Identification of molecular markers associated with fruit flavour and day-neutrality in strawberry (*Fragaria x ananassa*) using a microarray-based approach

A thesis submitted in total fulfilment of the requirements for the degree of Doctor of Philosophy

Mian Chee Gor
Master of Science (Plant Biotechnology)

School of Applied Sciences
College of Science, Engineering and Health
RMIT University
August 2014
Declaration

I certify that except where due acknowledgement has been made, the work is that of the author alone; the work has not been submitted previously, in whole or in part, to qualify for any other academic award; the content of the thesis is the result of work which has been carried out since the official commencement date of approved research program; any editorial work, paid or unpaid, carried out by a third party is acknowledged; and, ethics procedures and guidelines have been followed.

Mian Chee Gor

08/2014
I wish to express my greatest gratitude to my main supervisor, Professor Eddie Pang for his valuable guidance and advice throughout my PhD candidature. I would also like to thank my second supervisor Dr Nitin Mantri for his mentoring and support throughout the years.

I gratefully acknowledge School of Applied Sciences, RMIT University for awarding me with the RMIT International Research Scholarships and Horticulture Australia Limited for funding this project. The completion of this project would not be possible without their financial support.

I would also like to extend my sincere gratitude to Mr Sam Violi, Chair of Strawberry Australia and all the breeders at the research station in Wandin North for sharing their knowledge and assisting in experimental site maintenance. I am thankful to Dr Alexandra Olarte who helped me through many of the microarray experiments, and to Ms Chrishani Candappa and Ms Thish de Silva for their technical assistance in running the GC-MS experiments. I also wish to thank Dr Emily Gan and all my friends who helped proofread my thesis.

Thanks to all the staffs in the Department of Biotechnology and Environmental Biology for their assistance throughout my candidature and to all my lab mates and friends especially Aida Golneshin, Monica Mu and Aya Taki who cheered me up whenever I faced difficulties. Finally, special thanks to my parents and family members for their moral support throughout my study.
Abstract

Breeding strawberry (*Fragaria x ananassa*) with enhanced fruit flavour and day-neutrality traits is one of the top priorities of the Australian Strawberry Breeding Program. Although several genes involved in the biosynthetic pathways of key volatile compounds and molecular control of flowering response in strawberry have been identified, the development and application of molecular markers associated with these traits remain limited. This can be attributed to the complex genetic control and environmental influences on both these traits. The current research aims to identify molecular markers closely linked to genes controlling strawberry flavour and day-neutrality by combining the Subtracted Diversity Array (SDA) and the Bulked Segregant Analysis (BSA) approaches.

To develop a platform for molecular marker identification, a SDA was constructed using a broad subtraction approach between five strawberry genotypes and nine non-angiosperm species. Validation of the strawberry-specific SDA was performed using the tester and driver pools. Its efficacy was evaluated by fingerprinting 15 strawberry genotypes, as well as correlating the aroma profiles and genetic data of three commercial cultivars and two breeding lines. In addition, to identify molecular markers associated with fruit flavour, 50 F<sub>1</sub> progeny derived from 07-102-41 x Juliette cross were selected for BSA based on the extreme levels of key volatile compounds assessed. For the discovery of molecular markers associated with day-neutrality, the F<sub>1</sub> progeny derived from a day-neutral (DN) x short day (SD) and two DN x DN crosses were selected for BSA according to the extremes of flowering response. In both experiments, the extreme DNA bulks were hybridised individually onto the strawberry specific-SDA. Differential hybridisation patterns between the DNA bulks were analysed using Discriminant Function Analysis (DFA), Fisher’s ratio...
and Independent Samples t-Test to identify putative markers associated with fruit flavour and day-neutrality.

Subtracted genomic library construction resulted in a SDA containing 287 features specific to the strawberry genome. Hierarchical clustering demonstrated that the strawberry-specific SDA was able to correctly cluster 15 strawberry genotypes based on known pedigree information. The aroma profiles and the SDA 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 SDA may be capable of detecting markers associated with aroma compound production. The SDA-BSA approach successfully identified a set of the most predictive features that best discriminated between the extreme DNA bulks. DNA sequence analyses showed that FLP1D7, FLP1A7 and FLP3E12 features were linked to genes possibly involved in the biosynthesis of methyl butanoate, γ-dodecalactone and linalool. Moreover, a putative C/T SNP was observed for FLP1D7 feature, suggesting the presence of allelic variants in this locus.

In addition, a putative DNA marker, FLP2E11, was found to be closely linked to a cytokinin oxidase (CKXI) gene possibly involved in promoting flowering under non-inductive condition. A 12 bp deletion was observed in the FLP2E11 sequence cloned from the SD but not the DN genotypes. These results indicate that allelic variants are present in this locus. It is hypothesised that these FLP2E11 allelic variants are linked to the wild type and the low activity CKXI alleles, and that flowering strength may be controlled by the copy number of these CKXI alleles. A hypothetical model is proposed to
demonstrate the mechanism of flowering control regulated by CKX1 based on the results obtained from this study and previously documented cytokinin-induced flowering response in strawberry.

Recommendations for further study based on the DNA polymorphism found within FLP1D7 and FLP2E11 marker alleles are described. This is the first report on the application of SDA-BSA for the identification of molecular markers potentially associated with important genes controlling fruit flavour and day-neutrality in strawberry. This robust platform should prove useful for future marker-trait association studies in strawberry.
Thesis publications

Refereed Conference


Manuscript in preparation


Mian-Chee, G., Chrishani, C., Patabadi Maddumage Thishakya, dS., Nitin, M. & Edwin, P. 2014. Integrating Subtracted Diversity Array (SDA) and Bulked Segregant Analysis (BSA) to identify molecular markers associated with flavour trait in strawberry. (Publication type: Research article; Target journal: Theoretical and Applied Genetics; Impact factor: 3.658; Target size: 10 -15 pages)

Table of contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Declaration</td>
<td>ii</td>
</tr>
<tr>
<td>Acknowledgements</td>
<td>iii</td>
</tr>
<tr>
<td>Abstract</td>
<td>iv</td>
</tr>
<tr>
<td>Thesis publications</td>
<td>vii</td>
</tr>
<tr>
<td>Table of contents</td>
<td>viii</td>
</tr>
<tr>
<td>List of tables</td>
<td>xiv</td>
</tr>
<tr>
<td>List of figures</td>
<td>xvi</td>
</tr>
<tr>
<td>List of abbreviations</td>
<td>xix</td>
</tr>
</tbody>
</table>

CHAPTER 1. Literature Review                  |     |
1.1 INTRODUCTION                               | 1    |
1.1.1 The strawberry plant                    | 1    |
1.1.2 Botany                                   | 2    |
1.1.3 Origin of cultivated strawberry         | 3    |
1.1.4 Strawberry industry                     | 4    |
1.1.5 Industrial constraints                  | 5    |
1.2 STRAWBERRY FLAVOUR                         | 6    |
1.2.1 Flavour perception                      | 6    |
1.2.2 Volatile compounds                      | 7    |
1.2.3 Factors affecting strawberry aroma      | 8    |
    1.2.3.1 Pre- and post-harvest environments  | 8    |
    1.2.3.2 Maturity                            | 9    |
    1.2.3.3 Genotype                            | 10   |
1.2.4 Molecular studies for flavour biosynthesis | 11   |
    1.2.4.1 Esters                              | 12   |
    1.2.4.2 Furanones                           | 13   |
    1.2.4.3 Terpenes                            | 15   |
1.3 DAY-NEUTRALITY                             | 16   |
1.3.1 Photoperiodic flowering habits          | 16   |

viii
1.3.1.1 Sources of day-neutrality in strawberry 17
1.3.1.2 Inheritance of day-neutrality 18
1.3.2 Other factors affecting flowering response 21
  1.3.2.1 Temperature 21
  1.3.2.2 Influence of plant growth regulators on flowering 22
1.3.3 Molecular control of photoperiod-dependent pathway 24
1.4 MOLECULAR MARKER DEVELOPMENT IN STRAWBERRY 26
  1.4.1 Molecular markers 27
    1.4.1.1 Morphological markers and isozymes 27
    1.4.1.2 PCR-based markers 28
  1.4.2 Linkage map construction 29
  1.4.3 Marker-trait association 31
  1.4.4 Molecular marker development for strawberry flavour improvement 32
  1.4.5 Molecular marker development for day-neutrality in strawberry 35
  1.4.6 Array-based marker systems 37
    1.4.6.1 Single Feature Polymorphism (SFP) 39
    1.4.6.2 Diversity Array Technology (DArT) 40
    1.4.6.3 Diversity Suppression Subtractive Hybridisation Array 42
    1.4.6.4 Subtracted Diversity Array (SDA) 43
1.5 RATIONALE OF THE STUDY 46

CHAPTER 2. Construction, Validation and Evaluation of Strawberry-specific Subtracted Diversity Array (SDA) 49
2.1 INTRODUCTION 49
2.2 MATERIALS AND METHODS 53
  2.2.1 Plant material 53
  2.2.2 Construction of the strawberry-specific SDA 56
    2.2.2.1 Genomic DNA extraction and quantification 56
    2.2.2.2 Genomic DNA subtraction 56
    2.2.2.3 Subtracted DNA fragment cloning and screening 59
    2.2.2.4 Microarray probe preparation and SDA printing 60
  2.2.3 Validation of the strawberry-specific SDA 61
    2.2.3.1 Target sample preparation and biotin labelling 61
2.2.3.2 Hybridisation of biotin-labelled target DNA onto the strawberry-specific SDA

2.2.3.3 Scanning and image quantification

2.2.4 Analysis of the strawberry-specific SDA

2.2.4.1 Fingerprinting of 15 strawberry genotypes

2.2.4.2 Data analysis

2.2.4.3 Statistical analysis

2.2.5 Aroma profiling of strawberry genotypes

2.2.5.1 Solid Phase Microextraction (SPME)

2.2.5.2 Gas Chromatography-Mass Spectrometry (GC-MS)

2.2.5.3 GC-MS data analysis

2.2.6 Correlation between chemical and molecular data

2.2.6.1 Comparison of aroma profiles and SDA fingerprints

2.2.6.2 Correlation of branch point DNA markers with volatile compounds

2.2.7 Sequencing of branch point markers

2.3 RESULTS AND DISCUSSION

2.3.1 Genomic DNA extraction and quantification

2.3.2 Subtracted genomic DNA library construction

2.3.3 Strawberry-specific SDA validation

2.3.4 Microarray data analysis

2.3.5 Genetic relationships among strawberry genotypes

2.3.6 Aroma profile analysis of selected strawberry genotypes

2.3.6.1 Chemical and genetic relationships of strawberry genotypes

2.3.7 Putative DNA marker discovery

2.3.7.1 Identification of polymorphic DNA sequences

2.3.7.2 Correlation between branch point markers and key volatile compounds

2.3.8 Identity of putative DNA markers

2.4 CONCLUSIONS
CHAPTER 3. Identification of DNA Markers Associated with Flavour using Bulked Segregant Analysis (BSA) and Subtracted Diversity Array (SDA)

3.1 INTRODUCTION

3.2 MATERIALS AND METHODS
   3.2.1 Parental genotypes and segregating population
   3.2.2 Phenotyping of strawberry flavour
   3.2.3 Bulked Segregant Analysis (BSA)
      3.2.3.1 Generation of DNA bulks with extreme phenotypes
      3.2.3.2 SDA hybridisation, scanning and image quantification
   3.2.4 Statistical Analysis
      3.2.4.1 Discriminant Function Analysis (DFA)
      3.2.4.2 Fisher’s ratio
      3.2.4.3 Independent Samples t-Test
   3.2.5 DNA sequencing of selected features
   3.2.6 Putative DNA marker validation
      3.2.6.1 Primer design and PCR amplification
      3.2.6.2 Sequence alignment and determination of DNA polymorphism

3.3 RESULTS AND DISCUSSION
   3.3.1 Variability and distribution of volatile compounds in the F₁ population compared to the parental genotypes
   3.3.2 Statistical Analysis
      3.3.2.1 Discriminant Function Analysis (DFA)
      3.3.2.2 Fisher’s ratio
      3.3.2.3 Independent Samples t-Test
   3.3.3 Sequence identity of the putative DNA markers
   3.3.4 Size and DNA sequence polymorphism

3.4 CONCLUSIONS

CHAPTER 4. Identification of DNA Markers Associated with Day-neutrality using Strawberry-specific SDA

4.1 INTRODUCTION

4.2 MATERIALS AND METHODS
   4.2.1 Plant genotypes and segregating population
4.2.2 Scoring for flowering response
4.2.3 Bulked Segregant Analysis (BSA)
   4.2.3.1 Generation of the phenotypic extreme DNA bulks
   4.2.3.2 SDA hybridisation, scanning and image quantification
4.2.4 Statistical analyses
4.2.5 DNA sequencing and sequence analysis
4.2.6 Detection of DNA microstructural variation in the sequences of nuclear-specific features

4.3 RESULTS AND DISCUSSION
  4.3.1 Day-neutrality assessment in field environment
  4.3.2 Statistical Analysis
    4.3.2.1 Discriminant Function Analysis (DFA)
    4.3.2.2 Fisher’s ratio
    4.3.2.3 Independent Samples t-Test
  4.3.3 Sequence identity of putative DNA markers
  4.3.4 Size and DNA sequence polymorphism
  4.3.5 Proposed model for FLP2E11 linked to the low activity CKX1 allele

4.4 CONCLUSIONS

CHAPTER 5. Conclusions and Future Directions
5.1 CONCLUSIONS
5.2 FUTURE DIRECTIONS

BIBLIOGRAPHY

APPENDIX 1. Overview of the Suppression Subtractive Hybridisation (SSH) process
APPENDIX 2. Position of the 290 array features and 18 controls gridded on four meta-grids for the strawberry-specific SDA
APPENDIX 3. Configurations of the BioRobotics® Total Array System (TAS) Application Suite software v2.6.0.1
APPENDIX 4. Position of two subarrays, six technical replicates and four meta-grids on the strawberry-specific SDA
APPENDIX 5. Settings for ScanArray® Express v4.0
APPENDIX 6. Positioning of the grid onto the strawberry-specific SDA
APPENDIX 7. Pearson’s bivariate correlation among the most polymorphic features across 15 strawberry genotypes
APPENDIX 8. Pearson’s bivariate correlation between selected branch point markers and key volatile compounds
APPENDIX 9. Polymorphic DNA sequences corresponding to branch point markers found in the fingerprinting of strawberry genotypes
APPENDIX 10. Polymorphic DNA sequences corresponding to flavour traits in strawberry plants
APPENDIX 11. Polymorphic DNA sequences corresponding to day-neutrality traits in strawberry plants
APPENDIX 12. DNA sequence alignment for three clones derived from four strawberry parental genotypes
APPENDIX 13. Full DNA sequence alignment for all clones and parental genotypes compared to the original FLP2E11 sequence
List of tables

Table 1.1 Proposed inheritance of flowering habit in strawberry. 20
Table 1.2 Comparison of the array-based techniques. 38
Table 2.1 Strawberry genotypes and non-angiosperm species used in the subtracted genomic library construction and array validation. 55
Table 2.2 Spectrophotometric reading of purified DNA. 75
Table 2.3 The strawberry genotypes used in this study, their immediate parents and country of origin. 85
Table 2.4 Relative composition (%) of volatile compounds in the headspace of strawberry puree. 91
Table 2.5 Normalised mean signal-to-noise ratio (SNR) of nine features showing the most variability across 15 strawberry genotypes based on PCA and magnitude of variance. 99
Table 2.6 Putative identity of the most polymorphic DNA sequences searched against the Fragaria vesca draft genome (v1.1). 110
Table 3.1 Descriptive statistics for 23 volatile compounds analysed in the F1 population and in the parental genotypes ‘07-102-41’ and ‘Juliette’. 135
Table 3.2 Discriminant functions generated for each key volatile compound based on the canonical discriminant function coefficients of the most predictive features. 146
Table 3.3 Putative DNA markers selected by DFA and the classification results for both the training set and test set based on the discriminant functions generated for each key volatile compound. 148
Table 3.4 Discriminant functions generated for each key volatile compound using the features selected by the first DFA. 150
Table 3.5 The reduced set of putative DNA markers selected by DFA, the classification results for the original and the estimated proportion for any future dataset. 150
Table 3.6 List of top 10 features ranked in decreasing order based on their respective Fisher’s ratio for key volatile compounds. 153
Table 3.7  Group statistics and Independent Samples t-Test for the putative markers selected by DFA.

Table 3.8  Putative identity of the most discriminatory features searched against the *Fragaria vesca* draft genome (v1.1).

Table 4.1  Flowering response on a 1-4 scale based on their day-neutrality strength.

Table 4.2  The number of F$_1$ progeny collected for each DN x SD and DN x DN crosses based on flowering strength.

Table 4.3  Discriminant functions generated for each flowering response group based on the canonical discriminant function coefficients of the most predictive features.

Table 4.4  Classification results for the training set and test set based on the DFA-selected features for each flowering response group.

Table 4.5  Discriminant functions generated for each flowering response group using the features selected by the first DFA.

Table 4.6  The reduced set of putative DNA markers selected by DFA, the classification results for the original cases and the estimated proportion for any future dataset.

Table 4.7  List of top 10 features ranked in decreasing order based on the respective Fisher’s ratio for flowering response groups.

Table 4.8  Group statistics and Independent Samples t-Test for the putative DNA markers selected by DFA.

Table 4.9  Putative identity of the most discriminatory features searched against the *Fragaria vesca* draft genome (v1.1).
## List of figures

<table>
<thead>
<tr>
<th>Number</th>
<th>Figure Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 1.1</td>
<td>Strawberry fruits.</td>
<td>3</td>
</tr>
<tr>
<td>Figure 1.2</td>
<td>The last two steps in fruit ester biosynthesis.</td>
<td>12</td>
</tr>
<tr>
<td>Figure 1.3</td>
<td>Proposed biosynthetic pathway of 4-hydroxy-2,5-dimethyl-3(2H)-furanone (HDMF).</td>
<td>14</td>
</tr>
<tr>
<td>Figure 1.4</td>
<td>Biosynthesis of monoterpene linalool and sesquiterpene nerolidol.</td>
<td>15</td>
</tr>
<tr>
<td>Figure 1.5</td>
<td>A simplified diagram showing genes involved in the molecular control of photoperiod-dependent pathway in <em>Arabidopsis thaliana</em>.</td>
<td>25</td>
</tr>
<tr>
<td>Figure 1.6</td>
<td>Schematic of the procedure used for detecting SFPs and ELPs.</td>
<td>39</td>
</tr>
<tr>
<td>Figure 1.7</td>
<td>Schematic of DArT development.</td>
<td>41</td>
</tr>
<tr>
<td>Figure 1.8</td>
<td>Schematic of the Subtracted Diversity Array (SDA) development and fingerprinting of angiosperm species.</td>
<td>44</td>
</tr>
<tr>
<td>Figure 2.1</td>
<td>The workflow of strawberry-specific SDA construction and performance validation.</td>
<td>57</td>
</tr>
<tr>
<td>Figure 2.2</td>
<td>The data normalisation procedures performed to the raw SNR of each feature after image scanning and spot quantification.</td>
<td>66</td>
</tr>
<tr>
<td>Figure 2.3</td>
<td>Extraction of volatile compounds from strawberry purees using SPME method.</td>
<td>69</td>
</tr>
<tr>
<td>Figure 2.4</td>
<td>Electrophoresis of genomic DNA isolated from (a) 15 strawberry genotypes and (b) nine non-angiosperm species on a 1.0 % agarose/EtBr gel.</td>
<td>73</td>
</tr>
<tr>
<td>Figure 2.5</td>
<td>Results of adaptor ligation efficiency analysis on a 2.0 % agarose/EtBr gel.</td>
<td>77</td>
</tr>
<tr>
<td>Figure 2.6</td>
<td>Results of subtraction efficiency analysis on a 2.0 % agarose/EtBr gel.</td>
<td>78</td>
</tr>
<tr>
<td>Figure 2.7</td>
<td>Example of DNA inserts amplified from pGEM®-T Easy vector using nested PCR primer1 and nested PCR primer 2R on a 1.0 % agarose/EtBr gel.</td>
<td>79</td>
</tr>
<tr>
<td>Figure 2.8</td>
<td>Validation of strawberry-specific SDA based on the hybridisation patterns of (a) tester and (b) driver samples.</td>
<td>81</td>
</tr>
<tr>
<td>Figure</td>
<td>Description</td>
<td></td>
</tr>
<tr>
<td>--------</td>
<td>-------------</td>
<td></td>
</tr>
<tr>
<td><strong>Figure 2.9</strong></td>
<td>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.</td>
<td></td>
</tr>
<tr>
<td><strong>Figure 2.10a</strong></td>
<td>Hierarchical dendrogram generated using average-linkage-between-groups method and squared Euclidean distance, based on the aroma profiles of five strawberry genotypes.</td>
<td></td>
</tr>
<tr>
<td><strong>Figure 2.10b</strong></td>
<td>Hierarchical dendrogram using the average-linkage-between-groups method and squared Euclidean distance showing genetic relatedness of five strawberry genotypes based on their hybridisation patterns on the strawberry-specific SDA.</td>
<td></td>
</tr>
<tr>
<td><strong>Figure 2.11</strong></td>
<td>Two-dimensional plot of the principal component analysis (PCA) of 287 SDA features corresponding to 15 strawberry genotypes.</td>
<td></td>
</tr>
<tr>
<td><strong>Figure 2.12</strong></td>
<td>Dendrogram constructed using average-linkage-between-groups method and squared Euclidean distance, based on the eight most discriminatory features among 15 strawberry genotypes.</td>
<td></td>
</tr>
<tr>
<td><strong>Figure 2.13</strong></td>
<td>Significant correlations among the normalised mean SNR of branch point markers and relative composition of key volatile compounds.</td>
<td></td>
</tr>
<tr>
<td><strong>Figure 2.14</strong></td>
<td>Landmark of FLP1C6 on LG6:12886104..12886402.</td>
<td></td>
</tr>
<tr>
<td><strong>Figure 3.1</strong></td>
<td>Frequency distribution of key volatile compounds measured as relative peak areas in the ‘07-102-41’ x ‘Juliette’ progeny.</td>
<td></td>
</tr>
<tr>
<td><strong>Figure 3.2</strong></td>
<td>A three-way Venn diagram showing the putative DNA markers in the intersection of DFA (green), Fisher’s ratio (top 10 features; blue) and Independent Samples t-Test ($p &lt; 0.01$; red) for all the key volatile compounds tested.</td>
<td></td>
</tr>
<tr>
<td><strong>Figure 3.3</strong></td>
<td>Landmark of selected putative DNA markers (red arrows) mapped onto the <em>F. vesca</em> draft genome (v1.1) and location of genes associated with key volatile compounds (green arrows).</td>
<td></td>
</tr>
<tr>
<td><strong>Figure 3.4</strong></td>
<td>Phospholipid catabolic pathway and its relationship to ester volatile biosynthesis.</td>
<td></td>
</tr>
<tr>
<td><strong>Figure 3.5</strong></td>
<td>The proposed biosynthesis pathway of linalool from monoterpenes geraniol, nerol, $\alpha$- and $\beta$-pinene.</td>
<td></td>
</tr>
</tbody>
</table>
Figure 3.6  The PCR products amplified from both parental genotypes and the F1 progeny plants in two phenotypic extreme bulks using FLP1D7 (upper row) and FLP1A7 (bottom row) sequence-specific primers on a 2.0 % agarose/EtBr gel.

Figure 3.7  DNA sequence polymorphism between two parental genotypes.

Figure 4.1  An example of flowering response based on a 1-4 scale.

Figure 4.2  A three-way Venn diagram showing the putative DNA markers in the intersection of DFA (green), Fisher’s ratio (top 10 features; blue) and Independent Samples t-Test ($p < 0.01$; red) for all flowering response classes assessed.

Figure 4.6  Landmark of FLP2E11 on LG6:14315954..14316207.

Figure 4.7  The PCR products amplified from both parental genotypes and the F1 progeny plants using FLP2E11 sequence-specific primers on a 2.0 % agarose/EtBr gel.

Figure 4.8  The PCR products amplified from all parental genotypes using FLP2E11 sequence-specific primers (~256 bp) and T7 and Sp6 primers (~400 bp) on a 2.0 % agarose/EtBr gel.

Figure 4.9  Alignment of the parental variants for 05-069-63, 05-069-194, 01-061-311 and Juliette at FLP2E11 locus compared to the original DNA sequence derived from the strawberry-specific SDA.

Figure 4.10 Proposed model of flowering response controlled by copy number of the allelic variants of $CKX1$ allele in SD and DN strawberry genotypes.
**List of abbreviations**

- **°C**: degree Celsius
- **µg**: microgram
- **µL**: microlitre
- **µM**: micromolar
- **µm**: micrometre
- **AFLP**: amplified fragment length polymorphism
- **BA**: benzyladenine
- **BC1**: backcrossed 1 generation
- **bp**: base pair
- **BSA**: Bulked Segregant Analysis
- **CaMV35S**: cauliflower mosaic virus 35S promoter
- **cDNA**: complementary DNA
- **cM**: centimorgan
- **CoA**: coenzyme A
- **CO₂**: carbon dioxide
- **CTAB**: cetyltrimethylammonium bromide
- **Cy5**: cyanine-5
- **DArT**: Diversity Array Technology
- **DFA**: Discriminant Function Analysis
- **DIG**: digoxigenin
- **DMSO**: dimethyl sulfoxide
- **DNA**: deoxyribonucleic acid
- **DN**: day-neutral
- **dNTP**: deoxyribonucleotide triphosphate
- **dUTP**: deoxyuridine triphosphate
- **EB**: everbearing
- **EDTA**: ethylenediaminetetraacetic acid
- **ELP**: expression level polymorphism
- **EtBr**: ethidium bromide
- **G3PDH**: glyceraldehyde 3-phosphate dehydrogenase
- **GA**: gibberellin
PMT
photo multiplier tube

QTL
quantitative trait loci

RAD
restriction-site associated DNA

RAPD
random amplified polymorphic DNA

RFLP
restriction fragment length polymorphism

RNA
ribonucleic acid

RNA-Seq
RNA sequencing

rpm
revolutions per minute

s
seconds

SAM
shoot apical meristem

SCAR
sequence-characterised amplified region

SD
short day

SDA
Subtracted Diversity Array

SDS
sodium dodecyl sulphate

SFP
single feature polymorphism

SNP
single nucleotide polymorphism

SNR
signal-to-noise ratio

SPME
solid phase microextraction

SPSS
Statistical Package for the Social Sciences

SSC
saline sodium citrate buffer

SSH
suppression subtractive hybridisation

SSPE
saline-sodium phosphate-EDTA buffer

SSPE-T
SSPE-Triton X-100 buffer

SSR
simple sequence repeat

STS
sequence-tagged site

TBE
Tris/Borate/EDTA buffer

TOF
time-of-flight

Tris
tri(hydroxymethyl) aminomethane

TSS
total soluble solids

u
units

UV
ultra violet
CHAPTER 1
Literature Review

This review is aimed to provide an overview of the current state of knowledge regarding genetic improvement of strawberry, specifically for fruit flavour and day-neutrality. Firstly, the importance of developing new strawberry cultivars with enhanced fruit flavour and day-neutrality to the strawberry industry, particularly in Australia, is highlighted. Secondly, this review focuses on the volatile composition of strawberry flavour, factors that influence the production of these chemical constituents, and the biosynthetic pathways of the important biochemical classes contributing to strawberry flavour. Thirdly, this review summarises the categories of strawberry genotypes based on flowering habits, the sources of day-neutrality and the molecular control of flowering time. Finally, the development of molecular markers in strawberry is reviewed, followed by a section that highlights the array-based methods, with emphasis on the use of Subtracted Diversity Array (SDA) for molecular marker screening and marker-trait association studies. In this review, the gaps in the current understanding of molecular control of strawberry flavour and day-neutrality are identified, which lead to the rationale for this thesis study. The reviewed microarray-based approaches provide an opportunity to address some of the knowledge gaps.

1.1 INTRODUCTION

1.1.1 The strawberry plant

Strawberry, which belongs to the genus *Fragaria*, is a member of the family Rosaceae, that includes other fruit crops such as raspberry, blackberry, apple, peach and cherry (Folta and Davis, 2006). This genus comprises of 21 recognised species with different ploidy levels, including 12 diploids \((2n = 2x = 14)\), four tetraploids \((2n = 4x = 28)\), one hexaploid \((2n = 6x = 42)\), etc. The molecular control of flowering time in strawberry is an important aspect of genetic improvement, as it directly influences the production of fruits and thus the overall harvest. Moreover, the day-neutrality of strawberry fruit is a characteristic that allows for year-round production, which is crucial for the industry.

The review of genetic improvement in strawberry begins with an overview of the current knowledge on the volatile composition of strawberry flavour, including the identification and quantification of volatile compounds that contribute to the aromatic profile of strawberry. The biosynthetic pathways of the important biochemical classes involved in strawberry flavour are then discussed, providing insights into the molecular basis of these flavour traits.

The categorisation of strawberry genotypes based on flowering habits, the sources of day-neutrality, and the molecular control of flowering time is a critical aspect of genetic improvement. Understanding the genetic basis of flowering time and day-neutrality is essential for the development of new strawberry cultivars that can be grown under a variety of environmental conditions.

Molecular markers play a crucial role in strawberry genetics, allowing for the rapid identification of genetic traits of interest. The use of array-based methods, particularly Subtracted Diversity Array (SDA), has become a powerful tool for molecular marker screening and marker-trait association studies. The integration of these methods with genome-wide association studies (GWAS) offers a comprehensive approach to understanding the genetic basis of strawberry flavour and day-neutrality traits.

In conclusion, this review highlights the gaps in the current understanding of molecular control of strawberry flavour and day-neutrality, providing a rationale for the research directions presented in this thesis study. The reviewed microarray-based approaches offer promising avenues for addressing these knowledge gaps and advancing strawberry genetics.
6x = 42) and four octoploids (2n = 8x = 56). They are distributed predominantly in the Northern hemisphere (Davis et al., 2007; Hummer and Hancock, 2009). The most common diploid species, *Fragaria vesca* is native to northern Eurasia and North America (Staudt, 1989). Based on the estimation of Akiyama et al. (2001), it has a small genome at 200 Mb, which is only slightly larger than the genome size of *Arabidopsis thaliana* (Stewart, 2011). In contrast, the modern cultivated strawberry, *Fragaria x ananassa*, is an allo-octoploid with a complex genome composition (Sargent et al., 2009). The latest model proposed for octoploid strawberry, AAA′A′BBB′B′, implies the presence of up to four subgenomes (Bringhurst, 1990; Davis et al., 2007). According to Davis et al. (2006), *F. vesca* is likely to be the donor of A genome while a Japanese endemic species, *F. innumae* could be the contributor of B genome.

**1.1.2 Botany**

Strawberry is a perennial, herbaceous and low-growing plant (Davis et al., 2007). Although many *Fragaria* species are dioecious, most of the modern cultivars are hermaphroditic with complete flowers having both stamens and carpels. Once pollinated and fertilised, each pistil develops into a single-seeded fruit known as achene (Figure 1.1) (Chandler et al., 2012). These achenes are embedded in the epidermal layer of the receptacle and connected by fibrovascular strands to the interior of the receptacle. The edible red flesh is an aggregated fruit derived from the enlarged receptacle tissue of strawberry flower (Figure 1.1) (Darnell, 2003; Darrow, 1966; Perkins-Veazie, 2010). Besides self-fertilisation, strawberry plants also produce runners with multiple daughter plants as a means of vegetative propagation.
1.1.3 Origin of cultivated strawberry

The cultivated strawberry (*Fragaria x ananassa*) is derived from a chance hybridisation between two American octoploid progenitors during the mid-18th century when *F. virginiana* and *F. chiloensis* from North and South America, respectively, were brought to Europe (Darrow, 1966). *F. virginiana* is valued for its aroma while *F. chiloensis* produced larger and firmer fruits (Stewart, 2011). Cross-pollination produced hybrids exhibiting a combination of unique characteristics and was named *F. x ananassa*, which is the large-fruited dessert strawberry known today (Davis *et al.*, 2007; Hancock, 1999). Although *F. virginiana* and *F. chiloensis* have been used in the early years of strawberry improvement, most of the recent breeding efforts depend mainly on the selection between *F. x ananassa* clones, resulting in a narrow genetic base of cultivated strawberries (Dale...
and Sjulin, 1990; Sjulin and Dale, 1987). Therefore, the introgression of desired traits such as fruit quality, disease resistance and day-neutrality from the wild progenitors is a promising strategy that utilises genetic diversity perhaps missing from the cultivated strawberries (Chandler et al., 2012; Stewart, 2011).

1.1.4 Strawberry industry

Strawberry (Fragaria x ananassa) is one of the most economically important soft fruits cultivated in the world (Hummer and Hancock, 2009). They are highly favoured for their unique appearance, smell, taste and nutritional quality (Chandler et al., 2012; Folta and Davis, 2006). Moreover, they are also rich in antioxidants such as vitamin C (Capocasa et al., 2008) and anticarcinogens (i.e. ellagic acid) (Maas et al., 1991). These qualities have sustained and elevated the market need for strawberry fruits (Whitaker, 2011). In line with the growing demand in global consumption, the world strawberry production has increased gradually from 2.4 million tonnes in 1992 to 4.5 million tonnes in 2012, with the USA contributing to approximately 36% of total world production, followed by Spain, Japan, Republic of Korea and Poland (FAOSTAT 2014). At present, strawberries are grown all year round in Australia by utilising different day-neutral (DN) and short-day (SD) cultivars as well as the diversity of climates in different states, with Victoria and Queensland each contributing to 35% of the total strawberry production nationwide. While strawberry production in Australia focuses mainly on domestic fresh fruit marketing, exports do occur in a largely opportunistic basis. As of 2010/2011, the largest export markets are Singapore, followed by Thailand, New Zealand, Hong Kong, and Malaysia (Strawberries Australia Inc., 2012).
1.1.5 Industrial constrains

Currently, domestic consumption of strawberry in Australia is relatively low at 1.35 kg per capita compared to 6.4 kg per capita in the USA and 3.6 kg per capita in Japan (Strawberries Australia Inc., 2012). The inconsistency of strawberry quality including fruit flavour across the season has been recognised as one of the factors affecting purchasing decision. The changes in flavour could be attributed to genotypic and environmental effects. In a study on the flavour qualities of ‘Albion’ and ‘Juliette’ strawberry fruits, environment was identified as the major source of variation for 23 aroma compounds whereas genotype-by-environment (G x E) interactions were significant for 116 volatile compounds (Samykanno et al., 2013b). In addition, strawberry flavour is also affected by other factors such as cultivar, fruit maturity and postharvest storage condition (Ayala-Zavala et al., 2004; Forney et al., 2000; Ragaert et al., 2006).

In addition, domestic consumption may be promoted by increasing the total strawberry production. This can be achieved by growing more DN cultivars which allows flower formation in the long days of spring and continued fruit production throughout the summer and autumn (Shaw and Famula, 2005). However, most of the cultivated DN genotypes in Australia are heavily dependent on imported strawberries. For instance, three out of the four DN cultivars currently grown commercially in Victoria, i.e., Albion, San Andreas and Portola were sourced from California whereas only Melba was developed locally by the Victorian strawberry breeding program (Brevis, 2013). Unfortunately, spring frost and summer heat have seriously affected the performance of the Californian cultivars and caused 10 % – 15 % production loss in 2013/14 season. Whilst most of the Californian cultivars were damaged by rain in 2013/14 season, Melba produced a good crop throughout the season (S Violi [Strawberries Australia, Victoria, Aust.] 2013, pers. Comm.,
20 December 2013). Therefore, it is important to develop more local DN cultivars that are tolerant to extreme fluctuations of the Australia climate and also have supreme and consistent fruit flavour (Byrne, 2012).

This review focuses on the genetic improvement of strawberry flavour and day-neutrality. For a broader perspective on strawberry breeding, readers are referred to excellent reviews by Chandler et al. (2012), Davis et al. (2007); Folta et al. (2011); Folta (2013); Folta and Davis (2006); Hancock et al. (2008); Hummer and Hancock (2009) and Stewart (2011).

1.2 STRAWBERRY FLAVOUR

1.2.1 Flavour perception

Consumer perception of flavour involves integration of multiple sensory systems, including taste, aroma, colour, texture and mouth feel (Kader, 2008; Klee, 2010; Schwieterman et al., 2014). The foundation of flavour is determined by taste, which is defined as the balance between sweetness and sourness or acidity based on the sugar/acid ratio present in fruits (Kader, 2008). Strawberry usually contains more sugars (fructose, glucose and sucrose) than acids (citric or malic acid) (Kallio et al., 2000). With over 300 volatile compounds identified in strawberry (Latrasse, 1991; Zabetakis and Holden, 1997), it has a complex mixture of aromatic compounds that contribute towards sensory perception. A recent consumer survey revealed that overall liking was greatly influenced by sweetness and aroma intensity (Colquhoun et al., 2012), indicating the importance of these organoleptic properties in shaping strawberry flavour. The present study focuses on aroma as it is often considered to play a dominant role in strawberry flavour (Kader, 2008).
1.2.2 Volatile compounds

To date, the identification of volatile compounds, either qualitatively or quantitatively, relied on different volatile extraction and separation methods (Schwab et al., 2009). Commonly used methods include solid phase microextraction (SPME) in combination with gas chromatography coupled with mass spectrometry (GC-MS) (Jetti et al., 2007; Kafkas and Paydas, 2007). In addition to aroma profiling, the odorants contributing to strawberry aroma are often analysed using GC-Olfactometry along with sensory analysis for qualitative and quantitative evaluation of sensory attributes (Du et al., 2011; Fukuhalra et al., 2005; Gunness et al., 2009; Jouquand et al., 2008). Based on these methods, hundreds of volatile compounds have been detected and classified into different biochemical classes, such as esters, alcohols, aldehydes, furanones, terpenes, lactones and sulphur compounds (Dirinck et al., 1981; Latrasse, 1991; Pérez et al., 1992; Ulrich et al., 1997; Zabetakis and Holden, 1997). Of all the volatiles detected, only about 20 compounds have been determined to dominate the typical strawberry aroma based on sensory descriptive analysis and odour activity values (OAVs), which is defined as the volatile concentration divided by odour threshold values (Larsen et al., 1992). Methyl butanoate, ethyl butanoate, ethyl hexanoate, ethyl 3-methylbutanoate, hexyl acetate, linalool, furaneol, mesifuranne, and γ-dodecalactone have been shown to possess the highest OAVs (Du et al., 2011; Jetti et al., 2007).

In general, esters are the most abundant aroma active compounds and constitute 25 - 90 % of the total volatiles in ripe strawberry fruit. They contribute to the sweet and fruity notes of strawberry flavour based on sensory descriptors (Douillard and Guichard, 1990; Pyysalo et al., 1979; Schreier, 1980). For instance, methyl butanoate has been described as apple-like whereas methyl hexanoate, ethyl butanoate and ethyl hexanoate have been described
as pineapple-like notes (Du et al., 2011). In addition, lactones such as γ-decalactone and γ-dodecalactone, are cyclic esters which possess a sweet and peach-like note. The sweet and caramel-like notes in strawberries are produced from the two most intense odorants, 4-hydroxy-2,5-dimethyl-3(2H)-furanone (HDMF, furaneol) and 2,5-dimethyl-4-methoxy-3(2H)-furanone (DMMF, mesifuranne). Terpenes, such as linalool and geraniol have been determined as the contributors of floral and citrus-like note in strawberries. Apart from that, aldehydes such as (E)-2-hexenal, hexanal and (Z)-3-hexenol are responsible for green and fresh notes (Du et al., 2011; Fukuhara et al., 2005). A proper balance of these chemical constituents determines the unique and fruity flavour that characterises the fresh strawberry (Forney et al., 2000). Therefore, more studies are required to identify which compounds contribute to desirable aroma and their threshold concentrations for a pleasant flavour (Kader, 2008).

1.2.3 Factors affecting strawberry aroma

1.2.3.1 Pre- and post-harvest environments

Pre-harvest factors such as cultural practices and environmental conditions including light availability, temperature fluctuation, relative air humidity and rainfall are some of the main factors affecting overall quality of strawberry fruits, including sugar/acid ratio, volatile content and other health-related compounds (Hadi et al., 2013; Samykanno et al., 2013b; Tulipani et al., 2011). While most of the studies on cultural practices are focused on the other aspects of strawberry quality, such as fruit colour, total soluble solids (TSS), titratable acidity (TA) and total anthocyanin content (Wang et al., 2002; Roussos et al., 2009), a few other researchers have started to investigate the effect of environmental variation on the production of strawberry aroma. For instance, shading treatment on strawberry plants reduced the amount of volatiles (hexanal, hexenal, ethylmethyl butanoate
and methyl butanoate) present in the headspace of strawberry puree (Watson et al., 2002), suggesting that full sunlight is required for proper aroma development during strawberry ripening. A recent study also found that photosynthetically active radiation (PAR) has greater influence on volatile variability compared to other environmental parameters (Samykanno et al., 2013b).

Additionally, the effect of postharvest storage conditions such as the differences in temperatures and CO₂ concentration on strawberry volatiles has also been studied extensively (Ayala-Zavala et al., 2004; Berna et al., 2007; Larsen and Watkins, 1995b; Pelayo et al., 2003; Ragaert et al., 2006). In general, deterioration of flavour life was observed in CO₂-enriched atmosphere although fruit firmness was retained (Larsen and Watkins, 1995b; Pelayo et al., 2003). Storage at 10 °C was better in enhancing the production of aroma compounds compared to 0 °C, where, an increase in ethyl hexanoate, hexyl acetate, methyl acetate and butyl acetate was observed (Ayala-Zavala et al., 2004). Similarly, the concentration of methyl esters increased more rapidly compared to ethyl esters during storage at 15 °C, however, the changes in ester synthesis remain unexplained (Forney et al., 2000). One common finding from these studies was that off-flavour usually developed during postharvest storage when ethanol is rapidly converted into ethyl acetate due to yeast proliferation (Ragaert et al., 2006).

1.2.3.2 Maturity

Volatile compounds are produced as a consequence of physiological changes during fruit ripening (D'Ambrosio et al., 2013). It has been shown that volatile content increases as fruit ripens, with the concentration of methyl esters increasing approximately 7-fold compared to ethyl esters (Forney et al., 1996). While most of the esters and furanones
increased during ripening (Ito et al., 1990; Pérez et al., 1996a). C₆ alcohols decreased, which may explain the loss of green, immature odour as fruit ripens (Forney, 2001; Pérez et al., 1992; Ulrich et al., 1997). In addition, more than 280 volatile compounds have been identified in a non-targeted GC-MS study by analysing the seven developmental stages of strawberry. Most of the esters and alcohols reached the highest levels in red-turning and red-ripening fruits, resulting in the formation of fruity, sweet and peach-like flavour in ripe strawberries (Zhang et al., 2011). Taken together, these suggest that the concentration of good flavour compounds increases whereas the concentrations of the compounds responsible for green and grassy notes decreases as strawberries ripen.

1.2.3.3 Genotype

Despite the presence of environmental influence and physiological changes at different developmental stages, genetic component remains the basis of aroma variability (Baldwin, 2008). Strawberry aroma has been shown to be cultivar-dependent (Forney, 2001; Forney et al., 2000; Kovačević et al., 2008). For instance, the total volatile composition varies up to 35-fold among different cultivars, and the ratio of methyl/ethyl esters was cultivar-dependent (Forney et al., 2000). However, contrasting results were reported by Pelayo-Zaldívar et al. (2005), where the ratio of methyl/ethyl esters were affected by harvest date, indicating the presence of genotype-by-environment (G x E) interactions (Samykanno et al., 2013b). In addition, cultivar-specific compounds such as ethyl butanoate, methyl butanoate, γ-decalactone and 2-heptanone have been reported after comparing the aroma profiles of six strawberry cultivars (Larsen et al., 1992).

Similarly, volatile compounds are also species-dependent as diversity of aroma patterns has been observed in wild and cultivated strawberries (Dong et al., 2013; Ulrich et al.,
2007; Ulrich and Olbricht, 2013). The aroma intensities of wild types including *Fragaria virginiana*, *F. vesca* and *F. moschata* were higher compared to the cultivated *F. x ananassa*. Esters and terpenes were present at higher concentrations in modern cultivars. In contrast, ketones occurred at higher amounts in *F. vesca* (Ulrich et al., 2007). However, contrasting results were observed in other studies, where ester content was found to be higher in modern cultivars than in *F. vesca* (Dong et al., 2013; Ulrich and Olbricht, 2013). These discrepancies could possibly be due to different accessions and volatile extraction methods used in the studies. Species-specific ester compounds have also been identified. For example, 1-methyltridecyl acetate, myrtenyl acetate, trans-pinocarvyl acetate and ethyl dodecanoate were present only in *F. vesca* (Dong et al., 2013). Apart from esters, different terpene compounds such as terpinene-4-ol, α-terpineol and myrtenol were detected in *F. vesca* but not in *F. x ananassa* (Aharoni et al., 2004; Ulrich and Olbricht, 2013). These specific monoterpenes contribute to turpentine-like, herbaceous and woody odour which have a negative effect on sensory quality and may have been selected against during strawberry breeding, resulting in their loss in modern cultivars (Aharoni et al., 2004). The above findings suggest that careful selection of wild accessions as volatile donors may possibly help to regain flavour characteristics that have been lost due to modern breeding (Kader, 2008).

1.2.4 Molecular studies for flavour biosynthesis

The genes encoding enzymes for volatile biosynthesis in fruits have been identified by the use of molecular and biochemical techniques. Most of the studies have reported three main sources for volatile biosynthesis, which are fatty acids, amino acids and carbohydrates (Bood and Zabetakis, 2002; Paliyath et al., 2009; Schwab et al., 2008). This review
focuses on the biosynthetic pathways of three important biochemical classes contributing to strawberry aroma.

1.2.4.1 Esters

Volatile esters are produced from fatty acids and amino acids. For instance, degradation of free fatty acids results in aldehyde formation which can be converted into alcohols and their corresponding esters (Pérez et al., 1996b). In addition, strawberry peduncles immersed in a growth solution supplemented with L-isoleucine resulted in a 7-fold increase in 2-methylbutanoate esters and double production of 2-methylbutyl esters compared to control fruits (Pérez et al., 2002). The authors suggested that aminotransferase is required to catalyse the release of ester precursors before the last step of ester biosynthesis (Pérez et al., 2002; Pérez et al., 1992). The last two steps of ester formation have been identified. Aldehyde dehydrogenase (ADH) is responsible for the reduction of aldehydes to alcohols whereas alcohol acyltransferase (AAT) links alcohols to an acyl-CoA molecule to produce the corresponding esters (Figure 1.2) (Bood and Zabetakis, 2002; Schwab et al., 2009).

\[
\text{Aldehydes} \xrightarrow{\text{ADH}} \text{Alcohols + Acyl-CoA} \xrightarrow{\text{AAT}} \text{Esters}
\]

**Figure 1.2** The last two steps in fruit ester biosynthesis. ADH: Aldehyde dehydrogenase; AAT: Alcohol acyltransferase. Adapted and modified from Bood and Zabetakis (2002).

