Genetic Analysis of Reproductive and Nut Traits in Almond [*Prunus dulcis* (Mill.) D.A. Webb]

A thesis submitted to the University of Adelaide in fulfilment of the requirements for the degree of the Doctor of Philosophy

By

Wasala Adikari Shashiprabha Nilupuli Sridevi Tennakoon Goonetilleke

School of Agriculture, Food and Wine

Faculty of Sciences

The University of Adelaide

December 2016
TABLE OF CONTENTS

TABLE OF CONTENTS .................................................................................................................................................. i
ABSTRACT ....................................................................................................................................................................... vii
THESIS DECLARATION ..................................................................................................................................................... ix
ACKNOWLEDGEMENTS ...................................................................................................................................................... x
LIST OF ABBREVIATIONS ............................................................................................................................................. xii
LIST OF TABLES .............................................................................................................................................................. xvi
LIST OF FIGURES ............................................................................................................................................................ xviii
LIST OF APPENDICES ....................................................................................................................................................... xxiv

CHAPTER 1: Introduction .................................................................................................................................................. 1

CHAPTER 2: Literature review ......................................................................................................................................... 7
  2.1 Almond (Prunus dulcis) ......................................................................................................................................... 7
  2.2 Almond reproduction ............................................................................................................................................. 8
    2.2.1 Self-incompatibility and the almond S locus ................................................................................................. 8
      2.2.1.2 The almond S-RNase gene ...................................................................................................................... 10
      2.2.1.3 The almond SFB gene ............................................................................................................................ 13
      2.2.1.4 The long terminal repeat (LTR) retrotransposons ................................................................................. 15
    2.2.2 Self-fertility in almond ....................................................................................................................................... 16
      2.2.2.1 Deduced amino acid sequences and structure of the S-RNase ............................................................. 16
      2.2.2.2 Dual expression of S-RNase and epigenetic variation in almond ......................................................... 17
      2.2.2.3 Self-compatibility due to stylar part and pollen part mutations ......................................................... 17
      2.2.2.4 Other proteins involved in the self-incompatibility mechanism in almond ....................................... 18
    2.2.3 Characterisation of S locus alleles in almond ................................................................................................. 20
  2.3 Genetic marker discovery and construction of linkage maps in almond ............................................................ 23
    2.3.1 Application of molecular markers in almond improvement ........................................................................ 23
    2.3.2 Almond linkage maps ..................................................................................................................................... 23
    2.3.3 Next-generation sequencing .......................................................................................................................... 26
2.4 Nut traits in almond

2.4.1 Kernel sweetness/bitterness

2.4.2 Shell hardness

2.4.3 Other physical nut traits in almond

2.4.3.1 Geometric mean diameter ($D_p$), sperical index ($\phi$), kernel size and shape

2.4.4 Chemical properties of almond kernels

2.4.4.1 Vitamin E content

2.4.4.1.1 High performance liquid chromatography

2.4.4.2 Lipid components in almond

2.4.4.2.1 Gas chromatography

2.5 Research questions

2.6 Research goals

CHAPTER 3: Resequencing of the almond S locus from self-fertile and self-incompatible genotypes

3.1 Statement of Authorship

3.2 Abstract

3.2.1 Background

3.2.2 Results

3.2.3 Conclusions

3.2.4 Keywords

3.3 Introduction

3.4 Materials and methods

3.4.1 Plant materials and DNA extraction

3.4.2 Primer design and a suitable DNA polymerase to obtain large amplicons

3.4.3 Primer testing and amplification of the S locus

3.4.4 Library preparation and sequencing

3.4.5 Sequence data analysis

3.4.6 Intron–exon structure of the S-RNase gene

3.4.7 Distribution of LTR retrotransposons in the S locus
3.4.8 Phylogenetic relationships among the S-RNase and SFB alleles ........................................... 55
3.5 Results ................................................................................................................................................. 55
  3.5.1 Enzyme and buffer combination suitable for the S locus amplification ................................. 55
  3.5.2 Primer testing and PCR amplification of the S locus ............................................................... 56
  3.5.3 Library preparation and sequencing ............................................................................................ 56
    3.5.3.1 Strong and weak DNA bulks ................................................................................................. 56
    3.5.3.2 Sequence data analysis ........................................................................................................... 56
    3.5.3.3 Sequence variation and gene organisation in the S locus ....................................................... 62
  3.5.4 Intron-exon structure of the SLF, S-RNase and SFB genes ....................................................... 71
  3.5.5 Distribution of LTR retrotransposons in the S locus ................................................................. 71
  3.5.6 New sequence information .......................................................................................................... 73
    3.5.6.1 Strong and weak DNA bulks ................................................................................................. 73
    3.5.6.2 Sequence data analysis ........................................................................................................... 73
    3.5.6.3 Sequence variation and gene organisation in the S locus ....................................................... 78
  3.5.7 Phylogenetic analysis of the almond S locus .............................................................................. 74
3.6 Discussion ............................................................................................................................................. 77

