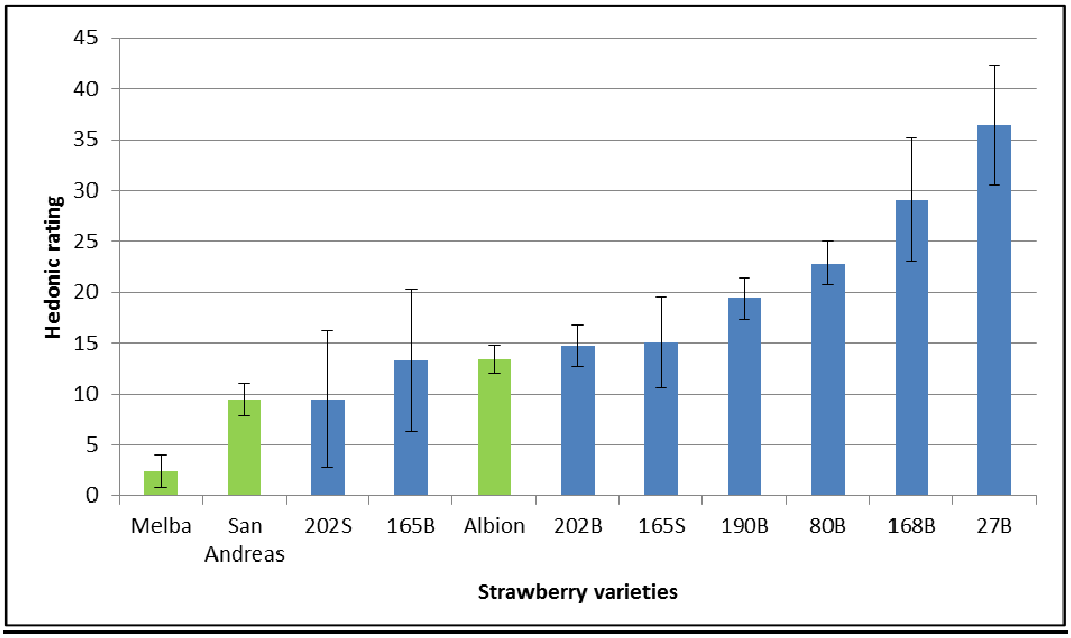


Fig 4. The "liking" ratings of three commercial varieties (green) and eight elite breeding lines (blue).

Molecular-Marker Discovery

Phenotyping of the progeny of the cross between Juliette and 07-102-41 for flavour volatiles was conducted as previously described. The GC-MS method detected between 50-120 compounds between the samples, of which descriptive statistics were derived for 23 of the most important volatile compounds (Appendix 4). The results for two of the important compounds, methyl butanoate and mesifuranne are included here in the body of the report (Fig. 5 and Fig. 6 respectively), but the reader is directed to Appendix 4 for the full set. It may be observed that although there was a significant difference between the parents in terms of methyl butanoate production (Fig. 5), there was negligible differences in terms of mesifuranne production (Fig. 6). This means that the prospect of discovering molecular markers for methyl butanoate was higher than that for mesifuranne in this cross, as the likelihood for marker discovery increases if the parental genotypes are sufficiently different for a particular phenotype (volatile compound). It was determined from the various graphs in Appendix 4 that methyl butanoate, an important ester, was a good candidate for marker discovery, and for the proof of concept for the microarray approach. This ester, along with several other compounds were targeted in subsequent experiments to discover the features on the strawberry-specific microarray that were highly correlated with their production. These experiments revealed several features (cloned fragments) which may possibly be candidate markers for a number of volatile compounds. These results are summarized in Table 2 below. The DNA sequences of these features were searched against the *Fragaria vesca* database to reveal their identity (Table 2).

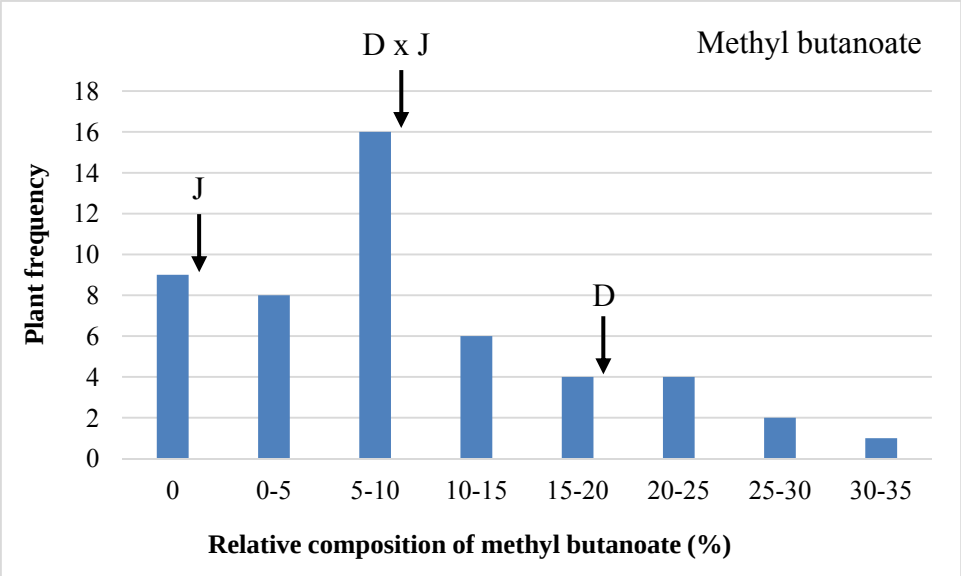
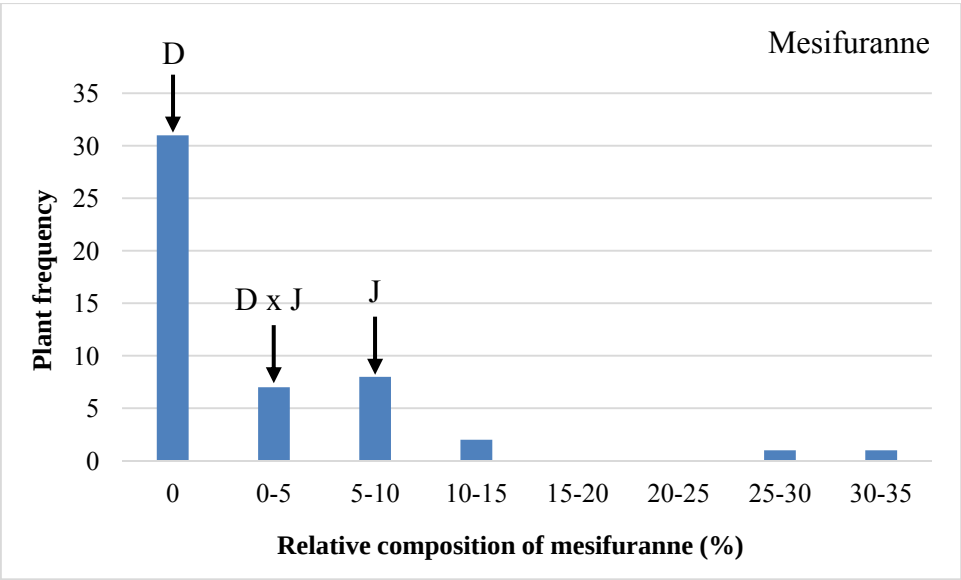


Figure 5. Frequency distribution of key volatile compounds measured as relative peak areas in the '07-102-41' x 'Juliette' progeny for methyl butanoate. The mean values of the parents and F_1 population are indicated by arrows (D: 07-102-41; J: Juliette; D x J, respectively.



(e)

Figure 6. Frequency distribution of key volatile compounds measured as relative peak areas in the '07-102-41' x 'Juliette' progeny for mesifuranne. The mean values of the parents and F_1 population are indicated by arrows (D: 07-102-41; J: Juliette; D x J, respectively.

Table 2. Putative identity of the most discriminatory features searched against the *Fragaria vesca* draft genome (v1.1). E-value regarded as significant if < $1e^{-5}$.

Clones	Length (bp)	Landmark or region	Sequence description	E-value	Specific to target
FLP1A7	442	LG6:21708323..21708764, scf0513196:589686-590127	Genomic DNA region on linkage group 6	0.0	γ-Dodecalactone
FLP1B3	343	gene32946 on scf0510865:52..396	NAD(P)H-quinone oxidoreductase subunit H, chloroplastic (similar to)	$5e^{-45}$	Ethyl hexanoate
FLP1D7	627	LG3:27624656..27625284, scf0513138:502060-502688	Genomic DNA region on linkage group 3	0.0	Methyl butanoate
		LG2:17544790..17544932, scf0513123:85522.. 85664	Genomic DNA region on linkage group 2	$1e^{-38}$	
FLP1D11	850	gene32967 on scf0510833:190..1040	ATP synthase subunit alpha, chloroplastic (similar to)	$1e^{-133}$	Ethyl butanoate
FLP2D11	670	scf0510759:1..513	N/A	0.0	(*E*)-Nerolidol
FLP3E12	539	scf0513205:141..680	N/A	0.0	Linalool
		LG3:20817890..20818104, scf0513118:718918-719132	Genomic DNA region on linkage group 3	$6e^{-83}$	

Subsequent DNA sequencing of Juliette, 07-102-41 for the region corresponding to the FLP1D7 clone revealed a C/T single-nucleotide polymorphism (SNP) between the Juliette (low methyl butanoate) and 07-102-41 (high methyl butanoate) indicating that this SNP could be a good marker for methyl butanoate. Ongoing studies are being performed on the F1 progeny and other lines to determine the reliability of this marker. A new study, commencing in 2016 will attempt to validate the potential markers listed for other volatile compounds in Table 2.

Apart from marker discovery, the strawberry-specific microarray was found to be very useful for DNA fingerprinting studies. A hierarchical dendrogram was generated based on the normalised mean signal-to-noise of all 287 microarray features (Figure 7). The relatedness of these genotypes was compared with known parentage information (Table 3). Hierarchical clustering identified two main clusters at a truncation point of 13, with Fraises Des Bois (the out-group control) showing the furthest genetic distance from the other octoploid strawberries as expected (Figure 7).

Table 3. Strawberry genotypes used in this study, their immediate parents and country of origin.

Genotypes	Source	Parents	Origin
Australian			
Adina	Wandin	Pajaro x 88-042-35	Victoria
Alinta	Wandin	Chandler x 88-011-30	Victoria
Juliette	Coldstream	Adina x 92-50-76	Victoria
Lowanna	Wandin	Selva x 89-064-1	Victoria
Melba	Coldstream	97-101-75 x 04-99-142	Victoria
USA			
Albion	Coldstream	Diamante x Cal 94.16-1	California
Camino Real	Coldstream	Cal 89.230-7 x Cal 90.253-3	California
Chandler	Digger's Club	Douglas x Cal 72.361-105	California
San Andreas	Coldstream	Albion x Cal 97.86-1	California
Japanese			
Hokowase	Digger's Club	Unknown	Japan
European			
Cambridge Rival	Digger's Club	Dorsett x Early Cambridge	UK
Fraises Des Bois	Digger's Club	N/A	France
Breeding lines			
07-102-41	Wandin	Unknown	Netherlands
07-095-35	Wandin	Unknown	Netherlands
04-069-91	Wandin	Unknown	Japan

This result indicated that the strawberry-specific microarray was capable of genotyping DNA polymorphism at the interspecific level in *Fragaria*. Two sub-clusters were resolved in Cluster I, in which the first sub-cluster represented a mixture of American (Chandler and Camino Real), Japanese (Hokowase) and European (Cambridge Rival) cultivars. The second sub-cluster grouped the Australian cultivars (Adina, Lowanna, Melba and Alinta). In Cluster II, an Australian cultivar (Juliette) was grouped together with three other breeding lines (07-102-41, 07-095-35 and 04-069-91) and they were separated from the American cultivars (Albion and San Andreas) (Figure 7).