While the research on \textit{ADH} gene is limited in strawberry (Wolyn and Jelenkovic, 1990), comprehensive studies have been performed on the characterisation of \textit{AAT} gene and its enzyme activities. A strawberry alcohol acyltransferase (\textit{SAAT}) gene was first identified using a cDNA microarray and its correlation with ester formation was confirmed using an
enzymatic assay (Aharoni et al., 2000). The SAAT enzyme showed maximum activity with aliphatic medium-chain alcohols, in which both acyl-CoA molecules and alcohol substrates are required for ester formation (Aharoni et al., 2000). Comparison between *F. vesca* alcohol acyltransferase (VAAT) and SAAT revealed that SAAT enzyme has a preference for C$_6$ to C$_{10}$ aliphatic alcohols whereas C$_6$ alcohols and n-butyl alcohols were good substrates for the VAAT enzyme (Beekwilder et al., 2004). This may explain the differences among the types of ester compound produced by *F. vesca* and *F. x ananassa* (Section 1.2.3.3). Similarly, heptanol, hexanol and pentanol were found to be the preferred substrates for three European varieties, two American cultivars and *F. vesca*, respectively (Olías et al., 2002). In addition, different AAT genes have been isolated from various *Fragaria* species and cultivars. For instance, *FcAAT1* and *FaAAT2* were isolated from *F. chiloensis* and *F. x ananassa* cv. Camarosa, respectively. Phylogenetic analyses using amino acid sequences show that *FaAAT2* was distantly related to *SAAT*, *VAAT* and *FaAAT1* (Cumplido-Laso et al., 2012; Gonzalez et al., 2009), suggesting that the diversity of ester volatiles produced in different strawberry species and cultivars may be controlled by genetic factors.

1.2.4.2 Furanones

The biosynthesis of 4-hydroxy-2,5-dimethyl-3(2H)-furanone (HDMF, furaneol) has been studied extensively due to its attractive flavour and high commercial value in food industries (Hauck et al., 2003; Pérez et al., 1996a; Zabetakis, 1997; Zabetakis et al., 1999; Zabetakis and Holden, 1996). D-fructose-1,6-diphosphate (FDP) has been proposed as the precursor for HDMF based on incorporation experiment of 15 radiolabelled precursors and substances, where D-[U-$^{14}$C]-fructose-1,6-diphosphate showed the highest incorporation rate into furanone structures (Roscher et al., 1998). FDP is transformed by an unknown
enzyme to 4-hydroxy-5-methyl-2-methylene-3(2H)-furanone (HMMF). The conversion of HMMF to HDMF is catalysed by an auxin-dependent putative quinone oxidoreductase (FaQR) (Raab et al., 2006), which was later renamed as enone oxidoreductase (FaEO) (Figure 1.3) (Klein et al., 2007; Schwab et al., 2009). HDMF is further metabolised into its methylated form, 2,5-dimethyl-4-methoxy-3(2H)-furanone (DMMF, mesifuranne) by *F. x ananassa* O-methyltransferase (FaOMT) (Figure 1.3) (Schwab et al., 2009; Wein et al., 2002). Transgenic studies show that *FaOMT* under the control of CaMV35S constitutive promoter resulted in a nearly complete loss of DMMF in strawberry fruits, indicating a significant impact of the enzyme on furanone production (Lunkenbein et al., 2006; Schwab et al., 2008).

![Proposed biosynthetic pathway of 4-hydroxy-2,5-dimethyl-3(2H)-furanone (HDMF). HMMF: 4-hydroxy-5-methyl-2-methylene-3(2H)-furanone; DMMF: 2,5-dimethyl-4-methoxy-3(2H)-furanone; FaEO: *Fragaria x ananassa* enone oxidoreductase; FaOMT: *F. x ananassa* O-methyltransferase. Adapted and modified from Schwab et al. (2009).](image)

**Figure 1.3** Proposed biosynthetic pathway of 4-hydroxy-2,5-dimethyl-3(2H)-furanone (HDMF). HMMF: 4-hydroxy-5-methyl-2-methylene-3(2H)-furanone; DMMF: 2,5-dimethyl-4-methoxy-3(2H)-furanone; FaEO: *Fragaria x ananassa* enone oxidoreductase; FaOMT: *F. x ananassa* O-methyltransferase. Adapted and modified from Schwab et al. (2009).
1.2.4.3 Terpenes

_Fragaria x ananassa Nerolidol synthase 1 (FaNES1)_ gene involved in the biosynthesis of terpenes was identified using a cDNA microarray (Aharoni _et al._, 2004). Characterisation of the gene product revealed that it is a biofunctional mono- and sesquiterpene synthase. It catalyses the formation of monoterpane linalool from geranyl diphosphate (GPP) and sesquiterpene nerolidol from farnesyl diphosphate (FPP) (Figure 1.4) (Aharoni _et al._, 2004; Schwab _et al._, 2009). Generally, sesquiterpenes are derived from the cytosolic mevalonate (MVA) pathway whereas monoterpenes originate from the plastidial methylerthritol phosphate (MEP) pathway (Bohlmann and Keeling, 2008; Martin _et al._, 2012). However, an _in vivo_ feeding experiment demonstrated that (S)-linalool and trans-(S)-nerolidol in _F. x ananassa_ are exclusively synthesised through the MVA pathway in cytosol (Hampel _et al._, 2006). This variation is attributed to the _NES1_ gene that encodes a protein truncated at its N terminus, where the plastid-targeting signal is lost. Therefore, in strawberry plants, _FaNES1_ utilises both GPP and FPP to produce linalool and nerolidol in cytosol (Aharoni _et al._, 2004).

![Figure 1.4](image)

**Figure 1.4** Biosynthesis of monoterpane linalool and sesquiterpene nerolidol. Reproduced from Aharoni _et al._ (2004).

_Pinene synthase_ is another gene involved in the biosynthesis of terpenes. An insertional mutation of this gene resulted in the loss of some olefinic monoterpenes in the cultivated
strawberry. Conversely, *F. vesca* retained a functional copy of *pinene synthase*, thus it is capable of producing α-pinene and β-myrcene (Aharoni *et al.*, 2004; Folta and Davis, 2006). These studies demonstrate that structural changes at the DNA level can cause a loss of function in important terpene synthase genes, resulting in diverse aroma profiles between wild and cultivated strawberry species.

1.3. DAY-NEUTRALITY

1.3.1 Photoperiodic flowering habits

Strawberries have been classified into different categories based on their photoperiodic flowering behaviours, including short-day or June bearing, everbearing and day-neutral (Stewart and Folta, 2010). The short-day (SD) strawberries are defined as the genotypes which initiate flower buds when the day length is shorter than 14 hours, with an optimum photoperiod around 12 hours (Darrow, 1936). Variation observed in the flowering behaviour of SD genotypes resulted in another class known as infra short-day, where longer day lengths (13.5 – 14 hours) and little chilling are required (Izhar, 1997). However, Stewart and Folta (2010) opined that this variation is an extension of the existing SD genotypes and could not be classified as a distinct flowering habit. Normally, the SD genotypes produce many flowers under optimal SD conditions but become vegetative (only runner formation occurs) under longer day lengths and higher temperatures (Durner, 1984; Piringer and Scott, 1964).

In contrast, everbearing (EB) genotypes promote flowering in the long days of summer (> 14 hours) (Darrow, 1966) and on newly formed runners for an autumn crop production (Stewart and Folta, 2010). Another distinct flowering class known as day-neutral (DN) is often confused with the EB genotypes. However, in contrast to the EB genotypes, DN
strawberries are truly day-length insensitive in the sense that they are not regulated by daylight regime (Ahmadi and Bringhurst, 1991; Durner, 1984; Whitaker, 2011). Apart from the difference in flowering behaviours, EB and DN genotypes also vary in runner formation. While EB genotypes are insensitive to photoperiod and temperature changes, runner initiation in DN genotypes is similar to SD genotypes. Moreover, DN cultivars have better heat tolerance and prolonged harvest seasons but they typically produce fewer runners compared to SD types (Chandler et al., 2012; Durner, 1984).

1.3.1.1 Sources of day-neutrality in strawberry

Generally, most of the Fragaria species are SD plants with a few naturally occurring exceptions. For instance, in the diploid strawberries, F. vesca var. vesca native to North America is a SD plant while the European strawberry, F. vesca var. semperflorens is the first exception noted with EB flowering habit (Darrow, 1966). Found in the European Alps, this Alpine strawberry appears to be a mutant of the common diploid strawberries. They were introduced into France and England in mid-1700 (Darrow, 1966). However, this mutant has less economic importance to modern strawberry breeding because inter-specific hybridisation is rare between diploid and octoploid species and the offspring produced are usually infertile (Folta and Davis 2006).

Similarly, the earliest cultivated octoploid strawberries (F. x ananassa) are all SD genotypes although they vary in photoperiod sensitivity (Darrow, 1966). Breeding for the everbearing flowering habit can be traced back to nearly two centuries. The European EB genotypes available today are the derivatives of Gloede’s seedling, a remontant strawberry introduced in France in 1866 (Ahmadi et al., 1990; Richardson, 1914). However, the genetic source of this European everbearer remains unclear. Conversely, the everbearing
trait in the American genotypes is derived from spontaneous mutation of a single gene in the SD cultivar ‘Bismark’ in New York in 1898. This mutation resulted in the production of ‘Pan-American’, the source of the everbearing behaviour in modern EB cultivars in the USA (Ahmadi et al., 1990; Stewart and Folta, 2010). A major achievement in strawberry breeding was the introduction of day-neutrality trait from *F. virginiana* ssp. *glauca* found in the Wasatch Mountains of Utah into the breeding program at University of California in 1954 (Bringhurst and Voth, 1984). Therefore, sources of both everbearing and day-neutrality traits from ‘Pan-American’ and glauca, respectively are present in the pedigree of the modern cultivars with continuous flowering habit (Stewart and Folta, 2010). Since then, these cultivars have created a major impact on strawberry production trends, especially in California (Hancock, 1999; Sakin et al., 1997).

1.3.1.2 Inheritance of day-neutrality

Inheritance of day-neutrality in strawberry has been studied extensively with no conclusive findings (Table 1.1). Day-neutrality in the diploid Alpine strawberry was first postulated to be controlled by a single locus (*S*), with the allele controlling perpetual flowering being recessive to the allele responsible for short-day flowering response (Brown and Wareing, 1965). A later study using a *F. vesca* var. *vesca* and *F. vesca* var. *semperflorens* testcross population further confirmed the simple inheritance model (*SFL*) of perpetual flowering response in the diploid strawberry (Albani et al., 2004). While monogenic inheritance of the European strawberry was convincing, the SD California *F. vesca* appeared to be controlled by three dominant genes based on the F2 segregation ratio (1:63) of a cross between the EB Alpine *F. vesca* and the SD California *F. vesca*. The authors suggested that three mutant recessive genes would be necessary to condition day-neutrality in the *F. vesca* found in California (Ahmadi et al., 1990).
Unlike the diploid strawberry, the inheritance of day-neutrality in the octoploid seems to be more complicated, although monogenic inheritance was found in some *F. x ananassa* genotypes (Table 1.1). An earlier research based on the selfing of the European everbearers, ‘St Antoine de Padoue’ and ‘Laxton’s Perpetual’ yielded a 3:1 ratio as expected from a dominant allele controlling perpetual flowering (Richardson, 1914). A later study confirmed that photo-insensitivity is controlled by a single dominant allele by evaluating up to 30,000 plants derived from crosses between the Californian DN and SD cultivars (Ahmadi *et al*., 1990). The monogenic dominant inheritance model was also supported by Sugimoto *et al*. (2005), where the chi-square test yielded an expected ratio of 1:1 and 3:1 from a Japanese EB ‘Ever Berry’ x SD ‘Toyonoka’ cross and the ‘Ever Berry’ selfed, respectively.

However, day-neutrality has also been reported as a polygenic trait in octoploid strawberry. The number of genes controlling day-neutrality ranged from one major dominant allele with other modifying genes (Clark, 1937) to two or three dominant complementary genes (Table 1.1) (Ourecky and Slate, 1967; Powers, 1954). These complicated results could be attributed to the different day-neutrality sources used, for example, EB ‘Pan-American’ and DN ‘glauca’ in contrast to the European EB source derived from Gloede’s seedling which resulted in monogenic dominant inheritance (Stewart and Folta, 2010). Apart from different day-neutrality sources used, the four dominant alleles with varying flowering strengths (*A’ > A > B > C*) proposed by Powers (1954) may also be interpreted as the allele dosage effect of genes conditioning day-neutrality (Serçe and Hancock, 2005a).
Table 1.1 Proposed inheritance of flowering habits in strawberry.

<table>
<thead>
<tr>
<th>Genetic inheritance</th>
<th>Number of genes</th>
<th>Genetic effect</th>
<th>Genotype</th>
<th>Ploidy</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Monogenic</td>
<td>1</td>
<td>Homozygous recessive (S)</td>
<td>Alpine F. vesca</td>
<td>2x</td>
<td>(Brown and Wareing, 1965)</td>
</tr>
<tr>
<td>Polygenic</td>
<td>3</td>
<td>Recessive</td>
<td>California F. vesca</td>
<td>2x</td>
<td>(Ahmadi et al., 1990)</td>
</tr>
<tr>
<td>Monogenic</td>
<td>1</td>
<td>Homozygous recessive (SFL)</td>
<td>Alpine F. vesca</td>
<td>2x</td>
<td>(Albani et al., 2004)</td>
</tr>
<tr>
<td>Monogenic</td>
<td>1</td>
<td>Dominant</td>
<td>European everbearer F. x ananassa</td>
<td>8x</td>
<td>(Richardson, 1914)</td>
</tr>
<tr>
<td>Polygenic</td>
<td>&gt;1</td>
<td>A major dominant allele with other modifying genes</td>
<td>F. x ananassa</td>
<td>8x</td>
<td>(Clark, 1937)</td>
</tr>
<tr>
<td>Polygenic</td>
<td>3</td>
<td>Three loci with four dominant alleles</td>
<td>F. x ananassa</td>
<td>8x</td>
<td>(Powers, 1954)</td>
</tr>
<tr>
<td>Polygenic</td>
<td>2</td>
<td>Two complementary dominant genes</td>
<td>F. x ananassa</td>
<td>8x</td>
<td>(Ourecky and Slate, 1967)</td>
</tr>
<tr>
<td>Monogenic</td>
<td>1</td>
<td>Dominant</td>
<td>F. x ananassa</td>
<td>8x</td>
<td>(Ahmadi et al., 1990)</td>
</tr>
<tr>
<td>Polygenic</td>
<td>&gt;1</td>
<td>More than one dominant locus</td>
<td>F. x ananassa and F. virginiana</td>
<td>8x</td>
<td>(Hancock et al., 2002)</td>
</tr>
<tr>
<td>Polygenic</td>
<td>&gt;1</td>
<td>More than one dominant locus</td>
<td>F. x ananassa selfed</td>
<td>8x</td>
<td>(Shaw, 2003)</td>
</tr>
<tr>
<td>Monogenic</td>
<td>1</td>
<td>Dominant</td>
<td>F. x ananassa</td>
<td>8x</td>
<td>(Sugimoto et al., 2005)</td>
</tr>
<tr>
<td>Polygenic</td>
<td>&gt;1</td>
<td>More than one dominant locus</td>
<td>F. x ananassa and F. virginiana</td>
<td>8x</td>
<td>(Serçe and Hancock, 2005a)</td>
</tr>
<tr>
<td>Polygenic</td>
<td>&gt;1</td>
<td>A major dominant locus with other minor loci</td>
<td>F. x ananassa</td>
<td>8x</td>
<td>(Shaw and Famula, 2005)</td>
</tr>
</tbody>
</table>
Interestingly, later studies performed using cultivated strawberries and interspecific hybrids between *F. virginiana* and *F. x ananassa* did not fit the one-gene model. Wide ranges in the percentage of DN progeny were observed in these studies, suggesting polygenic inheritance for DN trait in octoploid strawberries (Hancock *et al.*, 2002; Serçe and Hancock, 2005a; Shaw, 2003; Shaw and Famula, 2005). Compared to the monogenic inheritance suggested in some of the earlier studies, these variations could be explained by the application of different evaluation methods, locations and test environments (Shaw, 2003). Despite the contrasting results obtained to date, these studies seem to agree that day-neutrality in octoploid is dominant to short-day flowering response. It may also be hypothesised that day-neutrality in the cultivated strawberry is controlled by a small number of genes with potential allele dosage effects.

### 1.3.2 Other factors affecting flowering response

#### 1.3.2.1 Temperature

In addition to photoperiodism, temperature is the most widely studied factor controlling flower development in strawberry. The effects of interactions between photoperiod and temperature on flowering response have been reported (Bradford *et al.*, 2010; Darrow, 1936; Durner, 1984; Heidi, 1977; Ito and Saito, 1962; Nishiyama *et al.*, 2000; Okimura and Igarashi, 1996; Serçe and Hancock, 2005b). Under optimal photoperiod condition, a SD genotype will initiate flowering when the temperature is below 15 °C (Weebadde *et al.*, 2008). However, lower temperatures favour flowering in SD cultivars even under longer day length (Darrow, 1936). For instance, SD flowering was also observed at temperatures between 12 °C to 18 °C under photoperiods of 14, 16 and 24 hours (Heidi, 1977). In contrast, DN cultivars have a wider range of flowering permissive-temperatures, typically between 14 °C to 26 °C, although some exceptions have been reported (28 °C – 30 °C).
(Bradford et al., 2010; Durner, 1984; Nishiyama et al., 2000). In addition, vernalisation requirement is also noted as one of the factors affecting flowering response (Amasino and Michaels, 2010; Blázquez, 2000; Kim et al., 2009; Mouradov et al., 2002). Generally, 0 °C to 7 °C is required to break bud dormancy prior to flowering season (Piringer and Scott, 1964).

1.3.2.2 Influence of plant growth regulators on flowering

Apart from environmental variation, internal factors such as hormone signalling also play an important role in promoting flowering response. Gibberellin (GA) is the most widely studied plant growth regulator involved in promoting flowering and the GA-dependent pathway has been elucidated in Arabidopsis thaliana (Blázquez et al., 1998; Mutasa-Göttgens and Hedden, 2009; Wilson et al., 1992). For instance, a previous study reported that the GA-deficient (ga1-3) mutant of Arabidopsis thaliana never flowered under short day conditions unless treated with exogenous GA. Cold treatment was shown to promote flowering in the wild type and GA-insensitive mutant (gai) but not in ga1-3, suggesting that GA plays a role in initiating flowering in Arabidopsis (Wilson et al., 1992) and chilling is required to activate GA-regulatory network. This is in accordance to a previous study which reported that exogenous application of GA in strawberry could mimic the effect of long days and chilling (Tehranifar and Battey, 1996). It has been shown that the application of exogenous GA had a greater effect on the DN ‘Seascape’ than the SD ‘Camarosa’ and ‘Laguna’, where flowering was accelerated by reducing the amount of time needed for flower induction (Paroussi et al., 2002). However, the results were not very conclusive as only two different concentrations (50 mg/L and 200 mg/L) were tested in comparison to control (0 mg/L) and the endogenous levels of GA were not considered (Paroussi et al., 2002).
In addition to GA, the putative role of cytokinin has been demonstrated based on the changes in cytokinin fluxes during floral induction (Chen, 1991; Corbesier et al., 2003; Eshghi and Tafazoli, 2007; Lejeune et al., 1988) and the effect of exogenous application of cytokinin on flowering (Al-madhagi et al., 2012; Hoffman et al., 2009). For instance, the activity of endogenous cytokinin was shown to increase only during flower bud differentiation in lychee plants (Chen, 1991), indicating that cytokinin may not be required to break bud dormancy but is needed for flower development. More importantly, short day treatment was found to result in an increase of free cytokinins in the shoot tips and leaves but a decrease in the roots (Eshghi and Tafazoli, 2007). These results further confirmed the shoot-to-root-to-shoot signalling loop suggested by Bernier, Corbesier and Périlleux (2002).

Interestingly, the application of exogenous benzyladenine (BA) prolonged flower induction for more than two months compared to the untreated plants, resulting in “out-of-season” flowering in Protea, a type of flowering plant grown predominantly in South Africa (Hoffman et al., 2009). This result may imply that cytokinin-regulated flowering network is activated when photoperiod and temperature are not permissive for flowering.

The effect of the combination of gibberellins (e.g. GA₃ or GA₄+7) and cytokinins (e.g. benzyladenine, BA) on floral induction and development has also been reported based on exogenous application (Al-madhagi et al., 2012; Naor et al., 2004). However, different results were observed possibly due to the differences in hormone concentrations, time of application, environmental conditions and plant species used in the experiments. Nevertheless, flowering was induced compared to control plants, suggesting there is a crosstalk between GA and BA signalling pathways. Drawing together the findings of previous studies, it may be concluded that hormone signalling, especially gibberellins and
cytokinins, becomes obligatory and plays a major role when the photoperiods and temperatures are not favourable for flower induction.

### 1.3.3 Molecular control of photoperiod-dependent pathway

The molecular control of the photoperiod-dependent pathway was elucidated in *Arabidopsis thaliana* (Blázquez, 2000). Photoperiodism of flowering was triggered by the light-sensing photoreceptors known as phytochromes and cryptochromes involved in the circadian clock function (Kendrick and Kronenberg, 1994). Phytochromes exist in two photo-interconvertible isomeric forms: P<sub>r</sub> and P<sub>fr</sub> that absorb red and far-red light, respectively. At least five phytochromes (PHYA-E) and two cryptochromes (CRY1 and CRY2) that react to red light (600 – 700 nm) and blue light (400 – 500 nm), respectively have been identified in *A. thaliana*. (Lin, 2000). In response to far-red light, PHYA activates the circadian clock to regulate the transcription of the floral promoter CONSTANS (CO) in leaves (Figure 1.5) (Turck et al., 2008). In addition, cryptochromes are also known to induce flowering (Lin, 2000). Conversely, PHYB promotes the degradation of CO protein in response to red light (Figure 1.5) (Blázquez, 2000; Stewart and Folta, 2010). Under long day conditions, the *FLOWERING LOCUS T* (*FT*) gene is activated in leaves by CO protein (Suárez-López et al., 2001; Turck et al., 2008). The FT protein is then transported to shoot apical meristem (SAM) where it forms a complex with the bZIP transcription factor FD (Abe et al., 2005; Corbesier et al., 2007; Jaeger and Wigge, 2007). Together, the complex induces flowering by activating floral-meristem identity genes such as *APETALA1 (AP1), LEAFY (LFY)* and *SUPPRESSOR OF OVEREXPRESSION OF CONSTANS1 (SOC1)* (Amasino and Michaels, 2010; Bernier, 2011; D’Aloia et al., 2011; Liu et al., 2008).
Figure 1.5 A simplified diagram showing genes involved in the molecular control of photoperiod-dependent pathway in *Arabidopsis thaliana*. Adapted and modified from Blázquez (2000).
The effect of red light on flowering under long day condition has been evaluated by specifically directing the red light on the crown of a short day strawberry (Takeda et al., 2008). The results showed that red light treatment caused a significant reduction in flowering under short day conditions, and a high ratio of far-red light to visible light reaching the crown will induce floral bud formation under long day conditions (Takeda et al., 2008). These results suggest that a subset of the photoperiodic flowering pathway genes found in Arabidopsis may be present in strawberry. Several later studies confirmed that PHYA, CO, SOC1 and LFY exist in Fragaria vesca (Davis et al., 2010; Mouhu et al., 2009). Further, the CO gene has been isolated from Fragaria x ananassa. However, no accumulation of CO protein was observed in the DN strawberries, indicating that the regulatory networks in strawberry have unique molecular control (Stewart et al., 2007; Stewart and Folta, 2010).

1.4 MOLECULAR MARKER DEVELOPMENT IN STRAWBERRY

Advancement in molecular genetics and genomics has provided new tools, in the form of ‘molecular markers’ for discovering and tagging alleles and genes (Xu and Crouch, 2008). The application of molecular markers in plant breeding, known as marker-assisted selection (MAS), can improve the efficiency and precision of phenotype prediction (Collard et al., 2005; Collard and Mackill, 2008). Since conventional strawberry breeding process requires 8 – 10 years before a new cultivar can be released (Davis et al., 2007), it is anticipated that the application of molecular markers will accelerate the selection of breeding materials with desired traits and hence, varietal development. However, development of these tools in strawberry has been lagging behind other fruit crops in the Rosaceae family (Arús et al., 2003; Dirlewanger et al., 2004) and plant species such as tomato (Foolad and Panthee, 2012) and rice (Mackill, 2007), mainly due to the
complicated octoploid genome (Chandler et al., 2012). Nevertheless, several classes of molecular markers have been employed in strawberry and their applications have been discussed in a few reviews (Folta and Davis, 2006; Hokanson and Maas, 2001; Sargent et al., 2009; Whitaker, 2011).

1.4.1 Molecular markers
1.4.1.1 Morphological markers and isozymes
Before the advent of molecular marker technology, a few monomorphic morphological markers were used to identify linkage between phenotypic traits controlled by major genes (Sargent et al., 2009). For instance, the non-runnering or runnerless (r), perpetual flowering or semperflorens (s) and yellow/white fruit colour (c) loci were identified in the F₁ and BC₁ progenies by crossing the wild type Fragaria vesca with cultivated Alpine varieties ‘Baron Solemacher’ and ‘Bush White’ (Brown and Wareing, 1965). Other morphological markers reported include an arborea (arb) locus responsible for long-stemmed phenotype (Guttridge, 1973) and a pale-green leaf phenotype (pg) in the diploid F. vesca (Sargent et al., 2004). However, not many monogenic traits have been described in the octoploid Fragaria x ananassa (Davis et al., 2007; Sargent et al., 2009). In addition, a few isozyme markers such as phosphoglucoisomerase (PGI), leucine aminopeptidase (LAP) and phosphoglucomutase (PGM) have been reported in diploid and octoploid strawberries (Arulsekar and Bringham, 1981; Arulsekar et al., 1981; Bringham et al., 1981). These authors provided the first evidence that a single gene could be represented by four distinct loci, when the cultivated octoploid strawberry is diploidised (Davis et al., 2007; Sargent et al., 2009).
1.4.1.2 PCR-based markers

PCR-based marker systems, including random amplified polymorphic DNA (RAPD), amplified fragment length polymorphism (AFLP) and inter-simple sequence repeat (ISSR) have been used to detect DNA variation between *Fragaria* species and varieties (Sargent *et al.*, 2009). The primary use has been for cultivar identification (Arnau *et al.*, 2003; Degani *et al.*, 1998; Tyrka *et al.*, 2002; Žebrowska and Tyrka, 2003) and genetic diversity studies (Debnath *et al.*, 2008; Degani *et al.*, 2001; Graham *et al.*, 1996; Kuras and Korbin, 2004; Morales *et al.*, 2011; Nunes *et al.*, 2013). These arbitrary markers have limited utility due to difficulties in reproducing and interpreting fingerprints between laboratories, and lack of transferability across genotypes (Chandler *et al.*, 2012; Sargent *et al.*, 2009).

In contrast, simple sequence repeats (SSRs, also known as microsatellites) which are derived from primer sequences with known binding sites and amplicon sizes have proven to be more effective and reliable for understanding strawberry genetics (Folta *et al.*, 2011). The transferability of SSR markers was reported to be high (70 % to 100 % in range) between *Fragaria* species and subspecies (Davis *et al.*, 2006; Folta and Davis, 2006) but poor in other species from the closely-related Rosoideae genera (Lewers *et al.*, 2005). Apart from SSR markers, cleaved amplified polymorphic sequence (CAPS) that originate from known coding and intron sequences in strawberry-encoding genes (Konieczny and Ausubel, 1993) are also useful for marker-assisted selection in octoploid strawberries (Folta and Davis, 2006). Both SSR and CAPS markers are superior to arbitrary markers, and are particularly useful for cultivar identification and genetic diversity studies (Gil-Ariza *et al.*, 2009; Honjo *et al.*, 2011; Hosseini *et al.*, 2013; Kunihisa *et al.*, 2003, 2005; Njuguna *et al.*, 2011).
Moreover, gene-specific markers such as sequence-tagged site (STS) have been developed from ADH gene and intron length polymorphisms were detected between F. vesca parental genotypes (Davis and Yu, 1997). In addition, STS markers corresponding to five genes in the anthocyanin biosynthetic pathway have been developed. Interestingly, the Flavanone3-hydroxylase (F3H) has been shown to cosegregate with the c locus (yellow fruit colour), suggesting that a mutation in this gene may be responsible for the yellow fruits of diploid F. vesca spp. vesca ‘Yellow Wonder’ (Deng and Davis, 2001; Sargent et al., 2009).

1.4.2 Linkage map construction

All the markers described in Section 1.4.1 above have been used to develop genetic linkage maps. The first linkage map for diploid strawberry was developed using the F2 population derived from the cross, F. vesca spp. vesca ‘Baron Solemacher’ (BS) x F. vesca spp. americana ‘WC6’ (Davis and Yu, 1997). The application 75 RAPD markers, the STS marker for ADH gene, the Pgi-2 (phosphoglucoisomerase) and Sdh-1 (shikimate dehydrogenase) isozymes, and the yellow fruit colour (c) loci resulted in a 445 cM long genetic linkage map consisting of seven linkage groups (Davis and Yu, 1997). High levels of segregation distortion were noted (47 %), in all cases skewing towards the maternal grandparent ‘Baron Solemacher’. The authors speculated that the segregation distortion was caused by the maternal cytoplasm favouring the transmission of maternal alleles (Davis and Yu, 1997). Subsequently, six anthocyanin pathway genes were assigned using RAPD and morphological markers into five previously defined linkage groups using the F2 population of F. vesca ‘Yellow Wonder’ x F. nubicola. The F3H gene was found to be the likely candidate for c locus and was assigned to linkage group I (Deng and Davis, 2001).
Subsequently, a second generation diploid map based on an interspecific cross between *F. vesca* spp. *vesca* (FV) and *F. nubicola* (FN) was developed using 68 SSR markers, one STS marker, six gene-specific markers and three morphological markers (Sargent *et al.*, 2004). A 448 cM long genetic linkage map covering seven linkage groups ranging from 22.9 cM to 100.3 cM in length was produced. The *pale-green leaf* (*pg*) morphological marker was mapped onto linkage group VI. However, segregation distortion was observed at a high proportion of loci (54%) due to interspecific hybridisation and in some cases, self-incompatibility of *F. nubicola* (Sargent *et al.*, 2004). This map was later expanded by adding new microsatellite loci, resulting in a 182-marker enhanced linkage map, spanning 424 cM over seven linkage groups (Sargent *et al.*, 2006). While this FV x FN map is currently being refined (Folta and Davis, 2006; Sargent *et al.*, 2007), it is used as a reference map for diploid and octoploid strawberry. The map information is available at the Genome Database for Rosaceae (GDR), allowing a comparison with the recently sequenced *F. vesca* draft genome (Jung *et al.*, 2008; Sargent *et al.*, 2011).

The first octoploid strawberry (*F. x ananassa*) map was developed using the F₁ population of ‘Capitola’ x ‘CF1116’ and constructed with 789 AFLP markers using a two-way pseudo-test cross strategy (Lerceteau-Köhler *et al.*, 2003). Due to the difficulty in detecting repulsion phase linkage in polyploids, a two-step mapping procedure was employed, resulting in a female and male map at 1604 cM and 1496 cM in length, respectively (Davis *et al.*, 2007; Folta *et al.*, 2011; Lerceteau-Köhler *et al.*, 2003). Only 3.2% of the markers showed distorted segregation ratios. 28 male and 30 female linkage groups were produced, with the average linkage group length of 53.5 cM and 53.4 cM, respectively (Lerceteau-Köhler *et al.*, 2003; Sargent *et al.*, 2009). This map was further extended using a larger population consisting of 213 individuals and additional AFLP, SCAR and SSR markers.
Despite being incomplete, this map serves as a starting point to study octoploid chromosome behaviour (Rousseau-Gueutin et al., 2008). Comparative mapping between diploid and octoploid strawberry revealed high levels of colinearity between their genomes, suggesting disomic behaviour of allele segregation in cultivated strawberry (Rousseau-Gueutin et al., 2008; Sargent et al., 2009). Two other maps were recently developed to identify QTL controlling agronomic and fruit quality traits in cultivated strawberries, such as yield, anthocyanin, firmness, sugar and organic acid content (Lerceteau-Köhler et al., 2012; Zorrilla-Fontanesi et al., 2011). With the development of new dense octoploid linkage maps, more molecular markers associated with traits of interest can be discovered and applied in strawberry breeding program.

1.4.3 Marker-trait association

The major achievements in marker-trait association studies have been in the area of disease resistance in octoploid strawberries and perpetual flowering habit in diploid strawberries (See Section 1.4.4). Two molecular markers conferring disease resistance have been identified in cultivated strawberries. For instance, application of RAPD markers with Bulked Segregant Analysis (BSA) (Michelmore et al., 1991) has successfully identified a DNA region associated with Rfl1 locus controlling resistance to Phytophthora fragariae, the causative agent for red stele root rot disease (Haymes et al., 1997). However, these RAPD markers have limited application across different populations due to poor transferability and lack of specificity to different variants of pathogens (Van de Weg, 1997). The reproducibility of these arbitrary markers can be increased by converting them into sequence-characterised amplified region (SCAR) markers (Chandler et al., 2012; Sargent et al., 2009). For example, an AFLP marker linked to the Rca2 locus was identified in bulked cultivated material resistant to Colletotrichum acutatum, the cause of
strawberry anthracnose resistance. This AFLP marker was converted into a SCAR marker, where the segregation of the SCAR marker and disease resistance was detected in 81.4% of the accessions screened (Chandler et al., 2012; Lerceteau-Köhler et al., 2005).

As a conclusion for the current development of molecular markers in strawberry, it is worthwhile to note that most of the markers discovered to date are responsible for simple inheritance traits in diploid species. None of the markers have been associated with strawberry flavour and day-neutrality in cultivated strawberries as these complex traits are controlled by more than one gene or quantitative trait loci (QTL). Therefore, more advance biotechnological approaches such as QTL mapping, array-based and sequencing-based techniques are required for the genetic improvement of strawberry flavour and day-neutrality.

1.4.4 Molecular marker development for strawberry flavour improvement

Early studies on the molecular processes associated with strawberry fruit ripening and flavour development have been performed using a cDNA microarray, resulting in the identification of a strawberry alcohol acyltransferase (SAAT) gene (Aharoni et al., 2000; Aharoni and O’Connell, 2002). However, allelic diversity of the SAAT gene was not determined probably due to high DNA sequencing cost at the time of study. A recent study on the apple alcohol acyltransferase (MdAAT1) has successfully revealed 18 different SNP haplotypes generated from four SNPs identified within the apple collection. Haplotype H1 (C-A-C-A) was associated with strongly decreased ester concentrations whereas haplotype H8 (T-G-T-G) was associated with normal to elevated ester concentrations. These results suggested a putative functional causal relationship between MdAAT1 and ester production in apple (Dunemann et al., 2012). Therefore, it would be interesting to identify allelic
diversity of SAAT gene and examine its association with ester production in strawberry, with possibility of converting the polymorphic sites into PCR-based markers for rapid allelotyping.

In addition, the FaNES1 allele isolated by Aharoni et al. (2004) using the same cDNA microarray was recently used to detect the linalool phenotype in strawberry plants (Chambers et al., 2012). The study showed that the FaNES1 allele is absent in any diploid, tetraploid and hexaploid varieties but present in all but three octoploid genotypes, suggesting that FaNES1 allele possibly arose following octoploidisation (Chambers et al., 2012). However, unlike the SNP haplotypes identified within the MdAAT1 allele sequence, the FaNES1 allele can only be associated with the presence/absence of linalool but not with the levels of production in strawberry fruits. One possible explanation could be allele dosage effect, where the concentration of linalool is due to different copy numbers of the FaNES1 allele.

To date, only one octoploid linkage map has been developed to investigate the genetic control of strawberry flavour (Zorrilla-Fontanesi et al., 2012). Compared to the reference map generated by Rousseau-Gueutin et al. (2008), this map only spanned 63.8 % of the reference map, where 39 linkage groups with a cumulative length of 1,400 cM was produced. DNA sequence analysis revealed a 30bp deletion in the promoter sequence of the inactive allele of FaOMT gene. This polymorphism fully cosegregates with the presence/absence of mesifuranne, suggesting that FaOMT is the locus controlling variation in mesifuranne content (Zorrilla-Fontanesi et al., 2012). However, allelic-specific PCR has to be performed to confirm the association of this polymorphism with the production of
mesifuranne, ideally in another population other than the mapping population used in the study.

As most of the octoploid linkage maps developed are incomplete, next generation sequencing (NGS) techniques such as RNA-Seq have been recently used to identify loci controlling volatile compound production (Chambers et al., 2014; Sánchez-Sevilla et al., 2014). Although both studies have successfully identified a fatty acid desaturase (FaFAD1) gene conferring the production of γ-decalactone in strawberry fruits, only one study has reported the development of a PCR-based molecular marker which cosegregates with the presence of γ-decalactone in F₁ and BC₁ populations derived from a ‘Elyana’ x ‘Mara des Bois’ cross, as well as in other cultivars and wild Fragaria accessions (Chambers et al., 2014).

In addition, with the release of a diploid strawberry draft sequence in early 2011 (Shulaev et al., 2010), more attempts can be made to connect genes and the traits they control (Folta, 2013). While most of the flavour-related genes uncovered from the draft sequence have been previously characterised, the discovery of more molecular markers with easily traceable DNA patterns that are associated with a trait of interest are expected to be discovered in the near future (Whitaker, 2011). Nevertheless, molecular markers are not developed to replace chemical fingerprinting, but are more crucial for screening alleles controlling key volatile compounds and day-neutrality to ensure that the breeders do not lose the important alleles due to current breeding practises.
1.4.5 Molecular marker development for day-neutrality in strawberry

In an attempt to identify molecular markers linked to *SEASONAL FLOWERING LOCUS* (*SFL*) in the diploid *F. vesca*, ISSR markers were used to detect DNA polymorphisms in the BC$_1$ progeny derived from the seasonal-flowering *F. vesca* spp. *vesca* (*SFL/SFL*) and the perpetual-flowering *F. vesca* spp. *semperflorens* (*sfl/sfl*) (Albani *et al*., 2004). Three of the ISSR markers were successfully converted into SCAR markers, with SCAR1 and SCAR3 located at 3.0 cM and 1.7 cM at either side of the *SFL* locus, and SCAR2 mapped on the same locus as *SFL* (Albani *et al*., 2004). While these SCAR markers are tightly linked to *SFL* locus, their application in the cultivated strawberry breeding remains limited due to differences in the inheritance of day-neutrality in octoploid strawberries, which appear to be dominant (Shaw and Famula, 2005; Stewart and Folta, 2010).

In contrast to diploid strawberry, no successful molecular markers linked to day-neutrality in octoploid strawberry have been reported. The first published report on molecular markers associated with day-neutrality in octoploid strawberry was provided by Kaczmarska and Hortyński (2002). Only one RAPD marker was identified through BSA of the F$_1$ population segregating for 1:1 remontancy. However, genetic distance between the marker and photoperiodic insensitive was not defined in the study (Kaczmarska and Hortyński, 2002). In a later study, RAPD markers were used to identify molecular markers associated with everbearing (EB) trait in the segregating F$_1$ population derived from a cross between the everbearer ‘Ever Berry’ and June bearer ‘Toyonoka’. The closest markers found were mapped at 11.8 cM and 15.8 cM on either side of the EB gene (Sugimoto *et al*., 2005). Once again, the application of these markers is limited due to weak linkages and the reproducibility problem of RAPD markers (Stewart and Folta, 2010).
Thus far, only one linkage mapping approach has been reported for the identification of QTL for day-neutrality in octoploid strawberry (Weebadde et al., 2008). AFLP markers were used to construct a genetic linkage map using the population derived from a cross between the day-neutral (DN) ‘Tribute’ and the short day (SD) ‘Honeoye’. A consensus map of 1541 cM with 43 linkage groups was produced. The population was used to screen for flowering response in five locations. Eight QTL were detected but only one significant QTL accounting for 22% of the phenotypic variation in one of the locations was identified on LG 6-2, indicating that day-neutrality is likely to be a polygenic trait. Due to the relatively diffuse map, no molecular markers linked to major QTL controlling day-neutrality in octoploid strawberries could be identified (Weebadde et al., 2008).

The development of molecular markers has been hampered by the different sources of day-neutrality and everbearing traits inherited in the modern cultivars. Therefore, considerable efforts have been shifted to identifying homologs of Arabidopsis flowering genes in F. vesca (Mouhu et al., 2009). Although differential expression of the 25 selected genes was not observed between EB and SD genotypes, the putative flowering time genes APETALAI (API) and LEAFY (LFY) were co-regulated during early floral development. API transcript was specifically accumulated at the shoot apices of EB genotypes, suggesting the possibility of API as a marker for floral initiation (Mouhu et al., 2009).

More recently, two studies have reported that continuous flowering in F. vesca may be controlled by the TERMINAL FLOWERING 1 (TFL1) homolog, a flowering repressor. A 2-bp deletion in the first exon of TFL1 introduces a frame shift and thus promotes continuous flowering in F. vesca (Iwata et al., 2012; Koskela et al., 2012). However, these studies were again focused on the diploid F. vesca, and the DNA variation detected will have limited usefulness in the cultivated F. x ananassa. Although PHYTOCHROME A
(PHYA), CONSTANS (CO), SUPPRESSOR OF OVEREXPRESSION OF CONSTANSI (SOCI) and LFY have been identified in *F. x ananassa*, allelic variants associated with flowering habits have not been defined (Davis *et al.*, 2010). Therefore, the development of molecular markers based on these flowering time genes is crucial.

1.4.6 Array-based marker systems

Apart from the cDNA microarray studies performed by (Aharoni *et al.*, 2000), array-based marker systems have not been widely used for the genetic improvement of strawberry. In recent years, single nucleotide polymorphisms (SNPs) have become the marker of choice due to their abundance and uniform distribution throughout the genome (Gupta *et al.*, 2008). With the availability of microarrays, a 90,000 SNP array known as The Axiom® 90K Strawberry Array was announced by Affymetrix Inc. in October 2013. While the application and evaluation of this array requires some time before being adapted by the strawberry research community, other array-based marker systems including single feature polymorphism (SFP), Diversity Array Technology (DArT), diversity suppression subtractive hybridisation array and Subtracted Diversity Array (SDA) remain useful in identifying DNA markers for crop improvement. Table 1.2 provides a comparison of these array-based techniques.
Table 1.2 Comparison of the array-based techniques (Gupta et al., 2008; Niu et al., 2011a; Olarte, 2011).

<table>
<thead>
<tr>
<th>Parameter</th>
<th>SFP</th>
<th>DArT</th>
<th>Diversity SSH-array</th>
<th>SDA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sequence information required</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Restriction enzymes used</td>
<td>Yes if complexity reduction is employed</td>
<td>Yes (Usually one frequent cutter, <em>TaqI/BstNI/HaeIII</em> and one rare cutter, <em>PstI</em>)</td>
<td>Yes (One frequent cutter, <em>RsaI</em>)</td>
<td>Yes (Two frequent cutters, <em>AluI</em> and <em>HaeIII</em>)</td>
</tr>
<tr>
<td>Suppression subtractive hybridisation required</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Probe preparation</td>
<td>Oligonucleotide sequences are designed and synthesised</td>
<td>Selective amplified products of the digested DNA fragments</td>
<td>Subtracted DNA fragments</td>
<td>Subtracted DNA fragments</td>
</tr>
<tr>
<td>Target preparation</td>
<td>Labelled whole genome DNA</td>
<td>Selective amplified products of the digested DNA fragments</td>
<td>Restricted DNA fragments</td>
<td>Restricted DNA fragments</td>
</tr>
<tr>
<td>Polymorphism rate</td>
<td>N/A</td>
<td>3 – 27 %</td>
<td>40.6 - 46.8 %</td>
<td>10.5 – 68 %</td>
</tr>
<tr>
<td>Part of genome surveyed</td>
<td>Whole genome</td>
<td>Whole genome</td>
<td>Subtracted part of the genome</td>
<td>Subtracted part of the genome</td>
</tr>
<tr>
<td>Major application</td>
<td>(1) Marker-trait associations (2) Construction of molecular map followed by QTL interval mapping (3) Gene expression analysis</td>
<td>(1) Whole genome profiling (2) Diversity studies (3) Genetic and physical mapping (4) QTL identification</td>
<td>(1) Species identification (2) Inferring genetic relationships (3) Authentication of fresh and dried herbal material (4) Marker-trait associations</td>
<td></td>
</tr>
</tbody>
</table>
1.4.6.1 Single Feature Polymorphism (SFP)

Single feature polymorphism (SFP) refers to a collection of oligonucleotides or probes printed on the array. Based on the objective of the study, oligonucleotide probes/features can be prepared from genomic DNA or cDNA for sequence polymorphisms (SFP) or expression level polymorphism (ELP) detection, respectively (Figure 1.6) (Gupta et al., 2008).

Figure 1.6 Schematic of the procedure used for detecting SFPs and ELPs. Adapted and modified from Gupta et al. (2008).
SFP technology allows the detection of allelic variations such as SNPs and large deletions based on differences in hybridisation signals between genotypes assessed (Gupta et al., 2008). It has been used in combination with Bulked Segregant Analysis (BSA) for marker-trait associations in *Arabidopsis* (Hazen et al., 2005) and tomato (Zhu and Salmeron, 2007). However, its versatility in plants with complex genomes such as maize, wheat and barley may be limited. Although complexity reduction and gene enrichment methods through the use of cRNA have been applied, the detection of ELP was found to be more effective than SFP (Gore et al., 2007). Therefore, this technique may not be suitable to detect DNA polymorphism in polyploid plants such as the octoploid strawberry.

**1.4.6.2 Diversity Array Technology (DArT)**

Diversity Array Technology (DArT) is a high-throughput microarray platform that allows the genotyping of several hundred polymorphic loci distributed in a genome without prior sequence information (Jaccoud et al., 2001). However, the development of DArT is a cumbersome process. It involves the preparation of a “discovery array” from a pool of individuals representing the germplasm of interest, where genome complexity reduction is performed using a combination of rare (*PstI*) and frequent (*TaqI/BstNI/HaeIII*) cutting enzymes to reduce the level of repetitive DNA sequences (Figure 1.7) (Kilian et al., 2003). Polymorphic clones or DArT markers identified based on differential hybridisation signals between genomic representations of all genomes are then assembled into a “genotyping array”. The DNA polymorphisms detected depend on the type of restriction enzymes used in the complexity reduction method. Generally DArT is capable of detecting polymorphisms due to SNPs at the restriction sites of endonucleases, indels within restriction fragments (Jaccoud et al., 2001) and DNA methylation if a methylation-sensitive enzyme (e.g. *PstI*) is used (Kilian et al., 2003; Wittenberg et al., 2005).
**Figure 1.7** Schematic of DArT development. (A) Generation of a “discovery array”. (B) Generation of a “genotyping array”. Reproduced from Jaccoud *et al.* (2001).