CHAPTER 4: Marker design for the multi-allelic gametophytic self-incompatibility locus of almond ........................................................................................................................................ 83
  4.1 Statement of Authorship .................................................................................................................. 84
  4.2 Abstract ............................................................................................................................................. 86
  4.3 Introduction ....................................................................................................................................... 86
  4.4 Materials and methods .................................................................................................................... 88
    4.4.1 Plant materials and DNA extraction ........................................................................................... 88
    4.4.2 S allele sequencing and sequence data analysis ...................................................................... 90
    4.4.3 Primer design for S allele detection ........................................................................................... 91
    4.4.4 SNP genotyping ......................................................................................................................... 91
    4.4.5 Assessment of self-fertility ...................................................................................................... 92
  4.5 Results .............................................................................................................................................. 92
    4.5.1 The S-RNase and SFB allele sequences ................................................................................... 92
    4.5.2 Allele-specific primers to detect the S1, S3, S5, S7, S8, S9, S23 and S25 alleles of the S-
        RNase gene ..................................................................................................................................... 93
    4.5.3 Allele-specific primers to detect the S1 allele of the S-RNase gene ........................................ 93
Section 5.2: Construction of linkage maps for almond using four populations with a common parent

5.2.1 Introduction .................................................................................................................. 149
5.2.2 Materials and methods ............................................................................................. 151
  5.2.2.1 Plant materials ....................................................................................................... 151
  5.2.2.2 DNA extraction .................................................................................................... 151
  5.2.2.3 Polymorphic assay selection and population screen ............................................. 151
  5.2.2.4 Linkage maps for Nonpareil, Constantí, Tarraco and Vairo ................................ 151
  5.2.2.5 A composite map for Nonpareil ......................................................................... 152
  5.2.2.6 Marker order conservation within linkage groups of Nonpareil .......................... 152
5.2.3 Results ......................................................................................................................... 153
  5.2.3.1 Polymorphic marker detection and population screen ........................................ 153
  5.2.3.2 Linkage maps ....................................................................................................... 153
  5.2.3.3 Composite linkage map for Nonpareil ................................................................. 154
5.2.4 Discussion .................................................................................................................. 163

Section 5.3: Phenotyping and quantitative trait loci detection for nut and kernel traits in almond

5.3.1 Introduction ................................................................................................................. 165
5.3.2 Phenotypic evaluation ............................................................................................... 166
  5.3.2.1 Phenotypic evaluation of physical traits .............................................................. 167
  5.3.2.2 Phenotypic evaluation of almond chemical traits ................................................. 168
    5.3.2.2.1 Tocopherol extraction from almond kernel ...................................................... 168
    5.3.2.2.2 Tocopherol determination ............................................................................. 169
    5.3.2.2.3 Calibration curve preparation ....................................................................... 170
    5.3.2.2.4 Fatty acid determination .............................................................................. 171
  5.3.3 Statistical analysis ..................................................................................................... 171
  5.3.4 Quantitative trait loci detection ............................................................................... 172
  5.3.5 Results ...................................................................................................................... 172
    5.3.5.1 Trait means, heritability and correlation between kernel and nut physical traits... 172
5.3.5.2 Tocopherols and fatty acids in Nonpareil × Lauranne ........................................ 184
5.3.6 Quantitative trait loci ............................................................................................... 186
5.3.7 Discussion .................................................................................................................. 201
CHAPTER 6: General discussion ......................................................................................... 209
CHAPTER 7: Contributions to knowledge ........................................................................... 222
REFERENCES ...................................................................................................................... 225
APPENDIX 1: Supplementary materials of Chapter 3 ......................................................... 251
APPENDIX 2: Supplementary materials of Chapter 4 ......................................................... 254
APPENDIX 3: Supplementary materials of Chapter 5 ......................................................... 259
APPENDIX 4: A year in the life of an almond tree ................................................................. 315
S4.1 Dormancy .................................................................................................................... 315
S4.2 Bloom .......................................................................................................................... 316
S4.3 Full bloom ................................................................................................................... 317
S4.4 Pollination .................................................................................................................... 318
S4.5 Petal fall ....................................................................................................................... 318
S4.6 Post-petal fall ............................................................................................................... 318
S4.7 Nut growth and maturing ............................................................................................ 319
S4.8 Harvest ......................................................................................................................... 320
S4.9 Processing and storage ............................................................................................... 320
Almond is a perennial tree crop with a gametophytic self-incompatibility (SI) system. The SI system of almond is controlled by a multi-allelic locus, S, which is about 70,000 bp long. A nearly complete sequence for the entire S locus sequence has been available only for the S7 haplotype. In this research, next-generation sequencing technology was implemented to sequence the entire S locus simultaneously from 15 haplotypes. The results confirmed the accuracy of available S7 haplotype sequence, generated the entire S locus sequences for the S1, S7 and S6 haplotypes and generated partial S locus sequences for 11 other haplotypes (S3, S5, S8, S9, S13, S14, S19, S22, S23, S25 and S27). Comparisons among haplotype sequences revealed higher polymorphism in the region where the S-RNase and SFB genes are located and considerable differences in the number and locations of long terminal repeat retrotransposons.