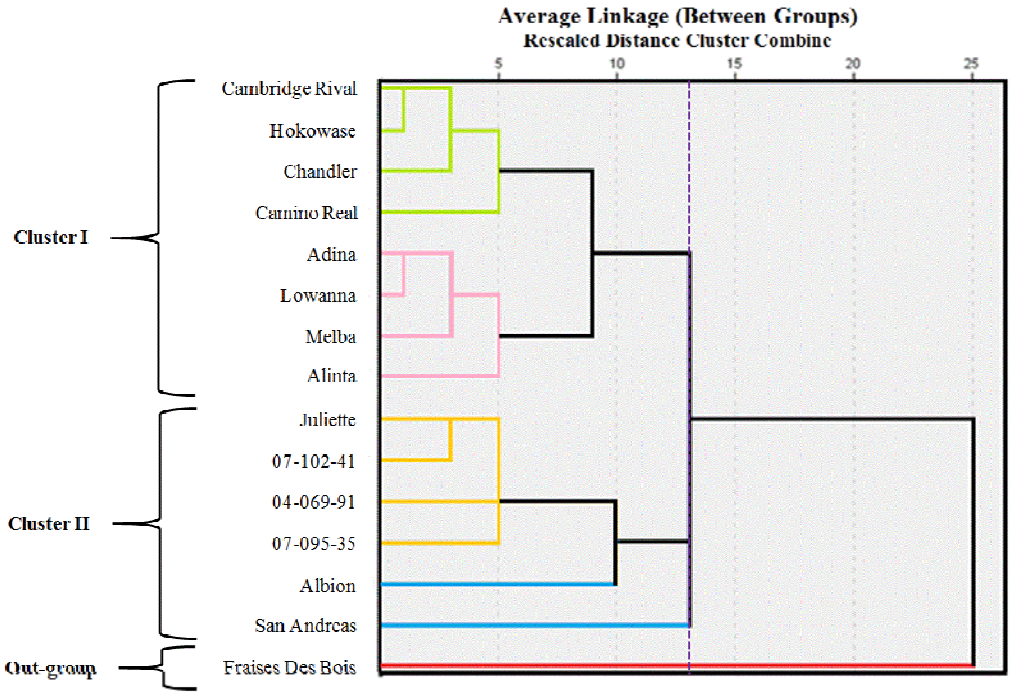


Figure 7. Hierarchical dendrogram using the average-linkage-between-groups method and squared Euclidean distance showing genetic relationships of 15 strawberry genotypes based on their hybridisation patterns on the strawberry-specific SDA. The steps of the dendrogram show the combined clusters and the values of the distance coefficients at each step; the values have been rescaled to numbers between 0 and 25, preserving the ratio of the distances between the steps.

Overall, the American cultivars were widely distributed in the dendrogram, suggesting diverse parentages for these cultivars. Camino Real and Chandler were grouped together in the first sub-cluster of Cluster I, inferring a common parent between these two cultivars. While located on different clusters, San Andreas, a direct descendant of Albion, was clustered next to Albion, confirming their close genetic relationship (Figure 7). In addition, all the Australian cultivars (Adina, Lowanna, Melba and Alinta) were clustered together except for Juliette (Figure 7). These Australian cultivars may share the same genetic background since they all have the American cultivars as one of their parents (Table 3). Furthermore, the breeding lines (07-102-41 and 07-095-35) which were known to contain some genetic information from the European strawberry were clustered together as expected.

Impacts

The impacts of this project are detailed in the Outputs section. In particular, we would like to highlight the development of new procedures and technology for fast-tracking the breeding process for new strawberry varieties. These models and methodologies may be further applied to other horticultural species, such as cherries, or apples where flavour characteristics are important, and where, as tree species, have long breeding cycles. We would also point out the capacity building, i.e. the training of PhD students for industry.

Consequences

The outputs from this study may be incorporated at any time into the National strawberry breeding program, and we are working closely with Industry to identify opportunities for marker-assisted breeding.

Cost/Benefit Analysis of Implementing Marker-Assisted Breeding for Strawberries

We do not intend to present a detailed cost-benefit analysis here, but we do need to point out:

1. The initial cost of marker development is inherently high.
2. The implementation of markers- a single PCR reaction (DNA test) used to cost about $2 in 1995, excluding labour costs. Twenty years later, the cost has dropped to less than 20 cents per reaction. The time required to prepare the samples has also been more than halved, thanks to new protocols which do not require initial DNA extraction, leading to significant labour cost savings.
3. Advancements in experimental methodology mean that plants do not have to be individually screened for a particular marker. Instead, we may construct pools of DNA from batches of e.g. 100 plants, screen each pool for the marker, and discard plants from any pool that do not possess the marker. For pools possessing the marker, we may re-pool the DNA from the 100 plants into e.g. 10 pools of 10 plants, and re-screen. Using this technique, we may drastically reduce the number of PCRs performed, and it is feasible to identify a single plant which possesses the marker, out of 10,000 plants, using only $100_{(\text{pools of 100 plants})} + 10_{(\text{pools of 10 plants})} + 10_{(\text{individual plant})} = 120$ PCR reactions.

Evaluation and Discussion

The project was conducted satisfactorily with several minor setbacks. Firstly, although the recruitment of the PhD students Ms MC Gor and Ms P. Oliver was problem-free, the recruitment of the third student for the GC-MS component of the project was difficult. The PhD candidate left the project after two months to pursue another PhD project, and recruitment had to be repeated. This caused a few delays at the beginning of the project. The retirement of the long-serving strawberry breeder, with whom we had developed the project, also introduced a few new challenges.

The methodologies developed for molecular marker discovery were the best we could achieve given the available technology at the beginning of the project. During the course of the project, RMIT University acquired two Next-Generation Sequencers (NGS). If these machines had been available when the project was being developed, we would have opted to use this technology rather than the older microarray-based platform which was employed in this project. Instead of gridding random subtracted fragments, we could have "simply" sequenced all the subtracted fragments, chose the ones that were associated with the biochemical pathways for the various volatile compounds, and then performed validation studies to assess their reliability as DNA markers. We remain hopeful, with additional funding, to perform this analysis.

At the time of writing, the field data collection was completed as stated in the project objectives. Dr MC Gor, who was the PhD student on the molecular marker component of the project finished her studies, and has graduated. She is currently preparing several manuscripts for publication, and will assist in the training of a new Honours student who will continue the marker discovery work in 2016. However the remaining two PhD students, Ms P. Oliver and Ms T. DeSilva are still finalizing the analysis of a large quantity of experimental data, and writing their respective theses. They are also cross-linking their studies with Dr Gor's results. Consequently, not all the data obtained during this three-year study may be presented in this Final Report. This would also be impractical, as each thesis is several hundred pages in length. These theses, however, will be available for perusal from the RMIT and Deakin University libraries in 2016. A PDF copy will also be made available upon request.

In terms of consultation, we worked, and continue to work closely with Strawberries Australia to ensure that our project objectives met with approval from Industry. Mr Sam Violi, the current Chairman of Strawberries Australia provided not only constant feedback on our project, but much of the strawberries that we analysed. We worked with the new strawberry breeders, Mr H.

Marsh and Mr. P. Brevis, both of whom had useful input into the direction of our project. For example, our initial aim was to only assess the parents and breeding lines identified in the original project proposal. After consultation, we expanded the assessments to all available commercial varieties, and also sampled for many more weeks per season than we had initially proposed. In this respect, we have exceeded the scope of our project. We also initiated studies into the development of molecular markers for day-neutrality (not reported here) at the suggestion of a number of industry partners.

The outcomes from this project demonstrated that strawberry varieties, whether Australian, or sourced from the USA varied in their desirability over the growing season. Some varieties appeared to be more stable than others in this respect. The consumer panels also preferred the latest elite lines (a few of them soon to be varieties?) over the existing varieties such as Albion or Melba. This is very encouraging, and indicated the Victorian strawberry breeder(s) was on the right track. We are in the process of correlating fruit attributes with "liking", though we are increasingly certain that it is the COMBINATION of attributes such as sweetness, floral, berry, candy etc that is responsible. The usefulness of the strawberry-specific microarray to the breeding program depends on direction of the new strawberry breeders. As stated previously, although an older platform compared with NGS, it remains very useful for DNA fingerprinting, germplasm diversity assessment, and for future marker discovery research. Our continued research on the discovery, and validation of DNA markers for flavor compounds should provide the breeders with several useful markers in the near future.

Recommendations

1. Continue the research investment into flavour and health-promoting compounds in strawberry. This may be in terms of consumer preferences, chemistry, clinical studies, or DNA marker research.
2. Implement the outcomes of this study into the strawberry breeding program.
3. Provide the continuity in research and employment opportunities for PhD graduates in the horticultural industry, perhaps by the initiation of a postdoctoral fellowship program.

Scientific Refereed Publications

Conference Paper

Gor, M.C., Samykanno, K., Mantri, N., Pang, E. and Marriott, P. (2014). Development of molecular markers associated with strawberry quality traits using a subtracted diversity array. *Acta Hortic*. **1049**, 343-348 DOI: 10.17660/ActaHortic.2014.1049.46

Intellectual Property/Commercialisation

No commercial IP generated

References

[1] Samykanno, K., Pang, E., Marriott, P.J. (2013). Chemical characterization of two Australian-grown strawberry varieties by using comprehensive tow-dimensional gas chromatography-mass spectrometry. *Food Chemistry* **141**(3), 1997-2005.

[2] Samykanno, K., Pang, E., Marriott, P.J. (2013). Genotypic and environmental effects on flavor attributes of "Albion" and "Juliette" strawberry fruits. *Scientia Horticulturae* **164**, 633-642

[3] Keast, R. S. J. (2009). Food quality perception. In E. Ortega-Rivas (Ed.), Processing Effects on Safety and Quality of Foods. Boca Raton: Taylor & Francis.

[4] Keast, R. S. J. and Lau, J. (2006). Culture-specific variation in the flavor profile of soymilks. *Journal of Food Science* **71**(8), 567-572.

[5] Pawliszyn, J. & Pawliszyn, J. (1997) *Solid phase microextraction: theory and practice*, vol. 61, Wiley-Vch New York

[6] Niu, L, Mantri, N, Li, C, Xue, C & Pang, E. (2011), 'Array-based techniques for fingerprinting medicinal herbs', *Chinese medicine*, vol. 6, no. 1, p. 18.

