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

ABSTRACT

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 S_7 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 S_7 haplotype sequence, generated the entire *S* locus sequences for the S_f, S_1 and S_8 haplotypes and generated partial *S* locus sequences for 11 other haplotypes (S_3, S_5, S_6, S_9, S_{13}, S_{14}, S_{19}, S_{22}, S_{23}, S_{25} and S_{27}). 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 (S_3, S_6, S_9, S_{13}, S_{19}, S_{22} and S_{25}) and the *SFB* gene (S_9, S_{23} and S_{27}).

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.

LIST OF ABBREVIATIONS

AH : amygdalin hydrolase

ADGH : amygdalin diglucosidase

BAM : binary alignment/map format

Bp : base pair

BWA : Burrows Wheeler Alignment

C : conserved region

Ca^{2+} : calcium ion

CDS : coding sequences

CIG : cross incompatibility group

CIGs : cross incompatibility groups

CO_2 : carbon dioxide

cv. : cultivar

CYP : cytochrome P450 monooxygenase

DdRAD : double digest restriction site associated DNA

DMGGBQ : 2,3-dimethyl -5-geranylgeranyl-I,4-benzoquinone

DMPBQ : 2,3-dimethyl -5-phytyl-I,4-benzoquinone

DNA : deoxyribonucleic Acid

EMBL : European Molecular Biology Laboratory

F_1 : filial 1 generation

FA : fatty acid

G : gram

Gb : gigabit

GBS : genotyping-by- sequencing

GC : gas chromatography

GDR : Genome database for Rosaceae

GT : glucosyltransferase

GSTs : glutathione S-transferases

H : hydrogen

HGA : homogentisic acid

HPLC : high performance liquid chromatography

HPPD : *p*-hydroxyphenylpyruvic acid dioxygenase

RHV : hypervariable region

HV : variable region

IGV : integrative genomics viewer

IN : integrase

ISSR : inter simple sequence repeat

ISW : in-shell weight

KASP™ : competitive allele-specific primer

Kb : kilo base

KS : kernel size

L : linoleic acid

LDL : low density lipoprotein

LG : linkage group

LINEs : long interspersed nuclear elements

LOD : likelihood of odds

LTRs : long terminal repeats

Mb : mega bases

MDL : mandelonitrile

Me : methyl

MGGBQ : 2-methyl-6-geranylgeranylplastoquinol

MIRA : Mimicking Intelligent Read Assembler

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 : phytyl 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

SFB : *S* haplotype-specific F-box

SFB : *S* haplotype-specific F-box gene

SH : shell hardness

SI : self-incompatibility

SINEs : short interspersed nuclear elements

SLF : *S* locus F-box

SLF : *S* locus F-box gene

SNP : single nucleotide polymorphism

SNPs : single nucleotide polymorphisms

SPM : stylar part mutation

S-RNASE : stylar-RNase

S-RNASE : stylar-RNase gene

SSR : simple sequence repeat

SW : shell weight

TE : transposable element

TIR : terminal inverted repeats

TMT : tocopherol methyltransferase

VCF : variant call format

VITE : genes for vitamin E biosynthesis

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

CHAPTER 1

Introduction

Almond [*Prunus dulcis* (Mill.) D.A. Webb] is a perennial tree crop with a gametophytic self-incompatibility (SI) system (De Nettancourt 1997). Its perennial growth habit, large tree size and long juvenile period hamper both the genetic improvement (breeding) and genetic analysis. Until recently, few genomic sequence resources have been available for almond, there has been limited application of molecular methods in almond breeding and relatively little is known about the genetic control of economically important traits.

In the research conducted for this thesis, next-generation sequencing (NGS) technology was employed to address six main objectives:

1. **Sequence multiple haplotypes of the *S* locus.**

The self-incompatibility system of almond is under the genetic control of a complex multi-allelic locus, *S*. This locus is about 70,000 bp long. It contains an *S* locus F-box gene (*SLF*), two specificity determination genes (*S-RNase* and *SFB*) and long terminal repeat retrotransposons (LTRs). Although several alleles from the *S-RNase* and *SFB* genes have been fully or partially sequenced, the entire *S* locus sequence has only been available for the S_7 haplotype from one cultivar, Nonpareil (Ushijima et al. 2003). The sequencing approach used in all of these analyses was limited to one sample in a reaction. In this research, the entire *S* locus was sequenced simultaneously from 48 cultivars and breeding selections, which carry 15 *S* haplotypes, using multiplexed samples in one instrument run.

2. Structural analysis of the *S-RNase* and *SFB* genes.

Of about 50 known *S* alleles, complete *S-RNase* and/or *SFB* sequences are available for only a few of these and partial sequences are available for a few others (Bošković et al. 2007; Channuntapipat et al. 2001; Hafizi et al. 2013; Halász et al. 2010; Rahemi et al. 2010). Analysis of structural variations of the *S-RNase* and *SFB* genes has been conducted only for those that have complete gene sequences. The sequencing approach applied here provided complete gene sequences for those for which partial sequences had been available and for several alleles for which no sequences had been available. These resources were used to analyse structural features of these genes.

3. Develop and apply methods to design high-throughput assays for screening of *S* alleles in almond.

S allele identification prior to crossing can increase the efficiency of almond breeding. Until now, *S* genotypes have mainly been detected using primers designed based on polymorphisms in introns of the *S-RNase* gene (Channuntapipat et al. 2005; Channuntapipat et al. 2001; Channuntapipat et al. 2003; Gu et al. 2015; Ma and Oliveira 2002; Ortega et al. 2005; Sánchez-Pérez et al. 2004; Tamura et al. 2000). These markers mainly rely on gel electrophoresis to detect polymorphisms. The main limitations of these markers are masking of the presence of one allele by the other (Channuntapipat et al. 2001) and similar size amplicons resulting from different *S* alleles (Sánchez-Pérez et al. 2004). Some of these limitations could be overcome by detecting polymorphisms other than length polymorphisms, such as single nucleotide polymorphism (SNP) and/or by using different marker detection techniques e.g., fluorescence-based marker detection system.

4. Construct sequence-based linkage maps for Nonpareil and Lauranne.

In almond, parents and progeny are highly heterozygous due to out-crossing. This necessitates mapping approaches that differ from those that have been widely used for self-pollinated plants. These mapping approaches have been applied for almond (Arús et al. 1994; Fernández i Martí et al. 2013; Font i Forcada et al. 2012; Tavassolian et al. 2010). Marker types that have been used for

linkage map construction in almond include restriction fragment length polymorphism (RFLP) markers (Joobeur et al. 1998), simple sequence repeat (SSR) markers (Aranzana et al. 2003; Dirlewanger et al. 2004), inter-simple sequence repeat (ISSR) markers, sequence characterised amplified region (SCAR) markers (Tavassolian et al. 2010) and single nucleotide polymorphism (SNP) markers (Wu et al. 2010; Wu et al. 2009). Next-generation sequencing approaches can make it possible to directly assay large numbers of sequence polymorphisms (Mammadov et al. 2012; Michael 2014; Reuter et al. 2015) without prior knowledge of the genome sequence. The genotyping-by-sequencing (GBS) (Elshire et al. 2011) is a highly regarded approach for preparing sequencing libraries that has been shown to be effective for plant and animal species (Bielenberg et al. 2015; Elshire et al. 2011; Etter et al. 2011; Guajardo et al. 2015; Hyma et al. 2013; Lu et al. 2013). Application of GBS in almond could make it possible to construct linkage maps from just one library preparation and from one instrument run.

5. **Develop uniplex fluorescence-based PCR assays for single nucleotide polymorphisms discovered through genotyping-by-sequencing and apply those assays to almond populations.**

For repeated genotyping of specific polymorphisms, it is useful to have simple marker assays. The GBS-based genetic maps of Nonpareil and Lauranne can provide a source of sequence polymorphisms for the design of uniplex fluorescence-based PCR assays. Those assays can then be used to detect polymorphisms in other almond populations in the University of Adelaide almond breeding program.

6. **Map QTLs that affect physical and chemical traits in almond nuts and kernels.**

Improving almond nut and kernel quality is an important objective in almond breeding. One aspect is to cater for consumer preferences by producing quality nuts with high nutrient levels, sweet and pleasing colour. Almond physical nut traits (kernel weight, kernel length, kernel shape, shell hardness, in-shell weight, shell weight, kernel thickness, geometric diameter and spherical index) are

important in standardisation of almond processing and mechanisation. Not much research has been conducted to genetically map these traits except for shell-hardness, kernel weight and in-shell weight (Arús et al. 1998; Fernández i Martí et al. 2013; Sánchez-Pérez et al. 2007).

Almond nuts contain high levels of tocopherols, monounsaturated, polyunsaturated and saturated fatty acids (Jenkins et al. 2008; Kodad and Socias i Company 2008; Zhu et al. 2015). The phenotypic evaluation methods for these chemical traits involve sophisticated instruments and expertise. Some of these methods might be replaced by marker-assisted selection if genomic regions that affect these traits can be mapped. A very little research has been conducted to map these traits in almond (Font i Forcada et al. 2012).

This thesis contains seven chapters.

CHAPTER 1 Introduction

CHAPTER 2 Literature review

CHAPTER 3 Resequencing of the almond *S* locus from self-fertile and self-incompatible genotypes: A report on research that sequenced the almond *S* locus of 15 *S* haplotypes implementing next generation sequence strategies, obtaining complete *S* locus sequence for self-compatible haplotype (S_f) and two self-incompatible haplotypes (S_1 and S_8), partial *S* locus sequences for another 11 *S* haplotypes (S_3, S_5, S_6, S_9, S_{13}, S_{14}, S_{19}, S_{22}, S_{23}, S_{25} and S_{27}), confirmation of the published S_7 haplotype sequence and discussion of the advantages and challenges of using NGS technologies to sequence multiple alleles of a large locus from a highly heterozygous species.

CHAPTER 4 Marker design for the multi-allelic gametophytic self-incompatibility locus of almond: A report on the application of SNP-based marker development strategies to the multi-allelic *S* locus. Markers were designed for detection of SI alleles (S_1, S_3, S_5, S_7, S_8, S_9, S_{23} and S_{25}) and to differentiate the S_f allele from SI alleles.

CHAPTER 5 Linkage and quantitative trait locus maps for almond

This chapter consists of 3 sections.

Section 5.1 Application of genotyping-by-sequencing to construct linkage maps for almond: A report on research that adapted the original GBS protocol for almond, constructed linkage maps for Nonpareil and Lauranne parents and compared the resulting maps with the peach genome using the peach genome sequence assembly.

Section 5.2 Construction of linkage maps for almond using four populations with a common parent: A report on application of markers designed based on polymorphisms detected in Nonpareil and Lauranne to progeny of three other crosses (Nonpareil × Constantí, Nonpareil × Tarraco and Nonpareil × Vairo) and construction of a composite map for Nonpareil based on genotypic data from all four populations.