There are about 50 known S alleles, of which one confers self-fertility. For some of these, complete or partial S-RNase and SFB sequences are available. Here, more complete sequences were generated for several alleles of the S-RNase gene (S3, S6, S9, S13, S19, S22 and S25) and the SFB gene (S9, S23 and S27).

In almond breeding, SI limits the parental combinations that can be used for crossing. Detection of S alleles prior to crossing would be beneficial. Until now, molecular detection of the S alleles has relied on detection of length polymorphisms in the S-RNase gene. Here, single nucleotide polymorphisms (SNPs) in the S-RNase and SFB genes were used in designing assays to distinguish among S alleles.

This thesis also reports on the construction of linkage maps for Nonpareil and Lauranne based on genotyping-by-sequencing (GBS) and on the design of uniplex assays for detection of SNPs.
detected by GBS. These assays were applied to additional Nonpareil × Lauranne progeny and to progeny from three other Nonpareil crosses (Nonpareil × Constantí, Nonpareil × Tarraco and Nonpareil × Vairo). Data from all four populations were used to generate a composite map for Nonpareil. Comparisons of marker positions detected for Nonpareil and Lauranne with positions in the peach genome confirmed high collinearity between the almond and peach genomes.

Quantitative trait loci analysis detected 23 genomic regions as affecting nut and/or kernel traits in Nonpareil × Lauranne. Nine and 14 QTLs were detected for Nonpareil and Lauranne, respectively. Of the kernel and nut traits mapped here, shell weight, kernel shape, tocopherol concentration, fatty acid concentration and oleic/linoleic ratio were mapped for the first time in almond. For shell hardness and oleic/linoleic ratio, markers were identified that could be useful for marker-assisted selection. Some of the QTLs related to fatty acid and tocopherol concentration were closely located to the genes that are known to be involved in the synthesis of fatty acids and/or tocopherols. Some of the sequence information generated here may be useful for designing primers to amplify these genes (or components of these genes) for resequencing from multiple almond genotypes.
THESIS DECLARATION

I certify that this work contains no material which has been accepted for the award of any other degree or diploma in my name, in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text. In addition, I certify that no part of this work will, in the future, be used in a submission in my name, for any other degree or diploma in any university or other tertiary institution without the prior approval of the University of Adelaide and where applicable, any partner institution responsible for the joint-award of this degree.

I give consent to this copy of my thesis when deposited in the University Library, being made available for loan and photocopying, subject to the provisions of the Copyright Act 1968.

The author acknowledges that copyright of published works contained within this thesis resides with the copyright holder(s) of those works.

I also give permission for the digital version of my thesis to be made available on the web, via the University’s digital research repository, the Library Search and also through web search engines, unless permission has been granted by the University to restrict access for a period of time.

..................................................  ..................................................
Signature                                           Date
ACKNOWLEDGEMENTS

I extend my sincere gratitude to my supervisors Prof. Diane E. Mather, Dr. Michelle G. Wirthensohn and Dr. Timothy J. March, for their guidance, support and encouragement throughout my PhD candidature. I am indeed grateful for their advice, thought-provoking discussions and ‘open door’ policy that allowed me to meet them whenever I needed guidance and help.

I am forever indebted to Prof. Diane Mather for her comments and suggestions on this thesis and on manuscripts, for helping me with improving scientific writing and for providing me feedback on drafts very quickly despite her busy schedule.

I appreciate Dr. Michelle Wirthensohn and Dr. Timothy March for their contributions and co-operation towards making this thesis a reality.

I thank my independent advisor Dr. Adam Croxford for his contribution and valuable advice on the almond self-incompatibility locus sequencing and my postgraduate coordinator Prof. Eileen Scott for her support. I also thank Dr. Ying Zhu for helping me with HPLC analysis and Associate Prof. Chris Ford and Dr. Daryl Mares for allowing me to use HPLC instruments in their laboratories.

I thank Dr. Julian Taylor for helping me with R and Dr. Jimmy Breen and Dr. Radoslaw Suchecki for valuable advice on sequence data analysis. A big thank to Mrs. Jana Kolesik for helping me with leaf sample collection and to Mr. Bart Van Gansbeke for assistance in taking photographs.

I thank Dr. Pere Arús, IRTA, Spain, for providing me the Nonpareil sequence and for making my visit to IRTA very enjoyable and useful.
I wish to acknowledge the eResearch computer facility, the University of South Australia and the Phoenix, the University of Adelaide, for providing supercomputer facilities, Horticulture Innovation Australia for research funding, Illumina for a research grant to resequence the almond self-incompatibility locus, the University of Adelaide for the Australian Postgraduate Award, Waite Analytical Services for fatty acid analysis and the Australian Genome Research Facility for sequencing the genotyping-by-sequencing library.