Acknowledgements

We wish to gratefully acknowledge the funding provided by HIA and RMIT University, without which this research would not be possible. We are very grateful to Mr Sam Violi, the Chairman of Strawberries Australia for all his guidance and practical assistance for this, and other projects. We are indebted to the strawberry breeders, Mr Hinga Marsh and Patricio Brevis who have assisted us, as well all the technical staff at the strawberry breeding farm at Wandin who have patiently assisted our efforts. In particular, we would like to thank Ms Karen Spence for all her assistance over the years. We also gratefully acknowledge the advice by Mr Jason Hingston, the IDO for Strawberries Australia over the course of the project.

Appendices

Appendix 1- Methods- Construction of the Strawberry-specific microarray (BS12006-Appendix1.docx)

Appendix 2- Methods- DNA Fingerprinting of 15 strawberry genotypes (BS12006-Appendix2.docx)

Appendix 3- Methods- Molecular-marker discovery (BS12006-Appendix3.docx)

Appendix 4- Results- Marker Discovery (BS12006-Appendix4.docx)

APPENDIX 1

Construction of the Strawberry-specific Microarray

1. Genomic DNA extraction and quantification

Genomic DNA was extracted from all the strawberry and non-angiosperm samples using the Qiagen™ DNeasy® Plant Mini Kit (Qiagen, Valencia, CA). Approximately 100 mg of fresh or frozen leaf tissues were ground into fine powder in a mortar and pestle with liquid nitrogen. Care was taken to keep the leaf powder frozen all the time. Subsequent steps were performed according to the manufacturer's manual. DNA intensity and integrity were determined by loading 4 µL of freshly extracted genomic DNA on 1.0 % TBE agarose gel. The gel was stained with ethidium bromide and visualised using the Molecular Imager® Gel Doc™ XR System (Bio-Rad, Hercules, CA). DNA concentration and purity was evaluated using a POLARstar Omega Microplate Reader (BMG LABTECH GmbH, Ortenberg, Germany). Briefly, 2 µL of purified DNA was loaded onto the micro-drop wells on the LVis Plate and subjected to spectrophotometric measurements. All measurements were performed in duplicate and the DNA concentration (ng/µL) as well as the 260/280 ratio were recorded.

2. Genomic DNA subtraction

Genomic DNA subtraction was performed using the PCR-Select™ cDNA Subtraction kit (Clontech, Mountain View, CA) based on Suppression Subtractive Hybridization (SSH) method with a few modifications. The workflow of SDA construction and performance validation is illustrated in Figure 1.

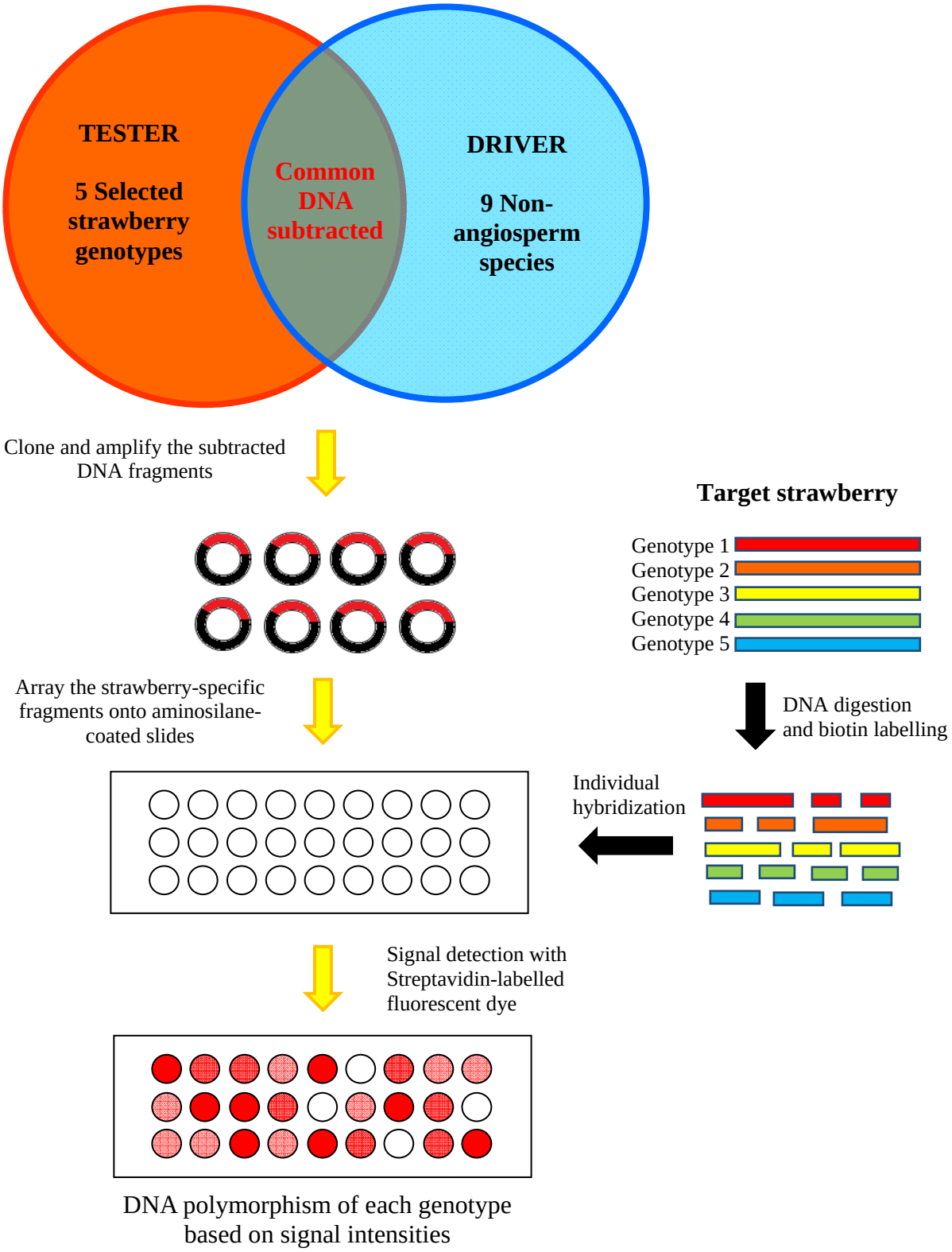


Figure 1 The workflow of strawberry-specific SDA construction and performance validation.

Prior to subtraction, genomic representations were prepared by pooling equal amount of DNA extracted from five strawberry genotypes and nine non-angiosperm species (Table 1) into tester and driver pools, respectively to a final amount of 4 µg. The pooled DNA samples were then fragmented overnight with 15 units of *Alu*I and *Hae*III restriction enzymes (NEB, Ipswich, MA) in 100 µL of digestion mixture. Purification of the digested products was achieved with phenol/chloroform extraction, divided into two portions and ligated individually with Adaptor 1 and Adaptor 2R. In addition, 0.34 ng of human skeletal muscle cDNA was added into the tester pool as a spike-in control to positively verify the efficiency of adaptor ligation.

The efficiency of adaptor ligation was performed by PCR amplification using G3PDH 3' and G3PDH 5' primers as per manufacturer's instructions and evaluated using a 2.0 % TBE agarose gel. The ligated products were then subjected to two rounds of tester-driver hybridisation. To perform the hybridisation step, excess driver was added to obtain a tester to driver ratio of 1:60. In addition, the spike-in control was removed by adding approximately 20.4 ng of human skeletal muscle cDNA into the driver pool.

Finally, subtracted DNA fragments (specific to strawberry) present in the tester but absent in the driver were selectively amplified by Suppression PCR and further enriched with nested PCR as described in the user manual. Subtraction efficiency was tested with PCR analysis using G3PDH 3' and G3PDH 5' primers and verified using a 2.0 % TBE agarose gel before ligation into T/A cloning vectors.

Table 1. Strawberry genotypes and non-angiosperm species used in the subtracted genomic library construction and array validation.

Plant Materials			**Sources**
Strawberry Genotypes	**Australian**	Adina	Wandin North
		Alinta	Wandin North
		Juliette	Coldstream
		Lowanna	Wandin North
		Melba	Coldstream
	USA	Albion	Coldstream
		Camino Real	Coldstream
		Chandler	Digger's Club
		San Andreas	Coldstream
	Japanese	Hokowase	Digger's Club
	European	Cambridge Rival	Digger's Club
	Breeding lines	07-102-41	Wandin North
		07-095-35	Wandin North
		04-069-91	Wandin North
	Wild Strawberry	Fraises Des Bois*	Digger's Club
Non-angiosperms	**Ferns**	*Dryopteris kuratae*	Bunnings Warehouse
		Dicksonia antarctica	RMIT Bundoora
		Asplenium australasicum	Bunnings Warehouse
		Blechnum tabulare	Bunnings Warehouse
	Conifers	*Wollemia nobilis*	RMIT Bundoora
		Cupressus macrocarpa	Bunnings Warehouse
		Juniperus communis	Medicinal Plant Herbarium, Southern Cross University
	Cycad	*Cycas revoluta*	Bunnings Warehouse
	Ginkgo	*Ginkgo biloba*	Digger's Club

**Fragaria vesca* (diploid)

3. Subtracted DNA fragment cloning and screening

The nested PCR products representing enriched subtracted fragments were purified using the QIAquick PCR Purification Kit (Qiagen, Valencia, CA), cloned into pGEM®-T Easy vector and transformed into *E. coli* JM109 competent cells (Promega, San Luis Obispo, CA) according to the manufacturer's instructions. The transformed cells were plated onto LB agar supplemented with 100 µg/mL ampicillin (PhytoTech, Shawnee Mission, KS), 40 µL of 0.1 M IPTG (Fermentas, Pittsburgh, PA) and 40 µL of 20 mg/mL X-GAL (Fermentas, Pittsburgh, PA). The plates were then incubated in dark at 37 °C for 16 hours.

After overnight incubation, the plates were kept at 4 °C to enhance blue colour formation until the colonies were ready for selection. A total of 331 white colonies were randomly selected and diluted in 100 µL of sterile milli-Q water in a 96-well PCR plate. The diluted colonies were boiled at 99 °C in the G-Storm GS1 thermal cycler (G-Storm ltd, Somerset, UK) for 10 min and centrifuged for 5 min at 3,000 rpm to collect the cell debris at the bottom of the plates. 1.5 µL of the clear supernatant was used as the DNA template in a 10 µL PCR mixture containing 2 µL of 5 X Green GoTaq® Flexi Buffer, 1 µL of 25 mM $MgCl_2$, 0.04 µL of each 10 µM Nested Primer 1 and Nested Primer 2R, 0.2 µL of 10 mM dNTP mix, 0.1 µL of 5 u/µL GoTaq® Flexi DNA Polymerase and 6.12 µL of sterile Milli-Q water. PCR amplification was done using the following thermal cycling conditions: initial denaturation at 94 °C for 5 min; followed by 30 cycles of denaturation at 94 °C for 30 s, annealing at 68 °C for 30 s, extension at 72 °C for 1.5 min; and a final extension at 72 °C for 5 min. The integrity and length of amplicons were determined from 1.0 % TBE agarose gel electrophoresis.