Section 5.3 Phenotyping and quantitative trait loci detection for nut and kernel traits in almond: A report on research that mapped QTLs for physical traits of nuts and kernels (kernel weight, kernel length, kernel shape, shell hardness, in-shell weight, shell weight, kernel thickness, length/width, thickness/length, thickness/width, geometric diameter and spherical index) and chemical traits (total tocopherol concentration, total fatty acid concentration and oleic/linoleic ratio).

CHAPTER 6 General discussion: A discussion on the significance of the research reported in this thesis, identification of limitations, suggestions for improvements and future research directions.

CHAPTER 7 Contributions to knowledge: A list of significant contributions to scientific knowledge from this research.

This thesis has four appendices:

APPENDIX 1: Supplementary materials of Chapter 3

APPENDIX 2: Supplementary materials of Chapter 4

APPENDIX 3: Supplementary materials of Chapter 5

APPENDIX 4: A year in the life of an almond tree

In the manuscript-style chapters (3 and 4), some changes have been made to provide a consistent format through-out the thesis. These include structure of the manuscript, organisation of the tables and figures and the consolidation of all references into a single list at the end.

CHAPTER 2

Literature review

2.1 Almond (*Prunus dulcis*)

Sweet almond [*Prunus dulcis* (Mill.) D.A. Webb] belongs to the family Rosaceae, sub-family Spiraeoideae and the genus *Prunus*. It is a diploid (2n = 16) and has a comparatively small genome of 270 Mb (Arumuganathan and Earle 1991; Bennett and Leitch 1997). Sweet almond originated from the intercrossing of *P. fenzliana* with other closely related wild almond species (*P. bucharica, P. kuramica, P. triloba,* and *P. webbii*) (Denisov 1998; Ladizinsky 1999). Wild populations of these species can be found in Western China, Iran, Israel, Armenia, Azerbaijan, Greece, Italy and Spain (Ladizinsky 1999; Martínez-Gómez et al. 2007; Rahemi et al. 2015). Wild almond species mainly consist of thorny shrubs with small leaves and bitter, round-shaped kernels (Ladizinsky 1999). Present-day cultivated almonds are trees with larger leaves and sweet, oval-shaped kernels. Almond differs from other cultivated members of the genus *Prunus* in that the seed (kernel) is the economically important product. In peach (*P. persica*), apricot (*P. armeniaca*) and sweet cherry (*P. avium*), the economically important product is the pulp (juicy mesocarp).

The USA is by far the largest producer of almonds in the world (FAOSTAT 2015) and accounted for 82% of global almond production in 2014/2015 (ABC 2015). These almonds are mostly grown in the Central Valley regions of California.

Australia is the second largest producer of almonds (FAOSTAT 2015). The Australian almond industry originated in South Australia and later expanded to other states. Almonds are grown in the Adelaide and Riverland regions of South Australia, the Riverina region of New South Wales, the

Sunraysia region of Victoria and the Swan region of Western Australia. Between 2001 and 2014, the total almond production area increased by more than fivefold. In 2014/2015, the total almond production in Australia was 85,000 tonnes, which accounted for 7 percent of global production, and 55,000 tonnes were exported, mainly to Europe and India.

In 2014/2015, domestic consumption of almond kernels in Australia reached 900 g per person (ABA 2015). Factors contributing to consumer preferences for almond may include flavour, health benefits and versatility of use. There is evidence that almond consumption helps to reduce LDL (low density lipoprotein) cholesterol levels in blood, improve heart health and prevent diabetes (Jambazian et al. 2005; Jenkins et al. 2008; Li et al. 2011; Wien et al. 2010)

2.2 Almond reproduction

Almond is a predominantly out-crossing species, with most cultivars being self-incompatible. The self-incompatibility (SI) system in the Rosaceae family is gametophytic: it involves recognition of the haploid genotype of the male gametophyte or pollen (De Nettancourt 1997). In commercial orchards, two or more compatible cultivars are grown together, with honey bees used as pollinators to ensure fruit set.

In gametophytic systems, pollen tube growth is usually arrested within the style and this involves contact between the pollen tube and the secretions by the cells of the transmitting tract. In the sporophytic SI system in the Brassicaceae, pollen tube growth is arrested at the stigma surface or soon after penetration. This involves contact between the pollen grain or emerging pollen tube and the secretions into the cell walls or on to the surface of the stigmatic papillae (De Nettancourt 1997).

2.2.1 Self-incompatibility and the almond *S* locus

Self-incompatibility prevents self-fertilisation by discriminating between self and non-self pollen grains. Pollen tubes from pollen grains that are genetically similar to pollinated plants are unlikely to

reach the ovule and effect fertilisation (De Nettancourt 1997). In almond, SI is under the genetic control of a complex multi-allelic locus (*S*), on linkage group 6 (LG6) (Ballester et al. 1998; Dicenta and Garcia 1993a). The *S* locus contains an *S-RNase* gene and two F-box genes (Ushijima et al. 2003; Ushijima et al. 1998). The *S-RNase* gene encodes glycoproteins with cytotoxic ribonuclease activity that is specifically expressed in pistils (Bošković et al. 1997; Ushijima et al. 1998). The *SFB* and *SLF* genes encode F-box proteins belonging to the ubiquitin ligase class (Deshaies 1999). One of the F-box genes, the *S* haplotype-specific F-box (*SFB*) gene is highly polymorphic and is specifically expressed in pollen. The other F-box gene, the *S* locus F-box (*SLF*) is less polymorphic and is expressed in both pollen and pistil. The *SFB* gene acts as the pollen determinant of SI whereas the *SLF* gene is not known to play any significant role in SI determination. Physically, the *SFB* and *S-RNase* genes are located in inverse orientation (Ushijima et al. 2003).

In almond, the distance between the *SFB* and *S-RNase* genes has been reported to range from 30 bp to 380 kb depending on the *S* haplotype (Ikeda et al. 2004; Ushijima et al. 2003; Yamane et al. 2003a; Yamane et al. 2003b). Each allele of the pistil *S-RNase* gene is co-inherited with the corresponding allele of the pollen *SFB* gene. The term '*S* allele' is often used to refer collectively to allelic combinations of the two genes. In contexts where it is important to distinguish between the two genes, the term 'allele' is used only for variants of individual genes and the term 'haplotype' for variants of the locus (McCubbin and Kao 2000).

After pollination, pollen grains germinate irrespective of their *S* haplotype or that of the pollinated plant. As pollen tubes grow through the pistil, they take up *S-RNases* irrespective of the *S-RNase* alleles by which those enzymes were encoded (Luu et al. 2000). Subsequently, however, 'non-self' S-RNases are inactivated while 'self' S-RNases are protected from inactivation. The recognition mechanism involves the ubiquitin/26s proteasome proteolytic system and the SFB protein, which contains an *S* haplotype-specific domain and an inhibitor domain (McCubbin and Kao 2000). This could be due to inhibition of the *S-RNase* activity (inhibitor model) (Luu et al. 2001; Sims and

Ordanic 2001), to degradation of non-self S-RNases (degradation model) (Hua et al. 2008; Qiao et al. 2004), or to blockage of S-RNase excretion into the cytosol of non-self pollen tubes (compartmentalisation model) (Goldraij et al. 2006; McClure et al. 2011).

One haplotype of the *S* locus has been fully sequenced: the S_7 haplotype of the almond cultivar, Nonpareil (Ushijima et al. 2003). The main structural features of the *S* locus of that haplotype are the *SLF*, *S-RNase* and *SFB* genes and long terminal repeat (LTR) retrotransposons (R). Within the S_7 haplotype, the gene order is *SLF* – *S-RNase* – *SFB*. The LTR retrotransposons are located downstream of the *SFB* gene (Fig. 2.1). There are four pairs of long terminal repeats (LTR 0 to LTR 3) and four retrotransposons (R0 to R3). In the S_7 haplotype, the R1, R2 and R3 sequences are inserted in the R0 sequence (Ushijima et al. 2003). The R3 sequence, which is longer than the R1 and R2 sequences, has an insertion of 2.5 kb which is similar to the 15 kb to 18 kb region of the S_7 haplotype. The LTR retrotransposons, R0, R1, R2 and R3 encode polyproteins (retro0, retro1, retro2 and retro3) (Ushijima et al. 2003). Comparative analysis of *S* haplotypes in *Prunus* has shown that differences in the *S* locus involve variation in the distance between the *S-RNase* and *SFB* genes, in the sequences of the *S-RNase*, *SFB* genes, retrotransposons and in the number of the *SLF* and *SFB* genes (Entani et al. 2003; Nunes et al. 2006; Rahemi et al. 2010; Romero et al. 2004; Vieira et al. 2010; Yamane et al. 2003a). These variations in the *S* locus may prevent recombination between the pistil *S-RNase* gene and pollen *SFB* gene, keeping the alleles of these two determinants of SI in linkage disequilibrium (Entani et al. 2003; Ushijima et al. 2003).

2.2.1.2 The almond *S-RNase* gene

Comparisons of amino acid sequences of *S-RNases* among *Prunus*, *Malus* species and Solanaceae have revealed five conserved regions (C1, C2, C3, RC4 or C4, C5) and one hypervariable region (RHV) (Broothaerts et al. 1995; Ioerger et al. 1990; Norioka et al. 1996; Sassa et al. 1997; Sassa et al. 1996; Ushijima et al. 1998). The locations of the five conserved regions are similar between the Rosaceae and the Solanaceae. Within and between these two families, there is high similarity in the

C1, C2, C3 and C5 regions. In contrast, sequences in the RC4 region seem to be specific to the Rosaceae, with no sequence similarity to the C4 region of the Solanaceae. In both families, this region includes a glycosylation site, which is responsible for attaching polysaccharides (glycans) to the S-RNase for proper folding and stability (Sassa and Hirano 1998; Ushijima et al. 1998).