Last but not least, I thank my family for being there for me at all times and friends for their constant support and encouragement.
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AH</td>
<td>amygdalin hydrolase</td>
</tr>
<tr>
<td>ADGH</td>
<td>amygdalin diglucosidase</td>
</tr>
<tr>
<td>BAM</td>
<td>binary alignment/map format</td>
</tr>
<tr>
<td>Bp</td>
<td>base pair</td>
</tr>
<tr>
<td>BWA</td>
<td>Burrows Wheeler Alignment</td>
</tr>
<tr>
<td>C</td>
<td>conserved region</td>
</tr>
<tr>
<td>Ca²⁺</td>
<td>calcium ion</td>
</tr>
<tr>
<td>CDS</td>
<td>coding sequences</td>
</tr>
<tr>
<td>CIG</td>
<td>cross incompatibility group</td>
</tr>
<tr>
<td>CIGs</td>
<td>cross incompatibility groups</td>
</tr>
<tr>
<td>CO₂</td>
<td>carbon dioxide</td>
</tr>
<tr>
<td>cv.</td>
<td>cultivar</td>
</tr>
<tr>
<td>CYP</td>
<td>cytochrome P450 monooxygenase</td>
</tr>
<tr>
<td>DdRAD</td>
<td>double digest restriction site associated DNA</td>
</tr>
<tr>
<td>DMGGBQ</td>
<td>2,3-dimethyl-5-geranylgeranyl-1,4-benzoquinone</td>
</tr>
<tr>
<td>DMPBQ</td>
<td>2,3-dimethyl-5-phytyl-1,4-benzoquinone</td>
</tr>
<tr>
<td>DNA</td>
<td>deoxyribonucleic Acid</td>
</tr>
<tr>
<td>EMBL</td>
<td>European Molecular Biology Laboratory</td>
</tr>
<tr>
<td>F₁</td>
<td>filial 1 generation</td>
</tr>
<tr>
<td>FA</td>
<td>fatty acid</td>
</tr>
<tr>
<td>G</td>
<td>gram</td>
</tr>
<tr>
<td>Gb</td>
<td>gigabit</td>
</tr>
<tr>
<td>GBS</td>
<td>genotyping-by-sequencing</td>
</tr>
<tr>
<td>GC</td>
<td>gas chromatography</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>GDR</td>
<td>Genome database for Rosaceae</td>
</tr>
<tr>
<td>GT</td>
<td>glucosyltransferase</td>
</tr>
<tr>
<td>GSTs</td>
<td>glutathione S-transferases</td>
</tr>
<tr>
<td>H</td>
<td>hydrogen</td>
</tr>
<tr>
<td>HGA</td>
<td>homogentisic acid</td>
</tr>
<tr>
<td>HPLC</td>
<td>high performance liquid chromatography</td>
</tr>
<tr>
<td>HPPD</td>
<td>p-hydroxyphenylpyruvic acid dioxygenase</td>
</tr>
<tr>
<td>RHV</td>
<td>hypervariable region</td>
</tr>
<tr>
<td>HV</td>
<td>variable region</td>
</tr>
<tr>
<td>IGV</td>
<td>integrative genomics viewer</td>
</tr>
<tr>
<td>IN</td>
<td>integrase</td>
</tr>
<tr>
<td>ISSR</td>
<td>inter simple sequence repeat</td>
</tr>
<tr>
<td>ISW</td>
<td>in-shell weight</td>
</tr>
<tr>
<td>KASP™</td>
<td>competitive allele-specific primer</td>
</tr>
<tr>
<td>Kb</td>
<td>kilo base</td>
</tr>
<tr>
<td>KS</td>
<td>kernel size</td>
</tr>
<tr>
<td>L</td>
<td>linoleic acid</td>
</tr>
<tr>
<td>LDL</td>
<td>low density lipoprotein</td>
</tr>
<tr>
<td>LG</td>
<td>linkage group</td>
</tr>
<tr>
<td>LINEs</td>
<td>long interspersed nuclear elements</td>
</tr>
<tr>
<td>LOD</td>
<td>likelihood of odds</td>
</tr>
<tr>
<td>LTRs</td>
<td>long terminal repeats</td>
</tr>
<tr>
<td>Mb</td>
<td>mega bases</td>
</tr>
<tr>
<td>MDL</td>
<td>mandelonitrile</td>
</tr>
<tr>
<td>Me</td>
<td>methyl</td>
</tr>
<tr>
<td>MGGBQ</td>
<td>2-methyl-6-geranylgeranylplastoquinol</td>
</tr>
<tr>
<td>MIRA</td>
<td>Mimicking Intelligent Read Assembler</td>
</tr>
</tbody>
</table>
MITEs: miniature inverted-repeat transposable elements
MPBQ: 2-methyl-6-phytylplastoquinol
MPBQ MT: 2-methyl-6-phytylplastoquinol methyltransferase
NAM: nested association mapping
NCBI: National Centre for Biotechnology Information
NGS: next-generation sequencing
O: oleic acid
ORF: open reading frame
PCR: polymerase chain reaction
PDP: phytol diphosphate
PH: prunasin hydrolase
PPM: pollen part mutation
PR: protease
QTL: quantitative trait locus
QTLs: quantitative trait loci
R: retrotransposons
RAD: restriction site associated DNA
RAPD: randomly amplified polymorphic DNA
Res: restriction enzymes
RFLP: restriction fragment length polymorphism
RH: RNase H
RT: reverse transcriptase
SAM: sequence alignment/map format
SAM: S-adenosyl methionine
S locus: self-incompatibility locus
SCAR: sequence characterised amplified region
Sf: self-fertility
<table>
<thead>
<tr>
<th>Acronym</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>SFB</td>
<td>S haplotype-specific F-box</td>
</tr>
<tr>
<td>SFB</td>
<td>S haplotype-specific F-box gene</td>
</tr>
<tr>
<td>SH</td>
<td>shell hardness</td>
</tr>
<tr>
<td>SI</td>
<td>self-incompatibility</td>
</tr>
<tr>
<td>SINEs</td>
<td>short interspersed nuclear elements</td>
</tr>
<tr>
<td>SLF</td>
<td>S locus F-box</td>
</tr>
<tr>
<td>SLF</td>
<td>S locus F-box gene</td>
</tr>
<tr>
<td>SNP</td>
<td>single nucleotide polymorphism</td>
</tr>
<tr>
<td>SNPs</td>
<td>single nucleotide polymorphisms</td>
</tr>
<tr>
<td>SPM</td>
<td>stylar part mutation</td>
</tr>
<tr>
<td>S-RNASE</td>
<td>stylar-RNase</td>
</tr>
<tr>
<td>S-RNASE</td>
<td>stylar-RNase gene</td>
</tr>
<tr>
<td>SSR</td>
<td>simple sequence repeat</td>
</tr>
<tr>
<td>SW</td>
<td>shell weight</td>
</tr>
<tr>
<td>TE</td>
<td>transposable element</td>
</tr>
<tr>
<td>TIR</td>
<td>terminal inverted repeats</td>
</tr>
<tr>
<td>TMT</td>
<td>tocopherol methyltransferase</td>
</tr>
<tr>
<td>VCF</td>
<td>variant call format</td>
</tr>
<tr>
<td>VITE</td>
<td>genes for vitamin E biosynthesis</td>
</tr>
</tbody>
</table>
# LIST OF TABLES