Glycerol stocks of the *E. coli* cultures were prepared to maintain the recombinant plasmids carrying the subtracted DNA fragments. 5 µL of each diluted colony was subcultured into 1 mL of LB broth containing 100 µg/mL ampicillin in a 96-well culture block. The culture block was incubated in an orbital shaker at 180 rpm and 37 °C for 18 hours. Subsequently, 50 % of the sterile glycerol was added into the bacteria culture to a final concentration of 15 % in a 96-well U-bottom plate and stored at -80 °C for further use.

2.2.2.4 Microarray probe preparation and SDA printing

Colony PCR revealed that 290 out of 331 white colonies picked from the subtracted genomic library showed a single band, ranging from 250 to 1000 bp. These subtracted DNA fragments were recovered from the *E. coli* culture glycerol stocks and re-amplified in 100 µL PCR mixture. The PCR products were precipitated overnight by adding 0.1 volume of 3M sodium acetate (pH 5.2) and 2 volumes of absolute ethanol. The resulting pellets after centrifugation were washed in 70 % ethanol, air dried and resuspended in 15 µL of 50 % DMSO printing buffer. DNA concentration of each PCR product was quantified using POLARstar Omega Microplate Reader (BMG LABTECH GmbH, Ortenberg, Germany) and adjusted to 250 ng/µL. 10 µL of each sample was transferred into a 384-well plate (Genetix, Hampshire, UK) together with nine negative controls, six positive controls, two printing controls (i.e. Cy-3 and Cy-5) and one spike-in controls, resulting in a 308-feature strawberry-specific SDA.

The positive controls including the subtracted PCR product, a housekeeping gene (beta-actin) and four strawberry-related genes (pectate lyase B, alcohol acyltransferase, alcohol dehydrogenase and sesquiterpene synthase) were PCR amplified from strawberry Albion DNA. Further, aromatase gene was included as a spike-in control to normalise systematic variation across slides. It was chosen as a spike-in control because it is derived from the

ovary of Murray River rainbowfish (*Melanotaenia fluviatilis*) (Shanthanagouda *et al*., 2012) and therefore not expected to cross-hybridise with any sequences in the strawberry genome.

The configuration of the microarray printing program was done using the TAS application suite. Two subarrays, each with six technical replicates, were printed onto Corning® GAPS™ II coated slides (Corning Incorporated, NY, USA) using a BioRobotics® MicroGrid II Compact array printing robot (Genomics Solutions, Ann Arbor, MI) at RMIT Bundoora. Each technical replicate consisted of 308 samples in a 11 x 7 format. The post-printing process was done by steaming the printed side surface for 5 seconds to rehydrate the DNA and snap-drying these slides with printed side facing up on a heating block at 100 °C for another 5 seconds. The spotted DNAs were then immobilised by UV-crosslinking for 10 minutes, baked at 80 °C for 3 hours to stabilise the interactions between the probes and aminosilane coating of the slides and stored in a clean desiccator in the dark.

4. Validation of the strawberry-specific SDA

4.1. Target sample preparation and biotin labelling

The strawberry-specific SDA validation was performed by hybridising the biotin-labelled target samples (tester and driver DNA pools) individually onto the slides. A single printed slide was hybridised with two biological replicates of each target sample, where one subarray corresponds to one biological replicate. In brief, 2 µg of DNA from each tester and driver DNA pool was digested with *Alu*I and *Hae*III and purified with QIAquick PCR Purification Kit (Qiagen, Valencia, CA). Approximately 200 ng of purified digested target DNA was labelled with Biotin-11-dUTP molecules using Biotin DecaLabel™ DNA Labeling Kit (Fermentas, Pittsburgh, PA) according to manufacturer's guidelines. The labelling reactions

were stopped by adding 1 µL of 0.5 M EDTA after 20 hours incubation at 37 °C and purified again with QIAquick PCR Purification Kit (Qiagen, Valencia, CA).

4.2. Hybridisation of biotin-labelled target DNA onto the strawberry-specific SDA

The SDA printed slides were pre-hybridised with a filtered sterilised buffer containing 1 % bovine serum albumin, 25 % formamide, 5 X SSC and 0.1 % SDS for 45 min at 42 °C. The slides were then rinsed with deionised water and dried with an air gun.

The biotin-labelled tester and driver DNA samples prepared in Section 2.2.3.1 were condensed from 50 µL to approximately 16 µL at 70 °C in the G-Storm GS1 thermal cycler. The concentrated DNA was then mixed with 17.5 µL of 2 X hybridisation buffer (5 X SSC, 0.2 % SDS, 50 % formamide); 0.5 µL of 5 µg/µL Human Cot1 DNA (Invitrogen, Carlsbad, CA), 5 µL of 10 mg/mL PolyA (Sigma-Aldrich, St. Louis, MO) and 0.5 µL of 10 mg/mL salmon sperm DNA (Sigma-Aldrich, St. Louis, MO). The DNA mixture was denatured at 100 °C for 2 minutes and immediately applied under a 22 x 25-mm lifter slip (Grale Scientific, Victoria, Australia) covering the printed areas on the slides. The slides were kept in a humidified hybridisation chamber and incubated overnight at 42 °C in a water bath. All hybridisations were performed with six technical replicates and two biological replicates, resulting in 12 data points per feature.

After hybridisation, the lifter slips were removed and the slides were washed twice in 1 x SSC and 0.1 % SDS at 45 °C for 5 min; followed by 0.1 x SSC and 0.1 % SDS at room temperature for 5 min; and a final wash of 0.1 x SSC at room temperature for 5 min. The biotinylated DNA was then labelled with FluoroLink^{TM} streptavidin-labelled Cy^{TM}5 dye (Amersham Pharmacia, Buckinghamshire, UK) and detection was performed as described by

Mantri *et al*. (2012). Firstly, the slides were washed for 5 min at room temperature in 6 x SSPE-T containing 300 mL of 20 X SSPE (175.3 g of NaCl, 27.6 g of $NaH_2PO_4.2H_2O$ and 7.4 g of EDTA), 50 µL of Triton X-100 and 700 mL of Milli-Q water. They were then immediately overlaid with 200 µL Biotin Detection Solution (200 µL 6 X SSPE-T, 0.8 µL of 25 µg/µL bovine serum albumin and 0.5 µL of 0.8µg/µL streptavidin-labelled Cy^{TM}5) under a 25 x 60-mm lifter slip and incubated at 37 °C for 1 hour. Finally, the slides were washed three times in 6 X SSPE-T, rinsed with deionised water and dried with an air gun before proceeding with scanning.

4.3. Scanning and image quantification

Slides were scanned in a ScanArray G_x Microarray Scanner (PerkinElmer, USA) and images were analysed using ScanArray Express® software (PerkinElmer, USA). The slides were first pre-scanned at 30 µm resolution with PMT gain at 70 % using the Cy-5 red laser at 633 nm. Once the printed area was identified, the slides were scanned at the higher resolution of 10 µm to improve sensitivity and a lower PMT gain at 55 % to reduce the background noise. To quantify the spots, a template consisting of the identifier for each feature was first uploaded onto the software. The grid was then manually adjusted onto the image according to the horizontal and vertical spacing between two spots. Detailed specifications are described in Appendix 5. The finely aligned images were registered to ensure correct spot recognition (Appendix 6). The signal intensities were then quantified using adaptive circle method and LOWESS normalisation. The quality and status of each feature was checked and flagged accordingly. Data filtering by manual flagging eliminated empty features (negative controls and unbound samples) and bad features (contaminated features or features with high background noise). The minimum signal-to-noise ratio (SNR) of 7 was used for quality measurement. The quantified data was exported to Microsoft Excel and only the good

features that passed all the quality control criteria were used to validate the strawberry-specific SDA.

4.4. Analysis of the strawberry-specific SDA

4.4.1. Fingerprinting of 15 strawberry genotypes

To evaluate the performance of the SDA, the DNA from the five strawberry genotypes (Albion, Juliette, 07-102-41, 07-095-35 and 04-069-91) comprising the tester pool was biotin-labelled and hybridised individually onto the array to determine their genetic relationships. In addition, the DNA from another nine genotypes (Camino Real, San Andreas, Chandler, Hokowase, Alinta, Lowanna, Melba, Adina and Cambridge Rival) and one woodland strawberry (Fraises Des Bois) were also hybridised onto the SDA to determine the ability of the SDA in fingerprinting strawberry genotypes not used in the initial library construction.

The hybridisation, detection, scanning and image quantification steps were performed as described in Section 2.2.3, except the target DNA labelling, which was modified as follows: approximately 400 ng of digested and purified target DNA was labelled with Biotin-11-dUTP molecules and divided into two portions for hybridisation. This modification prevented the biotin-labelled samples from being condensed in the PCR machine, which could introduce artefacts into the samples.

2.2.4.2 Data analysis

Raw data was obtained from the Excel sheet based on the method described in Section 2.2.3.3. The SNR of each feature was used for all further statistical analyses because it was considered to have the most accurate background correction. It is defined as the (Mean

Foreground - Mean Background) / (Standard deviation of Background) in the ScanArray Express® Microarray Analysis System User Manual (PerkinElmer, USA). All features were normalised between slides using the total intensity normalisation method as described by Olarte (2011). Data normalisation was carried out based on the following steps and summarised in Figure 2.2:

1. **Mean SNR across the technical replicates**

   The mean SNR of six technical replicates was calculated for each of the 290 spots in one single biological replicate.

2. **Normalisation factor**

   Normalisation factor was obtained using the equation below

   Normalisation factor = $\frac{A}{B}$

   where A = Mean SNR of all the features in all the six technical replicates and two biological replicates across 15 genotypes

   B = Mean SNR of all the features in all the six technical replicates for one genotype

3. **Data normalisation**

   Data normalisation was achieved by multiplying mean SNR of each feature with the normalisation factor in a single biological replicate.

4. **Mean SNR between biological replicates**

Mean of the two biological replicates was calculated using the normalised SNR for each of the 290 features, resulting in a fingerprint comprising of one value per feature per genotype.