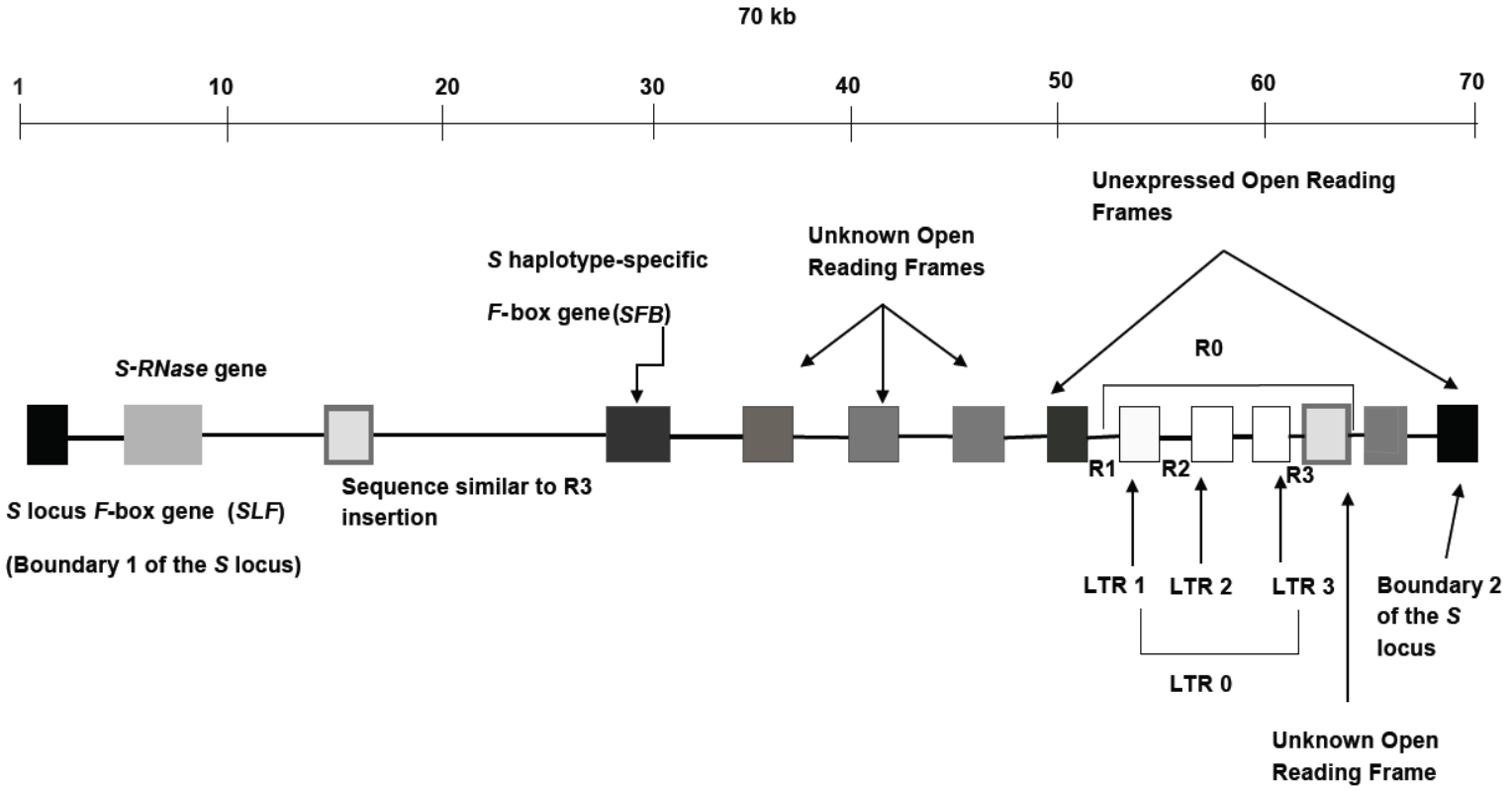


Fig. 2.1 A schematic diagram of the S_7 haplotype of almond *S* locus. This was drawn based on information reported by Ushijima et al. (2003).

The RHV region of *Prunus* (located between C2 and C3) corresponds with two hypervariable regions (HVa and HVb) that are known to mediate allele-specific pollen recognition in Solanaceae (Matton et al. 1997). It has been suggested that in *Prunus*, the haplotype specificity of most *S* alleles may be determined by the sequence variation in the RHV region (Ishimizu et al. 1998; Ushijima et al. 1998). However, there is evidence that RHV region in the *S* locus is not the only determinant of haplotype specificity in *Prunus* and *Pyrus* species. For example, the S_n*-RNase* and S_i*-RNase* of European pear (*Pyrus communis*) have identical RHV regions, yet S_i pollen can fertilise the ovules of plant that carry S_n allele (Zisovich et al. 2004). In sweet cherry (*Prunus avium*), S_6- and S_{24}*-RNases* sequences have identical RHV regions, yet maintain distinct *S* haplotypes (Wünsch and Hormaza 2003). In almond, another two variable regions (VR1 and VR2) have been identified between the RC4 and C5 region (Gu et al. 2012; Ortega et al. 2006) (Fig. 2.2).

2.2.1.3 The almond *SFB* gene

The protein encoded by the *SFB* gene of almond contains an F-box motif, two variable regions (V1 and V2) and two hyper variable regions (HVa and HVb) (Ikeda et al. 2004; Ushijima et al. 2003; Yamane et al. 2003a; Yamane et al. 2003b) (Fig. 2.3). The F-box motif, which is relatively conserved among *Prunus* species, is in the N-terminal region of the protein, while HVa and HVb are in the C-terminal region (Ikeda et al. 2004; Ushijima et al. 2003). The F-box motif is involved in forming the E3 ubiquitin ligase complex (Deshaies 1999) that can bind specifically to the target protein (S-RNase) for ubiquitination at the C-terminal region of the *SFB*. The HVa and HVb regions may be responsible for the discrimination between self and non-self S-RNases and polyubiquitination of the non-self S-RNases (Ikeda et al. 2004; Ushijima et al. 2003).

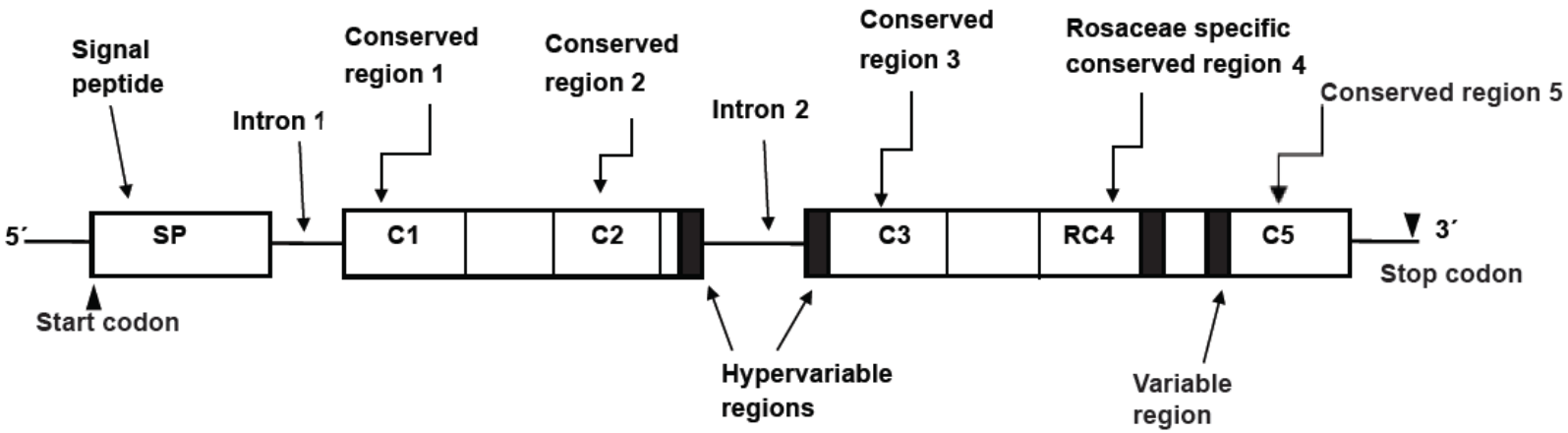


Fig. 2.2 A schematic diagram of the almond *S-RNase* gene. This was drawn based on the conserved regions described by Ortega et al. (2006) and Ushijima et al. (1998).

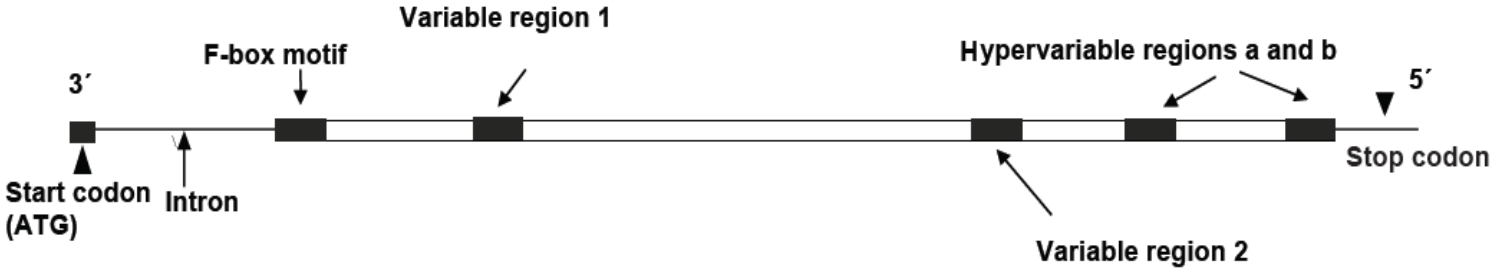


Fig. 2.3 A schematic diagram of the almond *SFB* gene. This was drawn based on the information reported by Ushijima et al. (2003).

2.2.1.4 The long terminal repeat (LTR) retrotransposons

Transposable elements (TEs) are able to insert themselves/make new copies of themselves into new locations within the genome (Goodier 2016; Havecker et al. 2004; Zhang et al. 2014). Based on the mechanism of transposition, TEs can be classified into class I (retrotransposons), which transpose through an RNA intermediate, and class II (DNA transposons), which transfer only *via* DNA. The major superfamilies of class I are *Ty1-copia* and *Ty3-gypsy* retrotransposons, while class II is comprised of TIR (terminal inverted repeat) and *Helitrons*, which are sometimes classified separately (Biémont 2010). Transposition can be autonomous or non-autonomous in both Class I and Class II TEs. Non-autonomous forms (e.g., MITEs–miniature inverted-repeat transposable elements, SINEs–short interspersed nuclear elements and LINEs–long interspersed nuclear elements) are quite prevalent among both retrotransposons and transposons (Havecker et al. 2004). A significant portion of plant genomes is constituted by LTR retrotransposons, which contain two long terminal repeats (LTRs) (Fig. 2.4) at their ends, generally 250 to 600 bp in length, flanking a 5 to 7 kb long internal protein-coding domain. Between the two LTRs, there are two open reading frames (ORFs): *gag* and *pol*. The *gag* ORF encodes the structural proteins that make up virus-like particles. The *pol* ORF encodes the enzymes required for reverse transcription and integration. In the *Ty1-copia* type, the enzymes are organised in the order of protease (PR), integrase (IN), reverse transcriptase (RT) and RNase H (RH) (Zhang et al. 2014). LTR retrotransposons replicate in a copy-and-paste manner. If this mode of transposition is not suppressed, retrotransposons can massively increase their copy numbers, resulting in a rapid expansion of genome size.

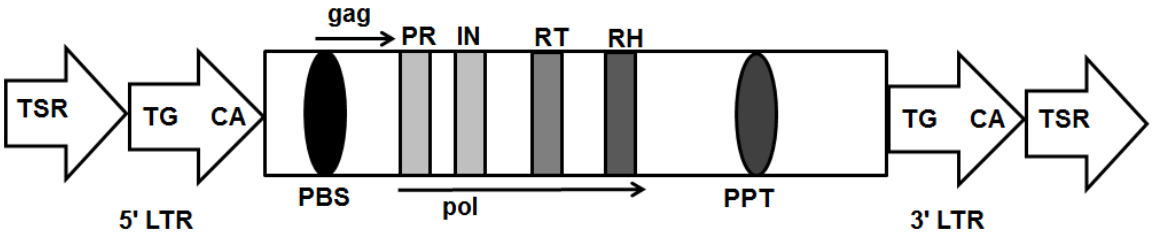


Fig. 2.4 Basic structure of a full-length LTR retrotransposon. This was drawn based on information reported by Zhang et al. (2014).