DArT has been used for genome-wide genotyping of plants with high levels of genome complexity, including barley (Wenzl *et al.*, 2004), wheat (Akbari *et al.*, 2006), Eucalyptus (Sansaloni *et al.*, 2010) and sugarcane (Heller-Uszynska *et al.*, 2011) as well as non-model organisms such as fern *Asplenium viride* and moss *Garovaglia elegans* (James *et al.*, 2008). Moreover, only one quantitative BSA has been successfully performed using DArT platform in barley, where the relative signal intensity from different bulks was shown to represent allele frequencies in complex samples (Wenzl *et al.*, 2007). However, one drawback of DArT is the complexity reduction method used, where low polymorphism rates are generated (3 % - 27 %) for most species tested (Akbari *et al.*, 2006; Heller-
Uszynska et al., 2011; James et al., 2008; Wenzl et al., 2004; Yang et al., 2006). This is because the diversity panels usually contain a large number of homologous or monomorphic sequences, where a majority of the features show equal hybridisation between organisms assayed (Li et al., 2006). Therefore, a more effective detection method for DNA polymorphism is required.

### 1.4.6.3 Diversity Suppression Subtractive Hybridisation Array

Diversity SSH-array is an alternative method to DArT. It employs a target enrichment technique called Suppression Subtractive Hybridisation (SSH) coupled with array-based hybridisation. SSH (Appendix 1) is initially used to selectively amplify target cDNA fragments (differentially expressed genes in one population) and simultaneously suppress non-target DNA amplification (genes expressed in both populations) (Diatchenko et al., 1996). The cDNA containing differentially expressed genes is referred as “tester” and the reference cDNA as “driver”. Briefly, the SSH process starts with the digestion of tester and driver DNA with RsaI, a four-base cutting restriction enzyme to generate shorter blunt end cDNA fragments. The tester cDNA is then divided into two samples and ligated with different adaptors at the 5’-ends of the cDNA. Excess RsaI digested driver is added to both the tester pools. Two rounds of hybridisation between tester and driver samples are performed to equalise and enrich the tester-specific cDNA fragments. This is followed by suppression PCR, where the tester-specific cDNA are selectively amplified and enriched, resulting in a subtracted cDNA library enriched with differentially expressed genes (Appendix 1) (Clontech, 2008).

The application of SSH was modified for the development of the subtracted genomic DNA library instead of subtracted cDNA library. For instance, pairwise DNA subtraction
between four *Dendrobium* species was performed using SSH and the resulting subtracted clones were hybridised onto a nylon membrane to produce a diversity suppression subtractive hybridisation array. Profiles of gDNA polymorphisms were obtained by hybridising DIG-labelled gDNA of different species onto the membrane. The average percentage of polymorphic clones based on pairwise comparisons of the four *Dendrobium* species was up to 42.4% (Li *et al.*, 2006). However, the requirement of multiple subtractions between species analysed has impeded further use of this technique, as it is costly and time-consuming to genotype a large number of species (Niu *et al.*, 2011a).

### 1.4.6.4 Subtracted Diversity Array (SDA)

Subtracted Diversity Array (SDA) was developed based on the modified diversity SSH-array, in which an alternative SSH was employed in combination with high-density microarray technology (Jayasinghe *et al.*, 2007). The SDA utilised *Alu*I and *Hae*III instead of *Rsa*I as in the diversity SSH-array. The first prototype array of 376 features was constructed using a broad subtraction approach by pooling the gDNA of 49 angiosperm species (tester) and subtracting this DNA from pooled gDNA of five non-angiosperm species (driver) (Figure 1.8) (Jayasinghe *et al.*, 2007). DNA fingerprinting of six main clades containing 28 species revealed that 68% of the features were polymorphic, and the SDA successfully discriminated species between Asterids, Rosids, Caryophyllids, Ranunculids, Monocots and Eumagnoliids based on the Angiosperm Phylogeny Group 2003 (APG, 2003; Jayasinghe *et al.*, 2007). Subsequently, the SDA technique has been refined to develop clade- and genus-specific arrays for genotyping closely related species, including an Asterid-specific SDA (Mantri *et al.*, 2012) and a *Salvia*- and *Echinacea*-specific SDA (Olarte, 2011).
The SDA technique possesses several advantages over the diversity SSH-array. Firstly, the preparation of genomic representations bypasses the need for making pairwise subtraction between species. As only one single subtraction is required, the SDA technique is more time and cost effective (Niu et al., 2011a). Secondly, unlike the diversity SSH-array that
limits the fingerprinting to genotypes used in the subtraction process, the SDA has higher discriminatory power (due to the pooling of a large number of angiosperm and non-angiosperm species) and allows genotyping of a wide range of species. For instance, the prototype SDA was able to differentiate species not included in the initial SDA development to clade level. For the species used in the preparation of genomic representations, genotyping was successful to the family levels, and to the species level with minor exceptions (Jayasinghe et al., 2009). Moreover, closely related species such as *Panax quinquefolius* and *P. ginseng* were successfully discriminated using the prototype SDA (Niu et al., 2011b).

In addition, high polymorphism rate (50 % - 68 %) is one of the main advantages of SDA compared to DArT (3 % - 27 %) (Niu et al., 2011a). This could be attributed to the SSH process which eliminated most of the highly conserved gDNA sequences between testers and drivers, resulting in a substantial enrichment of polymorphic sequences (Jayasinghe et al., 2007). Moreover, this target enrichment technique also circumvents the need of screening a large number of array features as in DArT, making SDA a more efficient and cost-effective genotyping method independent of sequence information (Jayasinghe et al., 2007). A previous DArT study on sugarcane demonstrated that the combination of SSH and DArT was able to generate higher polymorphism rate (13.3 %) compared to DArT only (9.78 %) (Heller-Uuszynska et al., 2011), suggesting that SDA may be superior to DArT in detecting DNA polymorphism between genotypes. Another important feature of the SDA is its ability to detect novel species-specific fragments which may be potential molecular markers for species identification (Mantri et al., 2012). For instance, chloroplast genes which have been used in DNA barcoding of plant species as well as sequences that resembled ISSR markers were found in the Asterid-specific SDA. These results indicated
that the SDA itself contained polymorphic features representing molecular markers which could be used for species fingerprinting and phylogenetic analysis of plants (Mantri et al., 2012). Moreover, the Salvia- and Echinacea-specific SDA constructed using a narrow subtraction approach also has the ability to discriminate closely related species. Apart from chloroplast genes, both SDA identified novel nuclear-specific molecular markers which correctly clustered Salvia species based on their geographical origins as well as the accessions of the same Echinacea species (Olarte et al., 2013a; Olarte et al., 2013b). These studies demonstrate the potential of the SDA to identify molecular markers associated with bioactive compound production. However, sequence characterisation of the corresponding K2 feature was not determined in the Salvia study, thus limiting its use as a molecular marker associated with the biosynthesis of tanshinones in Salvia species (Olarte, 2011). In addition, variation of signal strength of the features corresponding to lipophilic metabolite production between different Echinacea species implied more of a correlation with specific species rather than a correlation with the content of bioactive compounds (Olarte, 2011). These results indicate that further modification in the SDA methodology is required in order to identify molecular markers associated with traits of interest.

1.5 RATIONALE OF THE STUDY

Breeding strawberry with improved and consistent flavour over an extended period of time is a challenging task for the strawberry breeders (Chandler et al., 2012). Traditional breeding based on the selection of strawberry cultivars with higher yield, larger berry size, firmer fruits, longer shelf life, uniform shape and disease resistance (Davis et al., 2007) has unintentionally caused the loss of fruit flavour (Kader, 2008). One of the current difficulties facing breeders is the lack of a simple and robust tool to select for desired flavour volatiles or to maintain the unique flavour in the germplasm. Moreover, selection
for good flavour by tasting a large number of fruits in the field is impractical. Further, screening for volatile compounds using analytical chemistry methods is extremely costly and labour intensive (Klee, 2010; Whitaker, 2011). In addition, the development of day-neutral cultivars is hindered by the uncertainty of the inheritance of day-neutrality (Hancock and Weebadde, 2008). The assessment of these two polygenic traits is difficult due to confounding environmental factors. Phenotyping of fruit flavour and day-neutrality is time-consuming as it required mature plants for assessing flowering response and the aroma profiles of ripe fruits.

To address these obstacles, the best approach for success is to identify molecular markers associated with fruit flavour and day-neutrality as DNA-based markers are insensitive to environment and phenotyping can be performed without fully grown plants (Kumar et al., 2009; Mahajan and Gupta, 2012). Marker-assisted breeding provides an opportunity to track the alleles or genes through generations by using molecular markers associated with strawberry flavour and day-neutrality (Folta et al., 2011). However, genome complexity of the octoploid strawberry \((2n = 8x = 56)\) has seriously impeded the development of molecular markers in strawberry (Folta and Davis, 2006). Only two PCR-based markers have been developed for linalool (Aharoni et al., 2004) and \(\gamma\)-decalactone (Chambers et al., 2014). Moreover, the discovery of molecular markers closely linked to loci controlling day-neutrality in cultivated strawberry has been unsuccessful.

Consequently, a microarray-based approach known as Subtracted Diversity Array (SDA) (Jayasinghe et al., 2007) may serve as a robust platform for screening for DNA polymorphism in genotypes regardless of the ploidy levels of the species. However, one of the first attempts to associate bioactive compound production with species-specific
polymorphic features found using the Salvia- and Echinacea-specific SDA was of limited success (Olarte et al., 2013a; Olarte et al., 2013b). Therefore, other methods such as Bulked Segregant Analysis (BSA) may be used to complement the SDA in order to increase the chances of identifying molecular markers associated with strawberry flavour and day-neutrality. Considering the gaps in knowledge regarding the molecular control of fruit flavour and day-neutrality in strawberry, and the opportunities for further study identified in this review, the aims of this study were to:

1. Construct a strawberry-specific Subtracted Diversity Array (SDA) enriched with polymorphic DNA sequences.

2. Validate and evaluate the strawberry-specific SDA with 15 strawberry genotypes.

3. Utilise the strawberry-specific SDA and Bulked Segregant Analysis (BSA) approaches to identify molecular markers associated with fruit flavour.

4. Utilise the strawberry-specific SDA and Bulked Segregant Analysis (BSA) approaches to identify molecular markers associated with day-neutrality.

5. Characterise the DNA sequence of the polymorphic features or putative molecular markers identified in the studies.
CHAPTER 2
Construction, Validation and Evaluation of Strawberry-specific Subtracted Diversity Array (SDA)

2.1 INTRODUCTION

The advancement of genomic research and molecular marker development over the past few decades has greatly impacted the field of molecular plant breeding for crop improvement (Agarwal et al., 2008; Collard and Mackill, 2008). In strawberry (Fragaria x ananassa) breeding, DNA-based marker systems such as random amplified polymorphic DNA (RAPD), amplified fragment length polymorphism (AFLP), simple sequence repeat (SSR) and inter-simple sequence repeat (ISSR) have been widely used for cultivar identification, study of genetic diversity (Folta and Davis, 2006; Sargent et al., 2009; Whitaker, 2011) and linkage map construction (Davis and Yu, 1997; Lerceteau-Köhler et al., 2003). However, the disadvantages of these marker systems such as difficulty in scoring the allelic variants, poor reproducibility, prior sequence knowledge requirement and low-throughput have hampered the development of molecular markers associated with traits of interest (Rosyara, 2006). More recently, much attention has been given to single nucleotide polymorphisms (SNPs) due to their abundance and uniform distribution throughout a genome. Novel array-based marker systems including single feature polymorphisms (SFPs), Diversity Array Technology (DArT) and restriction-site associated DNA (RAD) have been developed to discover DNA polymorphism between genotypes, but not limited to SNP genotyping (Gupta et al., 2008). While the completion of Fragaria vesca physical map (Shulaev et al., 2010) promises to expedite marker development in octoploid strawberry, markers linked to loci controlling fruit flavour are yet to be discovered (Whitaker, 2011).
A few linkage maps have been generated for performing QTL studies on agronomic and fruit quality traits (Lerceteau-Köhler et al., 2003; Lerceteau-Kohler et al., 2004; Lerceteau-Köhler et al., 2012; Rousseau-Gueutin et al., 2008; Zorrilla-Fontanesi et al., 2011; Zorrilla-Fontanesi et al., 2012). Of these maps, only one QTL study on strawberry flavour has been performed, in which 70 QTL associated with 48 volatile compounds were detected. An unexpectedly large proportion of phenotypic variation (~ 30 – 93%) was explained by these QTL (Zorrilla-Fontanesi et al., 2012). However, the accuracy of the octoploid linkage maps remains a dispute since generating highly homozygous parents for the traits assayed is challenging as strawberry plants are sensitive to inbreeding depression. Several problems in linkage mapping of highly heterozygous plants have been reported by Maliepaard et al. (1997). For instance, markers may vary in the number of segregating alleles, markers may be dominant or codominant and the linkage phase of marker alleles is usually unknown (Maliepaard et al., 1997). This issue is mainly because of variation between the two parents (Kwon et al., 2013). Moreover, the redundancy of genetic information due to a large number of chromosomes makes it extremely difficult to identify subgenome-specific polymorphism and assign the unique DNA markers to linkage groups (Chandler et al., 2012). Therefore, there is a need to develop a platform with an ability to screen for predictive molecular markers associated with strawberry volatiles without the need for a mapping population.

Integration of Subtracted Diversity Array (SDA) (Jayasinghe et al., 2007) with phenotypic association analysis may be a potential discovery platform for marker-trait association studies. SDA is superior to other array-based techniques as it possesses higher levels of polymorphism compared to e.g., Diversity Array Technology (DArT), which utilises a complexity reduction approach (Heller-Uszynska et al., 2011; Jaccoud et al., 2001; Wenzl
et al., 2004). Moreover, it does not require multiple pair-wise subtractions between organisms to be genotyped as in the diversity Suppression Subtractive Hybridization (SSH) array (Li et al., 2006). It has been validated for its ability to genotype a wide range of flowering plants to clade, order and family level (Jayasinghe et al., 2009; Mantri et al., 2012). The discriminatory power of the SDA between closely related species within the same genus has also been recently confirmed (Niu et al., 2011b; Olarte et al., 2013a; Olarte et al., 2013b). It was predicted that the SDA could be used to develop potential molecular markers associated with chemical constituents in plants if appropriate phenotyping and genotyping was performed (Olarte et al., 2013b).

This chapter describes the construction of a strawberry-specific SDA by performing broad subtraction between strawberry genotypes and non-flowering plants including ferns, conifers, cycads and ginkgo. The rationale for this approach was to utilise the major differences between angiosperms and non-angiosperms as the baseline for this research. Strawberry is a flowering plant which produces seeds and fruits while gymnosperms such as conifers, cycads and ginkgo have no flowers or fruits but produce naked seeds for reproduction. Ferns are distantly related to angiosperms because they produce spores instead of seeds. Since aroma development only takes place during fruit maturation (Cumplido-Laso et al., 2012; Gonzalez et al., 2009; Pérez et al., 1996a; Zhang et al., 2011), this leads to the assumption that the loci controlling volatile compound synthesis are related to fruit formation. In theory, subtracting strawberry genomic DNA (gDNA) from the non-flowering plants gDNA could selectively capture DNA fragments related to fruit characteristic including aroma development. It was hypothesised that only the DNA fragments associated with flower, seed and fruit formation will remain in the subtracted pool while a large portion of highly conserved sequences between the strawberry genome
and non-flowering plants will be eliminated. This approach utilised the strawberry-specific SDA to find polymorphic sequences between 15 strawberry genotypes based on hybridisation intensities. The ability of the strawberry-specific SDA to identify molecular markers associated with fruit flavour was also examined by correlating the SDA data and aroma profiles of five selected strawberry genotypes.

The objectives of this study were to:

1. Construct and validate a strawberry-specific SDA enriched with strawberry DNA sequences associated with fruit and flavor development.
2. Evaluate the performance of the SDA by fingerprinting 15 strawberry genotypes.
3. Analyse the potential of the SDA for discovering markers associated with strawberry flavour.
2.2 MATERIALS AND METHODS

2.2.1 Plant material

Five strawberry (*Fragaria x ananassa*) genotypes including Californian (Albion) and Australian (Juliette) cultivars and three promising breeding lines (07-102-41, 07-095-35 and 04-069-91) were used as tester in the subtracted genomic library construction. Of the three breeding lines, 07-102-41 and 07-095-35 are known to have the genetic background of European strawberry whereas 04-069-91 may contain genetic information from the Japanese strawberry. This selection of genotypes ensured the inclusion of a broad range of genetic backgrounds to increase the discriminatory power of the Subtracted Diversity Array (SDA) (Jayasinghe et al., 2009). The commercial cultivars were collected from a strawberry farm in Coldstream, Victoria whereas the three breeding lines were sampled from the strawberry breeding station located in Wandin North, Victoria (Table 2.1) on 8th August 2011. In addition, 10 other strawberry genotypes were added onto the array to evaluate the ability of the SDA to fingerprint genotypes not included in the SDA development. Four Australian (Adina, Alinta, Lowanna and Melba), three Californian (Chandler, Camino Real and San Andreas), one Japanese (Hokowase), one European (Cambridge Rival) cultivars and one diploid strawberry (Fraises Des Bois) were sourced from different farms and an accredited nursery (Digger’s Club, Australia) (Table 2.1). For each genotype, a total of five young leaves were picked, sealed in zip-lock bags and immediately frozen on dry ice to avoid any degradation during transportation to the laboratory. They were then stored at -80 °C freezer until DNA isolation.

Nine non-angiosperms including four ferns (*Dryopteris kuratae, Dicksonia antarctica, Asplenium australasicum, Blechnum tabulare*), three coniferous trees (*Wollemia nobilis, Cupressus macrocarpa, Juniperus communis*), one cycad (*Cycas revoluta*) and one ginkgo
(Ginkgo biloba) served as driver for subtraction against the strawberry-specific tester pool. They were either purchased commercially from Bunnings Warehouse and Digger’s Club or collected from the Medicinal Plant Herbarium, Southern Cross University and around the RMIT Bundoora campus (Table 2.1). Leaf tissues were harvested from these plants on the day of DNA extraction.

Of all the strawberry genotypes listed in Table 2.1, three cultivars (Albion, Juliette and San Andreas) and two breeding lines (07-102-41 and 04-069-91) were chosen for chemical fingerprinting based on fruit availability for that sampling season. The fruits of these genotypes were harvested at the fully ripened stage on 27th November, 6th and 20th December of 2012 and immediately frozen on dry ice to terminate postharvest enzymatic reactions. They were kept at – 80 °C upon transportation to the laboratory.
Table 2.1 Strawberry genotypes and non-angiosperm species used in the subtracted genomic library construction and array validation.

<table>
<thead>
<tr>
<th>Plant Materials</th>
<th>Sources</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Strawberry Genotypes</strong></td>
<td></td>
</tr>
<tr>
<td>Australian</td>
<td></td>
</tr>
<tr>
<td>Adina</td>
<td>Wandin North</td>
</tr>
<tr>
<td>Alinta</td>
<td>Wandin North</td>
</tr>
<tr>
<td>Juliette</td>
<td>Coldstream</td>
</tr>
<tr>
<td>Lowanna</td>
<td>Wandin North</td>
</tr>
<tr>
<td>Melba</td>
<td>Coldstream</td>
</tr>
<tr>
<td>USA</td>
<td></td>
</tr>
<tr>
<td>Albion</td>
<td>Coldstream</td>
</tr>
<tr>
<td>Camino Real</td>
<td>Coldstream</td>
</tr>
<tr>
<td>Chandler</td>
<td>Digger’s Club</td>
</tr>
<tr>
<td>San Andreas</td>
<td>Coldstream</td>
</tr>
<tr>
<td>Japanese</td>
<td></td>
</tr>
<tr>
<td>Hokowase</td>
<td>Digger’s Club</td>
</tr>
<tr>
<td>European</td>
<td></td>
</tr>
<tr>
<td>Cambridge Rival</td>
<td>Digger’s Club</td>
</tr>
<tr>
<td>Breeding lines</td>
<td></td>
</tr>
<tr>
<td>07-102-41</td>
<td>Wandin North</td>
</tr>
<tr>
<td>07-095-35</td>
<td>Wandin North</td>
</tr>
<tr>
<td>04-069-91</td>
<td>Wandin North</td>
</tr>
<tr>
<td>Wild Strawberry</td>
<td></td>
</tr>
<tr>
<td>Fraises Des Bois*</td>
<td>Digger’s Club</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Non-angiosperms</th>
<th>Sources</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ferns</td>
<td></td>
</tr>
<tr>
<td>Dryopteris kuratae</td>
<td>Bunnings Warehouse</td>
</tr>
<tr>
<td>Dicksonia antarctica</td>
<td>RMIT Bundoora</td>
</tr>
<tr>
<td>Asplenium australasium</td>
<td>Bunnings Warehouse</td>
</tr>
<tr>
<td>Blechnum tabulare</td>
<td>Bunnings Warehouse</td>
</tr>
<tr>
<td>Conifers</td>
<td></td>
</tr>
<tr>
<td>Wollemia nobilis</td>
<td>RMIT Bundoora</td>
</tr>
<tr>
<td>Cupressus macrocarpa</td>
<td>Bunnings Warehouse</td>
</tr>
<tr>
<td>Juniperus communis</td>
<td>Medicinal Plant</td>
</tr>
<tr>
<td>Cycad</td>
<td>Southern Cross University</td>
</tr>
<tr>
<td>Cycas revoluta</td>
<td>Bunnings Warehouse</td>
</tr>
<tr>
<td>Ginkgo</td>
<td></td>
</tr>
<tr>
<td>Ginkgo biloba</td>
<td>Digger’s Club</td>
</tr>
</tbody>
</table>

*Fragaria vesca (diploid)
2.2.2 Construction of the strawberry-specific SDA

2.2.2.1 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.

2.2.2.2 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 as described in Olarte et al. (2013b). The workflow of SDA construction and performance validation is illustrated in Figure 2.1.
Figure 2.1 The workflow of strawberry-specific SDA construction and performance validation. Adapted and modified from Niu et al. (2011a).
Prior to subtraction, genomic representations were prepared by pooling equal amount of DNA extracted from five strawberry genotypes and nine non-angiosperm species (Table 2.1) 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 AluI and HaeIII 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.
2.2.2.3 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 MgCl₂, 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 as described in Section 2.2.2.3. 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 (Appendix 2).

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 (Appendix 3). 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 (Appendix 4). 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.

### 2.2.3 Validation of the strawberry-specific SDA

#### 2.2.3.1 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).

2.2.3.2 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™ streptavidin-labelled Cy®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 NaH₂PO₄.2H₂O 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 CyTM5) 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.

### 2.2.3.3 Scanning and image quantification

Slides were scanned in a ScanArray Gx 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.

2.2.4 Analysis of the strawberry-specific SDA

2.2.4.1 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

   \[
   \text{Normalisation factor} = \frac{A}{B}
   \]

   where
   
   \[A = \text{Mean SNR of all the features in all the six technical replicates and two biological replicates across 15 genotypes}\]
   \[B = \text{Mean SNR of all the features in all the six technical replicates for one genotype}\]

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.

2.2.5 **Aroma profiling of strawberry genotypes**

2.2.5.1 **Solid Phase Microextraction (SPME)**

To assess the ability of the SDA in finding DNA markers related to strawberry flavour, phenotyping of the aroma profiles of Albion, Juliette, San Andreas, 07-102-41 and 04-069-91 was performed. Volatile compounds were extracted from the fruit puree using Solid Phase Microextraction (SPME) method (Pawliszyn and Pawliszyn, 1997). Five large berries 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 2.3). 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 2.3** Extraction of volatile compounds from strawberry purees using SPME method.

### 2.2.5.2 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.

2.2.5.3 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:

\[ C_i = \frac{A_i}{A_t} \times 100\% \]

Where \( C_i \) = Content of a compound in the sample
\[ A_i = \text{Area of compound peak in the chromatogram} \]
\[ A_t = \text{Total area of the peaks in the chromatogram} \]

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.
2.2.6 Correlation between chemical and molecular data

2.2.6.1 Comparison of aroma profiles and SDA fingerprints

To compare the chemical and genetic relationship between genotypes assayed, two
dissimilarity dendrograms were generated with IBM SPSS Statistics v. 21 using the
average-linkage-between-groups method and Squared Euclidean distance. For hierarchical
cluster analysis based on aroma profile, mean relative compositions of all the volatile
compounds were employed for each strawberry genotype. Conversely, the normalised
mean SNR of the entire dataset was used to perform another hierarchical analysis based on
SDA fingerprints. The relationship between the aroma profile and genetic variability can
be deduced by comparing the clustering patterns of the dendrograms.

2.2.6.2 Correlation of branch point DNA markers with volatile compounds

To further investigate the ability of the SDA to identify DNA markers associated with
flavour trait, correlation analysis was performed between a subset of the SDA fingerprints
(i.e. the branch point DNA markers selected in Section 2.2.4.3) and the key volatile
compounds detected by the SPME-GC-MS. The mean relative compositions of the key
volatile compounds were correlated with the normalised mean SNR of the branch point
DNA markers using Pearson’s bivariate correlation (IBM SPSS Statistics v. 21) and linear
regression analysis (Microsoft Excel). Only the findings showing significant correlation
between the branch point DNA markers and volatile compounds were presented in the
results and discussion.

2.2.7 Sequencing of branch point markers

The *E. coli* cultures containing plasmids flanking DNA inserts corresponding to the branch
point DNA markers were recovered from their respective glycerol stocks. Plasmids were
isolated from each culture using QIAprep Spin Miniprep kit (Qiagen, Valencia, CA) following manufacturer’s manual. DNA inserts were amplified using the T7 and SP6 sequencing primers according to the protocol stated in the Promega Subcloning Notebook (Promega, San Luis Obispo, CA). Prior to sequencing, the quality and integrity of PCR products were examined on a 1.0 % TBE agarose gel to ensure that only a single band was present in the PCR products. Forward and reverse sequencing was performed for each branch point DNA marker by Macrogen Inc. (Korea). The DNA sequences were trimmed to remove the vector backbone and nested primer sequences used in the library construction. Similarity search was performed against the Fragaria vesca draft genome (v1.1) using PFR Strawberry Server at https://strawberry.plantandfood.co.nz/ (PFR, 2010) and reconfirmed with the nucleic acid and protein database at the National Center of Biotechnology Information (NCBI) using blastn and blastx programs (Altschul et al. 1997). The DNA sequences with an E-value < 1e^-5 were considered significant. Subsequently, genes located within 5 centiMorgan (cM) on either side of the putative DNA markers were manually searched using the same strawberry genome server based on previously mapped genes available in the Strawberry Genbank and the general RefSeq mRNA database. By assuming the genetic length of a chromosome as 100 cM (Kearsey and Pooni, 1996), the physical distance covering 5 cM was calculated following the equation below:

\[
\text{Physical distance (Mb)} = \frac{\text{Physical length of the linkage group (Mb)}}{100 \text{ cM}} \times 5 \text{ cM}
\]
2.3 RESULTS AND DISCUSSION

2.3.1 Genomic DNA extraction and quantification

The intensity and integrity of purified DNA from 15 strawberry genotypes and nine non-angiosperm species are illustrated in Figure 2.4. The results showed intact bands for sample 1 – 19 and sample 22 – 24, indicating the purified DNA was of good quality with no degradation. In contrast, only faint bands were observed on the gel for *Blechnum tabulare* (sample 20) and *Wollemia nobilis* (sample 21), thus (Figure 2.4) due to low DNA concentrations (Table 2.2).

![Figure 2.4](image)

Figure 2.4 Electrophoresis of genomic DNA isolated from (a) 15 strawberry genotypes and (b) nine non-angiosperm species on a 1.0 % agarose/EtBr gel. Lane M: 1 kb DNA ladder. Lane 1: Juliette; Lane 2: Alinta; Lane 3: Chandler; Lane 4: Lowanna; Lane 5: Albion; Lane 6: Melba; Lane 7: 07-102-41; Lane 8: 07-095-35; Lane 9: Adina; Lane 10: Fraises Des Bois; Lane 11: 04-069-91; Lane 12: San Andreas; Lane 13: Cambridge Rival; Lane 14: Hokowase; Lane 15: Camino Real; Lane 16: *Cycas revoluta*; Lane 17: *Asplenium australasicum*; Lane 18: *Ginkgo biloba*; Lane 19: *Juniperus communis*; Lane 20: *Blechnum tabulare*; Lane 21: *Wollemia nobilis*; Lane 22: *Dryopteris kuratae*; Lane 23: *Dicksonia antarctica*; Lane 24: *Cupressus macrocarpa*.

In general, many methods have been developed based on the Doyle and Doyle (1990) protocol to isolate quality DNA from various plant species. A modified CTAB method was
initially employed in this study to isolate DNA from strawberry leaves. However, this was unsuccessful as it resulted in purified DNA that was too viscous for pipetting due to the high polysaccharide and polyphenol content of the berry plants (Mercado et al., 1999; Porebski et al., 1997). Moreover, DNA extraction using the conventional CTAB method alone was also problematic for non-angiosperms such as ferns (Dempster et al., 1999); conifers (Bashalkhanov and Rajora, 2008; Telfer et al., 2013) and ginkgo (Maltas et al., 2011) as they contain high levels of secondary metabolites. The application of the purification columns (QiaShredder™ spin columns) and the DNA binding columns (DNeasy spin columns) in the Qiagen™ DNeasy® Plant Mini Kit (Qiagen, Valencia, CA) eliminated the need for phenol-chloroform extraction and alcohol precipitation steps and also efficiently removed contaminants such as proteins, polysaccharides and polyphenols from the leaves. DNA extraction with this kit yielded clean DNA free of colour and viscous substances.

Spectrophotometric assessment of the concentration and purity of the purified DNA is shown in Table 2.2. The DNA concentrations varied between different plants, with Cambridge Rival showing the highest concentration at 106.6 ± 3.9 ng/µL and Wollemia nobilis displaying the lowest concentration at 12.1 ± 0.6 ng/µL per 100mg of fresh tissues. The rest of the plants tested demonstrated different concentrations ranging from 14.0 ± 0.1 ng/µL to 101.5 ± 2.6 ng/µL. The low standard deviation between two replicates implied that the spectrophotometer readings were consistent and reliable. The 260/280 absorbance ratio in the range of 1.70 – 1.82 indicated that these samples were of high purity and can be used for the subtracted library construction.
<table>
<thead>
<tr>
<th>Sample</th>
<th>DNA concentration (ng/µL)</th>
<th>260/280 ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Albion</td>
<td>27.6 ± 4.3</td>
<td>1.70 ± 0.04</td>
</tr>
<tr>
<td>Juliette</td>
<td>56.1 ± 3.7</td>
<td>1.71 ± 0.06</td>
</tr>
<tr>
<td>07-102-41</td>
<td>87.3 ± 6.5</td>
<td>1.77 ± 0.04</td>
</tr>
<tr>
<td>07-095-35</td>
<td>38.3 ± 0.3</td>
<td>1.74 ± 0.01</td>
</tr>
<tr>
<td>04-069-91</td>
<td>67.3 ± 0.3</td>
<td>1.76 ± 0.01</td>
</tr>
<tr>
<td>San Andreas</td>
<td>80.2 ± 0.6</td>
<td>1.71 ± 0.01</td>
</tr>
<tr>
<td>Camino Real</td>
<td>18.3 ± 0.6</td>
<td>1.79 ± 0.04</td>
</tr>
<tr>
<td>Chandler</td>
<td>29.0 ± 0.6</td>
<td>1.75 ± 0.01</td>
</tr>
<tr>
<td>Cambridge Rival</td>
<td>106.6 ± 3.9</td>
<td>1.82 ± 0.01</td>
</tr>
<tr>
<td>Hokowase</td>
<td>62.9 ± 0.7</td>
<td>1.77 ± 0.01</td>
</tr>
<tr>
<td>Adina</td>
<td>16.0 ± 0.1</td>
<td>1.79 ± 0.01</td>
</tr>
<tr>
<td>Alinta</td>
<td>40.0 ± 6.8</td>
<td>1.74 ± 0.15</td>
</tr>
<tr>
<td>Lowanna</td>
<td>14.6 ± 0.1</td>
<td>1.78 ± 0.01</td>
</tr>
<tr>
<td>Melba</td>
<td>14.0 ± 0.1</td>
<td>1.72 ± 0.03</td>
</tr>
<tr>
<td>Fraises Des Bois</td>
<td>29.1 ± 2.3</td>
<td>1.73 ± 0.06</td>
</tr>
<tr>
<td><em>Dryopteris kuratae</em></td>
<td>38.7 ± 0.3</td>
<td>1.70 ± 0.01</td>
</tr>
<tr>
<td><em>Wollemia nobilis</em></td>
<td>12.1 ± 0.6</td>
<td>1.71 ± 0.06</td>
</tr>
<tr>
<td><em>Dicksonia antarctica</em></td>
<td>37.7 ± 0.1</td>
<td>1.70 ± 0.01</td>
</tr>
<tr>
<td><em>Cupressus macrocarpa</em></td>
<td>22.8 ± 0.1</td>
<td>1.80 ± 0.01</td>
</tr>
<tr>
<td><em>Cycas revoluta</em></td>
<td>99.9 ± 0.5</td>
<td>1.75 ± 0.01</td>
</tr>
<tr>
<td><em>Asplenium australasicum</em></td>
<td>101.5 ± 2.6</td>
<td>1.79 ± 0.01</td>
</tr>
<tr>
<td><em>Blechnum tabularae</em></td>
<td>16.9 ± 5.2</td>
<td>1.71 ± 0.06</td>
</tr>
<tr>
<td><em>Ginkgo biloba</em></td>
<td>27.5 ± 0.2</td>
<td>1.75 ± 0.01</td>
</tr>
<tr>
<td><em>Juniperus communis</em></td>
<td>33.4 ± 0.1</td>
<td>1.82 ± 0.01</td>
</tr>
</tbody>
</table>

The values represent mean DNA concentration ± standard deviation and absorbance ratio ± standard deviation.
2.3.2 Subtracted genomic DNA library construction

In this study, verification experiments were performed to determine the efficacy of the SSH at each stage. The adaptor ligation efficiency is shown in Figure 2.5. The intensity of PCR products using one gene-specific primer (G3PDH 3’ Primer and PCR Primer 1) and two gene-specific primers (G3PDH 3’ and 5’ Primers) differed by 2-3 fold, indicating that at least 25% of the DNA fragments contained different adaptors at both ends (Figure 2.5). It was concluded that the ligation was complete and the adaptor-ligated tester pool will allow selective amplification of strawberry-specific DNA sequences in subsequent suppression PCR steps.

In addition, the success of tester-driver subtraction could be deduced from secondary PCR amplification (Figure 2.6a) and the reduction of G3PDH (Figure 2.6b). The subtracted human skeletal muscle (Lane 3) and the PCR control subtracted cDNA (Lane 5) showed major bands corresponding to ΦX174/HaeIII fragments (Lane M), suggesting that the subtraction between tester and driver pools was complete (Figure 2.6a). The subtraction efficiency was further confirmed by comparing the abundance of spike-in G3PDH before and after subtraction. A faint band was observed after 28-33 cycles for the subtracted sample. A product was detected after 23 cycles for the unsubtracted sample and the intensity of the band gradually increased with the number of cycles (Figure 2.6b). This result indicated that the amount of excess driver DNA added into the tester pool was sufficient to suppress the common DNA sequences during the first hybridisation and further enriched the tester-specific DNA sequences in the second hybridisation.
Figure 2.5 Results of adaptor ligation efficiency analysis on a 2.0 % agarose/EtBr gel. Lane M1: ΦX174/HaeIII DNA ladder. Lane M2: 1kb DNA ladder. Lane 1: PCR products using Adaptor-1 ligated tester as the template and the G3PDH 3’ Primer and PCR 1 Primer; Lane 2: PCR products using Adaptor-1 ligated tester as the template and the G3PDH 3’ and 5’ Primers; Lane 3: PCR products using Adaptor-2R ligated tester as the template and the G3PDH 3’ Primer and PCR 1 Primer; Lane 4: PCR products using Adaptor-2R ligated tester as the template and the G3PDH 3’ and 5’ Primers; Lane 5: PCR products using Adaptor-1 ligated skeletal muscle cDNA as the template and the G3PDH 3’ Primer and PCR 1 Primer; Lane 6: PCR products using Adaptor-1 ligated skeletal muscle cDNA as the template and the G3PDH 3’ and 5’ Primers; Lane 7: PCR products using Adaptor-2R ligated skeletal muscle cDNA as the template and the G3PDH 3’ Primer and PCR 1 Primer; Lane 8: PCR products using Adaptor-2R ligated skeletal muscle cDNA as the template and the G3PDH 3’ and 5’ Primers; Lane 9: Negative control.
Figure 2.6 Results of subtraction efficiency analysis on a 2.0 % agarose/EtBr gel. (a) Secondary PCR products of the subtracted strawberry pool. Lane M: ΦX174/HaeIII DNA ladder; Lane 1: PCR products of subtracted experimental samples; Lane 2: PCR products of unsubtracted experimental control; Lane 3: PCR products of subtracted human skeletal muscle tester cDNA with 0.2 % ΦX174/HaeIII-digested DNA; Lane 4: PCR products of unsubtracted skeletal muscle tester cDNA ligated with both Adaptors 1 and 2R and containing 0.2 % ΦX174/HaeIII-digested DNA. Lane 5: PCR control subtracted cDNA. Lane 6: Negative control. (b) Reduction of spike-in G3PDH abundance in the experimental sample by PCR-Select subtraction on a 2.0 % agarose/EtBr gel. Lane M: ΦX174/HaeIII DNA ladder; Lanes 1 and 5: 18 cycles; Lanes 2 and 6: 23 cycles; Lanes 3 and 7: 28 cycles; Lanes 4 and 8: 33 cycles.

Following T/A cloning, the cloned secondary PCR products were detected by blue white screening and the DNA inserts were screened using colony PCR. Figure 2.7 presents an example of DNA inserts amplified from pGEM®-T Easy vector using nested PCR primer 1 and nested PCR primer 2R. Out of 331 randomly picked white colonies, 290 colonies
showed single DNA band ranging in size from 250 bp to 1,000 bp, suggesting 88% successful cloning rate. The remaining 41 colonies represented either clones with an insert size of more than 1 kb, unspecific PCR products or no DNA inserts. False positives were observed in these colonies due to non-functional LacZ gene or self-ligated vectors (Padmanabhan et al., 2011).

**Figure 2.7** Example of DNA inserts amplified from pGEM®-T Easy vector using nested PCR primer 1 and nested PCR primer 2R on a 1.0 % agarose/EtBr gel. Lane M: 1 kb DNA ladder. Lane 1-48: DNA inserts. Lane 4: DNA inserts more than 1 kb. Lane 21, 26 and 38: unspecific PCR products. Lane 13, 15 and 17: false positive colonies.
All of the 290 PCR products were deposited onto aminosilane coated slides as probes to generate the strawberry-specific SDA. The concentration of the DNA to be arrayed played an important role in ensuring successful probe-target hybridization. Corning Incorporated (NY, USA) recommended that a minimum of 200 ng/µL of arrayed materials be deposited onto aminosilane-coated slides. Similarly, 200 ng/µL was determined as the optimum DNA concentration by examining the signal reproducibility within experiments and among experiments in a previous study (Li et al., 2002). To avoid loss of sensitivity during hybridisation and image scanning (Hegde et al., 2000), DNA concentration was normalised to 250 ng/µL in this study.

2.3.3 Strawberry-specific SDA validation

The strawberry-specific SDA was validated by separately hybridising the gDNA pool of five different strawberry genotypes (tester) and the gDNA pool of nine non-angiosperm species (driver) onto the SDA. Figure 2.8 illustrates the different hybridisation patterns between the tester (a) and driver (b) targets. The printing controls and the spike-in control (shown in the yellow box) in both tester and driver samples exhibited high signal intensities. This suggested that printing and hybridisation was optimised between the probes and targets. More importantly, the subtracted PCR mixture only hybridised with the tester sample and not with the driver sample, revealing that the probes printed on the slides were tester-specific strawberry DNA sequences. In addition, most of the features produced different fluorescence intensity, indicating that the stringency wash and image scanning process was sufficient to distinguish between polymorphic DNA sequences (Figure 2.8a). It is also important to note that no positive hybridisation for negative controls was observed in both tester and driver targets, suggesting that no non-specific hybridisation occurred and increasing the stringency of the hybridisation and washing was not required.
**Figure 2.8** Validation of strawberry-specific SDA based on the hybridisation patterns of (a) tester and (b) driver samples. The features in the yellow boxes correspond to the positive controls. The features in the green, blue and purple circles shown in (b) represent the non-subtracted DNA fragments which were excluded in further data analysis.
In theory, a perfect DNA subtraction should result in no positive hybridisation between the probes and the driver target. However, three (FLP1G12, FLP4D5 and FLP4C2) out of 290 features showed hybridisation with driver target, suggesting 99.0 % subtraction success rate (Figure 2.8b). This could possibly be due to the random event that occurs during the second hybridisation step where a fragment of driver DNA happens to be present in the subtracted mixture (Clontech, 2008). However, considering that the abundance of the G3PDH gene in the spike-in human skeletal muscle cDNA was successfully reduced (Figure 2.6b), most of the driver DNA was eliminated during suppression PCR, resulting in only 1 % of non-subtracted DNA fragments in the secondary PCR products. These three non-subtracted DNA features were therefore discarded from any downstream analyses.

The high subtraction efficiency in this study was achieved by using 1:60 tester:driver ratio as recently described for the Echinacea-specific SDA (97.0 %) (Olarte et al., 2013b), which was an improvement over the Salvia-specific SDA (88.0 %) where 1:30 tester:driver ratio was used (Olarte et al., 2013a). The subtraction efficiency of strawberry-specific SDA is also comparable to the first prototype angiosperm SDA (97.0 %) and Asterids-specific SDA (99.6 %) (Jayasinghe et al., 2007; Mantri et al., 2012). Nevertheless, direct comparison should not be established between the strawberry-specific SDA with the Salvia-specific SDA because the current study utilised a broader subtraction approach whereas the Salvia-specific SDA was constructed using closely related species.

Interestingly, 28 out of 290 features did not hybridise with the strawberry-specific gDNA pool. One possible explanation is the ‘dilution effect’ reported by Jayasinghe et al. (2007) where DNA fragments with low frequency in the plant genome can remain undetected if the target gDNA pool consists of more than one plant genotype. When comparing the
tester and driver hybridisation results to the SDA profiles obtained from the hybridisation of a single strawberry genotype, these features were found to hybridise with either one or more genotypes (data not shown). This observation further supports the claim of ‘dilution effect’. It is important to note that these low copy number DNA sequences could represent potential cultivar or genotype-specific DNA markers, if not related to flavour trait in strawberry. It would be worthwhile to sequence these 28 features to determine their identities in the future. To avoid eliminating any useful subtracted array features, these features were retained during data analysis, resulting in a set of 287 potential polymorphic DNA sequences on the strawberry-specific SDA.

2.3.4 Microarray data analysis

There are three types of methods that may be used for data normalisation. The most commonly used method is based on all the genes on the array, followed by a set of housekeeping genes and finally the application of spike-in controls (Bowtell and Sambrook, 2003; Draghici, 2003). Due to the large dataset consisting of 51,660 features in total (287 features x 2 biological replicates x 6 technical replicates x 15 genotypes), the first normalisation approach based on all genes on the array was selected for analysis. It was chosen based on the assumption that “all” the genes in the strawberry genome were represented in the SDA (Draghici, 2003). Therefore, data normalisation was performed by taking the normalised mean SNR of all the features in all the hybridisations as outlined in Section 2.2.4.2.

Unlike most of the SDA-based (Jayasinghe et al., 2007; Jayasinghe et al., 2009; Mantri et al., 2012; Niu et al., 2011b) and DArT-based (Jaccoud et al., 2001; Sansaloni et al., 2010; Wenzl et al., 2004) genomic profiling studies, the current SDA data was not converted to
binary scores. This is mainly because it did not show a clear presence (1) or absence (0) of hybridisation patterns. Due to the high level of heterozygosity in octoploid strawberry, most of the features exhibited different signal intensities with respect to the copies of DNA fragments hybridised to each feature ranging from strong, intermediate to low. This result is similar to the recent *Salvia*-specific SDA study which revealed that application of dominant scoring to the dataset resulted in poor data reproducibility (Olarte *et al.*, 2013a). Moreover, a genomic polymorphism study on *Eucalyptus grandis* also suggested that direct comparison of signal intensity profiles may allow more accurate identification of individuals compared to binary scores in out-crossing species (Lezar *et al.*, 2004). To avoid eliminating informative features such as the co-dominant DNA markers, the normalised mean SNR was therefore used directly in all further analyses.

2.3.5 Genetic relationships among strawberry genotypes

A hierarchical dendrogram was generated based on the normalised mean SNR of all 287 microarray features (Figure 2.9). The relatedness of these genotypes was compared with known parentage information (Table 2.3). 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 2.9). This result indicates that the SDA is 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 2.9).

Table 2.3 The strawberry genotypes used in this study, their immediate parents and country of origin.

<table>
<thead>
<tr>
<th>Genotypes</th>
<th>Source</th>
<th>Parents</th>
<th>Origin</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Australian</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adina</td>
<td>Wandin</td>
<td>Pajaro x 88-042-35</td>
<td>Victoria</td>
</tr>
<tr>
<td>Alinta</td>
<td>Wandin</td>
<td>Chandler x 88-011-30</td>
<td>Victoria</td>
</tr>
<tr>
<td>Juliette</td>
<td>Coldstream</td>
<td>Adina x 92-50-76</td>
<td>Victoria</td>
</tr>
<tr>
<td>Lowanna</td>
<td>Wandin</td>
<td>Selva x 89-064-1</td>
<td>Victoria</td>
</tr>
<tr>
<td>Melba</td>
<td>Coldstream</td>
<td>97-101-75 x 04-99-142</td>
<td>Victoria</td>
</tr>
<tr>
<td><strong>USA</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Albion</td>
<td>Coldstream</td>
<td>Diamante x Cal 94.16-1</td>
<td>California</td>
</tr>
<tr>
<td>Camino Real</td>
<td>Coldstream</td>
<td>Cal 89.230-7 x Cal 90.253-3</td>
<td>California</td>
</tr>
<tr>
<td>Chandler</td>
<td>Digger’s Club</td>
<td>Douglas x Cal 72.361-105</td>
<td>California</td>
</tr>
<tr>
<td>San Andreas</td>
<td>Coldstream</td>
<td>Albion x Cal 97.86-1</td>
<td>California</td>
</tr>
<tr>
<td><strong>Japanese</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hokowase</td>
<td>Digger’s Club</td>
<td>Unknown</td>
<td>Japan</td>
</tr>
<tr>
<td><strong>European</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cambridge Rival</td>
<td>Digger’s Club</td>
<td>Dorsett x Early Cambridge</td>
<td>UK</td>
</tr>
<tr>
<td>Fraises Des Bois</td>
<td>Digger’s Club</td>
<td>N/A</td>
<td>France</td>
</tr>
<tr>
<td><strong>Breeding lines</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>07-102-41</td>
<td>Wandin</td>
<td>Unknown</td>
<td>Netherlands</td>
</tr>
<tr>
<td>07-095-35</td>
<td>Wandin</td>
<td>Unknown</td>
<td>Netherlands</td>
</tr>
<tr>
<td>04-069-91</td>
<td>Wandin</td>
<td>Unknown</td>
<td>Japan</td>
</tr>
</tbody>
</table>
Figure 2.9 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 (Table 2.3), was clustered next to Albion, confirming their close genetic relationship (Figure 2.9). Similarly, a good correlation between genetic distance and pedigree information in the USA cultivars based on RAPD was also demonstrated previously by Hancock et al. (1994) and Graham et al. (1996).