Table 2.1. Cross-incompatibility groups in almond .................................................. 21
Table 2.2. Shell hardness groups in almond ............................................................ 32
Table 2.3. Tocopherol and tocotrienol chemical compounds in vitamin E (Me: methyl group and H: hydrogen) .................................................................................. 35
Table 3.1. Almond cultivars and breeding selections used in this analysis ................. 48
Table 3.2. Percentage of DNA sequence identity among almond S haplotypes using the entire S locus sequences ................................................................. 62
Table 3.3. Percentage of nucleotide identity of the SLF gene in almond S alleles ........ 64
Table 3.4. Percentage of nucleotide identity of the S-RNase gene in almond S alleles .......... 67
Table 3.5. Percentage of nucleotide identity in the almond SFB alleles .......................... 69
Table 3.6. Percentage of nucleotide identity in the region between the S-RNase and SFB genes in almond S haplotypes ........................................................................... 72
Table 4.1. Populations used in this analysis ................................................................. 89
Table 4.2. Sets of PCR primers designed to provide KASP assays that distinguish among nine S alleles in almond, showing the fluorescence (FAM or HEX) emitted for each of the nine S alleles .................................................................................................................. 100
Table 4.3. Summary of results obtained from assessment of each of the 17 KASP assays on F1 progeny from two crosses, each showing the numbers of progeny for which HEX fluorescence, FAM fluorescence or both (HEX:FAM) were emitted ............................................ 104
Table 4.4. Fruit set evaluation for the progeny from the University Adelaide almond breeding program. Assay, population screened, fruit set percentage and mean fruit set percentage are shown ...... 106
Table 5.1.1. Total number of unique tags mapped to the peach genome sequence assembly using the data from the initial GBS library ................................................................. 124
Table 5.1.2. Sorted tag pairs for the almond genome using data from the initial GBS library ........ 126
Table 5.3.1. Means and heritability of physical traits of nuts and kernels assessed on nuts harvested in 2003 from 89 Nonpareil × Lauranne F1 progeny.

Table 5.3.2. Means of physical traits of nuts and kernels assessed on nuts harvested in 2015 from 95 Nonpareil × Constantí, 127 Nonpareil × Tarraco and 90 Nonpareil × Vairo F1 progeny.

Table 5.3.3. Pair-wise correlation coefficients for almond nut and kernel traits for Nonpareil × Lauranne F1 progeny in 2003.

Table 5.3.4. Means of fatty acids assessed on kernels from nuts harvested in 2015 from 180 Nonpareil × Lauranne F1 progeny.

Table 5.3.5. Summary of QTLs detected for physical nut and kernel traits in Nonpareil.

Table 5.3.6. Summary of QTLs detected for Lauranne physical nut and kernel traits.

Table 5.3.7. Summary of QTLs detected for chemical traits in Nonpareil and in Lauranne.

Table 5.3.8. Summary of QTLs detected in Constantí and Tarraco maps for the year 2015.