In *Prunus,* many TEs have been identified in *P. persica* (Tao et al. 2007), *P. cerasus* (Hauck et al. 2006; Yamane et al. 2003b), *P. armeniaca* (Halász et al. 2014) and *P. mume* (Ushijima et al. 2004). These have caused natural insertional mutations in those species that lead to breakdown of SI.

2.2.2 Self-fertility in almond

A few almond cultivars have been found to be self-fertile. For some of these, self-fertility has been attributed to an *S* allele designated S_f (Grasselly and Olivier 1976) that is considered to be dominant over other *S* alleles (Socias i Company 1990). The S_f allele in *P. dulcis* may have been introgressed from *P. webbii* (Bošković et al. 2007; Martinez-Gracía 2009) and transmitted into commercial S_f almond cultivars *via* the cultivar Tuono (S_1S_f) from the Puglia region of Italy. The self-fertile cultivars Blanquerna (S_8S_f), Lauranne (S_3S_f) and Tuono (S_1S_f) have very low level of *S-RNase* activity in the pistil and allow pollen tube elongation from all pollen grains (Alonso and Socias i Company 2005; Bosković et al. 1999; Fernández i Martí et al. 2014). This could be due to low transcription of the S_f-*RNase* (Fernández i Martí et al. 2010; Hanada et al. 2009). This S_f allele is sometime known as S_{fi} (S_f-inactive).

2.2.2.1 Deduced amino acid sequences and structure of the S-RNase

The S_f allele from Tuono (AF157009) has been sequenced (Ma and Oliveira 2002). In the deduced amino acid sequences of the S_f-*RNase*, there is an arginine (R) instead of a histidine (H) at the 4th position of the C2 region (Channuntapipat et al. 2001; Fernández i Martí et al. 2010; Hanada et al. 2009) and a phenylalanine (F) instead of leucine (L) at the 3rd position in the C5 region (Fernández i Martí et al. 2010; Hanada et al. 2009). Based on comparisons to other species, the histidine-to-arginine substitution seems more likely to be the cause of self-fertility than the phenylalanine-to-leucine substitution. In wild tomato (*Solanum peruvianum*), self-fertility has been attributed to loss of a histidine residue in the C2 region (Royo et al. 1994). In apricot there is an allele (S_2) with leucine in place of the same phenylalanine (Martinez-Gracía 2009) that does not confer self-fertility.

Three dimensional protein structures constructed by comparative modelling of the S_f (AB467371) and two SI (S_8 –AB481108 and S_{23} –AB488496) *S-RNases* of almond have shown that all three have six α strands and six β strands, and that the *S_f-RNase* has a long loop between the conserved domains RC4 and C5. This loop may also contribute to Sf determination in almond (Fernández i Marti et al. 2012).

2.2.2.2 Dual expression of *S_f-RNase* and epigenetic variation in almond

There are also some SI plants, mainly from the Majorca Island in Spain, for which the *S-RNase* DNA sequence is identical to that of self-fertile plants with the S_f allele (Bošković et al. 2007; Fernández i Martí et al. 2014; Kodad et al. 2009b). The SI cultivars that carry the S_f allele sequence (Fra Giulio Grande (Bošković et al. 2007), Ponç (Kodad et al. 2009b) and Vivot (Fernández i Martí et al. 2009; Fernández i Martí et al. 2010)) exhibit high levels of *S-RNase* transcripts in the stylar tissues. Therefore, the S_f allele in Ponç, Fra Giulio Grande and Vivot is known as active S_f (S_{fa}) (Kodad et al. 2009b).

Research on Vivot ($S_{23}S_{fa}$) and Blanquerna (S_8S_{fi}) has revealed that several cytosine (C) residues in the upstream region of the *S_f-RNase* gene promoter are methylated (5mCs) in Blanquerna but not in Vivot, which may suppress *S_f-RNase* gene expression in S_{fi} individuals. (Fernández i Martí et al. 2014).

2.2.2.3 Self-compatibility due to stylar part and pollen part mutations

In the Solanaceae and the Rosaceae, some cultivars that carry *S-RNase* alleles that normally confer SI are naturally self-compatible. Their self-fertility may be caused by mutations in the *S-RNase* gene (stylar part mutations (SPM)) or in the *SFB* gene (pollen part mutations (PPM)).

In the Solanaceae, a glycosylation site in the *S-RNase* gene may play an important role in *S* haplotype-specificity recognition and pollen rejection (Soulard et al. 2013). In woodland tobacco

(*Nicotina sylvestris*) individuals with a mutation in the glycosylation site have been found to be self-compatible, despite having a high level of *S-RNase* activity in pistils (Golz et al. 1998). However, in the Rosaceae, there is no record of the effect of a glycosylation site in self-fertility determination.

In the self-fertile Japanese pear variety Osa-Nijisseiki (S_2S_{4m}), which was found as a sport of the SI variety Nijisseiki (S_2S_4), there is no *S_4-RNase* gene in the *S* locus (Sassa et al. 1997). In the Italian self-fertile sweet cherry cultivar Kronio (S_5S_6), there is a premature stop codon in the upstream region of the hypervariable region (HVa) of the *SFB_5* allele. This produces a truncated SFB_5 protein that is unable to identify self-pollen for inhibition and allows pollen tube elongation from all pollen grains (Marchese et al. 2007). In the self-fertile Japanese apricot cultivar, Kensaki (S_fS_f) has a high level of *S-RNase* activity in pistils (Tao et al. 2002). Its self -fertility is due to by an insertion of 6.8 kb in the *SFB_f* gene (Ushijima et al. 2004). In peach, self-fertility conferred by the S_1 haplotype is also caused by a mutation in the *SFB* gene. This mutation is similar to a 5 bp insertion in the orthologous *SFB_6* allele of almond (Hanada et al. 2014), but is not known whether self-compatibility in almond also involves any *SFB* gene mutations.

In almond, the self-fertility of the cultivar Patalina (S_uS_n) has been attributed to inactivity of the S-RNase encoded by the null allele S_n. Yet the *Sn* coding sequence is identical to that of the S_2 allele carried by the SI cultivar Cristomorto (Bošković et al. 2007). Thus the self-fertility of Patalina may be mediated by some factor/s outside the coding region of the *S-RNase* gene (Bošković et al. 2007).

2.2.2.4 Other proteins involved in the self-incompatibility mechanism in almond

Proteins other than the S-RNase and SFB may act as modifying factors in the SI determination of almond (Gómez et al. 2015; Martínez-García et al. 2015). Research conducted using protein extracts from pistil and pollen tissues after controlled pollination of fully compatible and fully incompatible almond varieties Ferragnès (S_1S_3), Mono (S_5S_7), Titan (S_8S_{14}), Primorskyi (S_5S_9) and S_f selection A2-198 (S_fS_f) revealed that carbonic anhydrases, the ATPase β-subunits, enolase, cyclophilin, pectin

methylesterases, porins, malate dehydrogenase and quinine reductase may play important roles in pollen-pistil interactions (Martínez-García et al. 2015). Some of these proteins were up-regulated in incompatible responses, some were up-regulated in compatible responses and some were up-regulated in both compatible and incompatible responses. Carbonic anhydrases, which were up-regulated in self-incompatible responses, have been associated with the CO_2-induced breakdown of SI in field mustard (*Brassica campestris*) (O'Neill et al. 1988). Pectin methylesterases and porins were both up-regulated in compatible responses. Pectin methylesterases, which are located in the plasma membrane of the pollen tube, regulate pollen tube growth (Zonia and Munnik 2009). Porins have been found in voltage-dependant anion channels located on the outer mitochondrial membrane in almond pistils (Jiang and Ma 2003) and in the pollen tubes of *Picea meyeri* (Meyer's spruce). In spruce, porins are believed to play an important role in the maintenance of homeostasis during Ca^{2+} signalling and in balancing Ca^{2+} levels (Chen et al. 2009). Annexins, which were down-regulated in self-compatible responses, also regulate Ca^{2+} and phospholipid levels in the pollen tube. Co-localisation of high Ca^{2+} and annexin levels at the pollen tube can mediate the pollen tube growth inhibition. In SI varieties of common poppy (*Papaver rhoeas*), high Ca^{2+} in pollen tubes shortly after germination inhibits the elongation of incompatible pollen tubes (Franklin-Tong 1999; Franklin-Tong et al. 1997).

In another proteomic analysis conducted in almond, 28 proteins in pistils and 20 proteins in anthers were found to be differentially expressed between A2-198 ($S_{fi}S_{fi}$) and an SI individual, ITAP-1 ($S_{11}S_{fa}$) (Gómez et al. 2015). One of the up-regulated proteins in the pistil of A2-198 is a pathogenesis-related thaumatin-like protein that seems to be involved in pollen signal recognition. In self-incompatible Japanese pear varieties, this protein is highly accumulated in the style (Sassa and Hirano 1998). Glutathione S-transferases (GSTs), which were down-regulated in pistils of A2-198 are known to mediate the repression of self-pollen tube growth in self-incompatible varieties of Chinese cabbage (Wang et al. 2014), and to influence the pollen fertility level in upland cotton (Zhu et al. 2003). Another down-regulated protein in both pistils and anthers of A2-198 is glucan endo-1,3-beta-

glucosidase, which catalyses the hydrolysis of β-1,3-glucan in stigma walls to facilitate pollen tube entry (Suen et al. 2003).

2.2.3 Characterisation of *S* locus alleles in almond

Early research on cross compatibility relied mainly on controlled hybridisation followed by evaluation of fruit set. Evaluation of fruit set, assessment of pollen tube growth and breeding records were used to establish cross incompatibility groups (CIGs) (Tufts and Philip 1922). The progeny from a cross for which the two parents have no *S* alleles in common would consist of four CIGs. Each individual would be incompatible with itself and with other members of its CIG but compatible with all members of other CIGs. So far, almond cultivars have been assigned to 32 CIGs (CIG 0 to CIG XXXI) with CIG 0 consisting of unique genotypes (Table 2.1) (Bošković et al. 2003; Hafizi et al. 2013; Halász et al. 2010; Kester et al. 1994; Kodad et al. 2010b; Kodad and Socias i Company 2009; López et al. 2004; Mousavi et al. 2011; Rahemi et al. 2010).

Identification of *S* alleles prior to crossing allows almond breeders to plan crosses that will be reliably compatible. Before DNA-based methods for almond *S* allele detection were developed, the *S* genotypes of almond were mainly detected by analysing stylar ribonucleases (Bošković et al. 1997). Once DNA sequence information became available for some *S* alleles (Bošković et al. 2007; Channuntapipat et al. 2001; Channuntapipat et al. 2003; Halász et al. 2010; Kodad et al. 2010b; Mousavi et al. 2011), allele-specific primers were designed based on introns 1 and 2 of the *S-RNase* gene (Channuntapipat et al. 2003; Ortega et al. 2005; Sutherland et al. 2004).