In addition, all the Australian cultivars (Adina, Lowanna, Melba and Alinta) were clustered together except for Juliette (Figure 2.9). These Australian cultivars may share the same genetic background since they all have the American cultivars as one of their parents (Table 2.3). 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. The genotypes in the first sub-cluster of Cluster II including Juliette and the breeding lines (Figure 2.9) were coherent with their breeding period because they were bred around the same time (i.e. 2004 – 2008), as a result of crossing among the genotypes available in the Victorian strawberry breeding program. While lacking information on their pedigree, it could be deduced that they are genetically similar. Assessment of more European and Japanese genotypes would help in defining the genetic distance of the breeding lines.

In contrast, Hokowase and Cambridge Rival, originating from different countries were clustered together in the first sub-cluster of Cluster I. It is worthwhile to note that both Hokowase and Cambridge Rival are older cultivars from Japan and the UK, suggesting that
they may share a common ancestor. A research based on chloroplast DNA variations that
determined the maternal lineages of strawberry cultivars across the world further supports
this assumption (Honjo et al., 2009). The study showed that both Cambridge Prizewinner
and Hokowase possessed chloroplast DNA haplotype V corresponding to *F. virginiana*
from the northern hemisphere. As both Cambridge Rival and Cambridge Prizewinner
shared the same immediate parent, which is the Early Cambridge (Darrow, 1966), it was
believed that Cambridge Rival also has chloroplast DNA haplotype V. Moreover, Chandler
which possesses haplotype V (Honjo et al., 2009) was also included in this cluster (Figure
2.9), indicating that all the genotypes in the first cluster were possibly grouped based on
common progenitor species.

To the author’s knowledge, this is the first report on genetic relationships of Australian
cultivars compared to the USA, European and Japanese genotypes. Although it has been
shown that the modern cultivated strawberries have a narrow germplasm base (Sjulin and
Dale, 1987), these results confirm the small gene pool in local strawberry cultivars. It is
important to determine the pedigree of cultivars and breeding lines used in the breeding
program as inbreeding depression is observed in strawberry plants (Galletta and Maas,
1990; Melville et al., 1980; Niemirowicz-Szczytt, 1988). Crossing genetically closely
related parental lines can result in vulnerability to disease, pests and environmental stresses
(Graham et al., 1996; Honjo et al., 2009). As a result, there is a need to increase genetic
diversity of local cultivars by exploiting germplasm from other parts of the world.

It may be concluded that although the SDA was only represented by DNA sequences from
five strawberry genotypes, the overall results indicate that the array was able to correctly
cluster most of the genotypes based on known pedigree information or by comparison to
previous research. The performance of SDA was next validated using a subset of five genotypes (Albion, San Andreas, Juliette, 07-102-41 and 04-069-91) to determine the utility of the SDA for clustering these genotypes based on their aroma profiles.

2.3.6 Aroma profile analysis of selected strawberry genotypes

The volatile compounds and genetic variability of three strawberry cultivars (Albion, Juliette and San Andreas) and two breeding lines (07-102-41 and 04-069-91) were analysed by SPME coupled with GC-MS and SDA fingerprinting. The volatile compounds detected by SPME and GC-MS technique in the headspace of five strawberry genotypes showed diverse aroma patterns between genotypes, consistent with the findings of previous studies on strawberry (Dirinck et al., 1981; Du et al., 2011; Hakala et al., 2002). A total of 24 aroma compounds were identified, including 18 esters, one aldehyde, two terpenes, one furanone and two lactones (Table 2.4). Both breeding lines contained higher levels of esters (49.3 % - 67.5 %) than the commercial varieties (14.7 % - 29.7 %). These lines also contained similar or lower levels of terpenes (26.5 % - 46.2 %) compared to commercial cultivars (45.5 % - 66.8 %) (Table 2.4). Trace amounts of aldehyde (2.7 %), furanone (1.0 %) and lactone (0.3 %) were detected in the 04-069-91 breeding line which contained genetic background from the Japanese strawberry. Surprisingly, no aldehyde, furanone and lactones were found in the 07-102-41 breeding line (Table 2.4).

In contrast, the two USA cultivars, Albion and San Andreas were dominated by terpenes (45.5 % - 55.5 %), followed by esters (14.8 % - 29.7 %) and lactones (19.6 % - 26.7 %) (Table 2.4). This result is unforeseen because the USA cultivars have been shown to accumulate high levels of esters (Du et al., 2011; Jetti et al., 2007). Moreover, a recent study performed using comprehensive two-dimensional gas chromatography-mass
spectrometry (GC x GC/TOF-MS) has reported that Australian-grown Albion demonstrated higher levels of esters than Juliette at 48.9 % and 29.9 %, respectively (Samykanno et al., 2013a). Despite the methodology differences and genotypic variations, climatic conditions and postharvest environment should also be considered as factors contributing to these discrepancies (Du et al., 2011; Jouquand et al., 2008).

Similarly, Juliette was also characterised mainly by terpenes (66.8 %) and esters (22.0 %) with low levels of lactone (4.8 %). Among the three commercial cultivars, only Juliette and San Andreas were found to have low levels of furanone (1.1 % - 5.7 %) whereas a trace amount of aldehyde (1.3 %) was identified only in Albion (Table 2.4). The average relative compositions of individual aroma compounds in all of these volatile classes were used to determine the relationships between the genotypes as discussed in the next section.
Table 2.4 Relative composition (%) of volatile compounds in the headspace of strawberry puree.

<table>
<thead>
<tr>
<th>Compounds</th>
<th>Retention time (min)</th>
<th>Relative composition (%)</th>
<th>Albion</th>
<th>Juliette</th>
<th>San Andreas</th>
<th>07-102-41</th>
<th>04-069-91</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Ester</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Methyl butanoate</td>
<td>3.61</td>
<td>5.4 ± 2.3</td>
<td>nd</td>
<td>1.9 ± 0.9</td>
<td>15.4 ± 6.1</td>
<td>1.6 ± 0.8</td>
<td></td>
</tr>
<tr>
<td>Ethyl butanoate</td>
<td>4.96</td>
<td>4.8 ± 5.6</td>
<td>3.2 ± 1.4</td>
<td>1.8 ± 2.5</td>
<td>7.2 ± 1.9</td>
<td>5.5 ± 0.2</td>
<td></td>
</tr>
<tr>
<td>Isopropyl butanoate</td>
<td>5.84</td>
<td>0.4 ± 0.4</td>
<td>nd</td>
<td>0.1 ± 0.1</td>
<td>nd</td>
<td>nd</td>
<td></td>
</tr>
<tr>
<td>Ethyl 2-methylbutanoate</td>
<td>6.00</td>
<td>nd</td>
<td>1.5 ± 0.1</td>
<td>nd</td>
<td>nd</td>
<td>0.5 ± 0.7</td>
<td></td>
</tr>
<tr>
<td>Methyl hexanoate</td>
<td>7.79</td>
<td>1.8 ± 0.5</td>
<td>nd</td>
<td>2.1 ± 3.0</td>
<td>25.0 ± 1.6</td>
<td>7.1 ± 6.3</td>
<td></td>
</tr>
<tr>
<td>Ethyl isovalerate</td>
<td>6.11</td>
<td>nd</td>
<td>1.8 ± 0.1</td>
<td>nd</td>
<td>nd</td>
<td>nd</td>
<td></td>
</tr>
<tr>
<td>Butyl butanoate</td>
<td>9.61</td>
<td>1.7 ± 2.0</td>
<td>nd</td>
<td>nd</td>
<td>3.7 ± 1.2</td>
<td>nd</td>
<td></td>
</tr>
<tr>
<td>Ethyl hexanoate</td>
<td>9.68</td>
<td>10.6 ± 12.0</td>
<td>10.1 ± 1.2</td>
<td>7.0 ± 8.7</td>
<td>16.2 ± 1.3</td>
<td>19.4 ± 8.6</td>
<td></td>
</tr>
<tr>
<td>(Z)-Hex-3-enyl acetate</td>
<td>9.84</td>
<td>nd</td>
<td>0.7 ± 1.0</td>
<td>nd</td>
<td>nd</td>
<td>nd</td>
<td></td>
</tr>
<tr>
<td>Hexyl acetate</td>
<td>10.02</td>
<td>0.5 ± 0.5</td>
<td>1.5 ± 0.3</td>
<td>0.9 ± 0.1</td>
<td>nd</td>
<td>4.2 ± 3.3</td>
<td></td>
</tr>
<tr>
<td>(E)-Hex-2-enyl acetate</td>
<td>10.08</td>
<td>0.4 ± 0.5</td>
<td>2.5 ± 2.7</td>
<td>0.9 ± 0.2</td>
<td>nd</td>
<td>4.9 ± 5.0</td>
<td></td>
</tr>
<tr>
<td>Methyl octanoate</td>
<td>12.87</td>
<td>nd</td>
<td>nd</td>
<td>nd</td>
<td>nd</td>
<td>0.3 ± 0.5</td>
<td></td>
</tr>
<tr>
<td>Ethyl octanoate</td>
<td>14.59</td>
<td>nd</td>
<td>0.7 ± 1.0</td>
<td>nd</td>
<td>nd</td>
<td>2.9 ± 0.4</td>
<td></td>
</tr>
<tr>
<td>Hexyl butanoate</td>
<td>14.48</td>
<td>0.6 ± 0.5</td>
<td>nd</td>
<td>nd</td>
<td>nd</td>
<td>nd</td>
<td></td>
</tr>
<tr>
<td>(E)-Hex-2-enyl butanoate</td>
<td>14.55</td>
<td>0.4 ± 0.4</td>
<td>nd</td>
<td>nd</td>
<td>nd</td>
<td>nd</td>
<td></td>
</tr>
<tr>
<td>Octyl acetate</td>
<td>14.92</td>
<td>nd</td>
<td>nd</td>
<td>nd</td>
<td>nd</td>
<td>2.9 ± 2.0</td>
<td></td>
</tr>
<tr>
<td>Octyl butanoate</td>
<td>18.97</td>
<td>3.1 ± 3.6</td>
<td>nd</td>
<td>nd</td>
<td>nd</td>
<td>nd</td>
<td></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td></td>
<td>29.7</td>
<td>22.0</td>
<td>14.7</td>
<td>67.5</td>
<td>49.3</td>
</tr>
<tr>
<td><strong>Aldehyde</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(E)-Hex-2-enal</td>
<td>6.12</td>
<td>1.3 ± 0.2</td>
<td>nd</td>
<td>nd</td>
<td>nd</td>
<td>2.7 ± 1.0</td>
<td></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td></td>
<td>1.3</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>2.7</td>
</tr>
</tbody>
</table>

Table 2.4 continued on next page
Table 2.4 (continued)

<table>
<thead>
<tr>
<th>Compounds</th>
<th>Retention time (min)</th>
<th>Relative composition (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Albion</td>
<td>Juliette</td>
</tr>
<tr>
<td><strong>Terpenes</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Linalool</td>
<td>12.27</td>
<td>11.9±7.2</td>
</tr>
<tr>
<td>(E)-Nerolidol</td>
<td>22.57</td>
<td>33.6±4.1</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>45.5</strong></td>
<td><strong>66.8</strong></td>
</tr>
<tr>
<td><strong>Furanone</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mesifuranne</td>
<td>11.17</td>
<td>nd</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>0.0</strong></td>
<td><strong>5.7</strong></td>
</tr>
<tr>
<td><strong>Lactones</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>γ-Decalactone</td>
<td>20.74</td>
<td>16.9±13.6</td>
</tr>
<tr>
<td>γ-Dodecalactone</td>
<td>24.73</td>
<td>2.7±3.2</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>19.6</strong></td>
<td><strong>4.8</strong></td>
</tr>
<tr>
<td><strong>Overall total</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>96.1</strong></td>
<td><strong>99.3</strong></td>
</tr>
</tbody>
</table>

Data were presented in mean of relative compositions ± standard deviation.

nd: not detected
2.3.6.1 Chemical and genetic relationships of strawberry genotypes

Two hierarchical cluster analyses were performed based on the aroma profiles and SDA data for Albion, Juliette, San Andreas, 07-102-41 and 04-069-91. One main cluster was observed in the dendrogram generated from aroma profiles at the truncation point of 11 (Figure 2.10a). The main cluster consisted of Albion, San Andreas, Juliette and 04-069-91 and they were separated from 07-102-41 breeding line. This result indicates a unique aroma profile for 07-102-41 breeding line compared to others (Figure 2.10a). The volatile groups that distinguished this breeding line from the main cluster were esters and lactones, as 07-102-41 contained the highest levels of total esters (67.5 %) but no lactones were detected compared to other genotypes (Table 2.4).

Chemically, the 07-102-41 breeding line which contained the genetic background of European strawberry was most distant from other genotypes (Figure 2.10a), the most notable difference being the high levels of methyl butanote (15.4 %) and methyl hexanoate (25.0 %) and the absence of γ-dodecalactone (Table 2.4). Moreover, all the genotypes assessed showed higher levels of ethyl esters which comprised 52.2 % - 78.7 % of total esters compared to methyl esters (0.0 % - 27.4 %) except for 07-102-41, where the levels of methyl esters (59.9 %) were more than ethyl esters (34.6 %). These results further confirm the findings of previous studies which concluded that the ratio of methyl/ethyl esters was dependent on cultivar or genotype (Forney et al., 2000; Larsen and Poll, 1995; Pérez et al., 1992). For example, some varieties such as Hokowase, Kent, Senga Gigana and Annapolis contained more than 70 % of methyl esters (Dirinck et al., 1981; Forney et al., 2000) whereas Chandler and Configra were dominated by 60 % to 80 % of ethyl esters (Dirinck et al., 1981; Pérez et al., 1992). In addition, the presence/absence of individual ester compounds, methyl butanoate, isopropyl butanoate, ethyl 2-methylbutanoate, ethyl
isovalerate, methyl hexanoate, (Z)-hex-3-enyl acetate and ethyl octanoate were useful in differentiating between the commercial cultivars, as they were present in either the USA or the Australian genotypes (Table 2.4).

Parallel to the aroma profiling study, a genetic hierarchical cluster analysis based on the SDA data of the same five selected strawberry genotypes also identified one main cluster at the truncation point of 15 (Figure 2.10b). The main cluster was formed by Albion, Juliette, 07-102-41 and 04-069-91 and they were separated from San Andreas. Three of the genotypes (Albion, Juliette and 04-069-91) showed consistency in their genetic relatedness compared to the chemical cluster (Figure 2.10a). Both clusters showed that these genotypes were closely related chemically and genetically, suggesting that the chemical compositions may be predicted by their genetic profiles as revealed by the SDA. The present results suggest that it is reasonable to use the strawberry-specific SDA as a potential platform to screen for DNA markers associated with key volatile compounds.

Similarly, a recent study showed good correlation between a promising feature and the tanshinone content in Salvia miltiorrhiza using the Salvia-specific SDA. However, DNA polymorphism contributing to the variation of signal intensity among the population was not determined in this study. Further, the feature has to be mapped onto S. miltiorrhiza genome (Chen, 2010) to locate genes involved in tanshinone production before it can be confirmed as a good predictor (Olarte, 2011). Using an Echinaceae-specific SDA, the author also reported positive correlation between the hybridisation profiles of three features and the content of monoene alkamides in Echinacea purpurea. However, the signal strength of those features and the relative abundance of monoene alkamides did not share similar pattern of variation for other Echinacea species. Hence, those features were
determined as species-specific markers rather than the predictors for the content of monoene alkamides in different *Echinaceae* species (Olarte *et al.* 2013). These results suggest that method modifications are needed prior to SDA hybridisation to enrich the DNA sequences associated with the production of phytochemicals in target plant species. Additionally, more stringent statistical analyses are needed to eliminate false positive and increase the probability of discovering informative markers associated with the trait of interest.
Figure 2.10a Hierarchical dendrogram generated using average-linkage-between-groups method and squared Euclidean distance, based on the aroma profiles of five strawberry genotypes. 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.

Figure 2.10b Hierarchical dendrogram using the average-linkage-between-groups method and squared Euclidean distance showing genetic relatedness of five 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.
2.3.7 Putative DNA marker discovery

2.3.7.1 Identification of polymorphic DNA sequences

To determine the features (i.e. potential branch point markers) which account for the most variation across 15 genotypes, principal component analysis (PCA) was performed using the normalised mean SNR of all the 287 features. Figure 2.11 demonstrates that a high percentage of variation (92.1\%) can be explained by the first two components. The first component (X axis) is able to explain most of the variation, contributing to 88.8\% of the total variation, followed by the second component (Y axis) which accounted for 3.3\%.

Most of the features clustered close to zero, with a few features scattered distantly from the X axis. Since the first component is able to adequately explain most of the variation, only the eight most distant features (FLP1D5, FLP1C6, FLP1A2, FLP4D7, FLP2E1, FLP2E12, FLP1G3 and FLP1D11) from zero on the X axis were selected for further analyses.

The normalised mean SNR of these features were extracted from the dataset to determine their hybridisation intensities. The observed features displayed high mean and high variance across all genotypes (Table 2.5). It was discovered that features which exhibited low mean but high variance were neglected in the PCA analysis. In general, DNA sequences in a plant genome are assigned into low copy, intermediate repetitive and high repetitive classes. The low copy number sequences are usually unique sequences representing genes (Cullis, 2004). Therefore the features with low signal intensities could be low copy number DNA markers associated with flavour trait. Hence, it is beneficial to perform more than one statistical analysis in the process of identifying useful features. A second assessment was performed by calculating the variance of each feature across the 15 genotypes. $k$-means clustering revealed that only FLP1E7, FLP1D5 and FLP1C6 showed
the highest variances across the 15 genotypes in which the latter two features were selected by PCA as well.

Figure 2.11 Two-dimensional plot of the principal component analysis (PCA) of 287 SDA features corresponding to 15 strawberry genotypes.
Table 2.5 Normalised mean signal-to-noise ratio (SNR) of nine features showing the most variability across 15 strawberry genotypes based on PCA and magnitude of variance.

<table>
<thead>
<tr>
<th>Strawberry Genotypes</th>
<th>FLP1D5a,b</th>
<th>FLP1C6a,b</th>
<th>FLP1A2a</th>
<th>FLP4D7a</th>
<th>FLP2E1a</th>
<th>FLP2E12a</th>
<th>FLP1D11a</th>
<th>FLP1G3a</th>
<th>FLP1E7b</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cambridge Rival</td>
<td>829.1</td>
<td>653.0</td>
<td>662.9</td>
<td>499.2</td>
<td>699.3</td>
<td>800.1</td>
<td>1062.7</td>
<td>885.3</td>
<td>56.8</td>
</tr>
<tr>
<td>Hokowase</td>
<td>890.8</td>
<td>702.8</td>
<td>640.0</td>
<td>634.9</td>
<td>593.2</td>
<td>863.5</td>
<td>1100.8</td>
<td>849.9</td>
<td>61.2</td>
</tr>
<tr>
<td>Camino Real</td>
<td>704.4</td>
<td>790.8</td>
<td>551.3</td>
<td>652.9</td>
<td>750.0</td>
<td>841.6</td>
<td>1190.5</td>
<td>750.0</td>
<td>39.3</td>
</tr>
<tr>
<td>Alinta</td>
<td>808.5</td>
<td>650.5</td>
<td>603.2</td>
<td>766.3</td>
<td>578.6</td>
<td>882.2</td>
<td>956.5</td>
<td>788.4</td>
<td>36.4</td>
</tr>
<tr>
<td>Adina</td>
<td>474.0</td>
<td>540.0</td>
<td>461.7</td>
<td>597.7</td>
<td>693.6</td>
<td>586.5</td>
<td>957.5</td>
<td>654.2</td>
<td>37.5</td>
</tr>
<tr>
<td>Lowanna</td>
<td>639.5</td>
<td>551.8</td>
<td>554.6</td>
<td>627.5</td>
<td>633.6</td>
<td>634.8</td>
<td>1041.0</td>
<td>857.8</td>
<td>37.6</td>
</tr>
<tr>
<td>Melba</td>
<td>493.3</td>
<td>377.9</td>
<td>316.7</td>
<td>466.7</td>
<td>647.8</td>
<td>490.5</td>
<td>1071.8</td>
<td>656.5</td>
<td>33.3</td>
</tr>
<tr>
<td>Juliette</td>
<td>1069.6</td>
<td>716.2</td>
<td>769.0</td>
<td>859.7</td>
<td>894.5</td>
<td>562.9</td>
<td>1001.1</td>
<td>710.8</td>
<td>52.8</td>
</tr>
<tr>
<td>04-069-91</td>
<td>1211.3</td>
<td>875.4</td>
<td>858.8</td>
<td>862.2</td>
<td>672.6</td>
<td>572.5</td>
<td>782.4</td>
<td>562.4</td>
<td>47.1</td>
</tr>
<tr>
<td>07-102-41</td>
<td>985.7</td>
<td>766.6</td>
<td>660.0</td>
<td>703.7</td>
<td>772.3</td>
<td>687.0</td>
<td>883.1</td>
<td>586.6</td>
<td>334.3</td>
</tr>
<tr>
<td>San Andreas</td>
<td>735.0</td>
<td>633.2</td>
<td>719.5</td>
<td>418.3</td>
<td>684.5</td>
<td>578.8</td>
<td>981.6</td>
<td>470.0</td>
<td>46.8</td>
</tr>
<tr>
<td>Chandler</td>
<td>737.2</td>
<td>614.7</td>
<td>708.9</td>
<td>586.7</td>
<td>742.1</td>
<td>689.4</td>
<td>1120.1</td>
<td>640.9</td>
<td>41.3</td>
</tr>
<tr>
<td>Albion</td>
<td>1265.0</td>
<td>632.1</td>
<td>806.7</td>
<td>728.5</td>
<td>544.9</td>
<td>764.6</td>
<td>864.3</td>
<td>402.5</td>
<td>34.1</td>
</tr>
<tr>
<td>07-095-35</td>
<td>868.1</td>
<td>752.8</td>
<td>887.2</td>
<td>657.8</td>
<td>1042.0</td>
<td>789.2</td>
<td>1043.5</td>
<td>479.1</td>
<td>30.8</td>
</tr>
<tr>
<td>Fraises Des Bois</td>
<td>193.4</td>
<td>38.5</td>
<td>519.3</td>
<td>720.3</td>
<td>647.6</td>
<td>523.3</td>
<td>1046.9</td>
<td>836.9</td>
<td>38.0</td>
</tr>
<tr>
<td>Mean</td>
<td>793.7</td>
<td>619.7</td>
<td>647.6</td>
<td>652.2</td>
<td>706.4</td>
<td>684.4</td>
<td>1006.9</td>
<td>675.4</td>
<td>61.8</td>
</tr>
<tr>
<td>Variance</td>
<td>79799.7</td>
<td>39990.0</td>
<td>23530.7</td>
<td>16715.4</td>
<td>16057.3</td>
<td>17196.6</td>
<td>11321.5</td>
<td>23664.4</td>
<td>5760.2</td>
</tr>
</tbody>
</table>

a Features selected by PCA
b Features selected by magnitude of variance

Boxes shaded in yellow represent branch point markers corresponding to a specific genotype or genotypic group
All the nine features (FLP1D5, FLP1C6, FLP1A2, FLP4D7, FLP2E1, FLP2E12, FLP1G3, FLP1D11 and FLP1E7) selected by both PCA and magnitude of variance could serve as branch point markers associated with specific genotypes or genotypic groups. To further reduce the number of useful features or clones, Pearson’s bivariate correlation was performed to eliminate features with similar hybridisation patterns because they could represent redundant clones in the subtracted library. The findings revealed that there is a significant positive correlation between FLP1D5 and FLP1C6 ($r = 0.80, p < 0.01$) and FLP1D5 and FLP1A2 ($r = 0.77, p < 0.01$). Although FLP1D5 showed a strong correlation with FLP1C6 and FLP1A2, the latter two clones only weakly correlated with each other ($r = 0.60, p < 0.05$) (Appendix 7). To avoid losing important features, only FLP1D5 was removed from the list of polymorphic DNA sequences. Consequently, the eight selected features (FLP1C6, FLP1A2, FLP4D7, FLP2E1, FLP2E12, FLP1G3, FLP1D11 and FLP1E7) were used to generate a second hierarchical dendrogram as depicted in Figure 2.15. By examining their normalised mean SNR (Table 2.5), six features (FLP1C6, FLP1A2, FLP4D7, FLP2E1, FLP2E12 and FLP1E7) were manually placed at their corresponding branch point in the dendrogram (Figure 2.12) whereas no specific genotypes or genotypic groups could be assigned to two of the features (i.e. FLP1D11 and FLP1G3).
Figure 2.12 Dendrogram constructed using average-linkage-between-groups method and squared Euclidean distance, based on the eight most discriminatory features among 15 strawberry genotypes. 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.
Comparison between the dendrogram generated with the eight most polymorphic features (Figure 2.12) and the first dendrogram constructed with the whole SDA dataset (Figure 2.9) shows consistency in clustering most of the genotypes. Two main clusters were identified from the dendrogram at the truncation point of 13, with Fraises Des Bois remaining genetically most distant from the other octoploid strawberries. This result indicates that FLP1C6 could be the feature responsible for differentiating *Fragaria* at species level (Figure 2.12). Two sub-clusters were discerned from Cluster I, where Cambridge Rival, Hokowase and Camino Real remained clustered together in the first sub-cluster as observed in the original analysis while Chandler appeared to be more distantly related. The second sub-cluster was formed by Adina, Lowanna and Melba as in the first dendrogram but in this instance Alinta was relocated to the first sub-cluster (Figure 2.12). The results suggest that FLP2E12 and FLP1A2 could be features specific to the Australian and American/Japanese cultivars, respectively.

Cluster II contained two sub-clusters, with Juliette, 04-069-91 and 07-102-41 forming one sub-cluster and San Andreas, Chandler and Albion forming another sub-cluster. These genotypes were separated from 07-095-35 within Cluster II (Figure 2.12). While all these genotypes remained grouped together in the same cluster as observed in the original dendrogram (Figure 2.9), FLP4D7 may be more efficient in discriminating the American cultivars from the genotypes developed through the Victorian strawberry breeding program. Most importantly, the close genetic relationship between San Andreas and Albion is confirmed (Figure 2.12). Interestingly, FLP1E7 and FLP2E1 were found to be specific to 07-102-41 and 07-095-35, respectively. These findings demonstrate the importance of these branch point markers in understanding the genetic relatedness of the genotypes assessed (Figure 2.12).
Although some minor differences were observed between the two dendrograms (Figure 2.9 and Figure 2.12), most of the genotypes were clustered correctly based on known pedigree information (the American and Australian cultivars) and their places of origin (the genotypes developed through the Victorian strawberry breeding program). Moreover, the diploid strawberry, Fraises Des Bois, was as expected, genetically most distantly related to the other octoploid strawberries as shown in both dendrograms. Therefore, the other genotypes with unknown parentage information were expected to be classified correctly. It was suggested that the differences between the two dendrograms could possibly be due to limited number of genotypes used in the initial subtracted library construction. The SDA could be over-represented with the DNA sequences from only five strawberry genotypes (Albion, Juliette, 07-102-41, 07-095-35 and 04-069-91) (Olarte et al., 2013a). Nevertheless, the current findings suggest that the strawberry-specific SDA is a robust technique to screen for DNA polymorphisms between and within species in Fragaria. Hence, the SDA could be a potential platform to discover DNA markers associated with aroma compounds specific to a strawberry genotype through the branch point markers identified in this study. Correlation of these branch point markers with key volatile compounds will be discussed in the next section.
2.3.7.2 Correlation between branch point markers and key volatile compounds

The six branch point markers (FLP1A2, FLP1C6, FLP1E7, FLP2E1, FLP2E12 and FLP4D7) which showed unique hybridisation pattern to specific genotypes were used to establish correlation with nine key volatile compounds (i.e. methyl butanoate, methyl hexanoate, ethyl butanoate, ethyl hexanoate, linalool, (E)-nerolidol, mesifuranne, γ-decalactone and γ-dodecalactone). The results of linear regression and Pearson’s bivariate correlation with significant correlation between the branch point markers and key volatile compounds are shown in Figure 2.13 a – d and Appendix 8, respectively. The results showed that the FLP1E7 feature was positively correlated with methyl esters, i.e. methyl butanoate ($r = 0.93, p < 0.05$) and methyl hexanoate ($r = 0.97, p < 0.01$), in which 87 % and 93 % of the total variation in the normalised mean SNR of FLP1E7 can be explained by the linear relationships between the two variables, respectively (Figure 2.13 a & b). The strong association between FLP1E7 and methyl esters corroborated with the results from aroma profiles (Table 2.4) and normalised mean SNR (Table 2.5), suggesting that FLP1E7 is a putative positive marker specific to 07-102-41 which contained high levels of methyl butanoate and methyl hexanoate.
Figure 2.13 Significant correlations among the normalised mean SNR of branch point markers and relative composition of key volatile compounds. (a) Correlation between FLP1E7 and methyl butanoate. (b) Correlation between FLP1E7 and methyl hexanoate.

Figure 2.13 continued on next page
Figure 2.13 (continued)

Significant correlations among the normalised mean SNR of branch point markers and relative composition of key volatile compounds. (c) Correlation between FLP1C6 and ethyl hexanoate. (d) Correlation between FLP2E1 and linalool.

** Correlation is significant at the 0.01 level (2-tailed)

* Correlation is significant at the 0.05 level (2-tailed)
Similarly, FLP1C6 was positively correlated with ethyl hexanoate \((r = 0.93, \ p < 0.05)\), where 86% of the total variation in the normalised mean SNR of FLP1C6 was explained by the linear regression. The signal intensities of FLP1C6 were low in Albion and San Andreas, moderate in Juliette and high in 07-102-41 and 04-069-91. It is considered as a putative positive marker since the signal intensities increased proportionally with ethyl hexanoate levels in different strawberry genotypes (Figure 2.13c). Interestingly, this DNA sequence hybridised weakly with the diploid strawberry but showed high signal intensities in all the octoploid strawberries (Table 2.5). This finding further supports the studies that reported different aroma patterns between *Fragaria vesca* and *Fragaria x ananassa* (Aharoni *et al.*, 2004; Ulrich *et al.*, 2007; Ulrich and Olbricht, 2013). However, inconsistent results were obtained from the same research group. For instance, ethyl butanoate and ethyl hexanoate were detected at lower levels in *F. x ananassa* compared to *F. vesca* as shown in Ulrich *et al.* (2007) but contradictory results were reported in Ulrich and Olbricht (2013). This difference could probably be attributed to different accessions used in the analyses. Unfortunately, no direct comparison between the volatile levels in diploid and octoploid strawberries could be made in this study due to the unavailability of *F. vesca* fruits. Based on the results from Aharoni *et al.* (2004) and Ulrich and Olbricht (2013), it may be concluded that the high signal intensities of FLP1C6 were positively correlated with ethyl hexanoate production in the octoploid strawberries.

Conversely, FLP2E1 was negatively correlated with linalool \((r = -0.91, \ p < 0.05)\), where 83% of the total variation in the normalised mean SNR of FLP2E1 was explained by the linear regression and the other 17% remained unexplained. High signal intensity of FLP2E1 was found to be inversely correlated with linalool levels in different strawberry genotypes. Therefore, it could be a putative “negative” marker for linalool (Figure 2.13d).
No significant correlation was found between FLP1A2, FLP2E12 and FLP4D7 with any of the key volatile compounds, indicating that they could only be DNA sequences specific to a genotype or genotypic group (Figure 2.12). Characterisation of these six polymorphic DNA sequences may infer (1) the identity of putative DNA markers linked to the loci controlling economically important traits (e.g. strawberry flavour) and (2) branch point markers specific to a genotype or genotypic group.

2.3.8 Identity of putative DNA markers

Sequence editing was performed to ensure that only the regions flanked by nested primer 1 (5’-CCGGGCAGGT-3’) and nested primer 2R (5’-ACCTCGGCCG-3’) were used in the similarity search (Appendix 9). In cases where DNA fragments were inserted into the vector in a reverse orientation, reverse sequences corresponding to nested primer 1 (5’-ACCTGCCCGG-3’) and nested primer 2R (5’-CGGCCGAGGT-3’) were found in the sequencing results (Appendix 9). All the DNA sequences contained primer sequences corresponding to different adaptors at both ends, indicating that the subtraction process was efficient and the subtracted DNA fragments were amplified exponentially.

However, it was unexpected that two of the branch point markers, FLP1A2 and FLP4D7 displayed the same sequence length of 506 bp (Appendix 9). DNA sequence alignment performed using Clustal Omega at http://www.ebi.ac.uk/Tools/msa/clustalo/ (EMBL-EBI 2014) revealed a perfect match between FLP1A2 and FLP4D7, suggesting they are redundant clones. Surprisingly, they were not significantly correlated to each other based on Pearson’s bivariate correlation analysis \( r = 0.43, p > 0.05 \) (Appendix 7), indicating that they exhibited different hybridisation patterns (Table 2.5). Therefore, the image files for all the genotypes were retrieved from the microarray system to identify outliers which
contribute to the inconsistency in signal intensities. It was found that artifacts were present in one of the technical replicates of FLP1A2, causing a “split-spot” which produced low signal intensity relative to the normal feature. As a result, FLP1A2 was also removed from the list including its potential function as a branch point marker (Figure 2.12) while FLP4D7 was retained in the analysis.

The results of the similarity search against the *F. vesca* genome database are shown in Table 2.6. All the nuclear-specific clones (FLP1C6, FLP2E12 and FLP4D7) did not match any genes in the strawberry genome database. Instead, they were non-coding DNA fragments located on various chromosomes as expected from a subtracted genomic library. The full sequence of FLP4D7 displayed the highest similarity (E-value: 0.0) to a DNA region on linkage group 5 (LG5:9705764..9706269, scf0513135:696548..697053) when aligned to *F. vesca* draft genome v1.1. Likewise, 86% of FLP2E12 demonstrated high similarity (E-value: 1e-163) to a DNA region on linkage group 1 (LG1:16266963..16267424, scf0513175:274765..275226). Moreover, 54% of FLP1C6 presented significant similarity (E-value: 3e-51) to a DNA region on linkage group 6 (LG6:12886104..12886402, scf0513102:1224409..1224707). For the chloroplast-specific clone, 100% of FLP2E1 showed a high similarity (E-value: 1e-144) to part of the 16S ribosomal RNA gene (scf0510902:749..1013).
Table 2.6 Putative identity of the most polymorphic DNA sequences searched against the *Fragaria vesca* draft genome (v1.1). E-value regarded as significant if < $1 \times 10^{-5}$.

<table>
<thead>
<tr>
<th>Clones</th>
<th>Length (bp)</th>
<th>Landmark or region</th>
<th>Sequence description</th>
<th>E-value</th>
<th>Specific to target</th>
</tr>
</thead>
<tbody>
<tr>
<td>FLP1C6&lt;sup&gt;a,b&lt;/sup&gt;</td>
<td>555</td>
<td>LG6:12886104..12886402, scf0513102:1224409..1224707</td>
<td>Genomic DNA region on linkage group 6</td>
<td>$3 \times 10^{-51}$</td>
<td><em>F. x ananassa</em></td>
</tr>
<tr>
<td>FLP1E7&lt;sup&gt;b&lt;/sup&gt;</td>
<td>512</td>
<td>LG1:9965781..9965800, scf0512991:103282..103301</td>
<td><em>Arabidopsis thaliana</em> beta-amylase 6 mRNA, complete cds</td>
<td>0.089</td>
<td>07-102-41</td>
</tr>
<tr>
<td>FLP2E1&lt;sup&gt;a&lt;/sup&gt;</td>
<td>265</td>
<td>scf0510902:749..1013</td>
<td><em>Fragaria vesca</em> voucher P.M. McNutt 020 (FLDCI) 16S ribosomal RNA gene, partial sequence; tRNA-Ile (trnI) and tRNA-Ala (trnA) genes, complete sequence; and 23S ribosomal RNA gene, partial sequence; chloroplast</td>
<td>$1 \times 10^{-144}$</td>
<td>07-095-35</td>
</tr>
<tr>
<td>FLP2E12&lt;sup&gt;a&lt;/sup&gt;</td>
<td>540</td>
<td>LG1:16266963..16267424, scf0513175:274765..275226</td>
<td>Genomic DNA region on linkage group 1</td>
<td>$1 \times 10^{-163}$</td>
<td>First sub-cluster of Cluster I</td>
</tr>
<tr>
<td>FLP4D7&lt;sup&gt;a&lt;/sup&gt;</td>
<td>506</td>
<td>LG5:9705764..9706269, scf0513135:696548..697053</td>
<td>Genomic DNA region on linkage group 5</td>
<td>0.0</td>
<td>First sub-cluster of Cluster II</td>
</tr>
</tbody>
</table>

<sup>a</sup> Features chosen by PCA.

<sup>b</sup> Features revealed by magnitude of variance.
In contrast, only 4% of FLP1E7 showed a very low similarity (E-value: 0.089) to gene23897 (*Arabidopsis thaliana* beta-amylase 6 mRNA) located on LG1:9965781..9965800, scf0512991:103282..103301. Since the E-value falls outside the acceptable threshold (1e−5), this clone is highly unlikely to be a beta-amylase gene. Hence, this clone was searched against the NCBI database to confirm its similarity to known DNA sequences from other organisms. The result revealed that 22% of the DNA sequence showed low similarity (3e−7) to the mitochondrial minicircle pO sequence of *Beta vulgaris* with 88% identity match. Therefore, FLP1E7 is more likely a mitochondrial DNA sequence specific to 07-102-41.

Further analysis was performed for FLP1C6 which was positively correlated to ethyl hexanoate (Figure 2.13c) by searching for genes located within 5 cM on either side of the features to minimize linkage drag (Collard and Mackill, 2008). The physical distance was deduced from the length of the linkage group by assuming the genetic length of a chromosome as 100 cM (Kearsey and Pooni, 1996). The length of linkage group 6 is approximately 40 Mb according to the PFR Strawberry Server (PFR, 2010). Therefore, genes were searched approximately 2 Mb on either side of FLP1C6 (40 Mb / 100 cM x 5 cM = 2 Mb). Two genes corresponding to *F. x ananassa* cytosolic aldolase (*SCA1*, scf0513102:533154..535503, accession AF308587) and CTR-1 (constitutive triple response 1) like protein kinase (*CTR1*, scf0513102:476089..477184, accession AY538771) were found at approximately 0.70 Mb and 0.75 Mb downstream of FLP1C6, respectively (Figure 2.14). These two genes are likely to be involved in the biosynthetic pathway of ethyl hexanoate as revealed by the linear regression in Figure 2.13c.
Figure 2.14 Landmark of FLP1C6 on LG6:12886104..12886402. *SCA1*: Cytosolic aldolase. *CTR-1*: Constitutive triple response 1-like protein kinase. Red arrows indicate FLP1C6; green arrows specify the location of *SCA1* and *CTR1* genes.

Cytosolic aldolase (*SCA1*) is an enzyme involved in sugar metabolism during glycolysis and gluconeogenesis (Paliyath *et al*., 2009). Up-regulation of *SCA1* transcript is often accompanied by the induction of additional glycolytic and anaerobic alcohol fermentation enzymes (Andrews *et al*., 1994; Minhas and Grover, 1999; Schwab *et al*., 2001), providing ethanol as the substrate for the biosynthesis of ethyl esters such as ethyl acetate, ethyl butanoate and ethyl hexanoate (Aharoni *et al*., 2000; Larsen and Watkins, 1995a; Paliyath *et al*., 2009). Additionally, CTR-1 like protein kinase has been identified as a negative regulator in the ethylene signal transduction pathway (Bleecker, 1998; Clark *et al*., 1998; Kieber *et al*., 1993; Klee and Giovannoni, 2011). The regulatory role of ethylene in ester production has been proven in climacteric fruits such as Charentais Cantaloupe melons (Flores *et al*., 2002), tomato (Alexander and Grierson, 2002) and apple (Defilippi *et al*., 2005; Schaffer *et al*., 2007). However, a recent study showed that an increase in ethylene production was concomitant with the up-regulation of *CTR-1* transcript in ‘Camarosa’ strawberry, suggesting that CTR-1 may be a positive regulator in non-climacteric fruits (Sun *et al*., 2013). This statement is corroborated by a previous study which suggested that low concentrations of ethylene produced a characteristic pattern that resembled ripening in
climacteric fruits (Iannetta et al., 2006). Moreover, correlation illustrated in Figure 2.13c implies that the variation in the levels of ethyl hexanoate in those genotypes may be attributed to allelic variants of the FLP1C6 locus or the respective genes. To investigate these hypotheses in the future, the allelotypes of FLP1C6, SCAI and CTR-1 among the five strawberry genotypes assessed could be identified by direct amplicon sequencing. Association should be established by correlating the polymorphisc sites found within the FLP1C6 locus with the levels of ethyl hexanoate among the five strawberry genotypes (Dunemann et al., 2012).

The discovery of putative mitochondrial DNA sequences (FLP1E7) in relation to volatile compounds could probably be attributed to the mitochondrial localisation of nuclear-encoded enzymes involved in fruit flavour biogenesis. It has been reported that the enzyme activities of mitochondrial citrate synthase and mitochondrial malate dehydrogenase increases substantially during strawberry fruit ripening. The genes could be potential determinants of fruit flavour via regulation of the sugar:acid balance (Iannetta et al., 2004). However, the role of this putative mitochondrial DNA fragment in assisting the localisation of nuclear-encoded enzymes or metabolic pathways for the production of aroma compound precursor in the mitochondrial matrix cannot be confirmed in this study. It may potentially be an organelle marker specific to 07-102-41 (Figure 2.12, Table 2.5).

While most of the DNA barcoding studies in plant species were performed using chloroplast genes (e.g. matK and rbcL) and intergenic spacers (e.g. trnH-psbA and trnL-F) (Kress et al., 2005; Lahaye et al., 2008; Li et al., 2011; Skuza et al., 2013), the role of mitochondrial DNA markers in cultivar identification should not be discounted. A few studies have demonstrated the utilisation of mitochondrial DNA polymorphisms for cultivar identification, such as in perennial ryegrass (Sato et al., 1995), broad bean (Scallan
and Harmey, 1996) and Korean ginseng (Wang et al., 2009). Further, it has been reported that maternally inherited markers could show much better genetic diversity within a population (Petit et al., 2005). Therefore, the FLP1E7 could also be interpreted as a putative mitochondrial DNA marker for potential genotype identification.

In addition, FLP2E1, a branch point marker for 07-095-35, which showed a high similarity to ribosomal RNA genes in F. vesca chloroplast sequence (Table 2.6) was negatively correlated to linalool (Figure 2.13d). In plants, sesquiterpenes (e.g. nerolidol) are synthesized predominantly through the mevalonate pathway (MVA) in the cytosol, whereas monoterpenes (e.g. linalool) are derived from the methylerythritol phosphate (MEP) pathway in chloroplasts (Bohlmann and Keeling, 2008; Martin et al., 2012). Unlike other plant species, linalool biosynthesis in strawberry fruit occurs in the cytosol due to the action of the Fragaria x ananassa Nerolidol Synthase 1 (FaNESI) gene that encodes a truncated protein sequence at the N-terminus which has a plastid-targeting function (Aharoni et al., 2004). This could probably explain the decrease in FLP2E1 signal intensities in association with the increase of linalool levels since its production is exclusive to the cytosol. However, the involvement of chloroplast ribosomal RNA genes in fruit flavour biosynthesis remains unclear. In addition, the possibility that FLP2E1 may be a potential chloroplast-specific marker for 07-095-35 (Figure 2.12) should not be discounted.

FLP4D7, a branch point marker specific to a sub-cluster formed by Juliette, 04-069-91 and 07-102-41 (Figure 2.12), may represent putative DNA markers for genotype identification since it was not significantly correlated with any of the key volatile compounds. Its sequence identity was further confirmed by a similarity search against the NCBI database.
25% of the FLP4D7 sequence was similar to *Fragaria vesca* spp. *americana* clone fosmid 38H02 (E-value = 3e⁻³²). A 127-bp segment of FLP4D7 showed 86% match to the unclassified repeat on the 31,669 bp-long fosmid 38H02 (Pontaroli *et al.* 2009), suggesting that this DNA sequence may be a new family of repetitive DNA sequences specific to the *Fragaria x ananassa*. The differences in normalised mean SNR of the repetitive DNA sequence between genotypes (Table 2.5) may be due to copy number variation. Studies have revealed that repetitive DNA sequences may be present in different copies in different species (Valárik *et al.*, 2002), cultivars (Zhao *et al.*, 2011) and ploidy levels (Čížková *et al.*, 2013). For instance, *Radka5* corresponding to a repeated sequence differed 30-fold between *Musa balbisiana* and *M. acuminata* (Valárik *et al.*, 2002). Also, repetitive DNA sequences with copy number variability or sequence polymorphism have been proposed as suitable molecular markers for cultivar identification in grapevine (Thomas *et al.*, 1993) and for phylogenetic analysis in cucumber (Zhao *et al.*, 2011). Therefore, it can be concluded that FLP4D7 may not be a marker associated with volatile compounds; nevertheless, it may be useful for genotype identification.