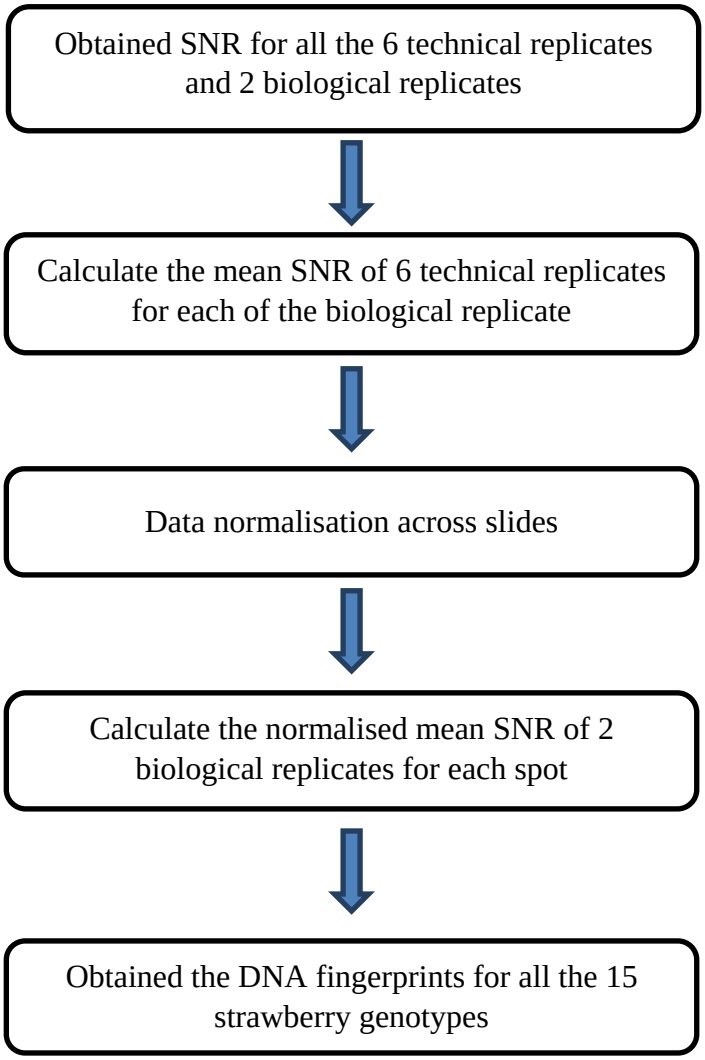


Figure 2.2 The data normalisation procedures performed to the raw SNR of each feature after image scanning and spot quantification.

2.2.4.3 Statistical analysis

After data normalisation, a series of statistical analyses were performed:

1. **Hierarchical clustering**

   Hierarchical cluster analysis was performed using the normalised mean SNR of the entire dataset (excluding spots that hybridised with the driver sample) to elucidate the genetic relationships of the 15 strawberry genotypes under study. The normalised mean values of each good feature were used as variables to construct a dissimilarity dendrogram with IBM SPSS Statistics v. 21 using the average-linkage-between-groups method and squared Euclidean distance.

2. **Principal component analysis (PCA)**

   The same dataset was analysed with PCA in Minitab v. 16 to identify the features that reveal maximum variability between the strawberry genotypes assessed. The 15 strawberry genotypes were used as variables in contrast to the hierarchical cluster analysis. A PCA score plot showing the proportion of variance explained by the first two components was obtained to identify the features that accounted for most of the variability found across the individual strawberry genotypes.

3. **Magnitude of variance**

   The magnitude of variance of the normalised mean SNR for each feature across the 15 strawberry genotypes was calculated to determine features with the highest variances between genotypes. This analysis was performed to identify useful features which were not detected by PCA. *k*-means clustering was used to partition the features into three clusters based on the high, intermediate and low variance values.

4. **Pearson's bivariate correlation**

   The features selected by PCA and the magnitude of variance were subjected to Pearson's bivariate correlation in IBM SPSS Statistics v. 21 to further reduce the

number of potential features containing DNA specific to each genotype and/or genotypic group. This analysis eliminated the features with similar hybridisation patterns due to redundancy of the subtracted library.

5. Branch point DNA marker identification

Finally, a set of features selected by PCA, magnitude of variance and Pearson's bivariate correlation which showed the highest variances across all genotypes tested was used to reconstruct a dissimilarity dendrogram. This set of features was manually placed on the dendrogram to identify their corresponding branch points based on their hybridisation patterns.

APPENDIX 2

Fingerprinting of 15 Strawberry Genotypes

To evaluate the performance of the SDA, the DNA from the five strawberry genotypes (Albion, Juliette, 07-102-41, 07-095-35 and 04-069-91) comprising the tester pool was biotin-labelled and hybridised individually onto the array to determine their genetic relationships. In addition, the DNA from another nine genotypes (Camino Real, San Andreas, Chandler, Hokowase, Alinta, Lowanna, Melba, Adina and Cambridge Rival) and one woodland strawberry (Fraises Des Bois) were also hybridised onto the SDA to determine the ability of the SDA in fingerprinting strawberry genotypes not used in the initial library construction.

The hybridisation, detection, scanning and image quantification steps were performed as described in Appendix 1, except the target DNA labelling, which was modified as follows: approximately 400 ng of digested and purified target DNA was labelled with Biotin-11-dUTP molecules and divided into two portions for hybridisation. This modification prevented the biotin-labelled samples from being condensed in the PCR machine, which could introduce artefacts into the samples.

2.2.4.2 Data analysis

Raw data was obtained from the Excel sheet based on the method described in Section 2.2.3.3. The SNR of each feature was used for all further statistical analyses because it was considered to have the most accurate background correction. It is defined as the (Mean Foreground - Mean Background) / (Standard deviation of Background) in the ScanArray

Express® Microarray Analysis System User Manual (PerkinElmer, USA). All features were normalised between slides using the total intensity normalisation method as described by Olarte (2011). Data normalisation was carried out based on the following steps and summarised in Figure 1.

1. **Mean SNR across the technical replicates**

   The mean SNR of six technical replicates was calculated for each of the 290 spots in one single biological replicate.

2. **Normalisation factor**

   Normalisation factor was obtained using the equation below

   Normalisation factor = $\frac{A}{B}$

   where A = Mean SNR of all the features in all the six technical replicates and two biological replicates across 15 genotypes

   B = Mean SNR of all the features in all the six technical replicates for one genotype

3. **Data normalisation**

   Data normalisation was achieved by multiplying mean SNR of each feature with the normalisation factor in a single biological replicate.

4. **Mean SNR between biological replicates**

Mean of the two biological replicates was calculated using the normalised SNR for each of the 290 features, resulting in a fingerprint comprising of one value per feature per genotype.

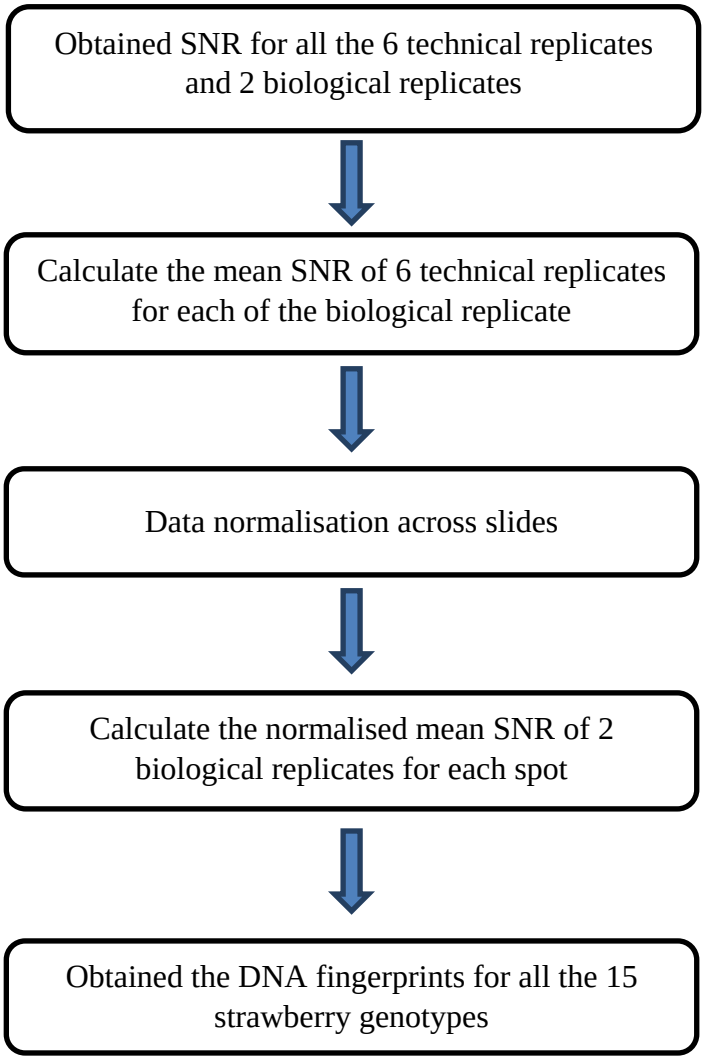


Figure 1. The data normalisation procedures performed to the raw SNR of each feature after image scanning and spot quantification.

Statistical analysis

After data normalisation, a series of statistical analyses were performed:

1. **Hierarchical clustering**

Hierarchical cluster analysis was performed using the normalised mean SNR of the entire dataset (excluding spots that hybridised with the driver sample) to elucidate the genetic relationships of the 15 strawberry genotypes under study. The normalised mean values of each good feature were used as variables to construct a dissimilarity dendrogram with IBM SPSS Statistics v. 21 using the average-linkage-between-groups method and squared Euclidean distance.

2. **Principal component analysis (PCA)**

The same dataset was analysed with PCA in Minitab v. 16 to identify the features that reveal maximum variability between the strawberry genotypes assessed. The 15 strawberry genotypes were used as variables in contrast to the hierarchical cluster analysis. A PCA score plot showing the proportion of variance explained by the first two components was obtained to identify the features that accounted for most of the variability found across the individual strawberry genotypes.

3. **Magnitude of variance**

The magnitude of variance of the normalised mean SNR for each feature across the 15 strawberry genotypes was calculated to determine features with the highest variances between genotypes. This analysis was performed to identify useful features which were not detected by PCA. *k*-means clustering was used to partition the features into three clusters based on the high, intermediate and low variance values.

4. **Pearson's bivariate correlation**

The features selected by PCA and the magnitude of variance were subjected to Pearson's bivariate correlation in IBM SPSS Statistics v. 21 to further reduce the number of potential features containing DNA specific to each genotype and/or

genotypic group. This analysis eliminated the features with similar hybridisation patterns due to redundancy of the subtracted library.

5. **Branch point DNA marker identification**

Finally, a set of features selected by PCA, magnitude of variance and Pearson's bivariate correlation which showed the highest variances across all genotypes tested was used to reconstruct a dissimilarity dendrogram. This set of features was manually placed on the dendrogram to identify their corresponding branch points based on their hybridisation patterns.

APPENDIX 3

Molecular-Marker Discovery

1. Parental genotypes and segregating population

A segregating population from a cross between an Australian cultivar 'Juliette' and '07-102-41' breeding line was chosen as the experimental material based on their different aroma profiles as described in Appendix 1. Juliette, a short day cultivar bred by the Victorian Department of Primary Industries (DPI) is a bright red strawberry which fruits in the early season (September) in Victoria, Australia. This cultivar produces fruits that are sweeter than other cultivars developed through the same breeding program (Victorian Strawberry Industry 2013). 07-102-41, a breeding line which contains genetic background from the European strawberry is also a short day strawberry but has dark red and very flavoursome fruits (data not shown). Fully ripe fruits and young leaves of 50 individual progeny plants from the F_1 population were collected over the summer of 2011/2012 according to the method described in Section 2.2.1 (page 53). Of these, 37 progeny plants had 07-102-41 as a maternal parent (07-102-41 x Juliette) and the remaining 13 progeny plants were collected from the reciprocal cross (Juliette x 07-102-41) where Juliette was used as the maternal parent.