Table S3.1. Primer sequences used for the amplification of the GBS library prior to Illumina sequencing.

Table S3.2. Allele-specific and common primer sequences of SNP-based assays.

Table S3.3. SNP-bearing sequences from Nonpareil that were used in comparative mapping with the peach sequence assembly.

Table S3.4. SNP-bearing sequences from Lauranne that were used in comparative mapping with the peach sequence assembly.
LIST OF FIGURES

Fig. 2.1 A schematic diagram of the $S_7$ haplotype of almond $S$ locus ......................................................... 12
Fig. 2.2 A schematic diagram of the almond $S$-$RNase$ gene ................................................................. 14
Fig. 2.3 A schematic diagram of the almond $SFB$ gene ................................................................. 14
Fig. 2.4 Basic structure of a full-length LTR retrotransposon ................................................................. 15
Fig. 2.5 An overview of the $S$-$RNase$ and $SFB$ gene sequences (200 blast hits) registered in the GenBank NCBI) aligned to the Nonpareil $S_7$ haplotype (AB081587) as the query using the NCBI BLASTN program ................................................................. 22
Fig. 2.6 The metabolic pathways for synthesis and catabolism of cyanogenic glucosides prunasin and amygdalin in almond ................................................................. 30
Fig. 2.7 Almond fruit: hull is attached to the almond nut (a), hull is removed from the shell (b), almond kernel is inside the shell (c) ................................................................. 31
Fig. 2.8 Almonds with different shell hardness groups: paper shelled almond (a), soft shelled almond (b), semi hard shelled almond (c), hard shelled almond (d) and stone shelled almond (e) ................................................................. 31
Fig. 2.9 Chemical structures of tocopherols and tocotrienols ................................................................. 35
Fig. 2.10 The tocopherol biosynthetic pathway in Arabidopsis thaliana ......................................................... 37
Fig. 3.1 Sequence variations observed in 48 almond cultivars and breeding lines used in this research ................................................................. 59
Fig. 3.2 Visualisation of a BAM file resulting from assembling the sequence reads from the almond cultivar, Mira ($S_7S$), using the BWA (Burrows-Wheeler) alignment tool ................................................................. 61
Fig. 3.3 Structure of the almond $S$ locus ................................................................. 63
Fig. 3.4 Alignment of deduced amino acid sequences of 15 $S$ alleles from the $SLF$ gene in almond ................................................................. 65
Fig. 3.5 Alignment of deduced amino acid sequences of 15 $S$ alleles from the $S$-$RNase$ gene in almond ................................................................. 68
Fig. 3.6 Alignment of deduced amino acid sequences of 15 $S$ alleles from the $SFB$ gene in almond ................................................................. 70

xviii
Fig. 3.7 Intron–exon structure of 15 almond S-RNase alleles from the sequences generated in this analysis.................................................................73

Fig. 3.8 Phylogenetic relationships based on 15 S alleles from the S-RNase gene in almond, using the deduced amino acid sequences from conserved region 1 (C1) to conserved region 5 (C5) of the S-RNase gene...............................................................75

Fig. 3.9 Phylogenetic relationships based on 15 almond S-RNase alleles, other Prunus, Malus, Pyrus, Antirrhinum and Solanaceae S-RNases, using deduced amino acid sequences from conserved region 1 (C1) to conserved region 5 (C5) of the S-RNase gene...............................................................76

Fig. 3.10 Phylogenetic relationships based on 11 almond SFB alleles...............................77

Fig. 4.1 Sequence alignment between conserved region 1 (C1) and conserved region 2 (C2) and the intron region 2 of nine S alleles of the S-RNase gene ...........................................................................95

Fig. 4.2 Sequence alignment between conserved region 3 (C3) and conserved region 4 (RC4) of nine S alleles of the S-RNase gene showing the positions at which primers were designed........96

Fig. 4.3 Sequence alignment of nine S alleles between conserved region 1 (C1) and conserved region 2 (C2) of the S-RNase gene...........................................................................................................97

Fig. 4.4 Sequence alignment of seven alleles of the SFB gene, showing the positions at which primers were designed.. ..................................................................................................................98

Fig. 4.5 Results obtained with fluorescence-based S allele markers........................................103

Fig. 5.1.1 Fragment size distributions for in-silico digestion of the peach genomic sequence with the restriction enzymes ApeKI, HpaII, PsiI and combinations of these enzymes........................................122

Fig. 5.1.2 Electrophoresis of GBS libraries resulting from different adapter concentrations ligated with 200 ng of DNA........................................................................................................123

Fig. 5.1.3 The relationship between the number of sequence reads and the number of unique tags obtained for each individual in the Nonpareil × Lauranne GBS library........................................123

Fig. 5.1.4 Comparison of unique tags and positions of SNPs (KASP markers) mapped to the peach genome sequence assembly (Pp).................................................................................................125
Fig. 5.1.5 Comparison of framework linkage maps constructed for Nonpareil linkage groups (NLG) 1 to 4 using genotyping-by-sequencing (GBS) data (TP codes), SSR markers and ISSR markers.