Similarly, FLP2E12, a branch point marker specific to a sub-cluster containing Cambridge rival, Camino real, Hokowase and Alinta (Figure 2.12) may also be a putative DNA marker for genotype identification. Blastx revealed that 85% of the FLP2E12 DNA sequence showed significant similarity (E-value = 4e⁻⁶⁴) with the uncharacterized protein LOC101295581 in *Fragaria vesca* ssp. *vesca*. Its amino acid sequence was 63% identical to Transposase_21 with a tnp2 family domain in *F. vesca*. Moreover, the CACTA element belonging to the second group of transposable elements (TEs) (Song *et al.*, 1998) found in this sequence further confirmed its identity as a putative transposon specific to *Fragaria x ananassa*. Previous studies have demonstrated the efficiency and potential use
of TE-derived markers in rice (Fukuchi et al., 1993) and sugarcane cultivar identification (Rossi et al., 2001). Interestingly, a previous paper reported the possibility of utilising the CACTA element to develop cultivar-specific transposon insertion-sequence characterized amplified regions (Ti-SCARs) for rapeseed (Lee et al., 2012). Hence, the high polymorphic structure of TEs or their insertion sites may serve as a source for molecular marker development leading to TEs-mediated fingerprinting for strawberry breeding.

2.4 CONCLUSIONS

A Subtracted Diversity Array (SDA) containing 287 strawberry-specific DNA sequences was successfully constructed using a combination of Suppression Subtractive Hybridisation (SSH) method and microarray technology. High subtraction efficiency (99%) was obtained based on a broad subtraction approach between the genomic DNA pool of five strawberry genotypes and nine non-angiosperm species. This strawberry-specific SDA may serve as a potential platform for a wide range of applications such as genotype identification and DNA marker discovery for marker-trait association studies.

Validation experiments performed on the strawberry-specific SDA confirmed its ability to fingerprint 15 strawberry genotypes originating from different countries based on the differential hybridisation patterns among these genotypes. The genetic relationships of these genotypes deduced from the first hierarchical cluster analysis was mostly consistent with pedigree information in published literature. This result is also consistent with another hierarchical dendrogram generated with the eight most polymorphic features across the 15 strawberry genotypes. Of these, three nuclear-specific features (FLP1C6, FLP2E12 and FLP4D7) and two organelle-specific features (FLP1E7 and FLP2E1) may possibly be
branch point markers specific to each genotype and/or genotypic group based on differences in signal intensities between genotypes.

In addition, variation in volatile compounds detected by GC-MS analysis indicated that each genotype has its own unique aroma profile, making selective breeding for desired aroma possible by cross hybridisation between two genotypes with distinctive aroma patterns. The patterns of relatedness observed in aroma profiles correlated with the genetic profiles generated by SDA, suggesting that there may be genetic basis for the biosynthesis of volatile compounds. A strong correlation was found between some branch point markers (FLP1C6, FLP1E7 and FLP2E1) and key volatile compounds (i.e. methyl butanoate, methyl hexanoate, ethyl hexanoate and linalool). The SCAI and CTRI genes situated at close proximity to the nuclear-specific FLP1C6 feature may be involved in producing precursors for the biosynthesis of ethyl hexanoate. Nevertheless, the genes responsible for the rate-limiting steps in ethyl hexanoate production may be located if the search is expanded to 10 cM on either side of the markers. These three features (FLP1C6, FLP1E7 and FLP2E1) may be potential DNA markers closely linked to loci controlling the biosynthesis of corresponding key volatile compounds. Once validated, these branch point markers that correlate with key volatile compounds could be turned into PCR-based DNA markers to allow for efficient cultivar fingerprinting for routine diagnostic purposes.

Therefore, it may be concluded that the strawberry-specific SDA is an efficient technique to capture polymorphic DNA loci from both nuclear and organelle genomes of strawberry genotypes, eliminating the need to screen for a large number of DNA markers. The strawberry-specific SDA also contained genetic information related to flavour development, making it feasible to detect DNA loci associated with key volatile
compounds without having to generate a mapping population. This study may provide tremendous amount of information for strawberry breeding programs directed at the improvement of aroma compound composition or other quality traits. The discovery of DNA markers associated with fruit flavour is discussed in Chapter 3.
CHAPTER 3
Identification of DNA Markers Associated with Flavour using Bulked Segregant Analysis (BSA) and Subtracted Diversity Array (SDA)

3.1 INTRODUCTION

It is generally accepted that the flavour quality of many fruits including strawberry have deteriorated due to current breeding practises (Bartoshuk and Klee, 2013; Klee, 2010). There is evidence that consumers are willing to pay more for better tasting fruits (Byrne, 2012). Hence, development of cultivars with improved flavour is becoming one of the main breeding priorities in strawberry producing countries. Strawberry aroma has been studied extensively in the past. Over 300 compounds corresponding to a complex mixture of esters, furanones, aldehydes, alcohols, terpenes, lactones and sulphur compounds have been identified (Dirinck et al., 1981; Latrasse, 1991; Pérez et al., 1992; Ulrich et al., 1997; Zabetakis and Holden, 1997). Of these, only about 20 main volatile compounds have been determined to dominate the typical strawberry aroma based on sensory descriptive analysis and their odour activity values (Jetti et al., 2007). For instance, the key esters including methyl butanoate, methyl hexanoate, ethyl butanoate and ethyl hexanoate are attributed to the fruity notes whereas the aldehydes such as (E)-hex-2-enal and hexanal account for the green and grassy notes in strawberry aroma. Furaneol and mesifuranne have been described as the two most important compounds contributing to the sweet and caramel-like flavour in ripe strawberry fruits (Du et al., 2011; Fukuhara et al., 2005; Jetti et al., 2007).

In recent years, strawberry flavour research has progressed from the chemical and sensory analyses to investigation of biosynthesis and genetic control of important volatile compounds (Bood and Zabetakis, 2002). For example, the SAAT (strawberry-specific
alcohol acyl transferase) and FaNES1 (*Fragaria x ananassa* nerolidol synthase 1) genes involved in the biosynthesis of esters and terpenes, respectively were identified using cDNA microarray technology (Aharoni et al., 2004; Aharoni et al., 2000). Besides, FaOMT (*Fragaria x ananassa* O-methyltransferase) and FaQR (*Fragaria x ananassa* quinone oxidoreductase) genes have been reported to be involved in the formation of mesifuranne and furaneol, respectively (Raab et al., 2006; Schwab et al., 2008; Wein et al., 2002). Although the advent of molecular biology has facilitated the identification of several genes involved in fruit ripening and flavour biogenesis, little is known about the genetic inheritance of key aroma compounds.

A few studies have recently reported the discovery of loci or gene candidates controlling production of key volatile compounds and their potential use as molecular markers. For instance, the presence/absence of FaNES1 allele could mostly predict the linalool-producing and non-producing cultivated and wild strawberry materials with a few exceptions (Chambers et al., 2012). A 30-bp indel in the promoter region of the FaOMT allele responsible for the high expression of mesifuranne content may be turned into important tool for strawberry breeding (Zorrilla-Fontanesi et al., 2012). Moreover, two recent studies have shown that a fatty acid desaturase FaFAD1 gene was correlated with the production of γ-decalactone in strawberry fruit (Chambers et al., 2014; Sánchez-Sevilla et al., 2014) and a PCR-based marker cosegregating with the phenotype was developed (Chambers et al., 2014). Since most of the Australian cultivars lack or contain very low levels of γ-decalactone (Samykanno, 2012), a functional marker may be used to facilitate the introgression of allele controlling γ-decalactone production from other germplasm into the Australian cultivars. However, the genetic improvement of strawberry flavour is relatively slow due to the complex genetic control of fruit flavour biogenesis. Hence, there
is a crucial need to develop molecular markers associated with the genes controlling other strawberry flavour compounds as a valuable tool for elite plant selection in the Australian strawberry breeding program. As explained in Chapter 2 (page 50), it is difficult to create a mapping population which will produce an appropriate segregation ratio among the progeny for octoploid strawberry as the parental genotypes are usually highly heterozygous (Maliepaard et al., 1997). This is because strawberry is sensitive to inbreeding depression (Niemirowicz-Szczytt, 1988), resulting in the lack of inbred lines to be used as parents for generating a mapping population. Consequently, an alternative approach called Bulked Segregant Analysis (BSA) (Michelmore et al., 1991) coupled with SDA was used in the current study to identify genetic determinants of fruit flavour in strawberry.

Bulked Segregant Analysis (BSA), developed by Michelmore et al. (1991) is a rapid method for detecting DNA markers linked to any specific gene in the genome without the need for inbred parents. This method involves screening two DNA bulks with contrasting phenotypic traits to identify polymorphic markers linked to the loci determining the trait of interest (Michelmore et al., 1991). Application of the BSA approach in conjunction with molecular markers such as RAPD and SSR have been widely used to understand the genetic control of certain agronomically important traits. For instance, identification of markers linked to drought resistance in maize (Quarrie et al., 1999), markers associated with linolenic and erucic acid levels in spring rapeseed (Rajcan et al., 1999), QTLs linked to pod and kernel traits in peanut (Gomez Selvaraj et al., 2009) and tagging of brown planthopper resistance genes in rice (Venkateswarlu et al., 2012). BSA has also been used to identify RAPD markers associated with Rfp1 genes conferring resistance to Phytophthora fragariae var. fragariae, the causal agent of red stele root rot in cultivated strawberry (Haymes et al., 1997). In addition, the application of BSA in combination with
DArT has been reported for the identification of the DNA region corresponding to pubescent leaf (mPub) alleles in barley (Wenzl et al., 2007). Therefore, BSA is considered a flexible strategy that is amenable to other marker systems, e.g. the hybridisation-based marker technologies.

This chapter explains the identification of DNA markers associated with some of the key volatile compounds in cultivated strawberry using an integrated approach of BSA and the strawberry-specific SDA described in Chapter 2. It was hypothesised that pooling of DNA from individual progeny plants exhibiting two extreme phenotypes, i.e. either high (H) or undetectable (L) levels of a key volatile compound, should homogenise other loci unrelated to the particular key volatile compounds in study. The bulking process could minimise the differences in genetic background thus enabling DNA sequences specifically related to that compound to be identified. The putative DNA markers were also sequenced to identify the genes located close to the markers which may be associated with the production of the respective key volatile compounds.

The objectives of this study were to:

1. Perform Bulked Segregant Analysis (BSA) based on the aroma profiles of 50 progeny plants resulting from a cross between Juliette and 07-102-41.
2. Screen for DNA polymorphism between two phenotypically extreme bulks using the strawberry-specific Subtracted Diversity Array (SDA).
3. Evaluate the potential of SDA for screening DNA markers associated with strawberry flavour by searching for genes closely linked to the markers which may potentially be involved in the biosynthesis of key volatile compounds.
3.2 MATERIALS AND METHODS

3.2.1 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 Chapter 2. 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 F1 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.

3.2.2 Phenotyping of strawberry flavour

Sample preparation, aroma profiling and data analysis for the 50 F1 progeny plants were performed as described in Section 2.2.5 (page 68). Target compounds for DNA marker development were chosen based on the parental aroma profiles as described in Table 2.4 (page 91) and their relative contribution to strawberry flavour as elucidated in a number of publications (Bood and Zabetakis, 2002; Forney et al., 2000; Hakala et al., 2002; Jetti et al., 2007; Pérez et al., 1992; Ulrich et al., 1997). 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 F₁ 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.

3.2.3 Bulked Segregant Analysis (BSA)

3.2.3.1 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 F₁ 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 according to Section 2.2.2.1 (page 56).

3.2.3.2 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 AluI and HaeIII 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 Section 2.2.3 (page 61) 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 (Fardin et al., 2007).

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 according to Olarte (2011) with a few modifications as follow:

1. **Normalisation factor**

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

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

   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.

3.2.4 Statistical Analysis

3.2.4.1 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 lamda. 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.

### 3.2.4.2 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.
3.2.4.3. 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.

3.2.5 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.
3.2.6 Putative DNA marker validation

3.2.6.1 Primer design and PCR amplification

DNA sequences showing significant similarity (E-value < $10^{-5}$) 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 MgCl₂, 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.

3.2.6.2 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.
3.3 RESULTS AND DISCUSSION

3.3.1 Variability and distribution of volatile compounds in the F1 population compared to the parental genotypes

Aroma profiling of the fruits from the F1 population derived from a cross between Juliette and 07-102-41 successfully identified 41 volatile compounds from the strawberry puree headspace. Of these, 23 volatiles (Table 3.1) have been described as aroma-active compounds in strawberry based on gas chromatography quantification, their odour activity values (OAVs) and sensory analysis (Du et al., 2011; Jetti et al., 2007). The remaining 18 volatile compounds not shown in Table 3.1 (ethyl propanoate, 4-methyl-2-pentanone, ethyl isobutanoate, methyl 2-methylbutanoate, ethyl tiglate, isobutyl butanoate, isopentyl 3-methylbutanoate, methyl octanoate, ethyl benzoate, (E)-hex-3-enyl butanoate, (Z)-hex-3-enyl butanoate, trans-2-hexenyl butanoate, ethyl octanoate, 3-methylbutyl hexanoate, hexyl hexanoate, (E)-cinnamyl acetate, (Z)-ethyl cinnamate and benzaldehyde) exhibited a population mean value lower than 0.1 % and were detected infrequently in the F1 population. Some of these compounds have been reported to have low broad-sense heritability values in strawberry (Samykanno, 2012) and apple (Rowan et al., 2009). Therefore, they were excluded for DNA marker development since environmental factors may have a greater effect in controlling the inheritance of these compounds.

The descriptive statistics of the 23 volatile compounds for the F1 population are shown in Table 3.1 in comparison to their parental genotypes. Overall, 14 compounds were detected in either one or both of the parents whereas nine compounds were not identified in any of the parental genotypes, but were detected in the F1 plants. Distinctive volatile patterns were observed between parents, with ‘07-102-41’ displaying higher levels of C4 and C6 esters and linalool while the levels of C7 and C8 esters, mesifuranne, (E)-nerolidol and
γ-dodecalactone were higher in the second parent ‘Juliette’ (Table 3.1). This result is in agreement with the study which purported that strawberry aroma patterns (i.e., the combination and intensity of volatile compounds) were genotype-dependent (Hakala et al., 2002). Furthermore, most of the compounds detected in both parents showed a significant difference in the levels of production except for hexyl acetate and (E)-hex-2-enyl acetate (Table 3.1). Nonetheless, diversity in relative composition could be inferred from the range between minimum and maximum values of the F₁ population (Table 3.1).

Among the 23 volatiles, eight compounds were considered to have the greatest impact on strawberry aroma due to their low threshold values and high OAVs (Du et al., 2011; Jetti et al., 2007). The frequency distributions of these eight key volatile compounds, including four esters (methyl butanoate, ethyl butanoate, methyl hexanoate, ethyl hexanoate), one furanone (mesifuranne), two terpenes (linalool and (E)-nerolidol) and one lactone (γ-dodecalactone) were illustrated in Figure 3.1 a - h. Most of the frequency distributions were highly or moderately skewed towards lower values or zero except for methyl hexanoate which approached a bimodal distribution (Figure 3.1c). This result is in accordance with the frequency distributions for aroma compounds analysed in strawberry (Olbricht et al., 2008); apple (Dunemann et al., 2009) and peach (Eduardo et al., 2013) except that none of the compounds found in this study have a frequency distribution skewed towards the higher values. This type of frequency distribution is typical of a trait with polygenic inheritance, indicating that volatile compound production is under the control of more than one gene (Dunemann et al., 2009).
Table 3.1 Descriptive statistics for 23 volatile compounds analysed in the F$_1$ population and in the parental genotypes ‘07-102-41’ and ‘Juliette’.

<table>
<thead>
<tr>
<th>Volatile compounds</th>
<th>F$_1$ population</th>
<th>07-102-41</th>
<th>Juliette</th>
<th>$t$-Test (for parents)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean$^a$</td>
<td>Min$^b$</td>
<td>Max$^b$</td>
<td>Mean$^c$</td>
</tr>
<tr>
<td><strong>Esters</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Methyl butanote</td>
<td>9.4</td>
<td>0.0</td>
<td>34.3</td>
<td>15.4</td>
</tr>
<tr>
<td>Methyl 3-methylbutanote</td>
<td>0.3</td>
<td>0.0</td>
<td>2.7</td>
<td>0.0</td>
</tr>
<tr>
<td>Ethyl butanate</td>
<td>4.8</td>
<td>0.0</td>
<td>23.2</td>
<td>7.2</td>
</tr>
<tr>
<td>Isopropyl butanate</td>
<td>0.5</td>
<td>0.0</td>
<td>3.1</td>
<td>0.0</td>
</tr>
<tr>
<td>Ethyl 2-methylbutanote</td>
<td>0.2</td>
<td>0.0</td>
<td>2.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Ethyl isovalerate</td>
<td>0.7</td>
<td>0.0</td>
<td>14.5</td>
<td>0.0</td>
</tr>
<tr>
<td>Isoamyl acetate</td>
<td>0.8</td>
<td>0.0</td>
<td>7.3</td>
<td>0.0</td>
</tr>
<tr>
<td>Methyl hexanoate</td>
<td>14.1</td>
<td>0.0</td>
<td>48.1</td>
<td>25.0</td>
</tr>
<tr>
<td>Butyl butanoate</td>
<td>0.25</td>
<td>0.0</td>
<td>1.9</td>
<td>3.7</td>
</tr>
<tr>
<td>Ethyl hexanoate</td>
<td>6.9</td>
<td>0.0</td>
<td>35.6</td>
<td>16.2</td>
</tr>
<tr>
<td>(Z)-Hex-3-enyl acetate</td>
<td>0.7</td>
<td>0.0</td>
<td>5.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Hexyl acetate</td>
<td>2.1</td>
<td>0.0</td>
<td>35.8</td>
<td>0.0</td>
</tr>
<tr>
<td>(E)-Hex-2-enyl acetate</td>
<td>2.5</td>
<td>0.0</td>
<td>14.8</td>
<td>0.0</td>
</tr>
<tr>
<td>Isopropyl hexanoate</td>
<td>0.2</td>
<td>0.0</td>
<td>1.2</td>
<td>0.0</td>
</tr>
<tr>
<td>Benzyl acetate</td>
<td>1.4</td>
<td>0.0</td>
<td>23.7</td>
<td>0.0</td>
</tr>
<tr>
<td>Hexyl butanoate</td>
<td>0.4</td>
<td>0.0</td>
<td>9.2</td>
<td>0.0</td>
</tr>
</tbody>
</table>

Table 3.1 continued on next page
**Table 3.1 (continued)**

<table>
<thead>
<tr>
<th>Volatile compounds</th>
<th>F&lt;sub&gt;1&lt;/sub&gt; population</th>
<th>07-102-41</th>
<th>Juliette</th>
<th>t-Test (for parents)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Min&lt;sup&gt;b&lt;/sup&gt;</td>
<td>Max&lt;sup&gt;b&lt;/sup&gt;</td>
<td>Mean&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
<tr>
<td><strong>Aldehydes</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hexanal</td>
<td>1.0</td>
<td>0.0</td>
<td>15.2</td>
<td>0.0</td>
</tr>
<tr>
<td>(E)-Hex-2-enal</td>
<td>7.5</td>
<td>0.0</td>
<td>63.7</td>
<td>0.0</td>
</tr>
<tr>
<td><strong>Furanone</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mesifuranne</td>
<td>3.2</td>
<td>0.0</td>
<td>32.4</td>
<td>0.0</td>
</tr>
<tr>
<td><strong>Terpenes</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Linalool</td>
<td>3.3</td>
<td>0.0</td>
<td>18.3</td>
<td>9.1</td>
</tr>
<tr>
<td>(E)-Nerolidol</td>
<td>7.7</td>
<td>0.0</td>
<td>36.9</td>
<td>17.4</td>
</tr>
<tr>
<td>α-Terpineol</td>
<td>0.9</td>
<td>0.0</td>
<td>10.0</td>
<td>0.0</td>
</tr>
<tr>
<td><strong>Lactones</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>γ-Dodecalactone</td>
<td>1.4</td>
<td>0.0</td>
<td>7.9</td>
<td>0.0</td>
</tr>
</tbody>
</table>

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

<sup>a</sup> Mean of all analysed F<sub>1</sub> individuals, based on three technical replicates per genotype

<sup>b</sup> F<sub>1</sub> individuals with the lowest (Min) and highest (Max) relative composition, mean from three technical replicates

<sup>c</sup> Mean and standard deviation (SD) from all technical replicates

<sup>d</sup> t-Test between the parents ‘07-102-41’ and ‘Juliette’

ns: not significant (<i>p > 0.05</i>); * Significant at 0.05 ≥ <i>p > 0.01</i>; ** Significant at 0.01 ≥ <i>p > 0.001</i>
Figure 3.1 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 F₁ 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 3.1 continued on next page
Figure 3.1 (continued)
Figure 3.1 (continued)
Figure 3.1 (continued)
Transgressive segregation, the phenomenon where the individuals in a segregating population exhibit phenotypes that are extreme compared to the parental genotypes (deVicente and Tanksley, 1993; Rieseberg et al., 1999; Rieseberg et al., 2003), was observed for all the compounds assessed except for (E)-nerolidol (Figure 3.1g). For the ester compounds such as ethyl hexanoate and ethyl butanoate, 86 % and 74 % of the F₁ progeny possessed levels of these compounds outside the levels of the two parental genotypes, respectively (Figure 3.1b and d). Similar frequency distribution was also noted for compounds that were missing in either one of the parents but present in most of the F₁ progeny. For instance, both the methyl esters were detectable in 07-102-41 breeding line but missing in Juliette. Like the parent 07-102-41, methyl butanoate and methyl hexanoate were identified at 82 % and 90 % of the F₁ progeny, respectively (Figure 3.1a and c). Moreover, hexyl acetate was detected in the parent Juliette and 62 % of the F₁ progeny, with the maximum levels approaching 24 x the amount observed in Juliette (Table 3.1).

In addition, only 4 % of the F₁ progeny possessed levels of linalool between the parental means. 12 % and 84 % of the F₁ plants showed levels of linalool outside the levels of the parental genotypes in the positive and negative directions, respectively (Figure 3.1f). The maximum level of linalool was found to be 3 x higher than in Juliette and 2 x higher than in 07-102-41 (Table 3.1). Likewise, γ-dodecalactone was detected only in Juliette and 24 % of the F₁ progeny (Figure 3.1h), with a population mean of 1.4 %. The highest level of γ-dodecalactone in F₁ progeny was almost 2 x more than the levels in Juliette. Parental mean was significantly different at \( p < 0.01 \) (Table 3.1). Similarly, mesifuranne was detected only in Juliette and 38 % of the F₁ progeny (Figure 3.1e). One possible explanation could be the low stability of furaneol, where its degradation is pH and temperature-dependent (Hirvi et al., 1980). In addition, furaneol is quickly methylated to
mesifuranne during fruit ripening (Lunkenbein et al., 2006; Pérez et al., 1996a). The maximum level of mesifuranne detected in the F1 progeny (32.4%) was approximately 6 x higher compared to Juliette (Table 3.1). Comparable results were also observed in several other studies on the aroma profiles of tomato (Zanor et al., 2009), strawberry (Olbricht et al., 2008); apple (Dunemann et al., 2009) and peach (Eduardo et al., 2013). Based on the frequency distributions of key volatile compounds assessed in this study, it can be concluded that transgressive segregation was observed in a number of compounds in the segregating population. The high levels of transgressive segregation may be due to the significant differences between parental means (Table 3.1), that is, the parental genotypes were phenotypically very different from each other (Bell and Travis, 2005).

In addition, a number of compounds displayed a mix of qualitative and quantitative inheritance as described by Eduardo et al. (2013), where more than 50% of the F1 progeny accumulated none or very low levels (less than 5% of the population mean) of a particular compound. For instance, butyl butanoate was not detected in Juliette and in up to 72% of the F1 progeny. The population mean for this compound was extremely low at 0.3% (Table 3.1). Four other ester compounds (ethyl 2-methylbutanoate, ethyl 3-methylbutanoate, (Z)-hex-3-enyl acetate and (E)-hex-2-enyl acetate) were not identified in 07-102-41 but found to be present in Juliette and in 10 – 38% of the progeny plants, with a population mean ranging from 0.2% to 2.5% (Table 3.1). These results are in accordance with those presented in a former study where the authors proposed that the genetic variation of volatile compounds (e.g. terpene levels in Thymus and Mentha), could be influenced by a combination of Mendelian inheritance and more complex epistatic interactions (Theis and Lerdau, 2003). Although it is not within the scope of this study to investigate the effects of the environment on the biosynthesis of volatile compounds, it
should be noted that both genetic factors and/or environmental conditions may affect strawberry aroma qualitatively and quantitatively (Eduardo et al., 2013).

Interestingly, a terpenoid (alpha-terpineol) which is usually found at higher concentrations in wild strawberry (*F. vesca*) accessions (Ulrich and Olbricht, 2013) was detected in 10 % of the F1 progeny with an extremely low population mean of 0.9 %. This type of volatile compound may be easily lost after two to three generations of cultivar breeding, and was termed the “funnel effect” in plant breeding by Ulrich et al. (1997). This could also be the reason why most cultivated strawberries are lacking the herbaceous impressions given by terpenoids as observed in the wild strawberries (Ulrich and Olbricht, 2013). Similar inheritance was observed in methyl anthranilate, where it has been regarded as a typical example of the “funnel effect” in cultivated strawberry because it was inherited by only one-third of the progeny derived from a cross between the cultivars Mieze Schindler and Elsanta (Olbricht et al., 2008).

Several other compounds which were not detected in the parental genotypes were characterised in the F1 population at different levels. For examples, methyl 3-methylbutanoate, isopropyl butanoate, isoamyl acetate, isopropyl hexanoate, benzyl acetate, hexyl butanoate, hexanal and α-terpineol were detected at low levels with a population mean ranging from 0.2 to 1.4 % (Table 3.1). In contrast, (E)-hex-2-enal which was missing in both parental genotypes was detected in 68 % of the F1 progeny, with a population mean of 7.5 % and maximum levels approaching 63.7 % (Table 3.1).

Overall, the eight key volatile compounds including methyl butanoate, ethyl butanoate, methyl hexanoate, ethyl hexanoate, mesifuranne, linalool, (E)-nerolidol and
γ-dodecalactone) were selected for DNA marker development based on: (1) their importance in the characterisation of strawberry aroma (Du et al., 2011; Hakala et al., 2002; Jetti et al., 2007) and (2) their significant differences in parental means (Table 3.1). In theory, the probability of detecting the alleles controlling the trait of interest is higher if the parental means are significantly different from each other. Furthermore, the volatile compositions of two Australian-grown cultivars, Albion and Juliette, have been evaluated extensively over ten weeks in one growing season. The inheritance analysis revealed that methyl butanote, ethyl butanoate, mesifuran and (E)-nerolidol possessed high broad sense heritability values (46.0 % - 74.4 %) (Samykanno, 2012). This result coincides with the genotype-by-environment analysis, where all selected compounds, except linalool, were found to be predominantly influenced by genotype. These findings also suggested that the yearly phenotypic changes of strawberry plants were not as great as seasonal variations (Samykanno et al., 2013). Therefore, we have greater confidence to identify molecular markers associated with the selected key volatile compounds. Although no F₂ population was available for analysis, it is plausible that the F₁ individuals with extreme phenotypes (i.e. much higher or not-detected) can be used as DNA materials in the subsequent Bulked Segregant Analysis (BSA) for the discovery of alleles segregating for the key volatile compounds.

3.3.2 Statistical Analysis

3.3.2.1 Discriminant Function Analysis (DFA)

For DFA, the technical replicates from the SDA data were pre-assigned into one of the two phenotypic extreme groups (‘H’ and ‘L’ bulks for each selected key volatile compound). A training set (original cases) comprising six technical replicates from the ‘H’ bulk and another six technical replicates from the ‘L’ bulk was used to compute a discriminant
function based on the stepwise method. Since there are only two grouping variables (i.e., the ‘H’ and ‘L’ DNA bulks), a single discriminant function which accounted for 100% of the variance in hybridisation intensity was generated for all key volatile compounds tested, suggesting that the DFA-selected putative DNA markers could efficiently differentiate the membership of the original cases in the two groups showing extreme phenotypes (Table 3.2). The most predictive variables/features which formed the discriminant function were identified based on the canonical discriminant function coefficients which explained the relative contribution of each feature to the discriminant function (Table 3.2). Wilks’ lambda was statistically significant for the discriminant function generated for methyl butanoate ($\chi^2 = 51.19$, $p < 0.001$), ethyl butanoate ($\chi^2 = 60.51$, $p < 0.001$), methyl hexanoate ($\chi^2 = 47.77$, $p < 0.001$), ethyl hexanoate ($\chi^2 = 59.18$, $p < 0.001$), mesifuranne ($\chi^2 = 63.31$, $p < 0.001$), linalool ($\chi^2 = 60.52$, $p < 0.001$), (E)-nerolidol ($\chi^2 = 50.92$, $p < 0.001$) and γ-dodecalactone ($\chi^2 = 48.97$, $p < 0.001$).
Table 3.2 Discriminant functions generated for each key volatile compound based on the canonical discriminant function coefficients of the most predictive features.

<table>
<thead>
<tr>
<th>Key volatile compounds</th>
<th>Canonical Discriminant Function Coefficients</th>
</tr>
</thead>
<tbody>
<tr>
<td>Methyl butanoate</td>
<td>( Y = -21.033 - 0.017(\text{FLP2E7}) + 0.348(\text{FLP1E7}) - 0.126(\text{FLP1D7}) - 0.005(\text{FLP1A2}) + 0.294(\text{FLP3H8}) - 0.061(\text{FLP3A2}) )</td>
</tr>
<tr>
<td>Ethyl butanoate</td>
<td>( Y = -149.182 + 0.120(\text{FLP1D11}) - 0.087(\text{FLP2D4}) + 0.068(\text{FLP1E4}) + 0.404(\text{FLP3F12}) + 0.553(\text{FLP3B9}) + 0.376(\text{FLP3A2}) )</td>
</tr>
<tr>
<td>Methyl hexanoate</td>
<td>( Y = -20.855 + 0.388(\text{FLP2F1}) + 0.128(\text{FLP2A11}) + 1.215(\text{FLP1E7}) - 0.163(\text{FLP1A1}) )</td>
</tr>
<tr>
<td>Ethyl hexanoate</td>
<td>( Y = -75.233 + 0.220(\text{FLP2G3}) - 0.017(\text{FLP2E1}) + 0.402(\text{FLP1B3}) - 0.136(\text{FLP1G2}) - 0.068(\text{FLP1D8}) + 0.565(\text{FLP4B3}) )</td>
</tr>
<tr>
<td>Mesifuranne</td>
<td>( Y = -58.447 - 1.215(\text{FLP2G4}) + 0.111(\text{FLP2E2}) + 2.183(\text{FLP2E6}) - 0.079(\text{FLP3C3}) - 1.279(\text{FLP3H11}) + 2.248(\text{FLP3B4}) )</td>
</tr>
<tr>
<td>Linalool</td>
<td>( Y = -100.778 + 0.884(\text{FLP1C1}) + 0.385(\text{FLP3H8}) + 0.095(\text{FLP3A1}) - 0.149(\text{FLP4C6}) - 0.323(\text{FLP3G6}) - 0.012(\text{FLP3E12}) )</td>
</tr>
<tr>
<td>(E)-Nerolidol</td>
<td>( Y = -112.283 + 0.272(\text{FLP2D11}) + 0.134(\text{FLP2A3}) - 0.114(\text{FLP2E4}) - 0.059(\text{FLP1G8}) - 0.038(\text{FLP1A11}) + 0.71(\text{FLP3H10}) + 0.181(\text{FLP3F10}) )</td>
</tr>
<tr>
<td>γ-dodecalactone</td>
<td>( Y = -60.350 + 0.028(\text{FLP2E1}) + 0.041(\text{FLP1E3}) - 0.075(\text{FLP1D11}) + 0.367(\text{FLP1A7}) + 0.101(\text{FLP3C3}) - 0.534(\text{FLP3F8}) + 0.947(\text{FLP3F12}) - 0.173 (\text{FLP3E8}) )</td>
</tr>
</tbody>
</table>
Of the DFA-selected putative DNA markers, seven (FLP1E7, FLP3H8, FLP3A2, FLP1D11, FLP3F12, FLP2E1 and FLP3C3) were found to be important discriminants of more than two key volatile compounds (Table 3.3). These may be markers corresponding to important compounds in the same chemical classes that share common biosynthetic pathways. Similar results were observed in a study performed by Bonamico et al. (2010), where at least one SSR marker was common between two to three phenotypic traits assessed in maize (Bonamico et al., 2010).

The classification results based on the discriminant functions showed that 100 % of all the original cases in the training set were correctly classified (Table 3.3). However, not all the new cases in the test set were found to be correctly classified. Only the cases for ethyl butanoate, mesifuranne and \( \gamma \)-dodecalactone were found to be perfectly predicted (100 %) by their respective discriminant functions (Table 3.3). In contrast, 91.7 % of the new cases were correctly classified for methyl butanoate and linalool. Likewise, 83.3 % of the new cases were correctly predicted for ethyl hexanoate and (\( E \))-nerolidol. Finally, 75.0 % of the new cases were correctly assigned for methyl butanoate (Table 3.3). The present results are consistent with other studies, where the correct classification ranged between 67 - 100 % for microsatellite markers associated with the geographic distribution of wild emmer wheat (Fahima et al., 2002) and 86 – 100 % for DFA-selected RFLP/SSR alleles associated with several rice agronomic traits (Zhang et al., 2005).
Table 3.3 Putative DNA markers selected by DFA and the classification results for both the training set and test set based on the discriminant functions generated for each key volatile compound.

<table>
<thead>
<tr>
<th>Key volatile compounds</th>
<th>DFA-selected markers*</th>
<th>Classification Results (%)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Training set</td>
<td>Test set</td>
</tr>
<tr>
<td>Methyl butanoate</td>
<td>FLP2E7, <strong>FLP1E7</strong>, FLP1D7, FLP1A2, <strong>FLP3H8, FLP3A2</strong></td>
<td>100.0</td>
<td>75.0</td>
</tr>
<tr>
<td>Ethyl butanoate</td>
<td><strong>FLP1D11</strong>, FLP2D4, FLP1E4, <strong>FLP3F12, FLP3B9, FLP3A2</strong></td>
<td>100.0</td>
<td>100.0</td>
</tr>
<tr>
<td>Methyl hexanoate</td>
<td>FLP2F1, FLP2A11, <strong>FLP1E7</strong>, FLP1A1</td>
<td>100.0</td>
<td>91.7</td>
</tr>
<tr>
<td>Ethyl hexanoate</td>
<td>FLP2G3, <strong>FLP2E1</strong>, FLP1B3, FLP1G2, FLP1D8, FLP4B3</td>
<td>100.0</td>
<td>83.3</td>
</tr>
<tr>
<td>Mesifuranne</td>
<td>FLP2G4, FLP2E2, FLP2E6, <strong>FLP3C3</strong>, FLP3H11, FLP3B4</td>
<td>100.0</td>
<td>100.0</td>
</tr>
<tr>
<td>Linalool</td>
<td>FLP1C1, <strong>FLP3H8</strong>, FLP3A1, FLP4C6, FLP3G6, FLP3E12</td>
<td>100.0</td>
<td>91.7</td>
</tr>
<tr>
<td>(E)-Nerolidol</td>
<td>FLP2D11, FLP2A3, FLP2E4, FLP1G8, FLP1A11, FLP3H10, FLP3F10</td>
<td>100.0</td>
<td>83.3</td>
</tr>
<tr>
<td>γ-Dodecalactone</td>
<td><strong>FLP2E1</strong>, FLP1E3, <strong>FLP1D11</strong>, FLP1A7, <strong>FLP3C3</strong>, FLP3F8, <strong>FLP3F12</strong>, FLP3E8</td>
<td>100.0</td>
<td>100.0</td>
</tr>
</tbody>
</table>

* Features in bold represent putative DNA markers that were detected in more than two key volatile compounds

A second DFA was performed on the selected putative DNA markers to further reduce the number of markers that contributed the most to group separation. A single discriminant function which accounted for 100% of the variance in hybridisation intensity was generated for all compounds tested, indicating that the reduced set of putative DNA markers could also efficiently differentiate the membership of the two groups with extreme phenotypes (Table 3.4). These features were also selected based on their canonical discriminant function coefficients (Table 3.4). Wilks’ lambda was statistically significant.
for the discriminant function generated for methyl butanoate ($\chi^2 = 25.519, p < 0.001$), ethyl butanoate ($\chi^2 = 35.520, p < 0.001$), methyl hexanoate ($\chi^2 = 18.412, p < 0.001$), ethyl hexanoate ($\chi^2 = 42.106, p < 0.001$), mesifuranne ($\chi^2 = 50.911, p < 0.001$), linalool ($\chi^2 = 11.871, p < 0.001$), (E)-nerolidol ($\chi^2 = 31.566, p < 0.001$) and $\gamma$-dodecalactone ($\chi^2 = 28.553, p < 0.001$).

In this study, the reduced sets of putative DNA markers possibly associated with the key volatile compounds may be used to predict the group membership of any future datasets. The number of putative DNA markers selected by the second DFA ranged from 1 – 4 (Table 3.5), depending on the key volatile compounds assessed. Only FLP1E7 was detected in more than two key volatile compounds, which is methyl butanoate (referred to FLP1E7MB) and methyl hexanoate (referred to FLP1E7MH). The classification of the original cases based on the new discriminant functions showed a slight increase in error rate, where only ethyl butanoate, mesifuranne and $\gamma$-dodecalactone were correctly classified (100 %) into the two phenotypic extreme groups (Table 3.5). The proportion of original cases being classified correctly ranged from 83.3 % to 95.8 % for the other key volatile compounds (Table 3.5). In addition, the range of new cases being predicted correctly by cross-validation was estimated at 79.2 % - 100.0 % depending on the key volatile compounds (Table 3.5).
### Table 3.4 Discriminant functions generated for each key volatile compound using the features selected by the first DFA.

<table>
<thead>
<tr>
<th>Key volatile compounds</th>
<th>Canonical Discriminant Function Coefficients</th>
</tr>
</thead>
<tbody>
<tr>
<td>Methyl butanoate</td>
<td>$Y = 0.096 + 0.007(FLP1D7) - 0.013(FLP1E7)$</td>
</tr>
<tr>
<td>Ethyl butanoate</td>
<td>$Y = -2.851 + 0.04(FLP1D11) + 0.20(FLP3A2) - 0.110(FLP3B9)$</td>
</tr>
<tr>
<td>Methyl hexanoate</td>
<td>$Y = -0.332 + 0.024(FLP1E7) - 0.012(FLP2A11)$</td>
</tr>
<tr>
<td>Ethyl hexanoate</td>
<td>$Y = -3.496 + 0.066(FLP1B3) - 0.005(FLP1G2)$</td>
</tr>
<tr>
<td>Mesifuranne</td>
<td>$Y = -1.563 + 0.101(FLP2E6) - 0.038(FLP2G4) - 0.034(FLP3H11)$</td>
</tr>
<tr>
<td>Linalool</td>
<td>$Y = -1.389 + 0.003(FLP3E12)$</td>
</tr>
<tr>
<td>(E)-Nerolidol</td>
<td>$Y = -3.321 - 0.004(FLP1G8) + 0.010(FLP2D11) + 0.007(FLP3F10)$</td>
</tr>
<tr>
<td>γ-Dodecalactone</td>
<td>$Y = -1.742 + 0.012(FLP1A7) + 0.02(FLP2E1) - 0.007(FLP3E8) - 0.003(FLP3F8)$</td>
</tr>
</tbody>
</table>

*Features in bold represent putative DNA markers that were detected in more than two key volatile compounds.

### Table 3.5 The reduced set of putative DNA markers selected by DFA, the classification results for the original cases and the estimated proportion for any future dataset.

<table>
<thead>
<tr>
<th>Key volatile compounds</th>
<th>DFA-selected markers*</th>
<th>Classification Results (%)</th>
<th>Original</th>
<th>Cross-validation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Methyl butanoate</td>
<td>FLP1E7MB, FLP1D7</td>
<td></td>
<td>95.8</td>
<td>95.8</td>
</tr>
<tr>
<td>Ethyl butanoate</td>
<td>FLP1D11, FLP3B9, FLP3A2</td>
<td></td>
<td>100.0</td>
<td>95.8</td>
</tr>
<tr>
<td>Methyl hexanoate</td>
<td>FLP2A11, <strong>FLP1E7MH</strong></td>
<td></td>
<td>87.5</td>
<td>87.5</td>
</tr>
<tr>
<td>Ethyl hexanoate</td>
<td>FLP1B3, FLP1G2,</td>
<td></td>
<td>95.8</td>
<td>95.8</td>
</tr>
<tr>
<td>Mesifuranne</td>
<td>FLP2G4, FLP2E6, FLP3H11</td>
<td></td>
<td>100.0</td>
<td>100.0</td>
</tr>
<tr>
<td>Linalool</td>
<td>FLP3E12</td>
<td></td>
<td>83.3</td>
<td>79.2</td>
</tr>
<tr>
<td>(E)-Nerolidol</td>
<td>FLP2D11, FLP1G8, FLP3F10</td>
<td></td>
<td>95.8</td>
<td>95.8</td>
</tr>
<tr>
<td>γ-Dodecalactone</td>
<td>FLP2E1, FLP1A7, FLP3F8, FLP3E8</td>
<td></td>
<td>100.0</td>
<td>87.5</td>
</tr>
</tbody>
</table>

*Features in bold represent putative DNA markers that were detected in more than two key volatile compounds.
A similar approach was employed in a previous study which identified AFLP markers associated with stress tolerance index in Sardari wheat ecotypes, where stepwise analysis was firstly used to identify the number of polymorphic markers and then further applied to reduce the number of markers and form classification models of up to 24 markers (Siosemarde et al., 2012). The present results are in accordance with several previous studies, where the rate of misclassification increased as the number of predictors decreased (Alwala, 2007; Siosemarde et al., 2012; Zhang et al., 2005). For instance, the study performed by Alwala (2007) showed that only 61.7% - 62.2% of correct classification was obtained based on five markers, and a minimum of 10 markers were needed to achieve more than 90% of correct classification (Alwala, 2007). Moreover, the higher the levels of correct classification, the stronger the association between marker and phenotype that could be inferred (Zhang et al., 2005). Therefore, the low cross-validation error rate produced in this study implied that the reduced sets of putative DNA markers selected by the second DFA possessed high predictive power and could be used for phenotypic grouping.

### 3.3.2.2 Fisher’s ratio

Fisher’s ratio was used as the second measurement to evaluate the linear discriminating power of the SDA features (Lohninger, 1999). In general, a larger Fisher’s ratio value indicates greater differences between the means of two extreme bulks. Hence, the features which showed very strong signal intensities in the ‘H’ DNA bulk but very weak signal intensities in the ‘L’ DNA bulk, or vice versa, will produce larger Fisher’s ratio values providing that the background noise inherent in the microarray system is minimum (sum of the variances of two extreme bulks). As a result, the putative DNA markers with higher
Fisher’s ratio could probably be best in discriminating between the two phenotypic extreme bulks.

The Fisher’s ratio calculated displayed a broad range of values (0.0 to 7.5) depending on different key volatile compounds. The features were then arranged in decreasing order of Fisher’s ratio and only the top 10 features are presented in Table 3.6. It was found that out of 20 putative DNA markers selected by DFA, only six (FLP1D7, FLP1D11, FLP1B3, FLP3E12, FLP2D11 and FLP1A7) features showed high Fisher’s ratio values (Table 3.6). The other DFA-selected putative DNA markers (Table 3.5) did not possess high Fisher’s ratio values, indicating that their group means may not differ significantly. This discrepancy could be attributed to the different algorithms used to compute the variables in DFA and Fisher’s ratio. In DFA, the function is calculated based on weighted combination of variables where group differences on the function are maximised (Howitt and Cramer, 2008) whereas Fisher’s ratio involved only a single variable discriminating two groups based on a simple mathematical equation (Section 3.2.4.2) (Lohninger, 1999). In order to confirm the differences of group means among these putative DNA markers, an Independent Samples t-Test was employed as the third measurement.
Table 3.6 List of top 10 features ranked in decreasing order based on their respective Fisher’s ratio for key volatile compounds.

<table>
<thead>
<tr>
<th>Ranking</th>
<th>Fisher’s ratio</th>
<th>Key Volatile Compounds</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Methyl butanoate</td>
<td>Ethyl butanoate</td>
</tr>
<tr>
<td>1</td>
<td>FLP1C8</td>
<td>FLP3C12</td>
</tr>
<tr>
<td>2</td>
<td>FLP1A4</td>
<td>FLP4D1</td>
</tr>
<tr>
<td>3</td>
<td>FLP1A10</td>
<td>FLP1F12</td>
</tr>
<tr>
<td>4</td>
<td>FLP1E11</td>
<td>FLP3E8</td>
</tr>
<tr>
<td>5</td>
<td>FLP1D7</td>
<td>FLP2C2</td>
</tr>
<tr>
<td>6</td>
<td>FLP1F1</td>
<td>FLP1D11</td>
</tr>
<tr>
<td>7</td>
<td>FLP1E1</td>
<td>FLP4C4</td>
</tr>
<tr>
<td>8</td>
<td>FLP2D7</td>
<td>FLP3C7</td>
</tr>
<tr>
<td>9</td>
<td>FLP2F4</td>
<td>FLP1C6</td>
</tr>
<tr>
<td>10</td>
<td>FLP2B5</td>
<td>FLP2F8</td>
</tr>
</tbody>
</table>

Cells shaded in yellow represent putative DNA markers which were also selected by DFA.
3.3.2.3 Independent Samples t-Test

SPSS procedures for Independent Samples t-Test provided two versions of output depending on whether the two sets of scores have similar or distinct variances. In cases where the variances for two phenotypic extreme groups were significantly unequal \((p < 0.05)\), a t-test for unequal variances was used (Howitt and Cramer, 2008). The results showed that the mean differences between the ‘H’ and ‘L’ DNA bulks for eleven putative DNA markers (FLP1D7, FLP1D11, FLP3A2, FLP3B9, FLP1E7MH, FLP1B3, FLP1G2, FLP2E6, FLP3E12, FLP2D11 and FLP1A7) selected by DFA were statistically significant \((p < 0.01)\) (Table 3.7). This result is also corroborated by six of the features (FLP1D7, FLP1D11, FLP1B3, FLP3E12, FLP2D11 and FLP1A7) as putative DNA markers which could be the predictor for their respective key volatile compounds.
Table 3.7 Group statistics and Independent Samples t-Test for the putative DNA markers selected by DFA.