2. Phenotyping of strawberry flavour

Sample preparation, aroma profiling and data analysis for the 50 F_1 progeny plants were performed as described in Appendix 1. Target compounds for DNA marker development were chosen based on the parental aroma profiles as described in Appendix 1. Application of these selection criteria resulted in the selection of eight compounds of interest including four esters (methyl butanoate, methyl hexanoate, ethyl butanoate and ethyl hexanoate), one

furanone (mesifuranne), two terpenes (linalool and (*E*)-nerolidol) and one lactone (γ-dodecalactone). Frequency distributions of the 50 F_1 progeny plants along with their parental means were generated (Microsoft Excel) to determine the plants showing the extremes of phenotype for the selected compounds. The number of plants with extreme phenotypes identified from the segregation patterns was selected for subsequent BSA.

3. Bulked Segregant Analysis (BSA)

3.1. Generation of DNA bulks with extreme phenotypes

The mean relative compositions of the volatile compounds were subjected to *k*-mean clustering analysis to confirm which individual progeny plants that fall into the two extreme clusters as identified from the frequency distribution mentioned above. Equal amounts of DNA from F_1 progeny plants showing high (H) or undetectable (L) levels of key volatile compounds were bulked into the respective 'H' and 'L' to a final quantity of 2 µg. The number of individuals in each bulk ranged from 3 to 27 plants depending on the key volatile compounds. Total genomic DNA for BSA was isolated from the leaves of individual plants using Qiagen™ DNeasy® Plant Mini Kit (Qiagen, Valencia, CA). DNA isolation and quantification were performed as described in Appendix 1.

3.2. SDA hybridization, scanning and image quantification

16 DNA bulks corresponding to the high and low extremes of eight key volatile compounds were individually digested with *Alu*I and *Hae*III restriction enzymes and purified using QIAquick PCR Purification Kit (Qiagen, Valencia, CA). Subsequently, approximately 400 ng from each DNA bulk was labelled with Biotin-11-dUTP molecules using Biotin DecaLabel™ DNA Labeling Kit (Fermentas, Pittsburgh, PA). SDA hybridisation, scanning and image quantification were performed according to the protocol described in Appendix 1 except in

this study the data normalisation protocol was modified using the spike-in control normalisation method instead of the total intensity normalisation method as described in Chapter 2. This is because a relatively small dataset is being compared between two DNA bulks with extreme phenotypes (287 features x 2 biological replicates x 6 technical replicates x 2 DNA bulks = 6888 features). In addition, the signal intensities of the features may be biased towards one of the two DNA bulks due to the extreme phenotypes observed. It has been shown that the application of spike-in control normalisation method is superior for low-density microarrays and it is necessary when the distribution of gene expression is asymmetric and biased towards up-regulated genes.

The signal-to-noise ratio (SNR) of all features for the two DNA bulks with extreme phenotypes were normalised against the mean SNR of a spike-in control, the aromatase gene obtained from Murray River rainbow fish (*Melanotaenia fluviatilis*) (Shanthanagouda *et al*., 2012). Using the spike-in control method, data normalisation between technical replicates as well as between hybridisations can be achieved. Data normalisation was performed as follows:

1. **Normalisation factor**

   Normalisation factor was obtained from SNR of spike-in control using the equation below:

   Normalisation factor = $\frac{A}{B}$

   where A = Mean SNR of the spike-in control in all the six technical replicates and two biological replicates for two phenotypic extreme bulks

   B = SNR of the spike-in control for a respective technical replicate

2. **Data normalisation**

   Data normalisation was achieved by multiplying the SNR for each of the 287 features with the normalisation factor for every technical replicate.

3. **Mean SNR between technical replicates**

   Mean of the normalised SNR for each feature was calculated from the six technical replicates.

4. **Mean SNR between biological replicates**

   Data for the two biological replicates were combined to obtain mean normalised SNR for each of the 287 features, resulting in a fingerprint comprising one value per feature per phenotype.

4. Statistical Analysis

4.1. Discriminant Function Analysis (DFA)

Discriminant Function Analysis (DFA) was performed for the eight key volatile compounds to identify a set of variables (i.e., the features) that best discriminate between the two groups (i.e., the 'H' and 'L' DNA bulks corresponding to a particular key volatile compound). The DFA was then employed to predict whether the selected set of features (also known as predictors) could be used to classify new cases (i.e., a new SDA dataset) into either the 'H' or 'L' phenotypic groups.

SDA data obtained from the 'H' and 'L' DNA bulks of a key volatile compound were subjected to DFA using the stepwise method in Discriminant Analysis (IBM SPSS Statistics v. 21) to select the markers that best differentiate between 'H' and 'L' phenotypic groups.

Grouping variables were (1) the 'H' DNA bulk; and (2) the 'L' DNA bulk. The independent variables were the normalised mean SNR of 287 features (excluding three features that hybridised with driver DNA) from the six technical replicates derived from the 'H' and 'L' DNA bulks. In this study, the six technical replicates (original cases) from the first biological replicate of a given phenotypic group were assigned as a training set (1) to predict the membership of the other six technical replicates (the new cases) from the second biological replicate, which is the test set (2). Similarly, a reciprocal analysis was performed by using the second biological replicate as the training set and the first biological replicate as the test set.

The best subset of variables/features were selected from the training set using Wilks' lambda method and the selection criteria using *F* probability values was set to default (Entry = 0.05, Removal = 0.10). Subsequently, the selected features were employed to construct and validate a discriminant function for each group using Fisher's classification function coefficients (Statistics → Function Coefficients → Fisher's). The efficiency of the discriminant functions in predicting group membership for any given case in the training set can be determined based on the prior probabilities of the case (Classify → Prior probabilities → All groups equal). The group assigned to each case by the discriminant functions were obtained by selecting 'Predicted group membership' box in the 'Save' tab. To predict the group membership for new cases (test set) using the discriminant function generated from the training set, value '1' was entered into the 'Selection Variable' box. The results are presented as casewise statistics and the percentage of the original and new cases being correctly classified (Display → Casewise results → Summary table). Another parameter considered was the means of the independent variables for each group (Statistics → Descriptives → Means).

Several outputs were generated including (1) group statistics (mean and standard deviation for each independent variable for both groups); (2) stepwise statistics; (3) summary of canonical discriminant functions and (4) classification statistics. Data interpretation was focused on output 2, 3 and 4. The selected features were shown in stepwise statistics, where the features were entered or removed based on *F* probability values. Based on the coefficient values for each selected feature, a set of linear combinations of features (also known as discriminant function) that best separate the extreme groups was generated for each volatile compound assessed. The total variance explained by the selected features was shown in the summary of canonical discriminant functions whereas the significance of the discriminant function generated was indicated by Wilk's lambda. Classification statistics showed the predicted group membership calculated based on the classification function coefficients. Both original and new cases were assigned to the group with the highest value. The proportion of correct classification was determined from the number of misclassified cases.

DFA analysis with the stepwise method was further used to reduce the number of markers selected based on all the twelve technical replicates of the selected features without assigning them into training or test set. The same parameters were used for this analysis except that the full dataset was used to calculate the discriminant function. Classification of cases was performed by selecting the '*leave-one-out*' option under the 'Classify' tab. Each case was cross-validated using the discriminant function calculated from all cases except the one being classified. The probability of misclassification was calculated from these results. The predicted group memberships for both original and cross-validated group cases were reported as the percentage of correct classification. Based on the accuracy of the predicted group membership, a set of features were determined as the putative DNA markers that could best

predict whether a strawberry plant will bear fruits with either high or undetectable levels of a specific key volatile compound.

4.2. Fisher's ratio

Fisher's ratio was employed to measure the linear discriminating power of the 287 features between the 'H' and 'L' DNA bulks for the eight key volatile compounds. It is defined as the magnitude of the mean differences in signal intensity (mean of the normalised SNR) between two extreme bulks as a proportion to the sum of the variances measured in the two extreme bulks, which is the background noise generated by the hybridisation experiment inherent in the microarray system (Lohninger 1999):

$$\text{Fisher's ratio} = \frac{(M_1 - M_2)^2}{(V_1 + V_2)}$$

where M_1 = Mean of the normalised SNR for each feature in the 'H' DNA bulk

M_2 = Mean of the normalised SNR for each feature in the 'L' DNA bulk

V_1 = Variance of the normalised SNR for each feature in the 'H' DNA bulk

V_2 = Variance of the normalised SNR for each feature in the 'L' DNA bulk

The features were arranged based on the descending value of Fisher's ratio. The features demonstrating the top ten highest Fisher's ratio values were arranged in descending order and compared to the features selected by DFA.

4.3. Independent Samples *t*-Test

Comparison between the features selected by DFA and Fisher's ratio revealed that not all the features selected by DFA displayed high Fisher's ratio values, indicating low discriminating power for some of the DFA-selected features. In order to eliminate the irrelevant features, the

normalised mean SNR of features selected by DFA were subjected to Independent Samples *t*-Test (IBM SPSS Statistics v. 21). The six technical replicates and the two biological replicates of the 'H' and 'L' DNA bulks were assigned as variable 1 and 2, respectively. Only the features showing significant differences between the 'H' and 'L' DNA bulks were retained for further analysis. The features which fulfilled all three criteria (DFA, Fisher's ratio and Independent Samples *t*-Test) were selected by generating a three-way Venn diagram and sent for DNA sequencing.

4.4. DNA sequencing of selected features

The *E. coli* cultures containing cloned subtracted DNA fragments corresponding to the putative DNA markers were recovered from their respective glycerol stocks. Preparation of PCR products for DNA sequencing was performed according to Section 2.2.7 (page 72). PCR products were sequenced bi-directionally using the T7 and SP6 primers by Macrogen Inc. (Korea). DNA sequence processing and similarity search were performed as described in Section 2.2.7 (page 72). Sequence identity of each feature was confirmed with the strawberry draft genome (v1.1) using the PFR Strawberry Server at https://strawberry.plantandfood.co.nz/ (PFR, 2010) and confirmed with the Genome Database for Rosaceae at http://www.rosaceae.org/tools/ncbi_blast (GDR, 2009). Genes located within 5 cM on either side of the nuclear-specific features were searched manually on the same linkage group using the PFR Strawberry Server.

4.5. Putative DNA marker validation

4.5.1. Primer design and PCR amplification

DNA sequences showing significant similarity (E-value $<1e^{-5}$) to *Fragaria vesca* nuclear sequences were chosen for primer design as they are most likely to be linked to the loci

controlling the synthesis of key volatile compounds in strawberry. Forward and reverse primers specific to the DNA sequences were designed using Clone Manager Suite v. 7.1 (Sci-Ed Software, Durham, NC). The designed primers were then synthesised by GeneWorks Pty Ltd, Hindmarsh, SA.