Fig. 5.1.6 Comparison of framework linkage maps constructed for Nonpareil linkage groups (NLG) 5 to 8 using genotyping-by-sequencing (GBS) data (TP codes), SSR markers and ISSR markers.

Fig. 5.1.7 Comparison of framework linkage maps constructed for Lauranne linkage groups (LLG) 1 to 4 using genotyping-by-sequencing (GBS) data (TP codes), SSR markers and ISSR markers.

Fig. 5.1.8 Comparison of framework linkage maps constructed for Lauranne linkage groups (LLG) 5 to 8 using genotyping-by-sequencing (GBS) data (TP code), SSR markers and ISSR markers.

Fig. 5.1.9 Examples of results with primer sets derived from GBS tag sequences: the WriPdK0007 primer set, which assays a SNP within tag TP18674 that is heterozygous (G:C) in Nonpareil and homozygous (C:C) in Lauranne.

Fig. 5.1.10 A linkage map for Nonpareil, constructed using genotypic data from SNP-based marker assays applied to 231 Nonpareil × Lauranne F1 progeny, with eight linkage groups labelled as NLG1 to NLG8.

Fig. 5.1.11 A linkage map for Lauranne, constructed using genotypic data from SNP-based marker assays applied to 231 Nonpareil × Lauranne F1 progeny, with eight linkage groups labelled as LLG1 to LLG8.

Fig. 5.1.12 Linkage maps constructed for the almond linkage group LG3 for Nonpareil (NLG3) and Lauranne (LLG3).

Fig. 5.1.13 Synteny and collinearity between almond genetic maps and the peach genome sequence.

Fig. 5.1.14 Relationships between genetic and physical distances for each linkage group of almond and the peach genome sequence.
Fig. 5.2.1 A schematic diagram showing the other parents with which Nonpareil cultivar has been crossed in the University of Adelaide almond breeding program. ................................................................. 150

Fig. 5.2.2 Venn diagrams showing the number of KASP markers that detected polymorphisms in the populations used in this analysis, for (a) markers that were designed based on Nonpareil heterozygosity and (b) markers that were designed based on Lauranne heterozygosity. .............................. 153

Fig. 5.2.3 A linkage map for Nonpareil, constructed using genotypic data from SNP-based marker assays applied to 349 Nonpareil × Constantí F1 progeny. ....................................................................................... 155

Fig. 5.2.4 A linkage map for Nonpareil, constructed using genotypic data from SNP-based marker assays applied to 207 Nonpareil × Tarraco F1 progeny. ....................................................................................... 156

Fig. 5.2.5 A linkage map for Nonpareil, constructed using genotypic data from SNP-based marker assays applied to 198 Nonpareil × Vairo F1 progeny. ....................................................................................... 157

Fig. 5.2.6 A linkage map for Constantí, constructed using genotypic data from SNP-based marker assays applied to 349 Nonpareil × Constantí F1 progeny. ....................................................................................... 158

Fig. 5.2.7 A linkage map for Tarraco, constructed using genotypic data from SNP-based marker assays applied to 207 Nonpareil × Tarraco F1 progeny. ....................................................................................... 159

Fig. 5.2.8 A linkage map for Vairo, constructed using genotypic data from SNP-based marker assays applied to 198 Nonpareil × Vairo F1 progeny. ....................................................................................... 160

Fig. 5.2.9 A linkage map for Nonpareil, constructed using genotypic data from SNP-based marker assays applied to Nonpareil × Constantí, Nonpareil × Lauranne, Nonpareil × Tarraco and Nonpareil × Vairo F1 progeny. ....................................................................................... 161

Fig. 5.2.10 Comparison of marker order within linkage groups of a composite Nonpareil map and four individual Nonpareil maps. ....................................................................................... 162

Fig. 5.3.1 Histograms depicting the phenotypic distribution of kernel weight and in-shell weight in the progeny of Nonpareil × Lauranne F1 population. ....................................................................................... 177

Fig. 5.3.2 Histograms depicting the phenotypic distribution of shell weight and shell hardness in the progeny of Nonpareil × Lauranne F1 population. ....................................................................................... 179
Fig. 5.3.3 Histograms depicting the phenotypic distribution of kernel weight and in-shell weight in the progeny of Nonpareil × Constantí (N × C), Nonpareil × Lauranne (N × L), Nonpareil × Tarraco (N × T) and Nonpareil × Vairo (N × V) F₁ populations in 2015.

Fig. 5.3.4 Histograms depicting the phenotypic distribution of shell weight and shell hardness in the progeny of Nonpareil × Constantí (N × C), Nonpareil × Lauranne (N × L), Nonpareil × Tarraco (N × T) and Nonpareil × Vairo (N × V) F₁ populations in 2015.

Fig. 5.3.5 Proportions of tocopherols in Nonpareil × Lauranne. Proportions of the tocopherol components (α-, β- and γ-) relative to the total tocopherol concentration in 180 progeny of Nonpareil × Lauranne F₁ population.