<table>
<thead>
<tr>
<th>Key volatile compounds</th>
<th>DFA-selected markers</th>
<th>H Mean</th>
<th>SD</th>
<th>L Mean</th>
<th>SD</th>
<th>t value</th>
<th>df</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Methyl butanoate</td>
<td>FLP1D7&lt;sup&gt;b&lt;/sup&gt;</td>
<td>248.23</td>
<td>84.05</td>
<td>578.61</td>
<td>291.75</td>
<td>-3.77</td>
<td>12.8</td>
<td>**</td>
</tr>
<tr>
<td></td>
<td>FLP1E7MB&lt;sup&gt;a&lt;/sup&gt;</td>
<td>242.69</td>
<td>91.59</td>
<td>192.24</td>
<td>97.77</td>
<td>1.31</td>
<td>22</td>
<td>ns</td>
</tr>
<tr>
<td>Ethyl butanoate</td>
<td>FLP1D11&lt;sup&gt;b&lt;/sup&gt;</td>
<td>279.03</td>
<td>85.56</td>
<td>1333.86</td>
<td>541.59</td>
<td>-6.66</td>
<td>11.6</td>
<td>**</td>
</tr>
<tr>
<td></td>
<td>FLP3A2</td>
<td>81.80</td>
<td>16.68</td>
<td>180.47</td>
<td>72.65</td>
<td>-4.59</td>
<td>12.2</td>
<td>**</td>
</tr>
<tr>
<td></td>
<td>FLP3B9</td>
<td>17.72</td>
<td>4.95</td>
<td>35.11</td>
<td>16.73</td>
<td>-3.45</td>
<td>12.9</td>
<td>**</td>
</tr>
<tr>
<td>Methyl hexanoate</td>
<td>FLP1E7MH&lt;sup&gt;a&lt;/sup&gt;</td>
<td>13.48</td>
<td>6.31</td>
<td>136.56</td>
<td>95.59</td>
<td>-4.45</td>
<td>11.1</td>
<td>**</td>
</tr>
<tr>
<td></td>
<td>FLP2A11</td>
<td>96.35</td>
<td>48.97</td>
<td>151.74</td>
<td>111.30</td>
<td>-1.58</td>
<td>15.1</td>
<td>ns</td>
</tr>
<tr>
<td>Ethyl hexanoate</td>
<td>FP1B3&lt;sup&gt;b&lt;/sup&gt;</td>
<td>44.43</td>
<td>14.18</td>
<td>142.83</td>
<td>35.33</td>
<td>-8.95</td>
<td>14.5</td>
<td>**</td>
</tr>
<tr>
<td></td>
<td>FLP1G2</td>
<td>340.87</td>
<td>113.77</td>
<td>640.48</td>
<td>293.72</td>
<td>-3.30</td>
<td>14.2</td>
<td>**</td>
</tr>
<tr>
<td>Mesifuranne</td>
<td>FLP2E6</td>
<td>90.28</td>
<td>26.69</td>
<td>18.70</td>
<td>8.90</td>
<td>8.81</td>
<td>13.4</td>
<td>**</td>
</tr>
<tr>
<td></td>
<td>FLP2G4</td>
<td>78.51</td>
<td>53.40</td>
<td>64.81</td>
<td>18.80</td>
<td>0.84</td>
<td>22</td>
<td>ns</td>
</tr>
<tr>
<td></td>
<td>FLP3H11</td>
<td>34.96</td>
<td>12.62</td>
<td>37.67</td>
<td>15.79</td>
<td>-0.47</td>
<td>22</td>
<td>ns</td>
</tr>
<tr>
<td>Linalool</td>
<td>FLP3E12&lt;sup&gt;b&lt;/sup&gt;</td>
<td>884.13</td>
<td>521.74</td>
<td>226.77</td>
<td>218.23</td>
<td>4.03</td>
<td>14.7</td>
<td>**</td>
</tr>
<tr>
<td>(E)-Nerolidol</td>
<td>FLP1G8</td>
<td>684.26</td>
<td>421.21</td>
<td>807.65</td>
<td>535.49</td>
<td>-0.63</td>
<td>22</td>
<td>ns</td>
</tr>
<tr>
<td></td>
<td>FLP2D11&lt;sup&gt;b&lt;/sup&gt;</td>
<td>493.63</td>
<td>156.16</td>
<td>303.39</td>
<td>130.25</td>
<td>3.24</td>
<td>22</td>
<td>**</td>
</tr>
<tr>
<td></td>
<td>FLP3F10</td>
<td>448.95</td>
<td>276.61</td>
<td>274.31</td>
<td>155.03</td>
<td>1.91</td>
<td>17.3</td>
<td>ns</td>
</tr>
</tbody>
</table>

Table 3.7 continued next page
Table 3.7 (continued)

<table>
<thead>
<tr>
<th>Key volatile compounds</th>
<th>DFA-selected markers</th>
<th>H</th>
<th>L</th>
<th>t-Test (for two bulks)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
</tr>
<tr>
<td>γ-Dodecalactone</td>
<td>FLP1A7(^b)</td>
<td>282.66</td>
<td>119.63</td>
<td>106.71</td>
</tr>
<tr>
<td></td>
<td>FLP2E1</td>
<td>1451.21</td>
<td>887.38</td>
<td>795.11</td>
</tr>
<tr>
<td></td>
<td>FLP3E8</td>
<td>317.99</td>
<td>214.62</td>
<td>314.93</td>
</tr>
<tr>
<td></td>
<td>FLP3F8</td>
<td>175.01</td>
<td>132.88</td>
<td>223.85</td>
</tr>
</tbody>
</table>

\(^a\) Putative DNA markers which were selected for more than two key volatile compounds

\(^b\) Putative DNA markers which displayed high Fisher’s ratio values

SD Standard deviation

** Significant at \(p < 0.01\)

ns: Not significant \((p > 0.05)\)
Consequently, a three-way Venn diagram was generated using the three statistical analyses to select a final set of putative DNA markers (Figure 3.2). Six features were identified at the intersection of DFA, Fisher’s ratio (top 10 features) and Independent Samples t-Test \((p < 0.01)\) (Figure 3.2). Among the ester compounds tested, FLP1D7, FLP1D11 and FLP1B3 were found to be putatively associated with methyl butanoate, ethyl butanoate and ethyl hexanoate, respectively. For terpene compounds, FLP3E12 and FLP2D11 were putatively correlated with linalool and \((E)\)-nerolidol, respectively. FLP1A7 was putatively linked to \(\gamma\)-dodecalactone. The identities of these putative markers were revealed by DNA sequencing and similarity search against the strawberry genome database is discussed in Section 3.3.3.

In general, features exhibiting significant differences of group means will be expected to yield a higher Fisher’s ratio value. However, the intersection between DFA and Independent Samples t-Test revealed five other features (FLP3A2, FLP3B9, FLP1E7MH, FLP1G2 and FL2E6) which differed significantly in their group means were not included in the top 10 features with the highest Fisher’s ratios (Figure 3.2). Their low Fisher’s ratio values could be explained by smaller differences between group means or greater sum of the variances of the two extreme bulks. Some of these features (FLP3A2, FLP3B9, FLP1E7MH and FLP2E6) displayed low signal intensities as their group means were relatively lower compared to the other features (Table 3.7). In contrast, the Fisher’s ratio value of FLP1G2 which showed high signal intensities was lowered by a large sum of the variances (data not shown). Moreover, the group means of nine other features (FLP1E7MB, FLP2A11, FLP2G4, FLP3H11, FLP1G8, FLP3F10, FLP2E1, FLP3E8 and FLP3F8) were statistically not significant \((p > 0.01)\). This result further supports the data obtained from
the Fisher’s ratio analysis, where all of these features were having low Fisher’s ratio values (Table 3.7).

![Venn diagram](image)

**Figure 3.2** A three-way Venn diagram showing the putative DNA markers in the intersection of DFA (green), Fisher’s ratio (top 10 features; blue) and Independent Samples t-Test ($p < 0.01$; red) for all the key volatile compounds tested.

### 3.3.3 Sequence identity of the putative DNA markers

Sequencing results revealed that all the DNA sequences comprised of two different adaptors at both ends (Appendix 10), suggesting that the suppression PCR was efficient in eliminating common DNA fragments between the tester and driver pools, and that these sequences are tester-specific. Sequence identities were obtained by searching against the *Fragaria vesca* draft genome (v1.1) using the PFR and GDR strawberry servers (GDR,
2009; PFR, 2010). Out of the six features sequenced, FLP1A7 and FLP1D7 appeared to be nuclear-specific whereas FLP1B3, FLP1D11, FLP2D11 and FLP3E12 were chloroplast-specific (Table 3.8).
Table 3.8 Putative identity of the most discriminatory features searched against the *Fragaria vesca* draft genome (v1.1). E-value regarded as significant if $< 1e^{-5}$.

<table>
<thead>
<tr>
<th>Clones</th>
<th>Length (bp)</th>
<th>Landmark or region</th>
<th>Sequence description</th>
<th>E-value</th>
<th>Specific to target</th>
</tr>
</thead>
<tbody>
<tr>
<td>FLP1A7</td>
<td>442</td>
<td>LG6:21708323..21708764, scf0513196:589686-590127</td>
<td>Genomic DNA region on linkage group 6</td>
<td>0.0</td>
<td>γ-Dodecalactone</td>
</tr>
<tr>
<td>FLP1B3</td>
<td>343</td>
<td>gene32946 on scf0510865:52..396</td>
<td>NAD(P)H-quinone oxidoreductase subunit H, chloroplastic (similar to)</td>
<td>$5e^{-45}$</td>
<td>Ethyl hexanoate</td>
</tr>
<tr>
<td>FLP1D7</td>
<td>627</td>
<td>LG3:27624656..27625284, scf0513138:502060-502688,</td>
<td>Genomic DNA region on linkage group 3</td>
<td>0.0</td>
<td>Methyl butanoate</td>
</tr>
<tr>
<td></td>
<td></td>
<td>LG2:17544790..17544932, scf0513123:85522..85664</td>
<td>Genomic DNA region on linkage group 2</td>
<td>$1e^{-38}$</td>
<td></td>
</tr>
<tr>
<td>FLP1D11</td>
<td>850</td>
<td>gene32967 on scf0510833:190..1040</td>
<td>ATP synthase subunit alpha, chloroplastic (similar to)</td>
<td>$1e^{-133}$</td>
<td>Ethyl butanoate</td>
</tr>
<tr>
<td>FLP2D11</td>
<td>670</td>
<td>scf0510759:1..513</td>
<td>N/A</td>
<td>0.0</td>
<td><em>(E)-Nerolidol</em></td>
</tr>
<tr>
<td>FLP3E12</td>
<td>539</td>
<td>scf0513205:141..680</td>
<td>N/A</td>
<td>0.0</td>
<td>Linalool</td>
</tr>
<tr>
<td></td>
<td></td>
<td>LG3:20817890..20818104, scf0513118:718918-719132</td>
<td>Genomic DNA region on linkage group 3</td>
<td>$6e^{-83}$</td>
<td></td>
</tr>
</tbody>
</table>
The nuclear-specific features were not a match to any genes found in the *F. vesca* genome database, indicating that they are non-coding DNA region possibly linked to genes associated with compound production. The full sequence of FLP1A7 and FLP1D7 showed 100% match (E-value: 0.0) with a DNA region on linkage group 6 (LG6:21708323..21708764, scf0513196:589686..590127) and linkage group 3 (LG3:27624656..27625284, scf0513138:502060..502688), respectively (Table 3.8). However, FLP1D7 could not be assigned to any linkage group when the sequence was searched using the genome browser from GDR. Instead, both PFR Strawberry Server and GDR revealed that FLP1D7 showed significant similarity (E-value: 1e-38) to a sequence in linkage group 2 (LG2:17544790..17544932, scf0513123:85522..85664). Therefore, the later result was chosen for further analysis.

For chloroplast-specific features, FLP1B3 showed significant similarity (E-value: 5e-45) to gene32946, a chloroplastic-like NAD(P)H-quinone oxidoreductase subunit H located on scf0510865:52..396. FLP1D11 presented high similarity (E-value: e-133) to gene32967, a chloroplastic-like ATP synthase subunit alpha positioned on scf0510833:190..1040. In contrast, 77% of FLP2D11 and the full sequence of FLP3E12 revealed 100% match (E-value: 0.0) with scf0510759:1..513 and scf0513205:141..680, respectively and they did not correspond to any genes in the chloroplast genome (Table 3.8). Interestingly, up to 40% of FLP3E12 displayed significant similarity (E-value: 6e-83) to a DNA region on linkage group 3 (LG3:20817890..20818104, scf0513118:718918..719132) (Table 3.8), suggesting that this putative DNA marker could probably be linked to genes which may have dual functions related to both cytosol and chloroplast.
Figure 3.3 Landmark of selected putative DNA markers (red arrows) mapped onto the *F. vesca* draft genome (v1.1) and location of genes associated with key volatile compounds (green arrows). a) FLP1A7 on LG6:21708323..21708764. *ACAT2*: acetyl-CoA acetyltransferase (cytosolic). b) FLP1D7 on LG2:17544790..17544932. *PLA*: Patatin-related phospholipase A; *Ers1*: Ethylene receptor. c) FLP3E12 on LG3:20817890..20818104. *CYP*: Putative cytochrome P450 superfamily protein.
To identify genes possibly involved in production of key volatile compounds, the genes situated within 5 cM on either side of the nuclear-specific (FLP1A7 and FLP1D7) and the potential nuclear-specific (FLP3E12) features were manually searched on the same linkage group using the PFR Strawberry Server (PFR, 2010). Figure 3.3a shows a gene corresponding to *Arabidopsis thaliana* cytosolic acetoacetyl-CoA thiolase II, also known as acetyl-CoA acetyltransferase (*ACAT2*) was discovered at approximately 1.6 Mb downstream of FLP1A7. Moreover, an *A. thaliana* Patatin-related phospholipase A (*PLA*) gene and a *F. x ananassa* ethylene receptor (*Ers1*) gene were found at 1.4 kb and 2.0 Mb downstream of FLP1D7, respectively (Figure 3.3b). Similarly, a *Zea mays* putative cytochrome P450 superfamily protein was found at approximately 0.2 Mb downstream of FLP3E12 (Figure 3.3c). These genes are either involved in lipid catabolism or the biosynthesis of plant secondary metabolites (Sánchez *et al.*, 2013).

FLP1A7 was positively correlated with γ-dodecalactone production in strawberry fruits according to the differences in signal intensities between the ‘H’ (282.66) and ‘L’ (106.71) bulks (Table 3.7). While the precise lactone biosynthetic pathway in plants remain elusive, it is clear that the formation of γ- and δ-lactones starts from β-oxidation of fatty acids (Schwab *et al.*, 2008). Based on the fatty acid degradation pathway in the KEGG database, the acetyl-CoA acetyltransferase (*ACAT2*) gene found at close proximity to FLP1A7 is one of the biosynthetic thiolases that catalyses the reverse reaction in the last step of beta-oxidation (Kanehisa and Goto, 2000; Kanehisa *et al.*, 2014). While the role of the *ACAT2* gene in controlling the levels of γ-dodecalactone formation could not be determined in this study, the differences in signal intensities between the ‘H’ and ‘L’ bulks may suggest the presence of allelic variants in the FLP1A7 locus closely linked to *ACAT2*. To test this hypothesis, direct amplicon sequencing was performed on the PCR product amplified from
the parental genotypes, 07-102-41 and Juliette to determine the sequence polymorphism of FLP1A7 feature (Section 3.3.4).

In contrast, FLP1D7 is negatively correlated to methyl butanoate based on the differences of hybridisation signal between the ‘H’ (248.23) and ‘L’ (578.61) bulks (Table 3.7). It has been reported that amino acids and lipids are the most likely precursors of ester formation (Beekwilder et al., 2004). Interestingly, the Patatin-related phospholipase A (PLA) strongly linked to FLP1D7 (Figure 3.3b), is one of the members in the phospholipase superfamily that hydrolyses the sn1 and/or sn2 position of the membrane phospholipid to release fatty acids (Scherer et al., 2010). Surprisingly, the discovery of ethylene receptor (Ers1) at close proximity to FLP1D7 (Figure 3.3b) coincides with the identification of an ethylene regulator, CTR-1 linked to FLP1C6 which is associated with ethyl hexanoate production (Figure 2.13c, page 105; Figure 2.14 page 111). These results raise an interesting question regarding the role of ethylene in regulating ester formation in non-climacteric fruits, which is yet to be resolved. It may be suggested that a low amount of ethylene accumulation during strawberry fruit ripening is sufficient to regulate the ethylene receptor for the activation of phospholipase A enzyme, resulting in the release of free fatty acids as precursors for methyl ester formation (Figure 3.4) (Paliyath et al., 2009). To investigate the presence of allelic variation at the FLP1D7 locus closely linked to PLA and Ers1, sequence polymorphism of FLP1D7 was again investigated through direct amplicon sequencing to determine allelic variation between the parental genotypes with distinct aroma profiles (Section 3.3.4).
Figure 3.4 Phospholipid catabolic pathway and its relationship to ester volatile biosynthesis. The genes (ethylene receptor and phospholipase) closely linked to FLP1D7 are highlighted in red whereas the ester volatiles are highlighted in green. Adapted and modified from Paliyath et al. (2009).
In addition, FLP3E12 was also positively correlated to linalool production based on the differences of hybridisation intensities between the ‘H’ (884.13) and the ‘L’ (226.17) bulks (Table 3.7). The putative cytochrome P450 superfamily protein (CYP) closely linked to FLP3E12 has been shown to play an important role in the biosynthesis of terpenoids as well as contributing to their structural diversity (Weitzel and Simonsen, 2013). It has been reported to catalyse the hydrolysis of α-pinene to myrtenol in wild strawberry (Aharoni et al., 2004) and the conversion of linalool to its hydroxylated or epoxidised products in Arabidopsis (Ginglinger et al., 2013). According to published literature, it is clear that nerolidol synthase 1 (NES1) is the enzyme that catalyses the formation of linalool from geranyl diphosphate (GPP) through the methylerthritol (MEP) pathway (Figure 3.5) (Aharoni et al., 2003; Aharoni et al., 2004; Nagegowda et al., 2008). Therefore, the CYP identified in this study suggests that the gene may be involved in the biosynthesis of linalool from different substrates (Figure 3.5). This suggestion is supported by a previous study that demonstrated the biotransformation of linalool from nerol and geraniol by liquid cultures of fungi such as Aspergillus niger and Penicillium rugulosum (Demyttenaere et al., 2000). In addition, linalool can also be synthesised from other monoterpenes such as α- and β-pinene through chemical synthesis (Sathikge, 2008). Based on these studies, the involvement of cytochrome P450 for linalool production in plants is proposed in Figure 3.5. However, the right candidate CYP gene responsible for linalool production could not be determined in this study since there are 127 cytochrome P450-families described so far (Weitzel and Simonsen, 2013). Therefore, it is important to design a gene expression experiment together with an enzymatic assay to determine which CYP gene is involved in the biosynthesis of linalool before proceeding to examine the allelic variation of this gene between the parental genotypes and the F1 progeny with different phenotypes.
**Figure 3.5** The proposed biosynthetic pathway of linalool from monoterpenes geraniol, nerol, α- and β-pinene. MEP: Methylerythritol phosphate pathway; IPP: Isopentenyl diphosphate; DMAPP: Dimethylallyl diphosphate; GPP: Geranyl diphosphate; GPPS: Geranyl diphosphate synthase; NES1: Nerolidol synthase 1. The proposed reactions of cytochrome P450 were labelled with red dotted line.
Nonetheless, it should not be neglected that FLP3E12 also showed a perfect match to a chloroplast region (Table 3.8). This result coincides with the identification of a chloroplast rRNA gene (FLP2E1) associated with linalool production as discussed in Chapter 2 (Figure 2.13d, page 105; Table 2.6, page 109). These results imply that the chloroplast may play a role in fruit flavour development (Klee and Tieman, 2013; Powell et al., 2012). Other chloroplast-specific features include FLP1B3, FLP1D11 and FLP2D11 (Table 3.8).

The chloroplastic-like NAD(P)H-quinone oxidoreductase subunit H and ATP synthase subunit alpha corresponding to FLP1B3 and FLP1D11, respectively are required for electron transfer in chloroplast during photosynthesis (Nashilevitz et al., 2010; Peng et al., 2011). Interestingly, the importance of chloroplast on fruit quality and the genetic control of fruit flavour have been described in two recent studies. The wild type tomato has a U locus encoding a Golden 2-like transcription factor (SlGLK2) which determines chlorophyll accumulation in developing fruit but possesses a dark-green shoulder that is not desired by consumers. The modern tomato, in contrast, are bred for uniform ripening (u) but produces fewer and smaller chloroplast, hence reduced soluble solids and overall fruit quality (Klee and Tieman, 2013; Powell et al., 2012). A recent study reported that overexpression of the GLK genes increases chlorophyll content, chloroplast number and denser thylakoid grana in green tomato fruits. Upon ripening, elevated carbohydrate/sugars as well as other desirable nutritional compounds such as carotenoid and lycopene were observed with the increase of chloroplast number (Nguyen et al., 2014). Therefore, it may be speculated that the F₁ progeny used in the ‘H’ and ‘L’ DNA bulks contained either different numbers of chloroplast or have two different types of chloroplast (i.e. chloroplast with and without denser thylakoid grana), resulting in the increased levels of volatile compounds (i.e. ethyl hexanoate, ethyl butanoate, (E)-nerolidol and linalool) corresponding to the chloroplast-specific features.
Similar to the results obtained in Chapter 2 (page 111), the putative DNA markers identified in this study are associated with genes involved in precursor production rather than the rate-limiting steps for key volatile compounds formation. A few possible explanations could be deduced for this finding. Firstly, flavour phenotyping was based on the harvest from only one season and one location. The detection of volatile compounds may be affected by environmental factors (Samykanno, 2012). Secondly, a relatively small sample size (50 F1 progeny plants) was used in this study. Thirdly, the DFA method employed was more appropriate for selecting an array of DNA markers highly associated with a trait and not a specific gene of interest (Bonamico et al., 2010). It is thus suggested that flavour phenotyping should be performed on a larger sample size in three successive years at different locations to account for environmental variation (Hakala et al., 2002). Nevertheless, two interesting putative DNA markers, FLP1A7 and FLP1D7 associated with genes involved in lipid catabolism were detected in this study. DNA size and sequence polymorphism of FLP1A7 and FLP1D7 are discussed below.

3.3.4 Size and DNA sequence polymorphism

In an attempt to link the phenotypic traits (i.e the levels of specific volatile compounds) and molecular data, two promising nuclear-specific features, FLP1D7 and FLP1A7 were selected for size and DNA sequence polymorphism detection using sequence-specific primers. The hypothesis is that if these features are true DNA markers associated with methyl butanoate and γ-dodecalactone production, size variation in PCR amplicons or microstructural polymorphism in DNA sequences should be observed between plants producing high and undetectable levels of the specific volatile compounds.
Figure 3.6 The PCR products amplified from both parental genotypes and the F₁ progeny plants in two phenotypic extreme bulks using FLP1D7 (upper row) and FLP1A7 (bottom row) sequence-specific primers on a 2.0 % agarose/EtBr gel. Lane 1M: 1 kb DNA ladder; Lane 1 & 18: 07-102-41 breeding line; Lane 2 & 19: Juliette; Lane 3 – 9: F₁ progeny plants not producing methyl butanoate; Lane 10 – 16: F₁ progeny plants producing high levels of methyl butanoate. Lane 20 – 30: F₁ progeny plants not producing γ-dodecalactone; Lane 31 – 33: F₁ progeny plants producing high levels of γ-dodecalactone. Lane 17 & 34: negative controls.
Figure 3.6 displays the results of PCR amplification from both parental genotypes (i.e. 07-102-41 and Juliette) and F₁ progeny plants derived from the two DNA bulks with extreme phenotypes. Amplification with sequence-specific primers yielded amplicons with expected size at approximately 392 bp and 316 bp, respectively. Multiple faint bands could be seen above the amplicons amplified from FLP1A7 primers, indicating that the specificity of the forward and reverse primers may not be optimum. Therefore, the 316 bp amplicons were gel purified prior to DNA sequencing. Another more interesting explanation would be the PCR products contained a mixture of alleles considering the strawberry plant is an octoploid. No size variation could be determined from the agarose gel, indicating that the presence of microstructural changes is not prominent enough to show a variation in amplicon size between the plants differing in the levels of corresponding volatile compounds. Therefore, the possibility of the presence of large indels in the DNA sequence is highly unlikely. Yet, the occurrence of SNPs could not be discounted.

DNA sequence analysis obtained from FLP1D7 amplification revealed that a putative C/T SNP was detected between 07-102-41 and Juliette, and that the DNA fragment printed on the SDA was derived from Juliette compared to the original DNA sequence cloned into pGEM®-T Easy vector (Figure 3.7). This result is in accordance with the SDA data, where the SNR of Juliette (i.e., 513.10) is nearly 1.4 x higher than that of 07-102-41 (i.e., 386.38) and the SNR of the ‘ND’ bulk (i.e., 467.90) is almost 1.7 x higher than in the ‘H’ bulk (i.e., 268.67). This is because the DNA sequences with a “T” at the SNP site would bind loosely to the SDA compared to DNA sequences carrying a “C”, generating an altered hybridisation pattern (Lamartine, 2006). In contrast, difficulty in interpreting the DNA sequence derived from FLP1A7 was encountered due to the overlapping of multiple peaks.
(chromatogram not shown), indicating the presence of secondary products even after gel purification.

Based on the aroma profiles of 07-102-41 and Juliette shown in Table 3.1, higher levels of methyl butanoate were detected in 07-102-41 (15.4 %) compared to Juliette (0.0 %). This result suggests that the FLP1D7 marker allele harbouring either a “T” or “C” nucleotide at the SNP site may be associated with high or low levels of methyl butanoate production, respectively. Therefore, it is important to design a set of diagnostic PCR-based primers harbouring the putative C/T SNP region to test the association of the putative C/T SNP with the production of methyl butanoate in the future. The method described by Tabone et al. (2009) would be a good starting point for this work, where the phenotype of strawberry plants could be predicted from the presence/absence of banding pattern or the number of bands/alleles seen on an agarose gel. Allele-specific PCR can be performed on the individuals comprised of the two DNA bulks with extreme phenotypes and other breeding materials to confirm the association between the FLP1D7 putative marker and methyl butanoate production.
Figure 3.7 DNA sequence polymorphism between two parental genotypes. (a) DNA sequence alignment of PCR products amplified from 07-102-41 and Juliette using FLP1D7-specific primers. The original FLP1D7 derived from pGEM®-T Easy vector was used as a reference sequence. (b) Chromatograms showing the putative C/T SNP between 07-102-41 and Juliette. The C/T SNP is highlighted in yellow.
3.4 CONCLUSIONS

Flavour phenotyping revealed a diverse aroma profile of the F₁ population derived from 07-102-41 x Juliette cross. Frequency distribution of the selected key volatile compounds were highly or moderately skewed towards lower values or zero, suggesting that volatile compound production in strawberry fruit is controlled by more than one gene. Transgressive segregation of the key volatile compounds provides an opportunity to identify DNA loci associated with the traits using BSA. Furthermore, individuals showing extreme phenotype (i.e. higher concentrations of favourable volatile compounds) can be used as breeding materials for cross hybridisation.

This study also demonstrated the possibility of using DFA as a new tool for DNA marker selection by generating a discriminative model to classify the high and low levels of a particular volatile compound based on the SDA hybridisation patterns between the two DNA bulks showing extreme phenotypes. The models generated could also be used to allocate unknown strawberry genotypes into groups with distinct performance. The results of DFA, Fisher’s ratio and Independent Samples t-Test revealed that three of the SDA features including FLP1A7, FLP1D7 and FLP3E12 may be associated with the production of γ-dodecalactone, methyl butanoate and linalool, respectively.

In addition, DNA sequencing of the FLP1D7 feature showed a putative C/T SNP between 07-102-41 and Juliette which differed in methyl butanoate content. The results suggested that allelic variants of FLP1D7 are possibly associated with different levels of methyl butanoate production. This putative DNA marker should be turned into diagnostic PCR-based markers and validated on the F₁ population and a larger germplasm including other strawberry cultivars and advanced breeding lines to confirm the association between
FLP1D7 and the biosynthesis of methyl butanoate. In conclusion, the application of SDA-BSA approach could be a good strategy for rapid identification of informative markers. However, further modifications are needed to fine-tune the SDA-BSA approach in order to identify molecular markers closely linked to genes involved in the rate-limiting steps of the biosynthesis of key volatile compounds.
CHAPTER 4
Identification of DNA Markers Associated with Day-neutrality using Strawberry-specific SDA

4.1 INTRODUCTION

Day-neutrality is probably the most agriculturally desirable trait for strawberry (Fragaria x ananassa) breeding. Most of the commercially available cultivars are short day (SD) plants that initiate flowering when the day-length is shorter than 14 hours (Darrow, 1966) and when temperature is low (12 °C – 18 °C) (Hartmann, 1947; Heidi, 1977; Mouhu et al., 2009). On the contrary, the day-neutral (DN) genotypes are photoperiod insensitive and may produce flowers as long as temperatures are between 4 °C to 29 °C (Durner, 1984; Hancock, 1999; Ruan et al., 2013). Recently, the focus of strawberry breeding has shifted to developing DN cultivars as this trait allows flower formation during the long days in spring thereby assuring continued fruit production throughout the summer and autumn (Shaw and Famula, 2005). Consequently, the majority of strawberry production in California is now dominated by DN cultivars that are well adapted to the Mediterranean climate with cooler summer temperatures (Stewart, 2011). However, the source of day-neutrality used in California has poor adaptability to Continental climate where the midsummer temperatures usually exceed 28 °C, thus limiting use in the Eastern United States and Canada (Weebadde et al., 2008).

To overcome this, researchers have recently tried to identify molecular markers associated with flowering habit. For instance, three inter-simple sequence repeat (ISSR) markers were found to be associated with the SEASONAL FLOWERING LOCUS (SFL) in F. vesca using Bulked Segregant Analysis (BSA). These markers were successfully converted into locus
specific SCAR markers (SCAR1 – SCAR3) and mapped using the *F. vesca* spp. *vesca x F. vesca* spp. *semperflorens* testcross population. SCAR1 and SCAR3 were mapped at 3.0 cM and 1.7 cM from the *SFL* gene, respectively while SCAR2 was located on the *SFL* itself (Albani et al., 2004). Further, in the RAPD genotyping of the F1 progeny of a cross between the Japanese everbearer ‘Ever Berry’ and June bearer ‘Toyonoka’, two RAPD markers (OPE07-1 and OPB05-1) were mapped at 11.8 cM and 15.8 cM on either side of the everbearing gene (Sugimoto et al., 2005). However, the practical application of these markers was limited in strawberry breeding either due to the differences of day-neutrality between the wild diploid and commercial octoploid varieties (Stewart and Folta, 2010) or lack of reproducibility of the RAPD markers (Paran and Michelmore, 1993). To date, only one QTL study on day-neutrality has been reported where linkage mapping was performed using Amplified Fragment Length Polymorphism (AFLP) markers on a cross between a DN ‘Tribute’ and SD ‘Honeoye’. A sufficient number of QTL with modest effects were uncovered, indicating that day-neutrality is likely under polygenic control. The authors also suspected that a major dominant gene underpinning the trait could possibly have been left out due to the relatively diffusive map (Weebadde et al., 2008).

In contrast to the ambiguity of genetic control of day-neutrality in strawberry, the photoperiod flowering pathway is well-defined in *Arabidopsis* model system (Amasino and Michaels, 2010; Blázquez, 2000; Mouradov et al., 2002). Although a subset of the strawberry ortholog to photoperiodic flowering pathway genes such as *PHYTOCHROME A* (*PHYA*), *CONSTANS* (*CO*), *SUPPRESSOR OF OVEREXPRESSION OF CONSTANS 1* (*SOC1*) and *LEAFY* (*LFY*) exists in *F. vesca* (Davis et al., 2010; Mouhu et al., 2009), the specific genes controlling the critical switch from vegetative to reproductive growth
remain to be discovered. It has been suggested that these candidates may provide a better understanding of flowering response in *Fragaria* spp. (Chandler *et al*., 2012). Further, despite the large influence of environmental stimuli (e.g. day-length and temperature), it is believed that genotype effects on photoperiod flowering is substantial (Stewart, 2011).

This chapter describes the application of Bulked Segregant Analysis (BSA) and strawberry-specific Subtracted Diversity Array (SDA) for rapid identification of DNA markers associated with day-neutrality in octoploid strawberry. It was hypothesised that the features showing distinct difference in signal intensity between the DN and SD bulks will contain genetic information related to day-neutrality. The sequence identity of the putative DNA markers was also determined to evaluate the effectiveness of the SDA-BSA approach for uncovering genetic determinants underpinning day-neutrality in strawberry.

The objectives of this study were to:

1. Perform Bulked Segregant Analysis (BSA) based on the different flowering response of F1 progeny derived from three segregating populations.
2. Screen for polymorphic features between the DN and SD bulks using the strawberry-specific SDA.
3. Determine DNA sequence polymorphisms of the putative DNA markers and identify genes closely linked to the markers which may be associated with day-neutrality in strawberry.
4.2 MATERIALS AND METHODS

4.2.1 Plant genotypes and segregating population

To perform Bulked Segregant Analysis (BSA), plant materials were obtained from three segregating populations: (1) DN ‘01-061-311’ x SD ‘Juliette’, (2) DN ‘01-061-311’ x DN ‘05-069-63’ and (3) DN ‘01-061-311’ x DN ‘05-069-194’ grown at the Strawberry Breeding Station, Wandin North, Victoria on 5th and 10th of January 2012. As mentioned in Chapter 2 and Chapter 3, Juliette is a short-day Australian cultivar highly preferred by consumers and has been used as a parent in many crosses in the Australian strawberry breeding program. In contrast, 01-061-311, 05-069-63 and 05-069-194 are three strong day-neutral breeding lines bred by the Victorian strawberry breeding program (Brevis, 2013). A total of 22 F1 progeny plants were randomly selected from the 01-061-311 x Juliette cross, in which 10 plants exhibited SD phenotype and 12 plants displayed different strengths of DN phenotype (see Section 4.2.2 below). Only plants exhibiting day-neutral phenotypes were collected from the two DN x DN crosses to increase the chances of day-neutrality allele detection, where 16 and 11 day-neutral F1 progeny plants were sampled from 01-061-311 x 05-069-63 and 01-061-311 x 05-069-194 crosses, respectively. Leaf samples were harvested from both parents and F1 progeny of these crosses and stored according to the method described in Section 2.2.1 (page 53).

4.2.2 Scoring for flowering response

The individual F1 progeny plants were scored for the presence of flowers in midsummer (05/01/2012 and 10/01/2012) according to the subjective scoring method described by Shaw (2003). A few modifications were applied to the flowering score as an adaptation to the flowering response observed in this study. Plants were assigned into four classes based on a scale from 1 to 4, with 1 representing the strongest day-neutral strength and
4 corresponding to a complete SD phenotype (Table 4.1). Each plant was also photographed for future reference, and to facilitate DNA bulking at a later stage.

**Table 4.1** Flowering response on a 1-4 scale based on their day-neutrality strength (Shaw 2003).

<table>
<thead>
<tr>
<th>Scale</th>
<th>Description of flowering response</th>
<th>Strength of day-neutrality</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Flower formation on runners</td>
<td>Strong day-neutral</td>
</tr>
<tr>
<td>2</td>
<td>2 or more recently emerged inflorescences</td>
<td>Intermediate day-neutral</td>
</tr>
<tr>
<td>3</td>
<td>Less than 2 recently emerged inflorescences</td>
<td>Weak day-neutral</td>
</tr>
<tr>
<td>4</td>
<td>No flowers or fruits</td>
<td>Short-day</td>
</tr>
</tbody>
</table>

### 4.2.3 Bulked Segregant Analysis (BSA)

#### 4.2.3.1 Generation of the phenotypic extreme DNA bulks

Genomic DNA was isolated from the frozen leaf tissues of individual F1 progeny plants using Qiagen™ DNeasy® Plant Mini kit (Qiagen, Valencia, CA) following the manufacturer’s instructions. Equal quantities of DNA from SD and DN individuals were pooled into four different bulks defined as ‘strong day-neutral (DN1)’, ‘intermediate day-neutral (DN2)’, ‘weak day-neutral (DN3)’ and ‘short-day (SD)’ bulks to a final amount of 2 µg. The number of individuals in each bulk ranged from 2 to 19 plants depending on the flowering response class.

#### 4.2.3.2 SDA hybridisation, scanning and image quantification

Four DNA bulks corresponding to DN1, DN2, DN3 and SD phenotypes were digested with *AluI* and *HaeIII* and purified using QIAquick PCR Purification Kit (Qiagen, Valencia, CA). The digested DNA was individually hybridized onto the strawberry-specific SDA. Biotin-labelling of the target samples, SDA hybridisation, image scanning and
quantification were performed according to the protocols described in Section 2.2.3 (page 61). The resulting SDA data was analysed using the spike-in control normalisation method explained in Section 3.2.3.2 (page 124). The normalised mean signal-to-noise ratio (SNR) for all the 287 features were used for subsequent statistical analyses in order to determine the DNA sequences associated with day-neutrality in strawberry.

4.2.4 Statistical analyses

Three types of statistical analyses were conducted as described in Section 3.2.4 (page 125). Discriminant Function Analysis (DFA) was performed by comparing the SDA data between (1) SD and DN1, (2) SD and DN2 and (3) SD and DN3 to identify a set of variables that best discriminate between the two phenotypic extreme DNA bulks. The selected predictors (variables) were also tested for their ability to classify new cases into the correct flowering response class. The data set was analysed using IBM SPSS Statistics v. 21 according to the method described in Section 3.2.4.1 (page 125). Based on the accuracy of group prediction, a set of features that contribute maximally to group separation were determined as the putative DNA markers capable of predicting if a strawberry plant exhibits day-neutral or short day phenotype.

The number of putative DNA markers was further narrowed by analysing the SDA dataset with Fisher’s ratio (Microsoft Excel) and Independent Samples t-Test (IBM SPSS Statistics v. 21) as described in Section 3.2.4.2 and 3.2.4.3 (page 128). A Venn diagram was generated based on the results of DFA, Fisher’s ratio and Independent Samples t-Test. The clones that fulfilled all three criteria were selected for sequencing to determine their sequence identity.
4.2.5 DNA sequencing and sequence analysis

Sample preparation for DNA sequencing was performed as described in Section 2.2.7 (page 72). PCR products were sequenced by Macrogen Inc. (Korea) using T7 and Sp6 primers. The resulting sequences were examined for sequence quality by visually inspecting the electropherogram and locating the sequences of nested primer 1 and nested primer 2R at both ends (Appendix 10). The trimmed DNA sequences were identified using the PFR Strawberry Server at https://strawberry.plantandfood.co.nz/ (PFR, 2010) and further confirmed through the NCBI nucleic acid and protein databases. DNA sequences showing E-value < 1e-5 were considered significant. Annotated genes situated 5 cM upstream and downstream of the putative DNA markers were identified according to Section 2.2.7 (page 72).

4.2.6 Detection of DNA microstructural variation in the sequences of nuclear-specific features

Forward and reverse primers were designed for each nuclear-specific putative DNA markers using Clone Manager Suite v. 7.1 (Sci-Ed Software, Durham, NC). Amplicon size variation was determined according to Section 3.2.6.1 (page 129). Subsequently, PCR amplification was performed with sequence-specific primers and proofreading AccuPrime™ Pfx DNA Polymerase (Invitrogen, NY, USA) using all the parental genotypes (i.e. Juliette, 01-061-311, 05-069-63 and 05-069-194) as DNA templates. PCR reaction and cycling parameters were set as described (Section 3.2.6.1, page 129).

Prior to DNA sequencing, PCR products were cloned using pGEM®-T Easy Vector Systems (Promega, San Luis Obispo, CA) according to the manufacturer’s protocol, resulting in four sets of transformations, one for each of the four parental genotypes. Three
white colonies were randomly selected from each transformation. Colony PCRs were performed as described in Section 2.2.2.3 (page 59). Plasmids were recovered from the *E. coli* and amplified using two sets of primers, i.e. T7 and Sp6 sequencing primers and sequence-specific primers. After a quality check by agarose gel electrophoresis, only PCR products amplified with T7 and Sp6 primers were purified using QIAquick PCR Purification Kit (Qiagen, Valencia, CA). Subsequently, the purified PCR products were sequenced by the Australian Genome Research Facility Ltd. (AGRF) using T7 and Sp6 primers and proofreading AccuPrime™ Pfx DNA Polymerase (Invitrogen, NY, USA). Sequence editing was completed by trimming the pGEM®-T Easy backbone vector beyond the sequence-specific primer binding sites. Sequence alignment and microstructural variation were performed according to Section 3.2.6.2 (page 131).
4.3 RESULTS AND DISCUSSION

4.3.1 Day-neutrality assessment in field environment

The flowering response of F\textsubscript{1} progeny plants was assessed in the field on the 5\textsuperscript{th} and 10\textsuperscript{th} of January 2012. In Victoria, the short-day (SD) cultivars usually flower in early October and continue until the end of December. They switch to vegetative growth when the day length becomes longer in summer. Therefore, any plants flowering after 1\textsuperscript{st} January were considered to be day-neutral (DN). As some developed flowers may have probably been initiated by SD conditions, only the newly emerged inflorescences were scored to avoid misclassification. Examples of DN and SD F\textsubscript{1} genotypes scored according to flowering strength are presented in Figure 4.1.

![Example of flowering response based on a 1-4 scale. 1: strong day-neutral; 2: intermediate day-neutral; 3: weak day-neutral; 4: short-day.](image)

**Figure 4.1** An example of flowering response based on a 1-4 scale. 1: strong day-neutral; 2: intermediate day-neutral; 3: weak day-neutral; 4: short-day.
Previous studies have demonstrated that individuals which produced flowers on runners during the same sampling season were considered as strong day-neutral genotypes (Serçe and Hancock, 2003). Individuals with varying flowering strength (i.e. the number of inflorescences) were observed and assigned as DN1, DN2 and DN3 (Table 4.2), indicating flowering response may possibly be a polygenic trait controlled by more than one gene (Serçe and Hancock, 2005a; Shaw, 2003; Weebadde et al., 2008). In contrast, individuals that produced a lot of runners without any flowers were classified as SD plants (Table 4.2). The characteristic of runner production in SD plants permits easy classification, as these genotypes usually only produce runners under long day and high temperature conditions (Bradford et al., 2010; Durner, 1984).

Table 4.2 The number of F1 progeny collected for each DN x SD and DN x DN crosses based on flowering strength.

<table>
<thead>
<tr>
<th>Crosses</th>
<th>DN1</th>
<th>DN2</th>
<th>DN3</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>DN x SD</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>01-061-311 x Juliette</td>
<td>0</td>
<td>6</td>
<td>6</td>
<td>10</td>
</tr>
<tr>
<td>DN x DN</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>01-061-311 x 05-069-63</td>
<td>1</td>
<td>8</td>
<td>7</td>
<td>0</td>
</tr>
<tr>
<td>01-069-311 x 05-069-194</td>
<td>1</td>
<td>5</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>2</td>
<td>19</td>
<td>18</td>
<td>10</td>
</tr>
</tbody>
</table>

DN1: Strong day-neutral
DN2: Intermediate day-neutral
DN3: Weak day-neutral
SD: Short-day

Despite the small subset of plants sampled, two strong DN F1 progeny plants were observed from the DN x DN crosses. In contrast, no F1 progeny showed extreme phenotype for strong day-neutrality (i.e. flower formation on runners) in the DN x SD cross (Table 4.2). The results showed that DN x DN hybridisation could yield more DN
genotypes with greater flowering strength compared to DN x SD cross, indicating that the sources of day-neutrality in 05-069-63 and 05-069-194 may be stronger than in 01-061-311, as they are more potent in producing DN progeny (Serçe and Hancock, 2005a). Similar results were previously reported, where 71 % of the progeny of a combined DN x DN cross were DN with greater flowering strength compared to only 58 % of DN progeny obtained from a DN x SD cross (Serçe and Hancock, 2005a). Nevertheless, the DN x SD crosses seem more appropriate to Australian strawberry breeders since the current breeding aim is to integrate the flavour attributes of SD genotypes into the DN strawberries. Therefore, the proportion of DN progeny in such a cross may be increased by using more potent DN genotype as one of the parents when making the crosses in order to improve the chances of detecting alleles segregating for day-neutrality.

4.3.2 Statistical Analysis

4.3.2.1 Discriminant Function Analysis (DFA)

Bulked Segregant Analysis (BSA) was performed to identify DNA markers associated with day-neutrality in the octoploid strawberry. Instead of pooling all the DN genotypes into one bulk regardless of their flowering strength, the DNA of F_1 progeny plants were pooled at equal concentration into four classes, DN1, DN2, DN3 and SD (Table 4.2). This was to avoid homogenisation of loci controlling day-neutrality due to gene dosage effect (Powers, 1954; Serçe & Hancock, 2005a). Based on the expression of extreme phenotypes, it was expected that the probability of finding potential DNA markers will be higher when comparing the signal intensities of the SD bulk with the DN1 bulk than the SD-DN2 and SD-DN3 comparisons.
The normalised mean signal-to-noise ratio (SNR) of each DNA bulk was used as input data for Discriminant Function Analysis (DFA). Six technical replicates from one of the biological replicates of both SD and DN bulks were used as training set (original cases) to develop a discriminant function based on stepwise method. A single discriminant function accounting for 100% of the variation was generated for all the three day-neutral classes assessed (Table 4.3). This result indicated that the putative DNA markers selected based on their canonical discriminant function coefficients could efficiently distinguish the group membership of the original cases (Table 4.3). Wilks’ lambda was statistically significant for the discriminant function generated for DN1 ($\chi^2 = 62.54$, $p < 0.001$), DN2 ($\chi^2 = 51.83$, $p < 0.001$) and DN3 ($\chi^2 = 51.48$, $p < 0.001$).

A range of putative DNA markers that could maximally separate the SD bulk from the three DN bulks, varying by the degree of flowering strength, were identified using the stepwise method (Table 4.4). The classification results based on the discriminant functions revealed 100% correct grouping for the original cases in the training set for all the day-neutral classes (Table 4.4). Of the 18 DFA-selected putative DNA markers, only FLP2E9 was common between weak DN and intermediate DN flowering response. This result indicated that some of the individuals assigned into the weak DN plants may have been misclassified as intermediate DN plants, or vice versa, during field assessment. The classification results further support this explanation, where only 66.7% of the new cases in the test set were correctly classified into either DN3 or SD (Table 4.4), suggesting that the SDA features (FLP2E2, FLP3H2, FLP3E5, FLP1E10, FLP3E7 and FLP2E9) selected by the DFA may not be the best predictor variables discriminating between the SD and DN2 flowering classes.
In contrast, the rate of correct classification for both strong DN and weak DN flowering response was higher compared to intermediate DN, where 83.3% of the new cases in the test sets were correctly predicted as strong DN, weak DN or SD (Table 4.4). These results also implied that all the six features selected by the DFA to build the predictive model for the SD-DN1 and SD-DN3 comparisons were good predictor markers. One possible explanation for a higher rate of correct classification for SD-DN1 and SD-DN3 comparisons was that the strong DN and weak DN plants could be easily identified in the field based on their phenotypes. However, it is more difficult to classify plants showing intermediate DN flowering response. The major sources of error in classifying the photo-insensitive and photo-sensitive genotypes have been clearly reported by Ahmadi et al. (1990). For instance, some of the late-fruiting DN genotypes that do not need much chilling may be classified as SD when scored in January (midsummer). Further, SD genotypes with minimum chilling requirements may also initiate flower bud formation in summer with a mild climate and be misclassified as DN (Ahmadi et al., 1990).
Table 4.3 Discriminant functions generated for each flowering response group based on the canonical discriminant function coefficients of the most predictive features.