To determine fragment size variation between the 'H' and 'L' DNA bulks, PCR amplification was performed on both the parental genotypes (Juliette and 07-102-41) and the individual progeny from the two DNA bulks showing extreme phenotypes using GoTaq® DNA polymerase (Promega, San Luis Obispo, CA). Briefly, 1.0 µL of genomic DNA (~ 50 ng) isolated from Juliette and 07-102-41 was used as DNA template in a 25 µL PCR reaction containing 5 µL of 5 X Green GoTaq® Flexi Buffer, 1.5 µL of 25 mM $MgCl_2$, 0.5 µL of each 10 µM sequence-specific forward and reverse primer, 0.5 µL of 10 mM dNTP mix, 0.5 µL of 5 u/µL GoTaq® Flexi DNA Polymerase and 15.5 µL of sterile Milli-Q water. PCR amplification was done in the G-Storm GS1 thermal cycler (G-Storm ltd, Somerset, UK) using the following thermal cycling conditions: initial denaturation at 94 °C for 3 min; followed by 30 cycles of denaturation at 94 °C for 30 s, annealing at 55 °C for 30 s, extension at 72 °C for 1 min; and a final extension at 72 °C for 10 min. The integrity of the PCR products and any size variation of the DNA fragments were determined using 2.0 % TBE agarose gel electrophoresis.

Subsequently, PCR amplification was performed on the parental genotypes using AccuPrime™ *Pfx* DNA Polymerase (Invitrogen, NY, USA) to detect any SNPs or indels in the DNA sequences between the bulks. AccuPrime™ *Pfx* DNA Polymerase (Invitrogen, NY, USA) is a proofreading DNA polymerase used to improve accuracy and prevent mispriming. For the purpose of DNA sequencing, the volume of PCR mixture was increased to 50 µL

containing 5 µL of 10 X AccuPrime™ *Pfx* mix, 0.2 µL of each 10 µM sequence-specific forward and reverse primer, 1 µL of DNA template (~50 ng), 0.4 µL of 2.5 u/µL AccuPrime™ *Pfx* DNA Polymerase and 43.2 µL of Milli-Q water. The thermal cycling parameters for the proofreading DNA polymerase were modified as follows: initial denaturation at 95 °C for 2 min; followed by 30 cycles of denaturation at 95 °C for 15 s, annealing at 55 °C for 30 s, extension at 68 °C for 1 min. No final extension was required. The integrity and length of PCR products were examined on a 1.0 % TBE agarose gel electrophoresis. PCR product purification was performed using Qiaquick PCR Purification Kit or Qiaquick Gel Extraction Kit (Qiagen, Valencia, CA) depending on the specificity of the primers.

4.5.2. Sequence alignment and determination of DNA polymorphism

All the PCR products were sequenced by Australian Genome Research Facility Ltd. (AGRF) using the sequence-specific forward and reverse primers. The forward and reverse DNA sequences were aligned for each parental genotype using the Clustal Omega Multiple Sequence Alignment function at https://www.ebi.ac.uk/Tools/msa/clustalo/ (EMBL-EBI 2014). Sequence editing was then performed by trimming the DNA regions at the beginning and the end of the sequence reads. The consensus DNA sequences between the parental genotypes were obtained. Microstructural DNA variations such as SNPs or indels within the marker were determined.

APPENDIX 4

Flavour Phenotypes for the F1 Progeny between Juliette and 07-102-41

Table 4.1 Descriptive statistics for 23 volatile compounds analysed in the F_1 population and in the parental genotypes '07-102-41' and 'Juliette'.

Volatile compounds	F_1 population			07-102-41		Juliette		*t*-Test (for parents)		
	Mean[a]	Min[b]	Max[b]	Mean[c]	SD[c]	Mean[c]	SD[c]	*t* value[d]	*df*	*p*[d]
Esters										
Methyl butanote	9.4	0.0	34.3	15.4	6.1	0.0	0.0	4.36	2.0	*
Methyl 3-methylbutanote	0.3	0.0	2.7	0.0	0.0	0.0	0.0	-	-	-
Ethyl butanote	4.8	0.0	23.2	7.2	1.9	3.2	1.4	3.38	5	*
Isopropyl butanote	0.5	0.0	3.1	0.0	0.0	0.0	0.0	-	-	-
Ethyl 2-methylbutanote	0.2	0.0	2.0	0.0	0.0	1.5	0.1	-12.49	5	**
Ethyl isovalerate	0.7	0.0	14.5	0.0	0.0	1.8	0.1	-10.48	5	**
Isoamyl acetate	0.8	0.0	7.3	0.0	0.0	0.0	0.0	-	-	-
Methyl hexanoate	14.1	0.0	48.1	25.0	1.6	0.0	0.0	26.98	2.0	**

Butyl butanoate	0.25	0.0	1.9	3.7	1.2	0.0	0.0	5.57	2.0	*
Ethyl hexanoate	6.9	0.0	35.6	16.2	1.3	10.1	1.2	6.05	5	**
(*Z*)-Hex-3-enyl acetate	0.7	0.0	5.0	0.0	0.0	0.7	1.0	-42.51	5	**
Hexyl acetate	2.1	0.0	35.8	0.0	0.0	1.5	0.3	-2.41	5	ns
(E)-Hex-2-enyl acetate	2.5	0.0	14.8	0.0	0.0	2.5	2.7	-1.73	3.0	ns
Isopropyl hexanoate	0.2	0.0	1.2	0.0	0.0	0.0	0.0	-	-	-
Benzyl acetate	1.4	0.0	23.7	0.0	0.0	0.0	0.0	-	-	-
Hexyl butanoate	0.4	0.0	9.2	0.0	0.0	0.0	0.0	-	-	-

Table .1 continued on next page

Table 4.1 (continued)

Volatile compounds	F_1 population			07-102-41		Juliette		*t*-Test (for parents)		
	Mean[a]	Min[b]	Max[b]	Mean[c]	SD[c]	Mean[c]	SD[c]	*t* value[d]	*df*	*p*[d]
Aldehydes										
Hexanal	1.0	0.0	15.2	0.0	0.0	0.0	0.0	-	-	-
(*E*)-Hex-2-enal	7.5	0.0	63.7	0.0	0.0	0.0	0.0	-	-	-
Furanone										
Mesifuranne	3.2	0.0	32.4	0.0	0.0	5.7	3.1	-4.43	3.0	*
Terpenes										
Linalool	3.3	0.0	18.3	9.1	0.8	6.2	1.1	9.95	5	*
(*E*)-Nerolidol	7.7	0.0	36.9	17.4	4.1	60.6	6.4	-11.23	5	**
α-Terpineol	0.9	0.0	10.0	0.0	0.0	0.0	0.0	-	-	-
Lactones										
γ-Dodecalactone	1.4	0.0	7.9	0.0	0.0	4.8	1.8	-6.55	3.0	**

All values are normalised peak areas in relative composition (%)

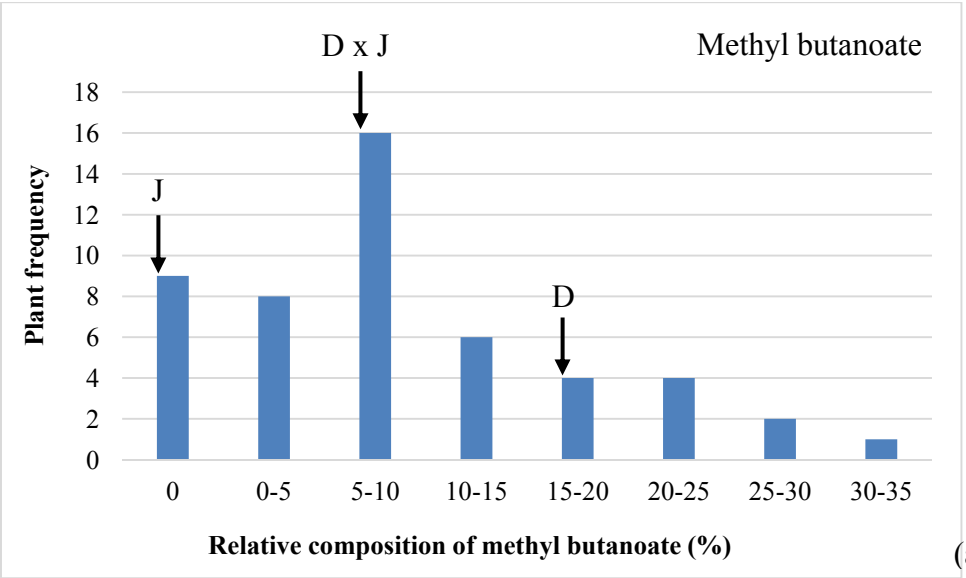
[a] Mean of all analysed F_1 individuals, based on three technical replicates per genotype

[b] F_1 individuals with the lowest (Min) and highest (Max) relative composition, mean from three technical replicates

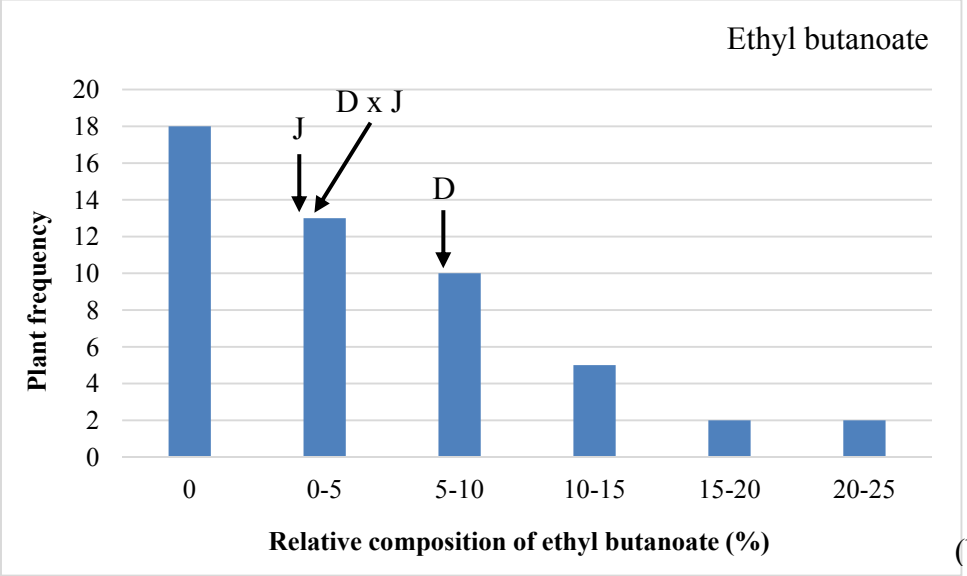
[c] Mean and standard deviation (SD) from all technical replicates

[d] *t*-Test between the parents '07-102-41' and 'Juliette'

ns: not significant ($p > 0.05$); * Significant at $0.05 \geq p > 0.01$; ** Significant at $0.01 \geq p > 0.001$