So far, about 50 *S* alleles have been identified in almond. In nucleotide databases such as NCBI (National Center for Biotechnology Information, EMBL (European Molecular Biology Laboratory) and GDR (Genome Database for Rosaceae), the only complete *S* locus sequence for almond is the one for the S_7 haplotype from cultivar Nonpareil (AB081587). For some other haplotypes, there are either complete or partial sequences derived from the *S-RNase* and *SFB* genes (Fig. 2.5).

Table 2.1. Cross-incompatibility groups in almond (Bošković et al. 2003; Kodad and Socias i Company 2009).

Group	*S* genotypes
I	S_7S_8
II	S_1S_5
III	S_5S_7
IV	S_1S_7
V	S_5S_8
VI	S_1S_8
VII	S_8S_{13}
VIII	S_1S_3
IX	$S_{7m}{}^*S_8$
X	S_7S_{14}
XIII	S_6S_8
XIV	S_1S_6
XV	S_2S_9
XVI	S_1S_{10}
XVII	S_2S_{10}
XVIII	$S_{12}S_{23}$
XIX	$S_{22}S_{23}$
XX	$S_{13}S_{27}$
XXI	S_8S_{12}
XXII	S_3S_{23}
XXIII	S_3S_5
XXIV	S_1S_9
XXV	S_1S_4
XXVI	S_7S_{13}
XXVII	$S_{11}S_{22}$
XXVIII	$S_{fm}{}^*S_{36}$
XXIX	S_1S_2
XXX	S_3S_4
XXXI	S_7S_{48}
0	S_5S_{22}, $S_{13}S_{22}$, $S_{22}S_{26}$, S_6S_{11}, S_8S_{21}, S_1S_{21}, S_8S_{22}, S_5S_{13}, S_4S_{12}, S_5S_{14}, S_6S_{22}, S_4S_{13}, $S_{12}S_{28}$, S_1S_{11}, S_1S_2, $S_{12}S_{22}$, $S_{13}S_{22}$, $S_{28}S_{29}$, $S_{24}S_{27}$, S_4S_{23}, $S_{10}S_{24}$, S_6S_{14}, $S_{21}S_{26}$, S_5S_6, $S_{23}S_{25}$, $S_{11}S_{12}$, $S_{22}S_{27}$, $S_{10}S_{23}$, $S_{10}S_{13}$, S_8S_{24}, S_1S_{18}, $S_1S_{34m}{}^*$, $S_{22}S_{34}$, $S_{12}S_{33}$, S_5S_9, S_8S_{23}, S_6S_{23}, S_3S_{25}, S_2S_3, $S_{11}S_{21}$, $S_{21}S_{23}$, S_1S_{23}, S_3S_9, S_1S_{22}, $S_{23}S_{27}$, S_8S_{14}, S_6S_7, S_8S_{31}, S_4S_{13}, S_5S_{10}, S_8S_{10}

* Mutation in the wild-type *S* allele.

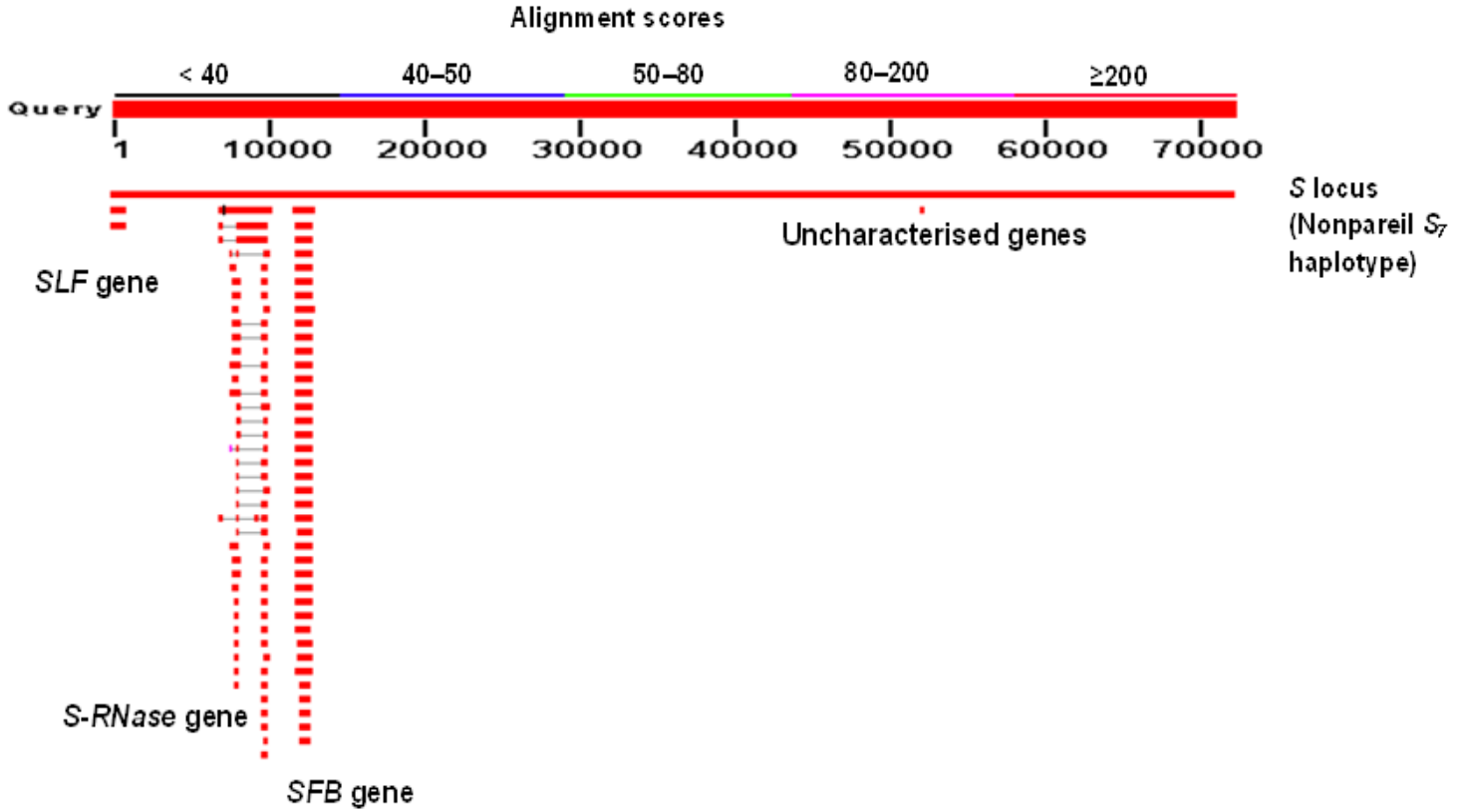


Fig. 2.5 An overview of the *S-RNase* and *SFB* gene sequences (200 blast hits) registered in the GenBank NCBI) aligned to the Nonpareil S_7 haplotype (AB081587) as the query using the NCBI BLASTN program.

2.3 Genetic marker discovery and construction of linkage maps in almond

2.3.1 Application of molecular markers in almond improvement

Most almond breeders share the objectives of producing more nutritious almonds with superior horticultural and nut traits while maintaining heterozygosity (Denisov 1998; Kester and Gradziel 1996; Kester et al. 1994; Kodad et al. 2010a; Wirthensohn and Sedgley 2002). Almond breeding mainly employs phenotypic selection of potential parents, crossing among those parents, vegetative propagation of progeny and phenotypic selection among progeny. This process is hampered by the prolonged juvenile period and perennial growth habit of almond. Any strategy that could permit early selection would have great potential to accelerate genetic gain in almond breeding. In some other crops with long juvenile periods, early selection (among parents and/or among juvenile progeny) has been achieved through the use of molecular markers (Grattapaglia and Sederoff 1994; Souza et al. 2013). To date, there has been relatively little implementation of molecular markers in almond breeding. One constraint to molecular marker development has been that the genetic research required to discover marker-trait associations is hampered by the same phenological factors that hamper almond breeding. Further, until recently, the development of molecular markers was constrained by the limited amount of sequence information that had been generated for almond.

Recently, the development of low-cost sequencing methods (Gao et al. 2012; Reuter et al. 2015) has made it possible to generate substantial amounts of genomic sequence information for almond (Koepke et al. 2013). It has also led to the development of genotyping-by-sequencing methods (Elshire et al. 2011) that could facilitate genetic analysis in almond.

2.3.2 Almond linkage maps

Linkage map construction requires analysis of the co-segregation of markers using a large number of individuals in a population. For most plant species, including almond, mapping populations have been derived from controlled crosses between two parents (Ballester et al. 2001; Joobeur et al. 2000; Sánchez-Pérez et al. 2007; Sánchez-Pérez et al. 2010). Due to the out-crossing nature of almond,

both the parents and the progeny of these populations are highly heterozygous (Kester et al. 1991; Martínez-Gómez et al. 2003). This necessitates mapping approaches that differ from those that have been widely used for self-pollinated species.

Several linkage maps of almond have been published, including one derived from an inter-specific cross between almond and peach (Joobeur et al. 1998). This almond-peach mapping population was derived by selfing a single plant (MB1-73) from a cross between almond (cv. Texas) and peach (cv. Earlygold) (T × E). The T × E map is considered as the reference map for all *Prunus* species, Many new markers have been added to it (Dirlewanger et al. 2004; Howad et al. 2005; Illa et al. 2011). The linkage maps that have been published for almond were constructed mainly using randomly amplified polymorphic DNA (RAPD), restriction fragment length polymorphism (RFLP), inter simple sequence repeat (ISSR), simple sequence repeat (SSR) and sequence characterised amplified region (SCAR) markers (Ballester et al. 1998; Dirlewanger et al. 2004; Dirlewanger et al. 1998; Fernández i Martí et al. 2013; Howad et al. 2005; Joobeur et al. 2000; Joobeur et al. 1998; Tavassolian et al. 2010).

One of the populations developed in the first year of the Australian almond breeding program (Wirthensohn and Sedgley 2002) consisted of a set of 182 F_1 progeny from a cross between Nonpareil and Lauranne (N × L). Nonpareil, which was developed in the USA, is the major cultivar grown in both Australia (ABA 2015) and the USA (ABC 2014). It is a self-incompatible cultivar, with sweet kernels and paper shelled characteristics. Lauranne, which originated in France, is an important cultivar grown in Europe. It has a self-fertile reproductive mechanism derived from its male parent Tuono, sweet kernels and hard shells (Grasselly 1972). The first N × L linkage map (Gregory et al. 2005) was developed using 93 F_1 plants. It had 36 polymorphic markers (RAPD, SSR, ISSR and SCAR) and seven linkage groups with a total length of 360.9 cM. With addition of more ISSR markers and some SCAR markers, more improved Nonpareil parental, Lauranne parental and N × L integrated maps were constructed. The Nonpareil parental map had 93 markers and a total length of

539 cM, while the Lauranne parental map had 97 markers and a total length of 534 cM (Tavassolian 2008). The N × L integrated map had 157 markers and a total length of 591.4 cM (Tavassolian et al. 2010). Although the average distance between markers of these maps ranged from 4 to 9 cM, each map had some larger intervals, including some longer than 25 cM.