<table>
<thead>
<tr>
<th>Flowering response</th>
<th>Canonical Discriminant Function Coefficients</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strong DN (DN1)</td>
<td>$Y = -31.350 - 0.224(FLP2E11) - 0.096(FLP2G4) + 0.207(FLP2D6) + 0.093(FLP2D12) - 0.380(FLP2B4) + 0.942(FLP3F4)$</td>
</tr>
<tr>
<td>Intermediate DN (DN2)</td>
<td>$Y = -37.449 + 0.333(FLP2E2) - 0.170(FLP2E9) + 0.080(FLP1E10) - 0.153(FLP3H2) + 0.457(FLP3E5) + 0.045(FLP3E7)$</td>
</tr>
<tr>
<td>Weak DN (DN3)</td>
<td>$Y = 5.291 + 0.158(FLP1A1) - 0.172(FLP2E9) + 0.308(FLP2A2) - 0.125(FLP3B3) - 0.741(FLP3B10) - 0.749(FLP3D4)$</td>
</tr>
</tbody>
</table>

Table 4.4 Classification results for the training set and test set based on the DFA-selected features for each flowering response group.

<table>
<thead>
<tr>
<th>Flowering response</th>
<th>DFA-selected markers*</th>
<th>Classification Results (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Training set</td>
</tr>
<tr>
<td>Strong DN (DN1)</td>
<td>FLP2E11, FLP3F4, FLP2D6, FLP2B4, FLP2D12, FLP2G4</td>
<td>100.0</td>
</tr>
<tr>
<td>Intermediate DN (DN2)</td>
<td>FLP2E2, FLP3H2, FLP3E5, FLP1E10, FLP3E7, <strong>FLP2E9</strong></td>
<td>100.0</td>
</tr>
<tr>
<td>Weak DN (DN3)</td>
<td>FLP1A1, FLP2A2, FLP3D4, FLP3B3, FLP3B10, <strong>FLP2E9</strong></td>
<td>100.0</td>
</tr>
</tbody>
</table>

* Features in bold represent putative DNA markers that were detected in more than one flowering response classes.
Several other authors also reported the effects of temperature, photoperiod, or a combination of both on strawberry flowering (Bradford et al., 2010; Ito and Saito, 1962; Serçe and Hancock, 2005b). Apart from these environmental variations, plant age also played a role in responding to photoperiod. It has been shown that under favourable flowering conditions (i.e. both day-length and temperature are optimal), both young and old plants will induce flowers at the same rate. On the contrary, only older or larger plants will form flower buds under unfavourable conditions (Ito and Saito, 1962; Stewart and Folta, 2010). All of these internal and external factors inevitably added to difficulties in the classification of flowering habits for strawberry plants. An earlier study reported a precautionous measurement taken where the DN plants were re-evaluated about one month later to reduce misclassification of late-flowering short day genotypes (Shaw, 2003). Therefore, similar confirmation should be performed in the future and reclassify the plants as SD genotypes if the absence of continued flowering is observed.

To further narrow the number of putative DNA markers, a second DFA was performed on the selected features listed in Table 4.4. Similar to the first DFA, a single discriminant function, which accounted for 100 % of the variation, was generated for all the flowering response classes (Table 4.5). A total of six features (FLP2E11, FLP2D12, FLP2E2, FLP1A1 and FLP3B3) were selected for all the flowering response classes based on their canonical discriminant function coefficients (Table 4.5). Wilks’ lambda was statistically significant for the discriminant function generated for DN1 ($\chi^2 = 35.27$, $p < 0.001$) and DN3 ($\chi^2 = 21.72$, $p < 0.001$). However, the Wilks’ lambda also indicated that the discriminant function generated for DN2 was less significant ($\chi^2 = 6.70$, $p < 0.05$). This outcome is expected since the classification result for the test set of DN2 based on the first discriminant function is less satisfying (Table 4.4). Therefore, the results indicate that the
DFA-selected features for DN1 (FLP2E11 and FLP2D12) and DN3 (FLP1A1 and FLP3B3) flowering response may be useful DNA markers to discriminate between the SD and DN strawberry genotypes.

**Table 4.5** Discriminant functions generated for each flowering response group using the features selected by the first DFA.

<table>
<thead>
<tr>
<th>Flowering response</th>
<th>Canonical Discriminant Function Coefficients</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strong DN (DN1)</td>
<td>( Y = -0.500 + 0.005(\text{FLP2E11}) - 0.007(\text{FLP2D12}) )</td>
</tr>
<tr>
<td>Intermediate DN (DN2)</td>
<td>( Y = -1.255 + 0.008(\text{FLP2E2}) )</td>
</tr>
<tr>
<td>Weak DN (DN3)</td>
<td>( Y = -0.224 + 0.004(\text{FLP1A1}) - 0.005(\text{FLP3B3}) )</td>
</tr>
</tbody>
</table>

**Table 4.6** The reduced set of putative DNA markers selected by DFA, the classification results for the original cases and the estimated proportion for any future dataset.

<table>
<thead>
<tr>
<th>Flowering response</th>
<th>DFA-selected markers</th>
<th>Classification Results (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Original</td>
</tr>
<tr>
<td>Strong DN (DN1)</td>
<td>FLP2E11, FLP2D12</td>
<td>95.8</td>
</tr>
<tr>
<td>Intermediate DN (DN2)</td>
<td>FLP2E2</td>
<td>70.8</td>
</tr>
<tr>
<td>Weak DN (DN3)</td>
<td>FLP1A1, FLP3B3</td>
<td>91.7</td>
</tr>
</tbody>
</table>

Unlike the first DFA, no common features were found for all the three flowering response classes and the numbers of putative DNA markers were reduced to one or two per class (Table 4.6). The classification results based on the newly selected features revealed that 70.8% - 91.7% of the original and new cases (obtained from cross-validation) were
correctly classified. Although the classification results for the original cases showed a slight increase in error rate, the misclassification rate for new cases decreased (Table 4.6) compared to the first DFA (Table 4.4). In contrast to previous studies that report an increased misclassification rate as the number of markers decrease (Alwala, 2007; Siosemarde et al., 2012; Zhang et al., 2005), the present results showed that a minimum number of two markers were efficient in discriminating the SD plants from weak DN and strong DN plants. This may however imply that not all the features selected through the first DFA are useful in predicting group membership of any given cases. As discussed in section 3.3.2.1, the application of a second DFA may help to reduce redundant markers and form classification models (Siosemarde et al., 2012).

4.3.2.2 Fisher’s ratio

In addition to the DFA, Fisher’s ratio was performed as an additional statistical tool to evaluate the linear discriminating power of the putative DNA markers (Lohninger, 1999). The top ten features with the highest Fisher’s ratio values are listed in Table 4.7. Of all the DFA-selected features, FLP1A1 and FLP2E11 ranked first and third, respectively (Table 4.7). This indicated a significant difference in their group means. The other features selected by DFA such as FLP3B3, FLP2E2 and FLP2D12 yielded lower Fisher’s ratios, suggesting that these features either have minimal differences between group means (i.e. SDA features with low signal intensities) or larger sum of the variances of two groups (i.e. greater variation between technical replicates).
Table 4.7 List of top 10 features ranked in decreasing order based on the respective Fisher’s ratio for flowering response groups.

<table>
<thead>
<tr>
<th>Ranking</th>
<th>Fisher’s ratio</th>
<th>Strong DN (DN1)</th>
<th>Intermediate DN (DN2)</th>
<th>Weak DN (DN3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>FLP1G3</td>
<td>FLP1A1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>FLP2G5</td>
<td>FLP4A8</td>
<td></td>
<td>FLP2G12</td>
</tr>
<tr>
<td>3</td>
<td><strong>FLP2E11</strong></td>
<td>FLP2D3</td>
<td></td>
<td>FLP3F7</td>
</tr>
<tr>
<td>4</td>
<td>FLP1B10</td>
<td>FLP4D1</td>
<td></td>
<td>FLP2G11</td>
</tr>
<tr>
<td>5</td>
<td>FLP3F7</td>
<td>FLP2F11</td>
<td></td>
<td>FLP3A5</td>
</tr>
<tr>
<td>6</td>
<td>FLP1B9</td>
<td>FLP2G5</td>
<td></td>
<td>FLP4B1</td>
</tr>
<tr>
<td>7</td>
<td>FLP2E4</td>
<td>FLP1G7</td>
<td></td>
<td>FLP2D3</td>
</tr>
<tr>
<td>8</td>
<td>FLP1G1</td>
<td>FLP3B1</td>
<td></td>
<td>FLP1C8</td>
</tr>
<tr>
<td>9</td>
<td>FLP1C4</td>
<td>FLP2C11</td>
<td></td>
<td>FLP2F5</td>
</tr>
<tr>
<td>10</td>
<td>FLP1H8</td>
<td>FLP3F7</td>
<td></td>
<td>FLP3G6</td>
</tr>
</tbody>
</table>

Cells shaded in yellow represent putative DNA markers which were also selected by DFA.

4.3.2.3 Independent Samples $t$-Test

To further confirm the differences between the group means for the DFA-selected features, an Independent Samples $t$-Test was employed. As described in Section 3.3.2.3 (page 153), a $t$-Test for unequal variances was used when the variances for two phenotypic extreme groups were significantly unequal ($p < 0.05$). The results revealed that the mean differences for FLP1A1, FLP2E11 and FLP2D12 were highly significant ($p < 0.01$) while FLP2E2 was significant ($p < 0.05$) (Table 4.8). It is interesting to note that a contrasting normalised mean SNR was obtained for FLP1A1 (i.e. strong hybridisation with SD bulk but no hybridisation with DN3 bulk), suggesting that FLP1A1 may be a DNA fragment specific to SD plants. In contrast, the group mean for FLP3B3 was not significantly different (Table 4.8). The results indicate that the application of DFA alone may not be sufficient to exclude unnecessary features and additional statistical measurements such as Fisher’s ratio and Independent Samples $t$-Test are required to supplement DFA data.
Table 4.8 Group statistics and Independent Samples t-Test for the putative DNA markers selected by DFA.

<table>
<thead>
<tr>
<th>Flowering response group</th>
<th>DFA-selected markers</th>
<th>Day-neutral</th>
<th>Short-day</th>
<th>t-test (for two bulks)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Mean&lt;sup&gt;b&lt;/sup&gt;</td>
<td>s</td>
<td>Mean&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>Strong DN (DN1)</td>
<td>FLP2E11&lt;sup&gt;a&lt;/sup&gt;</td>
<td>743.53</td>
<td>252.08</td>
<td>259.98</td>
</tr>
<tr>
<td></td>
<td>FLP2D12</td>
<td>223.23</td>
<td>146.47</td>
<td>433.96</td>
</tr>
<tr>
<td>Intermediate DN (DN2)</td>
<td>FLP2E2</td>
<td>85.71</td>
<td>48.94</td>
<td>232.53</td>
</tr>
<tr>
<td>Weak DN (DN3)</td>
<td>FLP1A1&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.27</td>
<td>0.35</td>
<td>713.71</td>
</tr>
<tr>
<td></td>
<td>FLP3B3</td>
<td>227.30</td>
<td>136.36</td>
<td>300.66</td>
</tr>
</tbody>
</table>

<sup>a</sup> Putative DNA markers which displayed high Fisher’s ratio values

<sup>b</sup> Mean signal-to-noise ratio (SNR) of two biological replicates and six technical replicates

s: Standard deviation

** Significant at <i>p < 0.01</i>

* Significant at <i>p < 0.05</i>

ns: Not significant (<i>p > 0.05</i>)
Consequently, the results of the DFA, Fisher’s ratio (top 10 features) and Independent Samples t-Test ($p < 0.01$) were combined by generating a three-way Venn diagram to select a final set of putative DNA markers (Figure 4.2). Only two features (FLP1A1 and FLP2E11) were identified in the intersection of the three statistical analyses, in which FLP1A1 was putatively associated with SD while FLP2E11 was putatively correlated to strong DN. The intersection between DFA and Independent Samples t-Test revealed two more features (FLP2D12 and FLP2E2) which differed significantly in their group means but were not included in the top 10 features with the highest Fisher’s ratios (Figure 4.2). Their low Fisher’s ratio values may be attributed to a smaller magnitude of differences in group means and larger sum of the variances of two groups, as previously mentioned. FLP3B3 is the only feature selected by DFA that did not fulfil the other two statistical analyses, indicating that it may not be a useful marker for genotyping DN plants. The other 23 features that possessed a high Fisher’s ratio did not overlap with any other criteria, suggesting that their group means were not statistically different. Therefore, these features were excluded from further analyses.

The application of the strawberry-specific SDA successfully identified two potential DNA markers associated with flowering response in strawberry. FLP2E11 is likely associated with day-neutrality whereas FLP1A1 is possibly related to the short day flowering habit. The identities of these putative markers were revealed by DNA sequencing and a similarity search against the *F. vesca* draft genome v1.1 was performed as discussed in Section 4.3.3.
Figure 4.2 A three-way Venn diagram showing the putative DNA markers in the intersection of DFA (green), Fisher’s ratio (top 10 features; blue) and Independent Samples $t$-Test ($p < 0.01$; red) for all flowering response classes assessed.

4.3.3 Sequence identity of putative DNA markers

DNA sequencing revealed that both the putative DNA markers contained the different SSH adaptors at both ends (Appendix 11), confirming that FLP1A1 and FLP2E11 are tester-specific sequences which were amplified exponentially. Similarity search against the *F. vesca* draft genome (v1.1) revealed that the full sequence of FLP1A1 matched 100% (E-value: 0.0) to a chloroplast region (Table 4.9). Moreover, it was found to be negatively correlated with day-neutrality based on the hybridisation intensitites of SD and DN3 bulks (Table 4.8). As the chloroplast sequences are usually represented as short scaffolds in the *F. vesca* draft genome, no genes closely linked to the putative marker could be identified from the same scaffold. Therefore, this feature was not considered for further analysis.
However, characterisation of this chloroplast-specific feature may provide useful information regarding the roles of chloroplast in controlling day-neutrality. Further analysis such as allelic variant determination should be performed in the future to confirm the association between FLP1A1 and day-neutrality in strawberry.

In contrast, FLP2E11 is a nuclear-specific feature, in which 99% of the sequence displayed significant similarity (E-value: 2e^{-72}) to a DNA region on linkage group 6 (LG6:14315954..14316207, scf0513185:155144..155397) (Table 4.9). This feature was positively correlated with day-neutrality based on the hybridisation intensities of SD and DN1 bulks (Table 4.8). Subsequent data analysis revealed four genes situated close to FLP2E11, including Arabidopsis thaliana embryo defective 2765, 6-phospogluconate dehydrogenase-like protein, cytokinin oxidase/dehydrogenase 1 (CKXI) and putative calcium-transporting ATPase 11. Of the four, CKXI is the most promising gene associated with day-neutrality. It was located at approximately 0.25 Mb downstream from FLP2E11 (Figure 4.6). As shown in Table 4.8, the normalised mean SNR of FLP2E11 was 3 x higher in the strong DN bulk (i.e. 745.33) than in the SD bulk (i.e. 259.98), suggesting that FLP2E11 and its closely linked CKXI are positively correlated with day-neutrality. Cytokinin oxidase/dehydrogenase encodes an enzyme that catalyses the breakdown of cytokinin (Bartrina et al., 2011). Thus it is reasonable to speculate that the CKXI in the DN genotypes may be the allele responsible for the production of CKX1 enzyme with weaker activity, since cytokinin accumulation is needed for floral transition (D’Aloia et al., 2011).
Table 4.9 Putative identity of the most discriminatory features searched against the *Fragaria vesca* draft genome (v1.1). E-value regarded as significant if $< 1 \times 10^{-5}$.

<table>
<thead>
<tr>
<th>Clones</th>
<th>Length (bp)</th>
<th>Landmark or region</th>
<th>Sequence description</th>
<th>E-value</th>
<th>Flowering response group</th>
</tr>
</thead>
<tbody>
<tr>
<td>FLP1A1</td>
<td>597</td>
<td>scf0510722:382..982</td>
<td>Chloroplast sequence</td>
<td>0.0</td>
<td>Weak DN</td>
</tr>
<tr>
<td>FLP2E11</td>
<td>256</td>
<td>LG6:14315954..14316207, scf0513185:155144..155397</td>
<td>Genomic DNA region on linkage group 6</td>
<td>$2 \times 10^{-72}$</td>
<td>Strong DN</td>
</tr>
</tbody>
</table>
Although cytokinins are involved in many aspects of plant growth and development, their roles in promoting flowering has only been recently elucidated (Bartrina et al., 2011; Bernier, 2011; D’Aloia et al., 2011). However, this hormone has yet to be integrated into the molecular pathways of flowering as defined in Arabidopsis thaliana (Amasino and Michaels, 2010; Blázquez, 2000; Mouradov et al., 2002). Interestingly, a recent study provided an initial glimpse of the role of cytokinin in regulating these flowering-time genes by using the Arabidopsis mutants for CONSTANS (CO), FLOWERING LOCUS T (FT), TSF (the FT parologue TWIN SISTER OF FT), SUPPRESSOR OF OVEREXPRESSION OF CONSTANS1 (SOC1) and the bZIP transcription factor FD (D’Aloia et al., 2011). The study showed that tsf-1 and soc1-2 single mutants did not respond to N\(^6\)-benzylaminopurine (BAP) treatment, indicating that TSF and SOC1 are required to initiate flowering in response to BAP. In contrast, the ft-10 mutant continues to form flower buds, suggesting that endogenous cytokinin is involved in another flowering route in Arabidopsis which bypasses FT but requires its parologue TSF (D’Aloia et al., 2011). Further, a few functional studies have been performed to determine the role of
cytokinin oxidase/dehydrogenase as a negative regulator for flowering. For instance, overexpression of \textit{CKX} gene has been shown to suppress flowering in \textit{Arabidopsis} plants (Werner et al., 2003). Conversely, an insertional mutagenesis study in \textit{Arabidopsis} showed that the \textit{ckx3 ckx5} double mutant produced significantly more flowers with a larger inflorescence meristem compared to the wild type and the single mutants (Bartrina et al., 2011). Collectively, these studies indicate the importance of cytokinin and \textit{CKX} genes in regulating flowering behaviour in the model plant.

Based on the current knowledge of flowering regulation in \textit{Arabidopsis}, it is believed that accumulation of CO protein in DN strawberry genotypes increases as the day length becomes longer resulting in flowering initiation. This is related to the fact that the CO protein is stabilised by light and degraded in darkness (Valverde et al., 2004). Surprisingly, no accumulation of CO protein was observed in some of the day-neutral lines tested in a previous study (Stewart et al., 2007), suggesting that flowering in strong DN genotypes may be regulated by other flowering promoters apart from CO protein (Stewart and Folta, 2010) such as cytokinin signalling proposed in this study. Since cytokinin is required to induce flowering, it is postulated that the \textit{CKX1} allele contained in the DN strawberry genotypes may be a low activity allele, allowing accumulation of cytokinin in the leaves and shoot apical meristem (SAM) under unfavourable flowering condition (long day) and subsequently promote flowering in summer (Corbesier et al., 2003). In contrast, SD genotypes may contain the wild type \textit{CKX1} allele. As a result, cytokinin degradation occurs at a faster rate, and hence flower initiation is repressed under long day condition. To test the hypothesis of allelic variation, PCR was performed on the parental genotypes using FLP2E11-specific primers as described in Section 4.3.4.
4.3.4 Size and DNA sequence polymorphism

Size variation in PCR amplicons or DNA sequence polymorphism should be observed between the SD and DN genotypes if the putative DNA marker FLP2E11 is strongly linked to the loci controlling day-neutrality. To test for this association, PCR amplification was performed on all parental genotypes (i.e. Juliette, 01-061-311, 05-069-63, 05-069-194) and F1 progeny plants (derived from SD, DN1, DN2 and DN3 bulks) using FLP2E11 sequence-specific primers. A single band present at approximately 256 bp was obtained as expected (Figure 4.7). No distinct size variation could be observed between the tested genotypes, however amplicon intensity differences between the genotypes were observed. Since the concentration of DNA template used for all genotypes was standardised to 50 ng, the differences in stoichiometry may suggest a potential allele dosage effect (Chambers et al., 2012).

As mentioned in Section 3.3.4 (page 169), multiple overlapping peaks were noted in the chromatogram of PCR products amplified using FLP1A7 sequence-specific primers, possibly due to a mixture of alleles present in the amplicons. To overcome this problem, the PCR products amplified from the four parental genotypes were cloned into pGEM®-T Easy vectors prior to DNA sequencing as an improvement to DNA sequence analysis to ensure only a single allele was sequenced per reaction. To allow the identification of different variant forms of FLP2E11, three colonies from each transformation were randomly selected. These colonies were designated as Ju_1, Ju_2, Ju_3, 311_1, 311_2, 311_3, 63_1, 63_2, 63_3, 194_1, 194_2 and 194_3.
Figure 4.7 The PCR products amplified from both parental genotypes and the F1 progeny plants using FLP2E11 sequence-specific primers on a 2.0 % agarose/EtBr gel. Lane M: 100 bp DNA ladder; Lane 1: 01-061-311; Lane 2: Juliette; Lane 3 – 12: F1 from SD bulk; Lane 13 – 18: F1 from DN3 bulk; Lane 19 – 24: F1 from DN2 bulk; Lane 25: 05-069-63; Lane 26: 05-069-194; Lane 27 – 28: F1 from DN1 bulk; Lane 29: Negative control.

As expected, amplification with sequence-specific primers yielded amplicons at approximately 256 bp, confirming the presence of FLP2E11 DNA inserts in the isolated plasmids (Figure 4.8). In addition, amplification with T7 and Sp6 primers resulted in intact and intense bands at approximately 400 bp, indicating that the quality of the PCR products was good for downstream sequencing (Figure 4.8). PCR products amplified using T7 and Sp6 primers were purified and sequenced to ensure that the full length of FLP2E11 fragment was captured.
Figure 4.8 The PCR products amplified from all parental genotypes using FLP2E11 sequence-specific primers (~ 256 bp) and T7 and Sp6 primers (~ 400 bp) on a 2.0 % agarose/EtBr gel. Lane M1: 100 bp DNA ladder; Lane M2: 1 kb DNA ladder; Lane 1 – 6: Juliette (Ju_1, Ju_2 and Ju_3); Lane 7 – 12: 01-061-311 (311_1, 311_2 and 311_3); Lane 13 – 18: 05-069-63 (63_1, 63_2 and 63_3); Lane 19 – 24: 05-069-194 (194_1, 194_2 and 194_3); Lane 25 – 26: Negative control.

The chromatograms obtained from these samples showed comparably less background noise than those in Section 3.3.4 (page 169). DNA sequence alignment was performed for the parental genotype clones. SNPs were detected at random between the three clones for all four parental genotypes (Appendix 12), supporting the occurrence of polymorphism among subgenomes in the octoploid genome (Chandler et al., 2012; Mithani et al., 2013). However, these types of polymorphisms are usually not informative and could be filtered. Additionally, a 12 bp deletion was discovered in the third clone of Juliette when Ju_3 was aligned to Ju_1 and Ju_2 (Appendix 12). These results indicated that at least two different variants of FLP2E11 are present in Juliette, and that Juliette is a heterozygous SD genotype.
Subsequently, DNA sequence alignment was conducted again by comparing all the sequences to the original FLP2E11 sequence recovered from \textit{E. coli} cultures used to construct the SDA array (Appendix 13). The original FLP2E11 feature printed on the SDA was most likely derived from DN genotypes (either Albion or 04-069-91) used in the subtracted library construction as it displayed high sequence homology with all other DN parental genotypes (Figure 4.9). Moreover, the third clone of SD Juliette contained a fragment with a 12 bp deletion (from nucleotide 82 to 93) which was present in all other clones sequenced. This may indicate that allelic variation is present in the FLP2E11 locus, and that the longer FLP2E11 marker allele may be linked to the low activity \textit{CKX1} allele whereas the shorter FLP2E11 marker allele may be linked to the wild type \textit{CKX1} allele. This study proposed that the copy number differences of the \textit{CKX1} variants could possibly be the cause of different flowering strength between SD and DN genotypes (Figure 4.9).

\begin{verbatim}
FLP2E11       GTGTTTCATGTATCCATGCTTCCAACTTATTTCCCAATCCTCTCATGTGCTTGAG  120
63_1         GTGTTTCATGTATCCATGCTTCCAACTTATTTCCCAATCCTCTCATGTGCTTGAG  120
63_2         GTGTTTCATGTATCCATGCTTCCAACTTATTTCCCAATCCTCTCATGTGCTTGAG  120
63_3         GTGTTTCATGTATCCATGCTTCCAACTTATTTCCCAATCCTCTCATGTGCTTGAG  120
194_1        GTGTTTCATGTATCCATGCTTCCAACTTATTTCCCAATCCTCTCATGTGCTTGAG  120
194_2        GTGTTTCATGTATCCATGCTTCCAACTTATTTCCCAATCCTCTCATGTGCTTGAG  120
194_3        GTGTTTCATGTATCCATGCTTCCAACTTATTTCCCAATCCTCTCATGTGCTTGAG  120
311_1        GTGTTTCATGTATCCATGCTTCCAACTTATTTCCCAATCCTCTCATGTGCTTGAG  120
311_2        GTGTTTCATGTATCCATGCTTCCAACTTATTTCCCAATCCTCTCATGTGCTTGAG  120
311_3        GTGTTTCATGTATCCATGCTTCCAACTTATTTCCCAATCCTCTCATGTGCTTGAG  120
Ju_1         GTGTTTCATGTATCCATGCTTCCAACTTATTTCCCAATCCTCTCATGTGCTTGAG  120
Ju_2         GTGTTTCATGTATCCATGCTTCCAACTTATTTCCCAATCCTCTCATGTGCTTGAG  120
Ju_3         GTGTTTCATGTATCCATGCTTCCAACTTATTTCCCAATCCTCTCATGTGCTTGAG  120

******** *** *** *** *** ** * * ** ** ** ** ** ** ** **

Figure 4.9 Alignment of the parental variants for 05-069-63, 05-069-194, 01-061-311 and Juliette at FLP2E11 locus compared to the original DNA sequence derived from the strawberry-specific SDA
\end{verbatim}
The SDA data (Table 4.8) revealed that the strong DN bulk (DN1) showed higher hybridisation intensity (743.53) than the SD bulk (259.98) for FLP2E11 feature, indicating that this putative marker is present in different copy number between the strong DN plants and the SD plants. Additionally, a wide range of flowering habits are observed in octoploid strawberries, suggesting that flowering strength is controlled by the copy number of different allelic variants (Powers, 1954; Serçe and Hancock, 2005a). Therefore, it is postulated that the DN genotypes may contain more copies of the longer FLP2E11 marker allele. These alleles bound strongly with the FLP2E11 probe due to high sequence homology, resulting in higher hybridisation intensity. In contrast, the marker allele amplified from SD Juliette bound loosely to the FLP2E11 feature due to a 12 bp deletion, and hence, easily removed by the stringency wash during post processing of the SDA slides. This concept is supported by the hybridisation pattern of Single Feature Polymorphism (SFP) which is also based on allelic variations, where target sequences with SNPs or large deletions will produce weaker signal intensities compared to target sequences that have a perfect match with probes (Gupta et al., 2008). Furthermore, previous studies have reported that mismatched targets possessed a barrier to DNA-DNA hybridisation (Binder et al., 2005; Peterson et al., 2002; Tawa and Knoll, 2004). The weak hybridisation intensity observed in SD bulk is possibly due to fewer copies of the longer FLP2E11 marker allele hybridising to the probe on the SDA.

These results provide the first evidence for the ability of the SDA in identifying DNA sequence polymorphism based on differences in hybridisation intensities. Further, the current findings also suggest that the strawberry-specific SDA may be used as a panel to identify molecular markers associated with other traits of interest including day-neutrality. As explained in Chapter 3, allele-specific markers flanking the deleted region are needed to
genotype both parental and F₁ progeny as an effort to confirm the association between FLP2E11 marker allele and day-neutrality.

In addition, the copy number of FLP2E11 can be determined by allele dosage genotyping using real-time PCR. An excellent work was performed on hexaploid persimmon to predict the Ast/ast genotype of different persimmon cultivars using a marker allele linked to the recessive ast allele (Akagi et al., 2009; Akagi et al., 2010; Kanzaki et al., 2009). The authors demonstrated the possibility of using a homozygous recessive cultivar as a control in real-time PCR to detect allele dosage of Ast and ast in other cultivars (Akagi et al., 2010). Allele dosage genotyping by real-time PCR could facilitate the selection of DN genotypes with more copies of the longer FLP2E11 marker allele as parents in the breeding program. Moreover, DNA sequence polymorphism of the CKX1 allele should also be characterised from both SD and DN genotypes for marker-trait association studies. For instance, a previous study has discovered that a mutation in the promoter region of the Os-CKX2 gene caused down-regulation of the OS-CKX2 gene in inflorescence meristem, leading to increased cytokinin content and the formation of more reproductive organs in rice (Ashikari et al., 2005). Therefore, it is very likely that a mutation either in the promoter or coding region of CKX1 has an effect on flowering habits in strawberries.
4.3.5 Proposed model for FLP2E11 linked to the low activity CKX1 allele

The changes in the levels of endogenous cytokinins during strawberry flower induction have been reported in previous studies (Eshghi and Tafazoli, 2007; Yamasaki and Yamashita, 1990). For instance, the levels of free cytokinin in the shoot tips and leaves of strawberry plants grown under inductive condition (short day, low temperature) were significantly higher than in the plants raised under non-inductive condition (long day, high temperature) (Eshghi and Tafazoli, 2007). The accumulation of more cytokinins in short day conditions could be explained by the degradation of cytokinins by light. A recent study further supports this explanation, where an increase in cytokinin oxidase activity was promoted by light during senescence of barley leaf segments (Schlüter et al., 2011).

Generally, the short day genotypes will stop flowering and start to produce runners in summer when the day is long. Therefore, cytokinin should be a negative regulator for runner production. It has been shown that cytokinin application on three SD cultivars significantly reduced the number of runners compared to control plants, indicating that cytokinin may be suppressing runner formation in SD strawberries (Momenpour et al., 2011). This is consistent with the results from a previous study which concluded that flowering inhibits runner formation in both SD and DN genotypes (Bradford et al., 2010). Based on the possible roles of cytokinin in flower and runner production, a model was proposed for the FLP2E11 putative marker linked to the CKX1 locus in controlling flowering response in strawberry plants (Figure 4.10).
Figure 4.10 Proposed model of flowering response controlled by copy number of the allelic variants of $CKX1$ allele in SD and DN strawberry genotypes. SAM: Shoot apical meristem. + $CKX1$: Wild type allele; - $CKX1$: Low activity allele.
The proposed model (Figure 4.10) explains the flower production in both SD and DN genotypes during spring and summer. When the day is shorter than night (spring), higher levels of free cytokinins are accumulated in the leaves and shoot apical meristem (SAM) (Eshghi and Tafazoli, 2007). Although the SD plants contain more copies of wild type \textit{CKX1} allele, the degradation of cytokinins is slower due to a longer period of darkness (Schlüter \textit{et al.}, 2011). For the DN plants which contain more copies of the low activity \textit{CKX1} allele, the levels of free cytokinins remain the same in the plants. Therefore, flowering is induced in both SD and DN strawberry genotypes in spring. In contrast, when the day is longer than night (summer), lower levels of free cytokinins are accumulated in the leaves and shoot apical meristem (Eshghi and Tafazoli, 2007). Due to more copies of the wild type \textit{CKX1} allele present in the SD plants, free cytokinins are degraded quickly in the plants. Consequently, flower formation in SD genotypes is suppressed under long days of summer causing runners to be vigorously produced. However, since the DN plants contain more copies of the low activity \textit{CKX1} allele, the levels of free cytokinins in the leaves and SAM are still sufficient to promote flowering. Therefore, only the DN strawberry genotypes continue to flower in summer.

4.4 CONCLUSIONS

Field assessment on the $F_1$ population of the DN x SD and DN x DN crosses revealed that the different flowering strengths of strawberry genotypes varied from short day to strong day-neutral according to flowering scores. DN x DN crosses were found to be more potent in producing progeny plants with extreme day-neutral phenotypes, such as flowering on runners. Two strong day-neutral progenies which produced flowers on runners obtained from the experimental samples could be good candidates for cross hybridisation in the strawberry breeding program.
In addition, the application of SDA-BSA approach has successfully identified a putative DNA marker (FLP2E11) which could be associated with the low activity CKX1 allele leading to cytokinin accumulation in DN genotypes. DNA sequence analysis successfully revealed a 12 bp deletion within the marker allele for FLP2E11. The longer and shorter FLP2E11 marker allele may be linked to the low activity CKX1 allele and the wild type CKX1 alleles, respectively. This study proposed that the wild type CKX1 allele may degrade cytokinin at a faster rate, thus leading to flowering restriction in SD genotypes. In contrast, the different flowering strength in the DN genotypes could possibly be attributed to the copy number of the wild type CKX1 allele, in which the strong DN genotypes will contain more copies of longer FLP2E11 marker allele linked to the wild type CKX1 allele than the intermediate and weak DN genotypes.

In conclusion, this study demonstrated the usefulness of the strawberry-specific SDA for the discovery of DNA markers which may be linked to loci controlling day-neutrality in strawberry. DNA sequence characterisation of the FLP2E11 putative DNA marker is needed to convert the 12 bp deletion sequence polymorphism into a PCR-based marker to test its association with day-neutrality using a wider range of strawberry germplasm. Once validated, the PCR-based marker can be a useful tool for predicting flowering strength in different strawberry breeding materials.
CHAPTER 5
Conclusions and Future Directions

5.1 CONCLUSIONS

A review of the current state of knowledge regarding the development of molecular markers associated with fruit flavour and day-neutrality in strawberry plants was examined (Chapter 1). One of the gaps identified in the review was that no markers associated with aroma compound production have been applied in marker-assisted selection (MAS) for the genetic improvement of strawberry flavour. Moreover, the development of molecular markers linked to the loci controlling day-neutrality in *F. x ananassa* has been unsuccessful. These obstacles could be attributed to the complex biosynthetic pathways of aroma compounds and the genome complexity of an octoploid; as well as limited understanding concerning flowering response and contrasting reports on the genetic inheritance of day-neutrality in cultivated strawberry. Subsequently, the opportunities of finding molecular markers associated with fruit flavour and day-neutrality without the need for a mapping population was highlighted by using the Subtracted Diversity Array (SDA) approach.

Screening for molecular markers based on DNA polymorphisms between genotypes in parallel required a platform containing a large number of sequence-specific probes. This requirement was fulfilled by constructing a SDA containing 287 strawberry-specific DNA probes based on a broad subtraction approach between the genomic DNA pools of five strawberry genotypes and nine non-angiosperm species (Chapter 2). The strawberry-specific SDA validation revealed a high subtraction efficiency (99 %), indicating that the subtracted pool was enriched with flowering and fruiting related DNA sequences from the...
strawberry genome (Section 2.3.3, page 80). Analysis of the strawberry-specific SDA hybridisation data by hierarchical clustering confirmed its ability to correctly cluster 15 strawberry genotypes based on the differential hybridisation patterns and known pedigree information (Section 2.3.5, page 86). One of the key findings from this study was that the comparison between the aroma profiles and SDA data of five selected strawberry genotypes produced highly similar hierarchical dendrograms, suggesting that the chemical compositions of different genotypes may be predicted by their genetic profiles (Section 2.3.6.1, page 95). Additionally, eight polymorphic features across 15 strawberry genotypes were identified, in which five features (FLP1C6, FLP4D7, FLP2E1, FLP2E12 and FLP1E7) were manually-assigned as branch point markers based on the differences of signal intensities between genotypes (Section 2.3.7.1, page 100). Of these, FLP1C6, FLP1E7 and FLP2E1 were found to be strongly correlated with ethyl hexanoate, methyl esters and linalool, respectively (Section 2.3.7.2, page 103). These findings suggested that the strawberry-specific SDA may be a potential platform for a wide range of applications such as genotype identification and DNA marker discovery for marker-trait association studies.

Chapter 3 described the identification of molecular markers associated with fruit flavour using the strawberry-specific SDA and the Bulked Segregant Analysis (BSA) approaches. The aroma profiles for 50 F₁ progeny from a cross between 07-102-41 and Juliette were obtained. Transgressive segregation was observed for most key volatile compounds, providing an opportunity to select progeny which exhibited high or undetectable levels of the compounds for BSA (Section 3.3.1, page 136). Hybridisation of DNA bulks with extreme phenotypes onto the strawberry-specific SDA identified six features, including FLP1D7, FLP1D11, FLP1B3, FLP3E12, FLP2D11 and FLP1A7 which were associated with methyl butanoate, ethyl butanoate, ethyl hexanoate, linalool, (E)-nerolidol and
γ-dodecalactone, respectively (Section 3.3.2, page 143). DNA sequencing and alignment with *Fragaria vesca* scaffolds revealed a few genes closely linked to FLP1D7, FLP1A7 and FLP3E12 features. However, these genes may be involved in supplying precursors rather than the rate-limiting steps for the key volatile compound formation. One possible explanation could be that the compositions of compounds detected varied according to environmental fluctuations, thereby affecting the pooling of DNA from individuals showing extreme phenotypes for BSA (Section 3.3.3, page 167). Nevertheless, a putative C/T SNP was found within FLP1D7 sequence amplified from the parental genotypes which allow further analysis to be performed in order to examine the association of FLP1D7 and methyl butanoate production (Section 3.3.4, page 170). These findings suggested that further modifications are needed to fine-tune the SDA-BSA approach in order to identify molecular markers closely linked to genes involved in the rate-limiting steps for key volatile compounds formation.

The discovery of molecular markers associated with day-neutrality was described in Chapter 4. Field assessment on the F₁ population of the day-neutral (DN) x short day (SD) and DN x DN crosses successfully classified the F₁ individuals into four different flowering response classes according to flowering strength, ranging from strong day-neutral to a complete short day behaviour. A key finding from this assessment was that DN x DN crosses were able to produce progeny plants with stronger flowering strength, such as flowering on runners compared to the DN x SD cross (Section 4.3.1, page 181). In addition, FLP2E11 has been identified as the most promising marker controlling day-neutrality in strawberry based on the results obtained from the integration of SDA-BSA strategy (Section 4.3.2, page 183). DNA sequencing revealed that a cytokinin oxidase/dehydrogenase (*CKX1*) gene was located at 0.25 Mb downstream of FLP2E11.
Since cytokinin is needed to promote flowering, the FLP2E11 marker allele may be linked to the low activity \textit{CKX1} allele, leading to cytokinin accumulation in DN genotypes (Section 4.3.3, page 196). A key finding from the DNA polymorphism analysis was that a 12 bp deletion within the marker allele amplified from Juliette was detected (Section 4.3.4, page 201). This finding suggested that the shorter sequence of FLP2E11 may be linked to the wild type \textit{CKX1} allele, resulting in flowering restriction in the SD genotypes. This finding also led to the development of a new hypothesis, where different flowering strength in both DN and SD genotypes could possibly be controlled by the copy number of the low activity \textit{CKX1} allele. A hypothetical model was proposed to illustrate the flowering control regulated by \textit{CKX1} locus in both SD and DN strawberries based on the results of the present study and previously published literature. This model forms the basis of future studies to either confirm or reject the mechanism proposed for flowering control in strawberry. In conclusion, this study successfully demonstrated the usefulness of the strawberry-specific SDA for the discovery of DNA markers associated with day-neutrality in strawberry.

5.2 FUTURE DIRECTIONS

Marker validation is essential to confirm the correlation between the putative markers found in this study with the traits of interest before they can be applied for marker-assisted breeding. In this study, FLP2E11 feature has been determined as the most promising putative marker associated with day-neutrality. The 12 bp deletion region found within the DNA sequence of FLP2E11 provide an opportunity to examine the association between this putative marker and day-neutrality. This validation can be achieved by designing allele-specific primers flanking the polymorphic sites which could hopefully result in a binary scoring system based on the presence/absence of the amplicon. Allelic diversity can
be assessed by allele-specific PCR using the individuals derived from the extreme bulks and a wider range of strawberry germplasm. In addition, functional markers can be identified by searching for polymorphic sites within the coding and regulatory regions of the CKX1 gene by amplicon sequencing to examine its association with phenotypic variation.

However, one should be aware that if the expression of traits was due to copy number differences or allele dosage effect, the binary scoring system will not be achieved unless the genotypes tested is a complete homozygote for the corresponding allelic variants. This limitation can be overcome by allele dosage genotyping using real-time PCR. The copy number of the FLP2E11 marker allele and the low activity CKX1 allele can be determined by using a known single copy allele as a reference point. The quickest and simplest way to address this question is to obtain the haploid genome for strawberry since only one copy of each allele will be present in the genome. Our RMIT research group has begun to extract DNA from the pollen grains of strawberry to be used as reference DNA. Once the amount of haploid DNA is sufficient to run the real-time PCR experiment, the copy number of the FLP2E11 and the low activity CKX1 allele in the cultivated strawberry in relative to the haploid genome could be estimated using the $2^{-\Delta\Delta Ct}$ method. It is hypothesised that the DN genotypes with stronger flowering strength will contain more copies of the low activity CKX1 alleles compared to the genotypes showing weak flowering strength, and vice versa for the SD genotypes. Once this hypothesis is tested empirically, the FLP2E11 marker can be used to accelerate the introgression of day-neutrality alleles into the current SD cultivars/breeding lines.
In summary, the findings from this thesis have enhanced the current techniques and knowledge regarding flavour improvement and the molecular control of day-neutrality in strawberry. The strawberry-specific SDA served as a robust platform for marker-trait association studies but may not be restricted to fruit flavour and day-neutrality. The broad subtraction approach employed in this study resulting in a strawberry-specific SDA enriched with flowering and fruiting related DNA sequences. Therefore, this array may also be useful for the identification of molecular markers associated with other flowering and fruiting related traits, for example, flower size, fruit colour, fruit shape and total soluble solids (TSS) content. The limited number of molecular markers associated with strawberry flavour has been improved through the identification of potential SDA features linked to genes possibly involved in the biosynthesis of key volatile compounds dominating the typical strawberry aroma. More interestingly, the discovery of the FLP2E11 marker allele linked to a *CKX1* locus has provided novel insights on the molecular control of flowering in strawberry. The hypothetical model signifies the copy number of the low activity *CKX1* allele as a critical switch controlling flowering response. This hypothesis provides a foundation for further studies to characterise and test the association of the marker allele as well as the *CKX1* locus with day-neutrality. Subsequently, validation of these putative markers and their corresponding alleles may facilitate the selection of breeding materials with the traits of interest, resulting in the development of new day-neutral cultivars with enhanced fruit flavour.
BIBLIOGRAPHY


Alwala, S 2007, 'Identification of molecular markers associated with resistance to *Aspergillus flavus* in maize', Louisiana State University.


Ayala-Zavala, JF, Wang, SY, Wang, CY & González-Aguilar, GA 2004, 'Effect of storage temperatures on antioxidant capacity and aroma compounds in strawberry fruit', *LWT-Food Science and Technology*, vol. 37, no. 7, pp. 687-95.


Bowtell, D & Sambrook, J 2003, DNA microarrays: A molecular cloning manual, Cold Spring Harbor Laboratory, N.Y.


Brevis, P 2013, Project BS11013: National Strawberry Varietal Improvement Program (Southern Node), Strawberries Australia Inc.


Byrne, DH 2012, 'Trends in fruit breeding', in ML Badenes & DH Byrne (eds), Fruit Breeding, Springer, pp. 3-36.
Capocasa, F, Diamanti, J, Mezzetti, B, Tulipani, S & Battino, M 2008, 'Breeding strawberry (Fragaria X ananassa Duch) to increase fruit nutritional quality', *Biofactors*, vol. 34, no. 1, pp. 67-72.


Davis, TM, DiMeglio, LM, Yang, R, Styan, SM & Lewers, KS 2006, 'Assessment of SSR marker transfer from the cultivated strawberry to diploid strawberry species:
functionality, linkage group assignment, and use in diversity analysis', *Journal of the American Society for Horticultural Science*, vol. 131, no. 4, pp. 506-12.


Dempster, E, Pryor, K, Francis, D, Young, J & Rogers, H 1999, 'Rapid DNA extraction from ferns for PCR-based analyses', *Biotechniques*, vol. 27, no. 1, pp. 66-8.

Demyttenaere, JC, del Carmen Herrera, M & De Kimpe, N 2000, 'Biotransformation of geraniol, nerol and citral by sporulated surface cultures of *Aspergillus niger* and *Penicillium sp*', *Phytochemistry*, vol. 55, no. 4, pp. 363-73.

Deng, C & Davis, T 2001, 'Molecular identification of the yellow fruit color (c) locus in diploid strawberry: a candidate gene approach', *Theoretical and applied genetics*, vol. 103, no. 2-3, pp. 316-22.


Foolad, MR & Panthee, DR 2012, 'Marker-assisted selection in tomato breeding', *Critical Reviews in Plant Sciences*, vol. 31, no. 2, pp. 93-123.


an adsorptive column and aroma dilution analysis', *Journal of the Japanese Society for Horticultural Science (Japan)*.


Hakala, MA, Lapveteläinen, AT & Kallio, HP 2002, 'Volatile compounds of selected strawberry varieties analyzed by purge-and-trap headspace GC-MS', *Journal of agricultural and food chemistry*, vol. 50, no. 5, pp. 1133-42.


Ito, H & Saito, T 1962, 'Studies on the flower formation in the strawberry plants I. Effects of temperature and photoperiod on the flower formation'.


Klee, HJ & Giovannoni, JJ 2011, 'Genetics and control of tomato fruit ripening and quality attributes', *Annual review of genetics*, vol. 45, pp. 41-59.


Lahaye, R, Van der Bank, M, Bogarin, D, Warner, J, Pupulin, F, Gigot, G, Maurin, O, Duthoit, S, Barraclough, TG & Savolainen, V 2008, 'DNA barcoding the floras of
biodiversity hotspots', *Proceedings of the National Academy of Sciences*, vol. 105, no. 8, pp. 2923-8.


Lerceteau-Köhler, E, Moing, A, Guérin, G, Renaud, C, Petit, A, Rothan, C & Denoyes, B 2012, 'Genetic dissection of fruit quality traits in the octoploid cultivated
strawberry highlights the role of homoeo-QTL in their control', *Theoretical and applied genetics*, vol. 124, no. 6, pp. 1059-77.


Li, F-W, Kuo, L-Y, Rothfels, CJ, Ebihara, A, Chiou, W-L, Windham, MD & Pryer, KM 2011, 'rbcL and matK earn two thumbs up as the core DNA barcode for ferns', *PloS one*, vol. 6, no. 10, p. e26597.