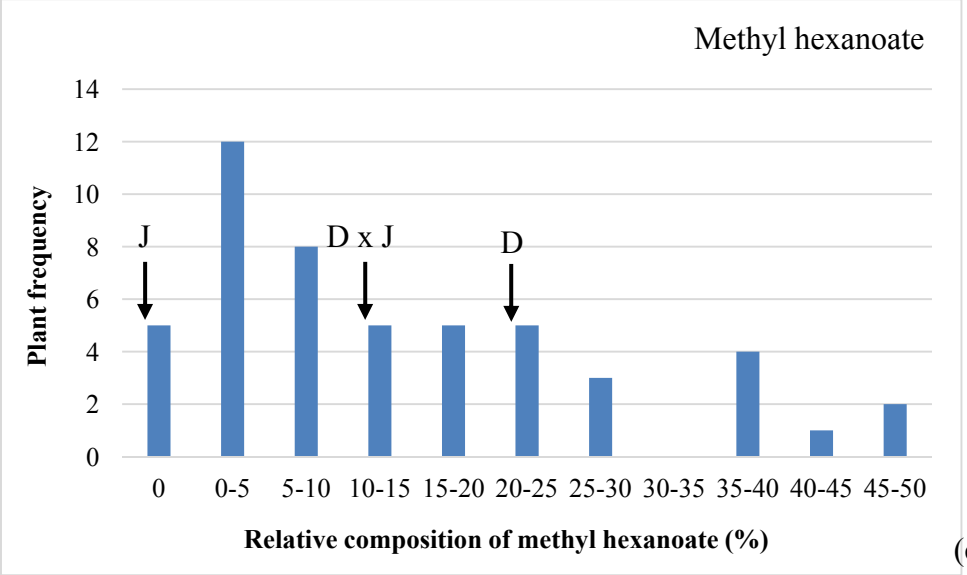
(a)



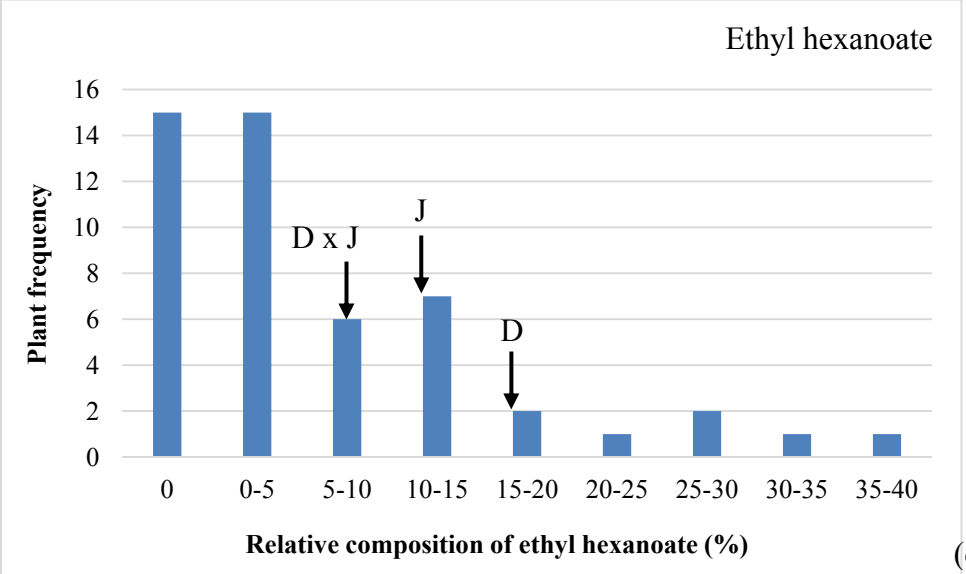
(b)

Figure 4.1 Frequency distribution of key volatile compounds measured as relative peak areas in the '07-102-41' x 'Juliette' progeny. The mean values of the parents and F_1 population are indicated by arrows (D: 07-102-41; J: Juliette; D x J, respectively). (a) methyl butanoate, (b) ethyl butanoate, (c) methyl hexanoate, (d) ethyl hexanoate, (e) mesifuranne, (f) linalool, (g) nerolidol and (h) gamma-dodecalactone. *x*-axis: relative composition (%), *y*-axis: plant frequency.

Figure 4.1 continued on next page

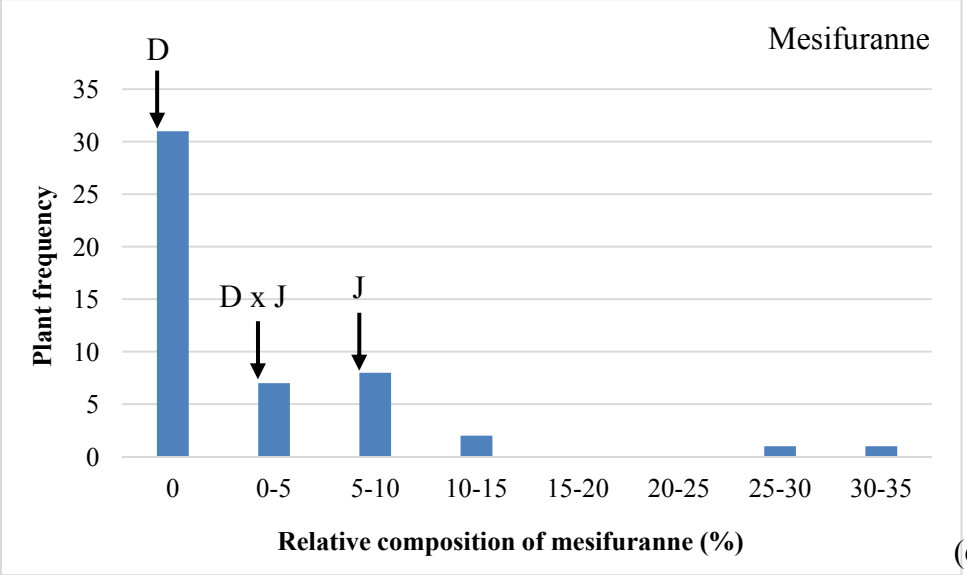


(c)

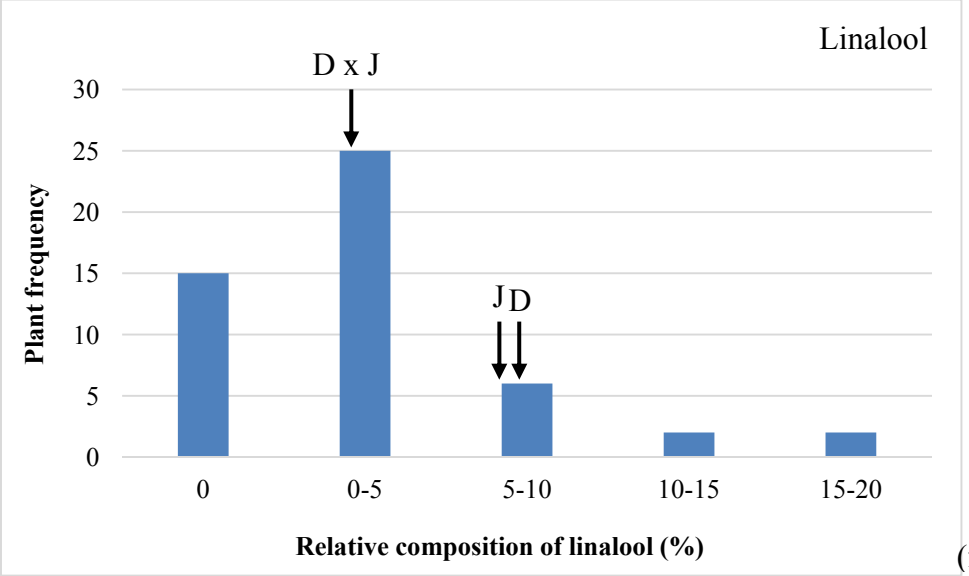


(d)

Figure 4.1 (continued)

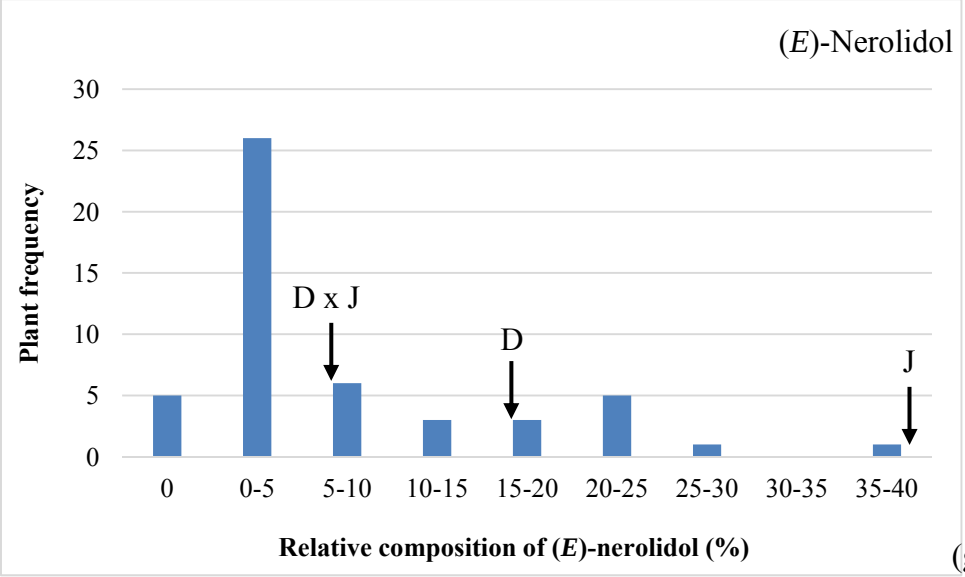


(e)

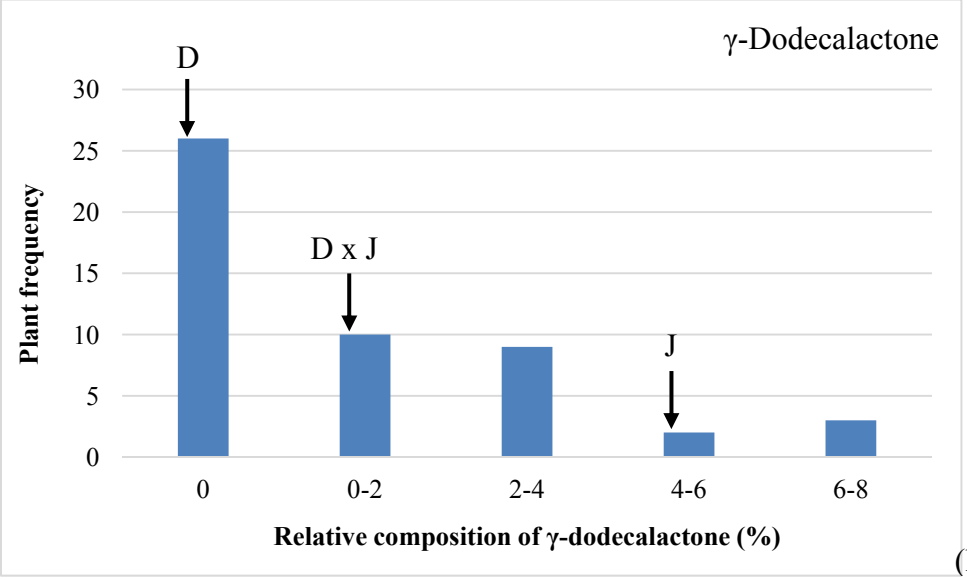


(f)

Figure 4.1 (continued)



(g)



(h)

Figure 4.1 (continued)