Development of high resolution genetic maps requires both large numbers of markers and large populations. In almond, the generation, maintenance and evaluation of large populations from particular individual crosses is limited by the perennial nature and large size of the trees. One possible alternative is to consider all of the progeny plants of an almond breeding program as one large multi-parent population, and apply genetic analysis approaches similar to those that have been developed for nested association mapping (NAM) populations in maize (Yu et al. 2008).

In the University of Adelaide almond breeding program, there are 8,100 F_1 plants derived from 39 different parents. Eleven parents have been commonly used as females and five of those can be considered as reference parents i.e. Carmel, Chellaston, Johnston, Nonpareil, and Somerton. Each of these cultivars has been crossed with several other male parents with a large number of progeny from each cross. Nonpareil has been used as the female parent in crosses with 15 male parents (Chellaston, Constantí, Ferraduel, Glorieta, Johnston, Lauranne, Mandaline, Marta, Somerton, Tarraco, Vairo, White, R1065, R1146 and 12-350) (Wirthensohn and Sedgley 2002). Current breeding germplasm includes a total of 2,300 progeny from Nonpareil crosses. With sufficient genotyping and phenotyping, these materials might be used as a NAM population. This will give not only a much larger population size to further improve genetic maps, but also would provide information how genes from one parent act in a range of genetic backgrounds.

In a collaboration that has been established among the University of Adelaide, Australia, the University of New England, Australia, Washington State University, USA and IRTA, Spain, the genome of the almond cultivar Texas (Mission) is being fully sequenced and resequencing is being

carried out for 39 other cultivars, including Nonpareil, Chellaston, Johnston and some other clones that have been used as parents in the University of Adelaide almond breeding program. Genetic mapping of Nonpareil × Lauranne and for the Nonpareil NAM could be valuable for anchoring the genomic contigs to genetic maps. Furthermore, the genomic sequence information from the consortium could be a source of candidate genes for QTLs mapped in these populations.

2.3.3 Next-generation sequencing

Next-generation sequencing (NGS) procedures make it possible to generate large amounts of genomic data quickly and cost effectively. Illumina, Roche and SOLiD are some of the sequencing platforms that are widely used (Bielenberg et al. 2015; Mascher et al. 2013; Rocher et al. 2015; Wendler et al. 2014) There are several protocols available for reduced representation library preparation. These include, restriction site associated DNA (RAD), genotyping-by-sequencing (GBS) and double digest restriction site associated DNA (ddRAD) (Baird et al. 2008; Elshire et al. 2011; Miller et al. 2007; Peterson et al. 2012; Poland et al. 2012). All of these protocols use restriction enzymes (REs) to digest the genome of interest. They differ from each with respect to the types and number of restriction enzymes used and the steps used in library construction.

Genotyping-by-sequencing is a simple, highly multiplexed, high-throughput genotyping approach, which is based on NGS technologies and was first applied to maize and barley (Elshire et al. 2011). It can be optimised for other species by selecting suitable restriction enzymes and adapter combinations to obtain the required depth of sequencing (Mitchell et al. 2012; Poland et al. 2012). The library preparation of the original GBS protocol described by Elshire et al. (2011) includes four steps. In the first step, the genomic DNA is restriction digested with methylation-sensitive rare cutting REs that form sticky ends. In the second step, ligation is carried out with two types of adapters: a common adapter and a barcode adapter. In the third step, DNA samples are pooled and PCR amplification is conducted with two primers which contain sequences that are complementary to the

adapters. The fourth step involves the fragment size analysis of the prepared libraries to verify the suitability for sequencing.

Poland et al. (2012) modified the original GBS protocol with a two-enzyme system that includes one rare cutter and one common cutter. This system uses barcoded forward adapters and a common reverse adapter. Every amplified fragment of the two-enzyme library carries both types of adapters. This two-enzyme approach can simplify the quantification of the library prior to sequencing. It also reduces complexity in large genomes (e.g., wheat).

The GBS protocol has been successfully used to genotype grasses (Elshire et al. 2011; Li et al. 2015; Lu et al. 2013; Poland et al. 2012) including finger millet (Bioinnovate-Africa 2014), peach (Bielenberg et al. 2015), sweet cherry (Guajardo et al. 2015), grapevine (Hyma et al. 2013), lentil (Wong et al. 2015), canola (Bayer et al. 2015), chickpea (Bayer et al. 2015; Kujur et al. 2015) and animals (De Donato et al. 2013; Mitchell et al. 2012).

2.4 Nut traits in almond

Catering to consumer demands by producing quality nuts that have high nutrient levels, sweet and pleasing colour and that meet industrial requirements has been an important objective in almond breeding (Gradziel and Martínez-Gómez 2013). Kernel sweetness, shell hardness, fatty acid, oil content and vitamin E (tocopherols) are considered as important nut traits in almond (Kester et al. 1977; Maguire et al. 2004; Romojaro et al. 1988). Physical parameters such as width, length, thickness, geometric diameter, spherical index in the kernel and the nut are also important in both almond harvesting and processing. Almonds are usually harvested mechanically using shakers and mechanical pickers are used to collect the fallen almonds. Almond processing facilities are equipped with various machines (prehullers, rollers, shell crackers and sorters). Almond physical parameters play important role in designing and standardisation of these machines (Fernández i Martí et al. 2013; Socias i Company et al. 2008).

2.4.1 Kernel sweetness/bitterness

Bitterness in almond is determined by the content of the cyanogenic diglucoside amygdalin (Dicenta and Garcia 1993b; Dicenta et al. 2007). In almond, enzymes and some of the genes involved in the amygdalin degradation pathway have been identified and characterised (Fig. 2.6). The sweetness or bitterness in almond kernel is determined by the genetic background of the mother plant (Kester and Gradziel 1996), and it has been demonstrated that there is no influence of pollinisers on this trait (Dicenta et al. 2000).

Prunasin is the precursor of amygdalin. It is synthesised in the tegument tissues (inherited from mother) of almond and both sweet and bitter cultivars seem to have the similar synthesis capacities. In bitter almonds, at the early developing stage, a transient accumulation of prunasin is observed in the tegument tissues (Franks et al. 2008). In the mid-development stage, amygdalin is observed in the nucleus and the endosperm, and in fully matured bitter almonds, it is only detected in embryos. In contrast, prunasin does not accumulate and amygdalin is not detectable in sweet almonds.

For amygdalin synthesis, prunasin has to be transported from tegument tissues to cotyledons (Sánchez-Pérez et al. 2008). In bitter almonds, a high level of prunasin is transported to cotyledons *via* outside the plasma membrane whereas in sweet almonds, it is transported through the cytoplasm during which most of the prunasin is degraded. The amount of prunasin available for amygdalin synthesis is mainly determined by the route of prunasin transport from tegument to cotyledons and distribution of amygdalin biosynthesis enzymes (prunasin hydrolase and mandelonitrile glucosyltransferase) in sweet and bitter almonds (Sánchez-Pérez et al. 2012; Sánchez-Pérez et al. 2008). Other than these factors, environmental conditions may also play a role in prunasin accumulation in almond kernels (Dicenta et al. 2007; Yildirim et al. 2014).

Almond bitterness is believed to be monogenic, conferred by a recessive allele (*sk)* with most individuals having a heterozygous genotype (Dicenta et al. 2007; Grasselly and Crossa-Raynaud 1980; Heppner 1923; Sánchez-Pérez et al. 2007).

On the T × E *Prunus* reference map, *Sk* was mapped at 32.9 cM on linkage group 5 (LG5) (Dirlewanger et al. 2004). In a cross between the sweet almond varieties R1000 (*Sksk*) and Desmayo Largueta (*Sksk*), this trait segregated in a ratio of 3 sweet (*SkSk or Sksk*) to 1 bitter (*sk*/*sk*). The Sk locus was mapped on LG5 in both parental maps (Sánchez-Pérez et al. 2007; Sánchez-Pérez et al. 2010) and was placed between 11 and 14.6 Mb on scaffold 5 of the peach genome assembly (Koepke et al. 2013).

Five genes that encode enzymes involved in amygdalin biosynthesis (*Gt1* (glucosyltransferase1), *Gt2* (glucosyltransferase2), *Gt3* (glucosyltransferase3), *Ah1* (amygdalin hydrolase1) *and Ph* (prunasin hydrolase)) have been mapped using the R × D cross and a T × E bin map, but none of them mapped on LG5 (Sánchez-Pérez et al. 2010).

Comparative sequences analysis of the *Sk* locus using sweet almond cultivars (Lauranne, Ramillete) and bitter almond selections from the CEBAS-CSIC almond breeding program in Murcia, Spain, identified 6304 polymorphisms in the *Sk* (sweet kernel) locus, of which 228 are codon changing. This provides candidate genes for the *Sk* locus (Koepke et al. 2013).

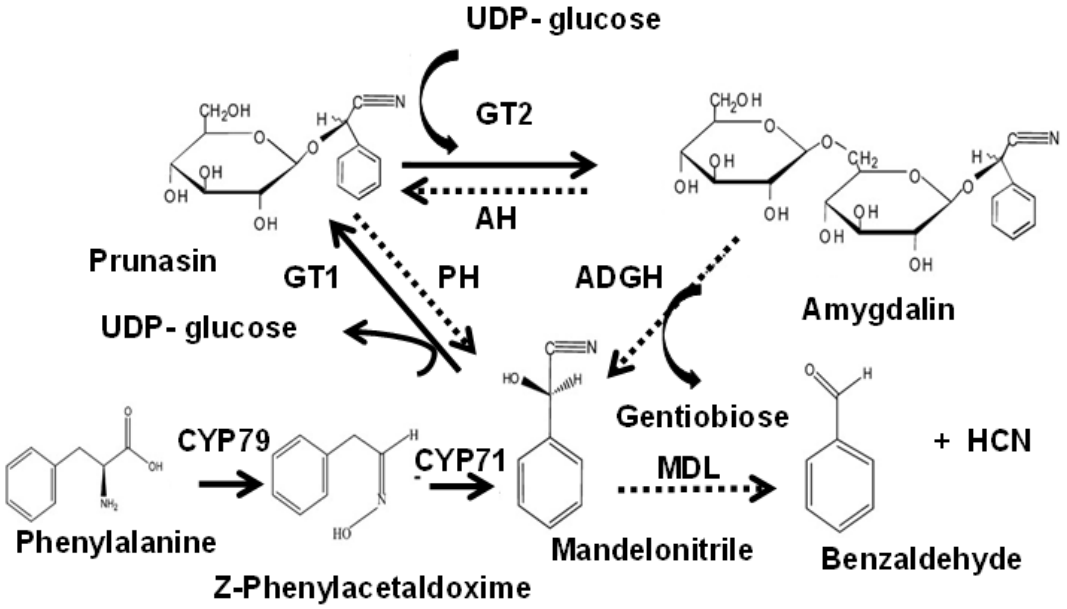


Fig. 2.6 The metabolic pathways for synthesis and catabolism of cyanogenic glucosides prunasin and amygdalin in almond. Biosynthetic enzymes (black lines) are: CYP79 and CYP71 (Cyt P450 monooxygenases), GT1 (UDPG-mandelonitrile glucosyltransferase) and GT2 (UDPG-prunasin glucosyltransferase). Catabolic enzymes (dashed lines) are: AH (amygdalin hydrolase), PH (prunasin hydrolase), MDL1 (mandelonitrile lyase) and ADGH (amygdalin diglucosidase). Adapted from Conn (1980) and Sánchez-Pérez et al. (2008).