Fig. 5.3.6 Proportions of the major fatty acids in Nonpareil × Lauranne. Proportions of oleic acid, linoleic acid, palmitic acid, stearic acid and vaccenic acid relative to the total fatty acid concentration in 180 progeny of Nonpareil × Lauranne F₁ population.

Fig. 5.3.7 A linkage map for Nonpareil, constructed using genotypic data from SNP-based marker assays applied to 231 Nonpareil × Lauranne F₁ progeny.

Fig. 5.3.8 A linkage map for Lauranne, constructed using genotypic data from SNP-based marker assays applied to 231 Nonpareil × Lauranne F₁ progeny.

Fig. 5.3.9 Comparisons of the positions of almond quantitative trait loci detected in Nonpareil (A to I, shaded in dark grey), Lauranne (J to W, shaded in light grey).

Fig. 5.3.10 Shell hardness percentages and their means for groups of Nonpareil × Lauranne F₁ progeny.

Fig. 5.3.11 Shell hardness percentages and their means for groups of Nonpareil × Lauranne F₁ progeny selected (favourable) to have paper shell traits.

Fig. 5.3.12 Genetic maps of linkage group 2 for Constantí (CLG2) and Tarraco (TLG2).

Fig. 5.3.13 Shell hardness percentages and their means for groups of Nonpareil × Constantí F₁ progeny (upper panel) and Nonpareil × Tarraco F₁ progeny (lower panel) defined based on their genotypes at markers in QTL regions for shell hardness detected on LG2.
Fig. 5.3.14 Oleic/Linoleic ratios and their means for groups of Nonpareil × Lauranne F₁ progeny defined based on their genotypes at markers in QTL regions for O/L ratio detected on LG1 (L and K) and LG6 (V). .......................................................................................................................... 201

Fig. 5.3.15 Oleic/Linoleic ratios and their means for groups of Nonpareil × Lauranne F₁ progeny elected (favourable) to have O/L ratio (> 2.5) .......................................................................................................................... 201

Fig. S1.1 Phylogenetic relationships among 15 almond S alleles from the S-RNase gene. .................. 251

Fig. S1.2 Phylogenetic relationships among the S-RNase alleles from Prunus, Malus, Pyrus, Antirrhinum species and Solanaceae .......................................................................................................................... 252

Fig. S1.3 Phylogenetic relationships among the almond SFB alleles using the bootstrap consensus tree. ........................................................................................................................................................................... 253

Fig. S2.1 A heat map showing the DNA level sequence identity of nine S-RNase alleles. ................. 254

Fig. S2.2 A heat map showing the DNA level sequence identity of seven SFB alleles. ....................... 254

Fig. S2.3 The S-RNase gene sequence of the S₃ allele from the almond cultivar, Lauranne. ............. 255

Fig. S2.4 The S-RNase gene sequence of the S₉ allele from the almond cultivar, Vairo..................... 256

Fig. S2.5 The S-RNase gene sequence of the S₂₅ allele from the almond cultivar, Johnston.............. 256

Fig. S2.6 The SFB gene sequence of the S₃ allele from the almond cultivar, Lauranne..................... 257

Fig. S2.7 The SFB gene sequence of the S₂₅ allele from the almond cultivar, Johnston..................... 258

Fig. S4.1 An almond tree in dormancy................................................................................................. 315

Fig. S4.2 Pink buds. An emerging flower bud (a), growing flower buds (b). ..................................... 316

Fig. S4.3 Popcorn stage ....................................................................................................................... 316

Fig. S4.4 A fully opened almond flower at full bloom. ...................................................................... 317

Fig. S4.5 Almond trees – at their full bloom stage. .......................................................................... 317

Fig. S4.6 Pollination in almond.................................................................................................. 318

Fig. S4.7 Flowers at petal fall stage................................................................................................. 318

Fig. S4.8 Flowers in post-petal fall stage........................................................................................ 319

Fig. S4.9 Fruit set in almond............................................................................................................ 319

Fig. S4.10 Nuts are ready to harvest............................................................................................... 320
LIST OF APPENDICES

APPENDIX 1: Supplementary materials of Chapter 3 .................................................. 251
Supplementary information S1.1 ........................................................................ 251
Supplementary information S1.2 ........................................................................ 252
Supplementary information S1.3 ........................................................................ 253
APPENDIX 2: Supplementary materials of Chapter 4 ................................................ 254
Supplementary information S2.1 ........................................................................ 254
Supplementary information S2.2 ........................................................................ 254
Supplementary information S2.3 ........................................................................ 255
Supplementary information S2.4 ........................................................................ 256
Supplementary information S2.5 ........................................................................ 256
Supplementary information S2.6 ........................................................................ 257
Supplementary information S2.7 ........................................................................ 258
APPENDIX 3: Supplementary materials of Chapter 5 .............................................. 259
Supplementary information S3.1 ........................................................................ 259
Supplementary information S3.2 ........................................................................ 260
Supplementary information S3.3 ........................................................................ 302
Supplementary information S3.4 ........................................................................ 309
Appendix 4: A year in the life of an almond tree .................................................... 315