Minhas, D & Grover, A 1999, 'Transcript levels of genes encoding various glycolytic and fermentation enzymes change in response to abiotic stresses', *Plant Science*, vol. 146, no. 1, pp. 41-51.


Morales, RGF, Resende, JTV, Faria, MV, Andrade, MC, Resende, LV, Delatorre, CA & Silva, PRd 2011, 'Genetic similarity among strawberry cultivars assessed by RAPD and ISSR markers', *Scientia Agricola*, vol. 68, no. 6, pp. 665-70.


Niemisrowicz-Szczytt, K 1988, 'Preliminary studies on inbreeding in strawberry *Fragaria x ananassa* Duch', paper presented to International Strawberry Symposium 265.


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


Olarte, A 2011, 'Array based genotyping of medicinal plants from the genera Salvia and Echinacea', RMIT University.


Pelayo, C, Ebeler, S & Kader, A 2003, 'Postharvest life and flavor quality of three strawberry cultivars kept at 5°C in air or air+ 20 kPa CO₂', Postharvest biology and technology, vol. 27, no. 2, pp. 171-83.


Rousseau-Gueutin, M, Lerceteau-Köhler, E, Barrot, L, Sargent, DJ, Monfort, A, Simpson, D, Arus, P, Guérin, G & Denoyes-Rothan, B 2008, 'Comparative genetic mapping between octoploid and diploid *Fragaria* species reveals a high level of colinearity between their genomes and the essentially disomic behavior of the cultivated octoploid strawberry', *Genetics*, vol. 179, no. 4, pp. 2045-60.


Ruan, J, Lee, YH & Yeoung, YR 2013, 'Flowering and fruiting of day-neutral and ever-bearing strawberry cultivars in high-elevation for summer and autumn fruit production in Korea', *Horticulture, Environment, and Biotechnology*, vol. 54, no. 2, pp. 205-20.


Samykanno, K 2012, 'Environmental effects on flavour development in Australian-grown strawberry varieties', RMIT University.


Samykanno, K, Pang, E & Marriott, PJ 2013b, 'Genotypic and environmental effects on flavor attributes of “Albion” and “Juliette” strawberry fruits', *Scientia horticulturae*, vol. 164, pp. 633-42.


Sansaloni, CP, Petroli, CD, Carling, J, Hudson, CJ, Steane, DA, Myburg, AA, Grattapaglia, D, Vaillancourt, RE & Kilian, A 2010, 'A high-density Diversity Arrays Technology (DArT) microarray for genome-wide genotyping in Eucalyptus', Plant Methods, vol. 6, no. 1, p. 16.


Serçe, S & Hancock, JF 2003, 'Assessment of day-neutrality scoring methods in strawberry families grown in greenhouse and field environments', *Turkish Journal of Agriculture and Forestry*, vol. 27, no. 4, pp. 191-8.


Skuza, L, Filip, E & Szućko, I 2013, 'Use of Organelle Markers to Study Genetic Diversity in Soybean'.


Ulrich, D, Komes, D, Olbricht, K & Hoberg, E 2007, 'Diversity of aroma patterns in wild and cultivated Fragaria accessions', Genetic resources and crop evolution, vol. 54, no. 6, pp. 1185-96.


Zabetakis, I 1997, 'Enhancement of flavour biosynthesis from strawberry (Fragaria x ananassa) callus cultures by Methylobacterium species', Plant cell, tissue and organ culture, vol. 50, no. 3, pp. 179-83.


APPENDIX 1

Overview of the Suppression Subtractive Hybridisation (SSH) process

Figure A1.1 Reproduced from Clontech PCR-Select™ cDNA Subtraction Kit User Manual.
APPENDIX 2

Position of the 290 array features and 18 controls gridded on four meta-grids for the strawberry-specific SDA

The first meta-grid containing 77 array features printed by the first pin:

<table>
<thead>
<tr>
<th>FLP2E11</th>
<th>FLP2F1</th>
<th>FLP2F4</th>
<th>FLP2F6</th>
<th>FLP2F9</th>
<th>FLP2F11</th>
<th>FLP2G1</th>
<th>FLP2G3</th>
<th>FLP2G5</th>
<th>FLP2G7</th>
<th>FLP2G9</th>
</tr>
</thead>
<tbody>
<tr>
<td>FLP2C9</td>
<td>FLP2C11</td>
<td>FLP2D3</td>
<td>FLP2D5</td>
<td>FLP2D7</td>
<td>FLP2D9</td>
<td>FLP2D11</td>
<td>FLP2E1</td>
<td>FLP2E3</td>
<td>FLP2E5</td>
<td>FLP2E7</td>
</tr>
<tr>
<td>FLP1H12</td>
<td>FLP2A3</td>
<td>FLP2A7</td>
<td>FLP2A9</td>
<td>FLP2A11</td>
<td>FLP2B1</td>
<td>FLP2B3</td>
<td>FLP2B5</td>
<td>FLP2B8</td>
<td>FLP2B11</td>
<td>FLP2C4</td>
</tr>
<tr>
<td>FLP1G1</td>
<td>FLP1G3</td>
<td>FLP1G5</td>
<td>FLP1G7</td>
<td>FLP1G9</td>
<td>FLP1G12</td>
<td>FLP1H2</td>
<td>FLP1H4</td>
<td>FLP1H6</td>
<td>FLP1H8</td>
<td>FLP1H10</td>
</tr>
<tr>
<td>FLP1E1</td>
<td>FLP1E3</td>
<td>FLP1E5</td>
<td>FLP1E7</td>
<td>FLP1E9</td>
<td>FLP1E11</td>
<td>FLP1F1</td>
<td>FLP1F3</td>
<td>FLP1F7</td>
<td>FLP1F9</td>
<td>FLP1F11</td>
</tr>
<tr>
<td>FLP1C1</td>
<td>FLP1C4</td>
<td>FLP1C6</td>
<td>FLP1C8</td>
<td>FLP1C11</td>
<td>FLP1D1</td>
<td>FLP1D3</td>
<td>FLP1D5</td>
<td>FLP1D7</td>
<td>FLP1D9</td>
<td>FLP1D11</td>
</tr>
<tr>
<td>FLP1A1</td>
<td>FLP1A4</td>
<td>FLP1A6</td>
<td>FLP1A8</td>
<td>FLP1A10</td>
<td>FLP1A12</td>
<td>FLP1B3</td>
<td>FLP1B5</td>
<td>FLP1B7</td>
<td>FLP1B9</td>
<td>FLP1B1</td>
</tr>
</tbody>
</table>

The second meta-grid containing 77 array features printed by the second pin:

<table>
<thead>
<tr>
<th>FLP2E12</th>
<th>FLP2F3</th>
<th>FLP2F5</th>
<th>FLP2F8</th>
<th>FLP2F10</th>
<th>FLP2F12</th>
<th>FLP2G2</th>
<th>FLP2G4</th>
<th>FLP2G6</th>
<th>FLP2G8</th>
<th>FLP2G10</th>
</tr>
</thead>
<tbody>
<tr>
<td>FLP2C10</td>
<td>FLP2D1</td>
<td>FLP2D4</td>
<td>FLP2D6</td>
<td>FLP2D8</td>
<td>FLP2D10</td>
<td>FLP2D12</td>
<td>FLP2E2</td>
<td>FLP2E4</td>
<td>FLP2E6</td>
<td>FLP2E9</td>
</tr>
<tr>
<td>FLP1G2</td>
<td>FLP1G4</td>
<td>FLP1G6</td>
<td>FLP1G8</td>
<td>FLP1G11</td>
<td>FLP1H1</td>
<td>FLP1H3</td>
<td>FLP1H5</td>
<td>FLP1H7</td>
<td>FLP1H9</td>
<td>FLP1H11</td>
</tr>
<tr>
<td>FLP1E2</td>
<td>FLP1E4</td>
<td>FLP1E6</td>
<td>FLP1E8</td>
<td>FLP1E10</td>
<td>FLP1E12</td>
<td>FLP1F2</td>
<td>FLP1F5</td>
<td>FLP1F8</td>
<td>FLP1F10</td>
<td>FLP1F12</td>
</tr>
<tr>
<td>FLP1C3</td>
<td>FLP1C5</td>
<td>FLP1C7</td>
<td>FLP1C10</td>
<td>FLP1C12</td>
<td>FLP1D2</td>
<td>FLP1D4</td>
<td>FLP1D6</td>
<td>FLP1D8</td>
<td>FLP1D10</td>
<td>FLP1D12</td>
</tr>
<tr>
<td>FLP1A2</td>
<td>FLP1A5</td>
<td>FLP1A7</td>
<td>FLP1A9</td>
<td>FLP1A11</td>
<td>FLP1B2</td>
<td>FLP1B4</td>
<td>FLP1B6</td>
<td>FLP1B8</td>
<td>FLP1B10</td>
<td>FLP1B12</td>
</tr>
</tbody>
</table>
The third meta-grid containing 77 array features printed by the first pin:

<table>
<thead>
<tr>
<th>FLP4D4</th>
<th>FLP4D5</th>
<th>FLP4D6</th>
<th>FLP4D7</th>
<th>50 %DMSO</th>
<th>50 %DMSO</th>
<th>50 %DMSO</th>
<th>50 %DMSO</th>
<th>Nested 2R</th>
<th>Nested 1</th>
<th>50 %DMSO</th>
</tr>
</thead>
<tbody>
<tr>
<td>FLP4B5</td>
<td>FLP4B7</td>
<td>FLP4B9</td>
<td>FLP4B11</td>
<td>FLP4C1</td>
<td>FLP4C3</td>
<td>FLP4C5</td>
<td>FLP4C7</td>
<td>FLP4C10</td>
<td>FLP4C12</td>
<td>FLP4D2</td>
</tr>
<tr>
<td>FLP3H6</td>
<td>FLP3H8</td>
<td>FLP3H10</td>
<td>FLP3H12</td>
<td>FLP4A2</td>
<td>FLP4A4</td>
<td>FLP4A7</td>
<td>FLP4A9</td>
<td>FLP4A11</td>
<td>FLP4B1</td>
<td>FLP4B3</td>
</tr>
<tr>
<td>FLP3F7</td>
<td>FLP3F9</td>
<td>FLP3F11</td>
<td>FLP3G1</td>
<td>FLP3G3</td>
<td>FLP3G5</td>
<td>FLP3G7</td>
<td>FLP3G10</td>
<td>FLP3G12</td>
<td>FLP3H2</td>
<td>FLP3H4</td>
</tr>
<tr>
<td>FLP3D7</td>
<td>FLP3D9</td>
<td>FLP3E1</td>
<td>FLP3E3</td>
<td>FLP3E5</td>
<td>FLP3E7</td>
<td>FLP3E9</td>
<td>FLP3E11</td>
<td>FLP3F1</td>
<td>FLP3F3</td>
<td>FLP3F5</td>
</tr>
<tr>
<td>FLP3B3</td>
<td>FLP3B5</td>
<td>FLP3B8</td>
<td>FLP3B10</td>
<td>FLP3C1</td>
<td>FLP3C3</td>
<td>FLP3C5</td>
<td>FLP3C7</td>
<td>FLP3C11</td>
<td>FLP3D2</td>
<td>FLP3D4</td>
</tr>
<tr>
<td>FLP2G11</td>
<td>FLP2H1</td>
<td>FLP2H5</td>
<td>FLP2H7</td>
<td>FLP3A1</td>
<td>FLP3A3</td>
<td>FLP3A5</td>
<td>FLP3A7</td>
<td>FLP3A9</td>
<td>FLP3A11</td>
<td>FLP3B1</td>
</tr>
</tbody>
</table>

The fourth meta-grid containing 77 array features printed by the second pin:

<table>
<thead>
<tr>
<th>pGEM-T</th>
<th>Beta-actin</th>
<th>PLB</th>
<th>AAT</th>
<th>ADH</th>
<th>SES</th>
<th>Cy-3</th>
<th>Cy-5</th>
<th>50 % DMSO</th>
<th>Aromatase</th>
<th>Subtracted product</th>
</tr>
</thead>
<tbody>
<tr>
<td>FLP4B6</td>
<td>FLP4B8</td>
<td>FLP4B10</td>
<td>FLP4B12</td>
<td>FLP4C2</td>
<td>FLP4C4</td>
<td>FLP4C6</td>
<td>FLP4C8</td>
<td>FLP4C11</td>
<td>FLP4D1</td>
<td>FLP4D3</td>
</tr>
<tr>
<td>FLP3H7</td>
<td>FLP3H9</td>
<td>FLP3H11</td>
<td>FLP4A1</td>
<td>FLP4A3</td>
<td>FLP4A6</td>
<td>FLP4A8</td>
<td>FLP4A10</td>
<td>FLP4A12</td>
<td>FLP4B2</td>
<td>FLP4B4</td>
</tr>
<tr>
<td>FLP3F8</td>
<td>FLP3F10</td>
<td>FLP3F12</td>
<td>FLP3G2</td>
<td>FLP3G4</td>
<td>FLP3G6</td>
<td>FLP3G9</td>
<td>FLP3G11</td>
<td>FLP3H1</td>
<td>FLP3H3</td>
<td>FLP3H5</td>
</tr>
<tr>
<td>FLP3D8</td>
<td>FLP3D10</td>
<td>FLP3E2</td>
<td>FLP3E4</td>
<td>FLP3E6</td>
<td>FLP3E8</td>
<td>FLP3E10</td>
<td>FLP3E12</td>
<td>FLP3F2</td>
<td>FLP3F4</td>
<td>FLP3F6</td>
</tr>
<tr>
<td>FLP3B4</td>
<td>FLP3B6</td>
<td>FLP3B9</td>
<td>FLP3B11</td>
<td>FLP3C2</td>
<td>FLP3C4</td>
<td>FLP3C6</td>
<td>FLP3C10</td>
<td>FLP3C12</td>
<td>FLP3D3</td>
<td>FLP3D5</td>
</tr>
<tr>
<td>FLP2G12</td>
<td>FLP2H3</td>
<td>FLP2H6</td>
<td>FLP2H9</td>
<td>FLP3A2</td>
<td>FLP3A4</td>
<td>FLP3A6</td>
<td>FLP3A8</td>
<td>FLP3A10</td>
<td>FLP3A12</td>
<td>FLP3B2</td>
</tr>
</tbody>
</table>

PLB: Pectate lyase B; AAT: Alcohol acyl transferase; ADH: Alcohol dehydrogenase; SES: Sesquiterpene synthase
APPENDIX 3

Configurations of the BioRobotics® Total Array System (TAS)
Application Suite software v2.6.0.1

The parameters used for the SDA printing were configured as follow:

**OPTION**

**Tool Type**
- Group: 2 Microspot (384 well)
- Tool: 2 x 1 configuration

**Pin refill frequency**

Spots per source visit: 85
Calculation for both flavour and day-neutrality array =

\[
\text{[(12 spots/slide x No. of slides to print)] + 25 pre-spots} \]

**Wash frequency**

Always wash before pins refills

**SOURCE**

**Microplate options**
- Microplate group: Generic
- Microplate type: 384 well low profile
- No. plates: 1
- No. samples: 308

**Last plate**

154/192

(154 is the number of sources on a 384-well plate per pin)

**Source loading**

Hold 1 plate at a time

**Source action**

Dwell
**TARGET**

**EDIT PATTERN**

Size 11 x 7 (Number of spots in a meta-grid)
Pitch 0.295 mm (Distance between centers of spots)
Format Standard

**ADAPTER PLATE AND SLIDE LAYOUT**

Targets 10 (No. of pre-spotting slides + No. of real slides)
Edit layout Adapter layout: 30 vertical slides
No. of copies fill: 10 Slides on tray 1
Slide layout Top margin: 10.49 mm
Bottom margin: 14.32 mm
Left margin: 4.40 mm
Right margin: 4.40 mm
X- and Y-spacing adjusted to fit: 0.20 mm
Slide size: 25 x 75 mm
Tool array width: 7.74 x 2.06 mm

**TARGET ACTION**

Delay before spotting 0.000 s
Target height 0.1 mm
Dwell time 0.000 s
Multiple strikes 1
Pre-spotting 25 spots
## EDIT SOFT TOUCH

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Specification</th>
</tr>
</thead>
<tbody>
<tr>
<td>Soft touch</td>
<td>Target height 0.1 mm</td>
</tr>
<tr>
<td>Soft touch distance</td>
<td>1.000 mm</td>
</tr>
<tr>
<td>Speed</td>
<td>4.0 mm/s</td>
</tr>
<tr>
<td>Climate</td>
<td>For DMSO buffer</td>
</tr>
<tr>
<td></td>
<td>Target humidity at 60 %</td>
</tr>
<tr>
<td></td>
<td>Minimum humidity at 36 %</td>
</tr>
<tr>
<td>Bath 1 and 2</td>
<td>Use both baths for 3 s</td>
</tr>
<tr>
<td>Action</td>
<td>Wiggle 0.3 mm</td>
</tr>
<tr>
<td>Behaviour</td>
<td>0.0 mm</td>
</tr>
<tr>
<td>MWS</td>
<td>Used main wash station for 1 cycle</td>
</tr>
<tr>
<td></td>
<td>Entire wash cycle 2 times</td>
</tr>
</tbody>
</table>
APPENDIX 4

Position of two subarrays, six technical replicates and four meta-grids on the strawberry-specific SDA

Each subarray was composed of six technical replicates. Each technical replicate contained four meta-grids printed in 11 x 7 format.

<table>
<thead>
<tr>
<th>Meta-grid 1</th>
<th>Meta-grid 2</th>
<th>Meta-grid 3</th>
<th>Meta-grid 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Technical replicate 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Technical replicate 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Technical replicate 3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Technical replicate 4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Technical replicate 5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Technical replicate 6</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Subarray 1

<table>
<thead>
<tr>
<th>Meta-grid 1</th>
<th>Meta-grid 2</th>
<th>Meta-grid 3</th>
<th>Meta-grid 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Technical replicate 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Technical replicate 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Technical replicate 3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Technical replicate 4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Technical replicate 5</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Subarray 2
APPENDIX 5

Settings for ScanArray® Express v4.0

The parameters used for spot quantification in the ScanArray® Express v4.0 were as follow:

**Subarrays**

Number of rows of subarrays: 6
Number of columns of subarrays: 4
Rotation (degrees): -0.08
Horizontal pin spacing (mm): 4.5
Vertical pin spacing (mm): 4.5
(The meaning of subarrays represents technical replicates in this context)

**Spots**

Horizontal spot spacing, center to center (µm): 295.000
Vertical spot spacing, center to center (µm): 293.000
Rows of spots per subarray: 7
Columns of spots per subarray: 11
Spot diameter (µm) : 190
APPENDIX 6

Positioning of the grid onto the strawberry-specific SDA

Misaligned features were manually adjusted for optimal spot recognition. Suspicious features were manually flagged and filtered to reduce variability during data analysis. Flagged spots (X) represent features with low quality or signal intensities.
Pearson's bivariate correlation among the most polymorphic features across 15 strawberry genotypes

Table A6.1 Pearson bivariate correlation of the nine most polymorphic features chosen by PCA (SPSS v. 21) and magnitude of variance across the 15 strawberry genotypes. Significant correlations were found between FLP1D5 and FLP1C6 ($r = 0.80, p < 0.01$) and FLP1D5 and FLP1A2 ($r = 0.77, p < 0.01$), suggesting they have similar hybridization patterns and may represent redundant clones.

<table>
<thead>
<tr>
<th>Feature</th>
<th>Pearson Correlation</th>
<th>FLP1C6</th>
<th>FLP1A2</th>
<th>FLP4D7</th>
<th>FLP2E1</th>
<th>FLP2E12</th>
<th>FLP1D11</th>
<th>FLP1G3</th>
<th>FLP1E7</th>
<th>FLP1D5</th>
</tr>
</thead>
<tbody>
<tr>
<td>FLP1C6</td>
<td></td>
<td>1</td>
<td>.602</td>
<td>.245</td>
<td>.309</td>
<td>.482</td>
<td>-.241</td>
<td>-.309</td>
<td>.237</td>
<td>.795</td>
</tr>
<tr>
<td>Sig. (2-tailed)</td>
<td></td>
<td>.018</td>
<td>.380</td>
<td>.263</td>
<td>.069</td>
<td>.387</td>
<td>.263</td>
<td>.396</td>
<td>.000</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td></td>
</tr>
<tr>
<td>FLP1A2</td>
<td>.602</td>
<td>1</td>
<td>.430</td>
<td>.406</td>
<td>.274</td>
<td>-.407</td>
<td>-.500</td>
<td>.046</td>
<td>.766</td>
<td></td>
</tr>
<tr>
<td>Sig. (2-tailed)</td>
<td></td>
<td>.018</td>
<td>.110</td>
<td>.134</td>
<td>.324</td>
<td>.132</td>
<td>.058</td>
<td>.869</td>
<td>.001</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td></td>
</tr>
<tr>
<td>FLP4D7</td>
<td>.245</td>
<td>.430</td>
<td>1</td>
<td>.111</td>
<td>.067</td>
<td>-.468</td>
<td>-.013</td>
<td>.110</td>
<td>.468</td>
<td></td>
</tr>
<tr>
<td>Sig. (2-tailed)</td>
<td></td>
<td>.380</td>
<td>.110</td>
<td>.693</td>
<td>.812</td>
<td>.078</td>
<td>.963</td>
<td>.697</td>
<td>.079</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td></td>
</tr>
<tr>
<td>FLP2E1</td>
<td>.309</td>
<td>.406</td>
<td>.111</td>
<td>1</td>
<td>-.034</td>
<td>.178</td>
<td>-.248</td>
<td>.133</td>
<td>.099</td>
<td></td>
</tr>
<tr>
<td>Sig. (2-tailed)</td>
<td></td>
<td>.263</td>
<td>.134</td>
<td>.693</td>
<td>.903</td>
<td>.525</td>
<td>.372</td>
<td>.636</td>
<td>.725</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td></td>
</tr>
<tr>
<td>FLP2E12</td>
<td>.482</td>
<td>.274</td>
<td>.067</td>
<td>.034</td>
<td>1</td>
<td>.243</td>
<td>.168</td>
<td>.023</td>
<td>.339</td>
<td></td>
</tr>
<tr>
<td>Sig. (2-tailed)</td>
<td></td>
<td>.069</td>
<td>.324</td>
<td>.812</td>
<td>.903</td>
<td>.383</td>
<td>.549</td>
<td>.936</td>
<td>.216</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td></td>
</tr>
<tr>
<td>FLP1D11</td>
<td>-.241</td>
<td>-.407</td>
<td>-.468</td>
<td>.178</td>
<td>.243</td>
<td>1</td>
<td>.514</td>
<td>-.311</td>
<td>-.541</td>
<td></td>
</tr>
<tr>
<td>Sig. (2-tailed)</td>
<td></td>
<td>.387</td>
<td>.132</td>
<td>.078</td>
<td>.525</td>
<td>.383</td>
<td>.050</td>
<td>.260</td>
<td>.037</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td></td>
</tr>
<tr>
<td>FLP1G3</td>
<td>-.309</td>
<td>-.500</td>
<td>-.013</td>
<td>-.248</td>
<td>.168</td>
<td>.514</td>
<td>1</td>
<td>-.112</td>
<td>-.455</td>
<td></td>
</tr>
<tr>
<td>Sig. (2-tailed)</td>
<td></td>
<td>.263</td>
<td>.058</td>
<td>.963</td>
<td>.372</td>
<td>.549</td>
<td>.050</td>
<td>.691</td>
<td>.088</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td></td>
</tr>
</tbody>
</table>

Table A6.1 continued on next page
Table A6.1 (continued)

<table>
<thead>
<tr>
<th></th>
<th>FLP1C6</th>
<th>FLP1A2</th>
<th>FLP4D7</th>
<th>FLP2E1</th>
<th>FLP2E12</th>
<th>FLP1D11</th>
<th>FLP1G3</th>
<th>FLP1E7</th>
<th>FLP1D5</th>
</tr>
</thead>
<tbody>
<tr>
<td>FLP1E7 Pearson Correlation</td>
<td>.237</td>
<td>.046</td>
<td>.110</td>
<td>.133</td>
<td>.023</td>
<td>-.311</td>
<td>-.112</td>
<td>1</td>
<td>.220</td>
</tr>
<tr>
<td>Sig. (2-tailed)</td>
<td>.396</td>
<td>.869</td>
<td>.697</td>
<td>.636</td>
<td>.936</td>
<td>.260</td>
<td>.691</td>
<td>.430</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td></td>
</tr>
</tbody>
</table>

|       | FLP1D5 Pearson Correlation | .795** | .766** | .468   | .099   | .339    | -.541   | -.455  | .220   |
|      Sig. (2-tailed)       | .000   | .001   | .079   | .725   | .216    | .037    | .088   | .430   |
|      N                     | 15     | 15     | 15     | 15     | 15      | 15      | 15     | 15     |

* Correlation is significant at the 0.05 level (2-tailed).
** Correlation is significant at the 0.01 level (2-tailed).
APPENDIX 8

Pearson’s bivariate correlation between selected branch point markers and key volatile compounds

Table A7.1 Significant positive correlations were found between FLP1E7 with methyl butanoate ($r = 0.93$, $p < 0.05$) and methyl hexanoate ($r = 0.97$, $p < 0.01$) and FLP1C6 with ethyl hexanoate ($r = 0.93$, $p < 0.05$) whereas significant negative correlation was found between FLP2E1 with linalool ($r = -0.91$, $p < 0.05$).

<table>
<thead>
<tr>
<th></th>
<th>Methyl butanoate</th>
<th>Methyl hexanoate</th>
<th>Ethyl hexanoate</th>
<th>Linalool</th>
<th>FLP1E7</th>
<th>FLP1C6</th>
<th>FLP2E1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Methyl butanoate</td>
<td>Pearson Correlation</td>
<td>.920*</td>
<td>.347</td>
<td>.241</td>
<td>.932*</td>
<td>.086</td>
<td>-.038</td>
</tr>
<tr>
<td>Sig. (2-tailed)</td>
<td>.027</td>
<td>.568</td>
<td>.696</td>
<td>.021</td>
<td>.891</td>
<td>.952</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Methyl hexanoate</td>
<td>Pearson Correlation</td>
<td>.581</td>
<td>.108</td>
<td>.965**</td>
<td>.419</td>
<td>.152</td>
<td></td>
</tr>
<tr>
<td>Sig. (2-tailed)</td>
<td>.027</td>
<td>.304</td>
<td>.863</td>
<td>.008</td>
<td>.482</td>
<td>.807</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Ethyl hexanoate</td>
<td>Pearson Correlation</td>
<td>.568</td>
<td>.378</td>
<td>.394</td>
<td>.926*</td>
<td>.010</td>
<td></td>
</tr>
<tr>
<td>Sig. (2-tailed)</td>
<td>.568</td>
<td>.304</td>
<td>.531</td>
<td>.512</td>
<td>.024</td>
<td>.988</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Linalool</td>
<td>Pearson Correlation</td>
<td>.241</td>
<td>.378</td>
<td>1</td>
<td>-.073</td>
<td>.075</td>
<td>-.911</td>
</tr>
<tr>
<td>Sig. (2-tailed)</td>
<td>.696</td>
<td>.863</td>
<td>.531</td>
<td>.907</td>
<td>.905</td>
<td>.032</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>FLP1E7</td>
<td>Pearson Correlation</td>
<td>.932*</td>
<td>.394</td>
<td>-.073</td>
<td>1</td>
<td>.251</td>
<td>.299</td>
</tr>
<tr>
<td>Sig. (2-tailed)</td>
<td>.021</td>
<td>.008</td>
<td>.512</td>
<td>.9097</td>
<td>.684</td>
<td>.625</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>FLP1C6</td>
<td>Pearson Correlation</td>
<td>.086</td>
<td>.926*</td>
<td>.075</td>
<td>.251</td>
<td>1</td>
<td>.247</td>
</tr>
<tr>
<td>Sig. (2-tailed)</td>
<td>.891</td>
<td>.482</td>
<td>.024</td>
<td>.905</td>
<td>.684</td>
<td>.689</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td></td>
</tr>
</tbody>
</table>

Table A7.1 continued on next page
Table A7.1 (continued)

<table>
<thead>
<tr>
<th>FLP2E1</th>
<th>Pearson Correlation</th>
<th>Methyl butanote</th>
<th>Methyl hexanoate</th>
<th>Ethyl hexanoate</th>
<th>Linalool</th>
<th>FLP1E7</th>
<th>FLP1C6</th>
<th>FLP2E1</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>-.038</td>
<td>.152</td>
<td>-.010</td>
<td>-.911*</td>
<td>.299</td>
<td>.247</td>
<td>1</td>
</tr>
<tr>
<td>Sig. (2-tailed)</td>
<td></td>
<td>.952</td>
<td>.807</td>
<td>.988</td>
<td>.032</td>
<td>.625</td>
<td>.689</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td></td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
</tbody>
</table>

*. Correlation is significant at the 0.05 level (2-tailed).

**. Correlation is significant at the 0.01 level (2-tailed).
APPENDIX 9

Polymorphic DNA sequences corresponding to branch point markers
found in the fingerprinting of strawberry genotypes

Primer sequences in forward orientation

Nested primer 1: 5’-CCGGGCAGGT-3’
Nested primer 2R: 5’-ACCTCGGCCG-3’

Primer sequences in reverse orientation

Nested primer 1: 5’-ACCTGCCCGG-3’
Nested primer 2R: 5’-CGGCCAGAGGT-3’

>FLP1A2 (506 bp)
CCGGGCAGGTCCGGACCTCCAACTGAAAAATAGTTTTAAAAACCGATGACACGCAAACAAAAAGATCGTAT
AAAAAGTTTTACGCAAAATAACCACTCGAATATAGGGGACTCTATAGGGGAAAATAATGGAAAAAT
TGCAATTGACCGAACTACGTACAGCAGAAATTTACGGAATATGAGTATATTTTTAACTCGGAT
GTATTTCACATTTCTCGTCATATCTTTATTTATCGACTTTTATAGTATTTCTGAGGTTAGGTTGT
TTTGGAAACGGAACATTTACAACTTATAACCTTTATAGTATTTTACATAGATATGTGTGTTTTTAA
ACCGCCGAGGT

>FLP1C6 (555 bp)
CCGGGCAGGTCCTGAGAGTATTTCCCGAATCACCCTACCGGGCTAGGAAGTAGATATTTTC
AGAATTACAGAAGGAAAGGGAAAGAACTAAGCAGCGAAATTTTAACTTGAGATTTTAACTTTAGATA
ACGGGCACCTCGGCAGG

>FLP1E7 (512 bp)
CCGGGCAGGTCAGGAAAGCGATCCTGACATTTCTTTGCTACGTAGAGCAACATTTTGGGCTAGAGTTCGA
AAAGATTAAGGAAAGAACTAAGCAGCGAAATTTTAACTTTGAGATTTTAACTTTAGATA
ACGGGCACCTCGGCAGG

260
APPENDIX 10

Polymorphic DNA sequences corresponding to flavour traits in strawberry plants

**Primer sequences in forward orientation**

Nested primer 1: 5’-\textbf{CCGGCAGGTT}-3’

Nested primer 2R: 5’-\textbf{ACCTCGGCCG}-3’

**Primer sequences in reverse orientation**

Nested primer 1: 5’-\textbf{ACCTGCCCCG}-3’

Nested primer 2R: 5’-\textbf{CGGCCCCG}-3’

\textgreater FLP1A6 (442 bp)

\textbf{CCGGCAGGTT}\textbf{CTCAAGGAAAGACCTACATGGTAGGTACACATGGTGCACCCACCAACCAACATCGATGGTCT}
\textbf{TCTTCGAGTTACTCAGATGGTAGTCAATGGGGCCCAATAGGTTATTTACACC}
\textbf{AGGGATGAATAGTTATGGCAAAATCTGCAACAATGAGGTTAGGATTATTACCG}
\textbf{ACCTGGCG}

\textgreater FLP1B3 (343 bp)

\textbf{CCGGCAGGTT}\textbf{CTCAAGGAAAGACCTACATGGTAGGTACACATGGTGCACCCACCAACCAACATCGATGGTCT}
\textbf{TCTTCGAGTTACTCAGATGGTAGTCAATGGGGCCCAATAGGTTATTTACACC}
\textbf{AGGGATGAATAGTTATGGCAAAATCTGCAACAATGAGGTTAGGATTATTACCG}
\textbf{ACCTGGCG}

\textgreater FLP1D7 (627 bp)

\textbf{CCGGCAGGTT}\textbf{CTCAAGGAAAGACCTACATGGTAGGTACACATGGTGCACCCACCAACCAACATCGATGGTCT}
\textbf{TCTTCGAGTTACTCAGATGGTAGTCAATGGGGCCCAATAGGTTATTTACACC}
\textbf{AGGGATGAATAGTTATGGCAAAATCTGCAACAATGAGGTTAGGATTATTACCG}
\textbf{ACCTGGCG}
>FLP1D11 (850 bp)

CCGGGCAGGT

>FLP2D11 (670 bp)

CCGGGCAGGT

>FLP3E12 (539 bp)

CCGGGCAGGT
APPENDIX 11

Polymorphic DNA sequences corresponding to day-neutrality traits in strawberry plants

Primer sequences in forward orientation
Nested primer 1: 5'-CCGGGCAGGT-3'
Nested primer 2R: 5'-ACCTCGGCCG-3'

Primer sequences in reverse orientation
Nested primer 1: 5'-ACCTGCGCCGG-3'
Nested primer 2R: 5'-CGCCCGAGGT-3'

>FLP1A1 (597 bp)
CCGGGCAGGT CCAATGAATTGGGATTTCCCTATTGAGGTTTTACAGCGTCATCTCTTGAGACCAAGGGATCACGACATGAGGATGCTTTCGCCCTATTGAGCAAGGGATCTCTGAGCTTACCAGAATCTATTTTCTCTTGAGAACATGCAAGCTTACTAACAGATCGAATCTTTTCAGCTTTTTCCCTATTGAGGAGAACTGCTTATAGGAGATCTTAACTCTACATCTGAGGTACCTCGGCCG

>FLP2E11 (256 bp)
CCGGGCAGGT CTAATCGTCACTTCTATCACTTACTTCTCTCAGGTGTATTTCCCTCTGAGATGATACACAAGGTTTCTGAGAATGCAATGAGGTTTTCCGATTTCAGTTCTCCATCCAGAGTTCTACGCGCGGGGGAGATGCAAGGAGGCTTACTGAGCACGTAGGAGACCTCGGCCG
APPENDIX 12

DNA sequence alignment for three clones derived from four strawberry parental genotypes

FLP2E11 primer binding sites

Forward primer: 5’-CTAGTTGGCTCTCTTACCTATC-3’
Reverse primer: 5’-GC CACCAAGTAGAGGAAG-3’

> 05-069-63

<table>
<thead>
<tr>
<th>Clone</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>63_1</td>
<td>CTAGTTGGCTCTCTTACCTATCCTTAGCTTTGCTCCGGAGTTGTCAGGATACATAAT</td>
</tr>
<tr>
<td>63_2</td>
<td>CTAGTTGGCTCTCTTACCTATCCTTAGCTTTGCTCCGGAGTTGTCAGGATACATAAT</td>
</tr>
<tr>
<td>63_3</td>
<td>CTAGTTGGCTCTCTTACCTATCCTTAGCTTTGCTCCGGAGTTGTCAGGATACATAAT</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Clone</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>63_1</td>
<td>GTGTTCATGTATCCATCTTCGGCAAGTGATTATTGCTCTCTCTCTGAAAG</td>
</tr>
<tr>
<td>63_2</td>
<td>GTGTTCATGTATCCATCTTCGGCAAGTGATTATTGCTCTCTCTCTGAAAG</td>
</tr>
<tr>
<td>63_3</td>
<td>GTGTTCATGTATCCATCTTCGGCAAGTGATTATTGCTCTCTCTCTGAAAG</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Clone</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>63_1</td>
<td>CAGCGATTAGCTGAGAGACTATTGTCCCTAGAAGGGAGACGTTTCAGATTCTAGA</td>
</tr>
<tr>
<td>63_2</td>
<td>CAGCGATTAGCTGAGAGACTATTGTCCCTAGAAGGGAGACGTTTCAGATTCTAGA</td>
</tr>
<tr>
<td>63_3</td>
<td>CAGCGATTAGCTGAGAGACTATTGTCCCTAGAAGGGAGACGTTTCAGATTCTAGA</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Clone</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>63_1</td>
<td>CGAAGGAGTGATGCTTAGAGGAGGATTTGTCCCTAGAAGGGAGACGTTTCAGATTCTAGA</td>
</tr>
<tr>
<td>63_2</td>
<td>CGAAGGAGTGATGCTTAGAGGAGGATTTGTCCCTAGAAGGGAGACGTTTCAGATTCTAGA</td>
</tr>
<tr>
<td>63_3</td>
<td>CGAAGGAGTGATGCTTAGAGGAGGATTTGTCCCTAGAAGGGAGACGTTTCAGATTCTAGA</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Clone</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>63_1</td>
<td>CACCAAGTAGAGGAAG</td>
</tr>
<tr>
<td>63_2</td>
<td>CACCAAGTAGAGGAAG</td>
</tr>
<tr>
<td>63_3</td>
<td>CACCAAGTAGAGGAAG</td>
</tr>
</tbody>
</table>
> Juliette

Ju_1  CAGTTGGCTCTCTTACCTATCGGTCTGGCCCTCGGAGTTTTGTTCCAGGATACATAAT
Ju_2  CAGTTGGCTCTCTTACCTATCGGTCTGGCCCTCGGAGTTTTGTTCCAGGATACATAAT
Ju_3  CAGTTGGCTCTCTTACCTATCGGTCTGGCCCTCGGAGTTTTGTTCCAGGATACATAAT

Ju_1  GTGTTTCATGTATCCATGCTTCGCAAGTATATTGC
Ju_2  GTGTTTCATGTATCCATGCTTCGCAAGTATATTGC
Ju_3  GTGTTTCATGTATCCATGCTTCGCAAGTATATTGC

Ju_1  CAGCCGATT
Ju_2  CAGCCGATT
Ju_3  CAGCCGATT

Ju_1  CG
Ju_2  CG
Ju_3  CG

Ju_1  CACCAAGTAGAGGAAG
Ju_2  CACCAAGTAGAGGAAG
Ju_3  CACCAAGTAGAGGAAG

Ju_1  CACCAAGTAGAGGAAG
Ju_2  CACCAAGTAGAGGAAG
Ju_3  CACCAAGTAGAGGAAG

*******************
# APPENDIX 13

Full DNA sequence alignment for all clones and parental genotypes compared to the original FLP2E11 sequence

<table>
<thead>
<tr>
<th>Clone</th>
<th>Sequence Alignment</th>
<th>Comparison</th>
</tr>
</thead>
<tbody>
<tr>
<td>FLP2E11</td>
<td>CTAGTTGGCTCTCTTCACCTATCGGTTAGTTTTTCCTCCGGAGTTGTCTAGGATACATAAT</td>
<td>60</td>
</tr>
<tr>
<td>63_1</td>
<td>CTAGTTGGCTCTCTTCACCTATCGGTTAGCTTTGCCTCCGGAGTTGTCCAGGATACATAAT</td>
<td>60</td>
</tr>
<tr>
<td>63_2</td>
<td>CTAGTTGGCTCTCTTCACCTATCGGTTAGCTTTGCCTCCGGAGTTGTCCAGGATACATAAT</td>
<td>60</td>
</tr>
<tr>
<td>63_3</td>
<td>CTAGTTGGCTCTCTTCACCTATCGGTTAGTTTTGCCTCCGGAGTTGTCCAGGATACATAAT</td>
<td>60</td>
</tr>
<tr>
<td>194_1</td>
<td>CTAGTTGGCTCTCTTCACCTATCGGTTAGCTTTGCCTCCGGAGTTGTCCAGGATACATAAT</td>
<td>60</td>
</tr>
<tr>
<td>194_2</td>
<td>CTAGTTGGCTCTCTTCACCTATCGGTTAGCTTTGCCTCCGGAGTTGTCCAGGATACATAAT</td>
<td>60</td>
</tr>
<tr>
<td>194_3</td>
<td>CTAGTTGGCTCTCTTCACCTATCGGTTAGCTTTGCCTCCGGAGTTGTCCAGGATACATAAT</td>
<td>60</td>
</tr>
<tr>
<td>311_1</td>
<td>CTAGTTGGCTCTCTTCACCTATCGGTTAGCTTTGCCTCCGGAGTTGTCCAGGATACATAAT</td>
<td>60</td>
</tr>
<tr>
<td>311_2</td>
<td>CTAGTTGGCTCTCTTCACCTATCGGTTAGCTTTGCCTCCGGAGTTGTCCAGGATACATAAT</td>
<td>60</td>
</tr>
<tr>
<td>311_3</td>
<td>CTAGTTGGCTCTCTTCACCTATCGGTTAGCTTTGCCTCCGGAGTTGTCCAGGATACATAAT</td>
<td>60</td>
</tr>
<tr>
<td>Ju_1</td>
<td>CTAGTTGGCTCTCTTCACCTATCGGTTAGCTTTGCCTCCGGAGTTGTCCAGGATACATAAT</td>
<td>60</td>
</tr>
<tr>
<td>Ju_2</td>
<td>CTAGTTGGCTCTCTTCACCTATCGGTTAGCTTTGCCTCCGGAGTTGTCCAGGATACATAAT</td>
<td>60</td>
</tr>
<tr>
<td>Ju_3</td>
<td>CTAGTTGGCTCTCTTCACCTATCGGTTAGCTTTGCCTCCGGAGTTGTCCAGGATACATAAT</td>
<td>60</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Clone</th>
<th>Sequence Alignment</th>
<th>Comparison</th>
</tr>
</thead>
<tbody>
<tr>
<td>FLP2E11</td>
<td>GTGTTTCATGTATCCATGCTTCGCAAGTATATTGCCAATCCTTCTCATGTGCTTGTTGAG</td>
<td>120</td>
</tr>
<tr>
<td>63_1</td>
<td>GTGTTTCATGTATCCATGCTTCGCAAGTATATTGCCGATCCTTCTCATGTGCTGGAAGAG</td>
<td>120</td>
</tr>
<tr>
<td>63_2</td>
<td>GTGTTTCATGTATCCATGCTTCGCAAGTATATTGCCGATCCTTCTCATGTGCTGGAAGAG</td>
<td>120</td>
</tr>
<tr>
<td>63_3</td>
<td>GTGTTTCATGTATCCATGCTTCGCAAGTATATTGCCGATCCTTCTCATGTGCTGGAAGAG</td>
<td>120</td>
</tr>
<tr>
<td>194_1</td>
<td>GTGTTTCATGTATCCATGCTTCGCAAGTATATTGCCGATCCTTCTCATGTGCTGGAAGAG</td>
<td>120</td>
</tr>
<tr>
<td>194_2</td>
<td>GTGTTTCATGTATCCATGCTTCGCAAGTATATTGCCGATCCTTCTCATGTGCTGGAAGAG</td>
<td>120</td>
</tr>
<tr>
<td>194_3</td>
<td>GTGTTTCATGTATCCATGCTTCGCAAGTATATTGCCGATCCTTCTCATGTGCTGGAAGAG</td>
<td>120</td>
</tr>
<tr>
<td>311_1</td>
<td>GTGTTTCATGTATCCATGCTTCGCAAGTATATTGCCGATCCTTCTCATGTGCTGGAAGAG</td>
<td>120</td>
</tr>
<tr>
<td>311_2</td>
<td>GTGTTTCATGTATCCATGCTTCGCAAGTATATTGCCGATCCTTCTCATGTGCTGGAAGAG</td>
<td>120</td>
</tr>
<tr>
<td>311_3</td>
<td>GTGTTTCATGTATCCATGCTTCGCAAGTATATTGCCGATCCTTCTCATGTGCTGGAAGAG</td>
<td>120</td>
</tr>
<tr>
<td>Ju_1</td>
<td>GTGTTTCATGTATCCATGCTTCGCAAGTATATTGCCGATCCTTCTCATGTGCTGGAAGAG</td>
<td>120</td>
</tr>
<tr>
<td>Ju_2</td>
<td>GTGTTTCATGTATCCATGCTTCGCAAGTATATTGCCGATCCTTCTCATGTGCTGGAAGAG</td>
<td>120</td>
</tr>
<tr>
<td>Ju_3</td>
<td>GTGTTTCATGTATCCATGCTTCGCAAGTATATTGCCGATCCTTCTCATGTGCTGGAAGAG</td>
<td>120</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Clone</th>
<th>Sequence Alignment</th>
<th>Comparison</th>
</tr>
</thead>
<tbody>
<tr>
<td>FLP2E11</td>
<td>CAGCCGATTAGTTTGCAAGAGGAGTCGGTTCAGATTCTCTTGC</td>
<td>180</td>
</tr>
<tr>
<td>63_1</td>
<td>CAGCCGATTAGTTTGCAAGAGGAGTCGGTTCAGATTCTCTTGC</td>
<td>180</td>
</tr>
<tr>
<td>63_2</td>
<td>CAGCCGATTAGTTTGCAAGAGGAGTCGGTTCAGATTCTCTTGC</td>
<td>180</td>
</tr>
<tr>
<td>63_3</td>
<td>CAGCCGATTAGTTTGCAAGAGGAGTCGGTTCAGATTCTCTTGC</td>
<td>180</td>
</tr>
<tr>
<td>194_1</td>
<td>CAGCCGATTAGTTTGCAAGAGGAGTCGGTTCAGATTCTCTTGC</td>
<td>180</td>
</tr>
<tr>
<td>194_2</td>
<td>CAGCCGATTAGTTTGCAAGAGGAGTCGGTTCAGATTCTCTTGC</td>
<td>180</td>
</tr>
<tr>
<td>194_3</td>
<td>CAGCCGATTAGTTTGCAAGAGGAGTCGGTTCAGATTCTCTTGC</td>
<td>180</td>
</tr>
<tr>
<td>311_1</td>
<td>CAGCCGATTAGTTTGCAAGAGGAGTCGGTTCAGATTCTCTTGC</td>
<td>180</td>
</tr>
<tr>
<td>311_2</td>
<td>CAGCCGATTAGTTTGCAAGAGGAGTCGGTTCAGATTCTCTTGC</td>
<td>180</td>
</tr>
<tr>
<td>311_3</td>
<td>CAGCCGATTAGTTTGCAAGAGGAGTCGGTTCAGATTCTCTTGC</td>
<td>180</td>
</tr>
<tr>
<td>Ju_1</td>
<td>CAGCCGATTAGTTTGCAAGAGGAGTCGGTTCAGATTCTCTTGC</td>
<td>180</td>
</tr>
<tr>
<td>Ju_2</td>
<td>CAGCCGATTAGTTTGCAAGAGGAGTCGGTTCAGATTCTCTTGC</td>
<td>180</td>
</tr>
<tr>
<td>Ju_3</td>
<td>CAGCCGATTAGTTTGCAAGAGGAGTCGGTTCAGATTCTCTTGC</td>
<td>180</td>
</tr>
</tbody>
</table>