2.4.2 Shell hardness

The fruit of almond is a drupe consisting of the outer hull (exocarp), leathery mesocarp and inner shell (endocarp) and the seed. The shell acts as a protective cover for the almond kernel (Figs. 2.7 and 2.8). It consists of two layers (inner and outer) which are connected by vascular tissues. Different degrees of lignification can be found in those layers. The layers can remain attached to each other during the nut development, but sometimes, the inner and outer layers separate from each other. During nut dehiscence, the outer layer may be removed with the hull, resulting in an almond with poor sealing (Kester and Gradziel 1996).

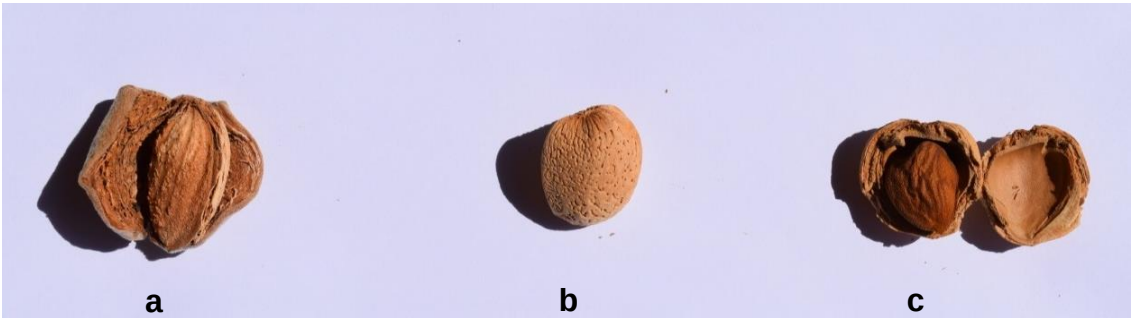


Fig. 2.7 Almond fruit: hull is attached to the almond nut (a), hull is removed from the shell (b) and almond kernel is inside the shell (c).

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).

Shell hardness is generally calculated according to Rugini (1986) as follows using the weight of the almond kernels divided by total weight of almond kernels and their shells. Shell hardness of almond is usually noted as a percentage:

Shell hardness = kernel weight/(kernel weight + shell weight)

Because shell hardness indicates how much kernel weight is contained within almond nuts, it is used to categorise almonds and is an important trait for growers and processors. Almond nuts are categorised into five groups according to shell hardness percentages (Table 2.2).

Table 2.2. Shell hardness groups in almond.

Shell hardness group	Shell hardness percentage (%)
Stone	≤ 24
Hard	25 – 34
Semi hard	35 – 44
Soft	45 – 54
Paper shell	≥ 55

Preference with respect to shell hardness varies among markets. Hard shells are less prone to insect attacks, pathogen infections and damage by birds and they are preferred in European markets. Other markets such as India and the Middle East prefer soft or paper shells which can be easily removed (Socias i Company et al. 2008). Phenotypic assessment to distinguish among the shell hardness groups is time consuming. This trait has sometimes been treated as a qualitative character and sometimes as a quantitative character with high heritability. Grasselly and Crossa-Raynaud (1980) attributed its genetic control to a single gene (*Dd*), with hard shell dominant over soft shell in a ratio of 3 (hard shell) to 1 (soft shell). This locus has been assigned to LG2 on the T × E (Arús et al. 1998) and R × D maps (Sánchez-Pérez et al. 2007). When Arús et al. (1998) treated this character as a quantitative trait, two QTLs were mapped using a Ferragnès × Tuono population. On both parental maps, a major-effect QTL was mapped on LG2 and a minor-effect QTL was mapped on LG8.

2.4.3 Other physical nut traits in almond

Other physical parameters such as width, length, thickness, geometric mean diameter and spherical index of the kernel and the nut are important traits for almond harvesting, processing and storage. Kernel size is often expressed using length, width and thickness. These parameters are established during the first growth phase of nut development in the spring and completed by summer (Kester and Gradziel 1996). Most of the physical properties in almond are affected by moisture availability during the nut development (Aydin 2003; Mohsenin 1970). At harvest the moisture content in the nut is around 10 to 20% (fresh weight basis). After harvest, the nuts are left to dry on the orchard floor for up to 10 days until the moisture content drops between 5 and 8%.

2.4.3.1 Geometric mean diameter (D_p), spherical index (Φ), kernel size and shape

The geometric mean diameter of almond nut/kernel is calculated by using the measurements of kernel or nut length (L), width (W) and thickness (T) (Mohsenin 1970).

$$D_p = (LWT)^{1/3}$$

The spherical index is calculated as follows:

$$\Phi = \frac{(LWT)^{1/3}}{L}$$

Kernel size (KS), calculated by multiplying length, width and thickness of the kernel as below:

$$KS = L \times W \times T$$

Kernel shape (KSH), calculated by dividing the width of the kernel by its length as follows:

$$KSH = \frac{W}{L}$$

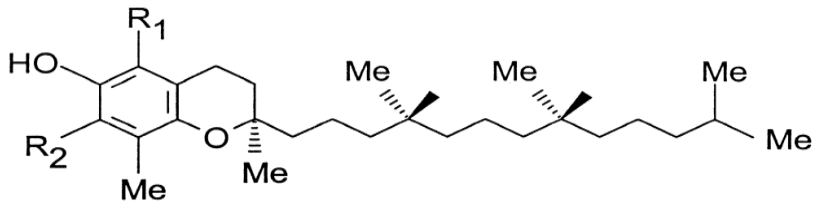
QTLs for these physical properties have been mapped on six linkage groups using the Vivot × Blanquerna almond population with most QTLs affecting multiple traits (Fernández i Martí et al. 2013).

2.4.4 Chemical properties of almond kernels

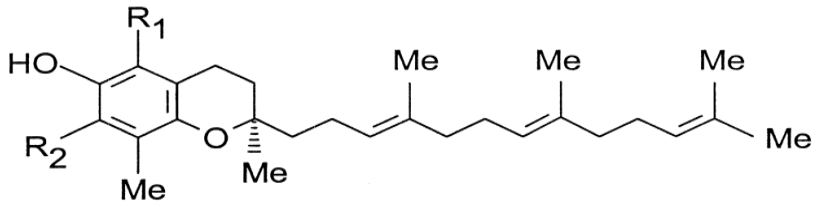
2.4.4.1 Vitamin E content

Vitamin E refers to a group of compounds that include both tocopherols and tocotrienols. They are essential nutrients for human health. They can be synthesised only by plants and other photosynthetic organisms (Grusak and DellaPenna 1999). Several epidemiological studies have shown that vitamin E has a potential to reduce risk of cancer, other degenerative diseases and cardiovascular diseases in humans (Jia et al. 2006; Steinmetz and Potter 1996; Wien et al. 2010). High levels of vitamin E have been reported to enable prolonged storage of seeds by conferring resistance against lipid rancidification (Kodad et al. 2010a).

The eight chemical compounds in vitamin E are lipid-soluble antioxidants: α-, β-, γ- and δ-tocopherol and α-, β-, γ- and δ-tocotrienol (are collectively known as tocol) (Fig. 2.9). The four members of each type differ from each other in the numbers and positions of methyl groups on the phenolic ring (Table 2.3). The most biologically active form of vitamin E is α-tocopherol (2 to 50 times more activity than any of the other three tocopherols because of the affinity of the mammalian hepatic α-tocopherol transfer protein for α-tocopherol). Tocol contents and compositions vary among plant tissues. For example, fresh photosynthetic tissues contain between 10 and 50 μg/g of tocopherols and tocotrienols with a high percentage of α-tocopherol whereas oil extracts from seeds contain from 300 to > 2000 μg/g of tocopherols and tocotrienols). In most oilseed crops, α-tocopherol is present as a minor component, e.g. 7% in soybean oil (DellaPenna and Mène-Saffrané 2011).



Tocopherols



Tocotrienols

Fig. 2.9 Chemical structures of tocopherols and tocotrienols. Downloaded and adapted from (http://www.ncbi.nlm.nih.gov/pccompound)

Table 2.3. Tocopherol and tocotrienol chemical compounds in vitamin E (Me: methyl group and H: hydrogen).

Tocopherol structure	**Tocotrienol structure**	**Residual groups**	
		R1	**R2**
α-tocopherol	α-tocotrienol	Me	Me
(5,7,8-trimethyl tocol)	(5,7,8-trimethyl tocotrienol)		
β-tocopherol	β-tocotrienol	Me	H
(5, 8-dimethyl tocol)	(5, 8-dimethyl tocotrienol)		
γ-tocopherol	γ-tocotrienol	H	Me
(7, 8-dimethyl tocol)	(7, 8-dimethyl tocotrienol)		
δ-tocopherol	δ-tocotrienol	H	H
(8-monomethyl tocol)	(8-monomethyl tocotrienol)		

Research using the model plant *Arabidopsis thaliana* has revealed that two different pathways are involved in tocopherol synthesis. In tocopherol synthesis, the cytosolic aromatic amino metabolism is involved in head group synthesis and the plastidic deoxyxylulose 5-phosphate pathway is involved in hydrophobic tail synthesis (DellaPenna 2005). A fully saturated tail in tocopherols is derived from phytyl diphosphate (PDP). The important steps in head group synthesis are the production of homogentisic acid (HGA) from *p*-hydroxyphenylpyruvic acid (HPP) by *p*-hydroxyphenylpyruvic acid dioxygenase (HPPD). HGA is converted to 2-methyl-6-phytylplastoquinol (MPBQ) and 2-methyl-6-geranylgeranylplastoquinol (MGGBQ). MPBQ and MGGBQ act as substrates for either tocopherol cyclase or MPBQ methyltransferase (MPBQ MT). MPBQ MT adds a second methyl group to MPBQ to form 2,3-dimethyl-5-phytyl-1,4-benzoquinone (DMPBQ) or to MGGBQ to form 2,3-dimethyl-5-geranylgeranyl-1,4-benzoquinone (DMGGBQ). Tocopherol cyclase converts MPBQ and DMPBQ to δ- and γ-tocopherols. The γ-tocopherol methyltransferase (γ-TMT) adds a methyl group to C-6 of the chromanol ring, converting δ- and γ-tocopherols and tocotrienols to β- and α-tocopherols. Genes encoding the committed steps in tocopherol biosynthetic pathway (*VITE1-5*) have been identified (Fig. 2.10) (DellaPenna 2005; DellaPenna and Mène-Saffrané 2011; Gilliland et al. 2006).

In almond, α-tocopherol is the major component of vitamin E, accounting for more than 90% of the tocopherol (Kodad et al. 2006; Zhu et al. 2015). The tocopherol content is affected by genotype and by environment (Kodad et al. 2006; Maguire et al. 2004; Zhu et al. 2015).

Vitamin E content in almond oil ranges from 14.4 to 55.3 mg/100 g (Kornsteiner et al. 2006; Madawala et al. 2012). Breeders are making efforts to improve vitamin E levels in almond. Tocopherol and tocotrienol in almond can be extracted using alkaline assisted, pressurised liquid or supercritical fluid extraction techniques. Generally, determination of tocopherol levels in plants is carried out with high performance liquid chromatography (HPLC), which is time consuming, labour-intensive and requires specific instruments and expertise. With improved understanding of the

genetic control of tocopherol concentration, it might be possible to replace some of this phenotyping with molecular marker assays.

QTL analysis of the content of tocopherol compounds (α-, γ- and δ-) using a Vivot × Blanquerna population has revealed multiple QTLs with small effect. Five of these affected α-tocopherol, four affected γ-tocopherol and two affected δ-tocopherol. (Font i Forcada et al. 2012).

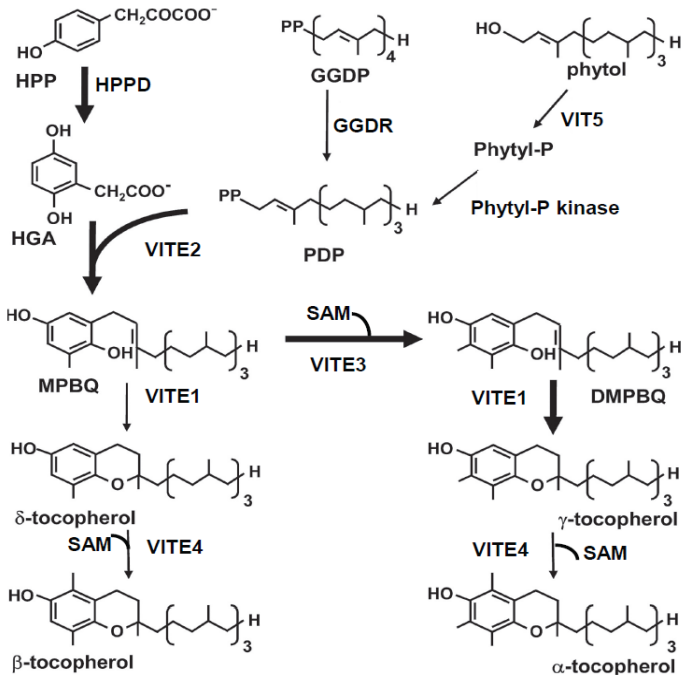


Fig. 2.10 The tocopherol biosynthetic pathway in *Arabidopsis thaliana*. Enzymes involved in are shown as VITE1-5. HPP (hydroxyphenylpyruvate), GGDP(geranylgeranyl-diphosphate), Phytyl-P (phytyl-monophosphate) PDP (phytyl-diphosphate), HGA (homogentisic acid), MPBQ (2-methyl-6-phytyl-1,4-benzoquinol), HPPD(hydroxyphenylpyruvate dioxygenase), DMPBQ (2,3-dimethyl-6-phytyl-1,4-benzoquinol), GGDR (geranylgeranyl-diphosphate reductase), SAM (*S*-adenosyl methionine) VITE1 (tocopherol cyclase) VITE2 (homogentisic acid phytyltransferase) VTE3 (2-methyl-6-phytyl-1,4-benzoquinol methyltransferase) VITE4 (γ-tocopherol methyltransferase) and VITE5 (phytol kinase). Adapted from Gilliland et al. (2006) and DellaPenna (2005).

2.4.4.1.1 High performance liquid chromatography

HPLC is a technique widely used in analytical chemistry to separate, identify and quantify components in a sample. A typical HPLC system consists of an autosampler, pumps, a column and detectors. The autosampler brings the sample vials to an injector which loads the samples into the stream of mobile phase that runs through the column. Each component in the sample interacts differently with the absorbent material in the column. This causes different flow rates for the different components. This leads to the elution of the components as they flow out of the column. Separation of tocopherols and tocotrienol can be performed on either normal-phase or reversed-phase columns (Font i Forcada et al. 2012; Kamal-Eldin et al. 2000; Kornsteiner et al. 2006; Zhu et al. 2015). In normal-phase columns, the stationary phase is polar (consisting of silica or of organic moieties with cyano amino functional groups) and the mobile phase is non-polar (consisting of a non-aqueous solvent such as hexane or heptane). In reversed-phase columns, the stationary phase is non-polar (consisting of octadecyl carbon (C18)-bonded silica) and the mobile phase consists of a polar aqueous solvent (Kamal-Eldin et al. 2000). The detectors generally used for tocopherol and tocotrienol determination could be DAD (diode array detector), FLD (fluorescence detector) or both (Font i Forcada et al. 2012; Kamal-Eldin et al. 2000; Zhu et al. 2015).

2.4.4.2 Lipid components in almond

The lipid content in 100 g of almond meal has been observed to range from 41 to 63 g (Madawala et al. 2012; Zhu et al. 2015). Of the various fatty acids in almond kernels, oleic (58 to 71%), linoleic (15 to 29%) and palmitic acids (5 to 7%) are the most abundant while stearic (1 to 3%), palmitoleic acids (3 to 5%) and vaccenic acid (1 to 2%) are also present in minor amounts (Gallier et al. 2012; Zhu et al. 2015). Comparison of fatty acid composition in almond with other nuts (walnut, hazelnut, peanut and macadamia), indicated that almond has a unique fatty acid (FA) profile with comparatively high levels of monounsaturated fatty acids (oleic (C18:1) and palmitoleic (C16:1), polyunsaturated fatty acids (linoleic (C18:2) and linolenic (C18:3), saturated fatty acids (palmitic (C16:0) and stearic (C18:0)

and phytosterols (β-sitosterol) (Jenkins et al. 2008; Kodad and Socias i Company 2008; Maguire et al. 2004; Sabate and Hook 1996).

Both oleic (O) and linoleic (L) acids have the ability to prevent cardiovascular diseases in humans (Vezvaei and Jackson 1995). High ratios of oleic to linoleic acid (O/L ≥ 2.5) reduce oil rancidification, improving stability during storage and transportation (Kodad et al. 2009a; Kodad and Socias i Company 2008). Almond cultivars differ with respect to their oil contents and FA compositions (Romojaro et al. 1988). Environmental conditions also play a major role determining the FA and oil content of almond (Socias i Company et al. 2008; Zhu et al. 2015). To determine the oil and the FA composition in almond gas chromatography (GC) is widely used (Font i Forcada et al. 2012; Gallier et al. 2012; Zhu et al. 2015).

Genetic analysis of oil content and FA composition in almond using a Vivot × Blanquerna population detected one QTL that affected oleic, linoleic, palmitic, stearic and palmitoleic acids, one QTL that affected linoleic and oleic acids, plus individual QTLs with major effects on palmitic acid, palmitoleic acid or oil content (Font i Forcada et al. 2012).

2.4.4.2.1 Gas chromatography

Gas chromatography (GC) is an analytical technique that is widely used to separate and to analyse compounds in the gas phase. A typical chromatography instrument consists of sample injection port, a column, carrier gas flow control equipment, ovens and heaters, a detector and an integrator chart recorder. Components of the samples are dissolved in an organic solvent and vaporised before mixing with an inert carrier gas (often helium, nitrogen, hydrogen or argon). The inert gas (mobile phase) goes through a column (a microscopic layer of liquid or polymer on an inert solid support (silica) inside a piece of glass or tube). Substances with less solubility in the liquid will elute faster than the substances with greater solubility while moving through the column.

In GC, a sample injection port is used to introduce the sample into the column head. The vaporisation chamber vaporises the sample and the sample vapours mix with the carrier gas to transport into the column. An oven with a thermostat is used to maintain the desired temperature of the column throughout the entire operation. A detector generates signals (peaks) as substances of the mixture elute from the column. The most commonly used detectors in GC are flame ionization detector, thermal conductivity detector or electron capture detector.

2.5 Research questions

Analysis of the literature cited above led to the following research questions:

1. What is the sequence diversity of *S* alleles present in almond breeding materials in Australia?
2. Can marker assays be designed for efficient differentiation of the S_f allele from SI alleles and/or for efficient differentiation among SI alleles?
3. Can almond linkage maps be improved by the application of NGS technologies to bi-parental populations?
4. What are the positions and effects of QTLs for physical traits of almond nuts and kernels and for the chemical composition (vitamin E (α-, β- and γ- tocopherol) and fatty acid content) of almond kernels?
5. Can the crosses made in the Australian almond breeding program using Nonpareil as the female parent and 15 other clones as male parents be treated as a NAM population to improve genetic and QTL maps and to investigate how genes from one parent act in a range of genetic backgrounds?

2.6 Research goals

To address these questions, the following approaches were taken in the research that is reported in this thesis:

1. Application of a next-generation sequencing approach to resequence the *S* alleles from almond cultivars and breeding selections.
2. Identification of which *S* alleles are present in the University of Adelaide almond breeding germplasm.
3. Development of molecular marker assays to differentiate the S_f allele from SI alleles.
4. Development of molecular marker assays to distinguish among SI alleles using SNPs in the *S-RNase* and *SFB* genes.
5. Application of genotyping-by-sequencing to construct a sequence-based genetic linkage map using a Nonpareil × Lauranne F_1 population.
6. Exploitation of other Nonpareil progeny to construct a high-resolution genetic map for Nonpareil.
7. Mapping of QTLs that affect the vitamin E (α-, β- and γ-tocopherol) content of almond kernels.
8. Mapping of QTLs that affect the fatty acid composition (oleic, linoleic and palmitic acid) of almond kernels.
9. Mapping of QTLs that affect physical nut qualities, such as kernel weight, kernel length, kernel shape, shell hardness, in-shell weight, kernel thickness, geometric diameter and spherical index.

CHAPTER 3

Resequencing of the almond *S* locus from self-fertile and self-incompatible genotypes

Shashi N. Goonetilleke[1], Adam E. Croxford[1], Michelle G. Wirthensohn[1], Timothy J. March[1] and Diane E. Mather[1]

[1]School of Agriculture, Food and Wine, Waite Research Institute, The University of Adelaide, Australia.

Corresponding author: diane.mather@adelaide.edu.au

Tel: +61 8 8313 7156

Fax: +61 8 8312 7102

SG: shashi.goonetilleke@adelaide.edu.au

AC: adam.croxford@adelaide.edu.au

MW: michelle.wirthensohn@adelaide.edu.au

TM: timothy.march@adelaide.edu.au

3.1 Statement of Authorship

Title of Paper	Resequencing of the almond S locus from self-fertile and self-incompatible genotypes
Publication Status	☐ Published ☐ Accepted for Publication ☐ Submitted for Publication ☑ Unpublished and Unsubmitted work written in manuscript style
Publication Details	This chapter will be prepared as a manuscript for submission to a refereed journal.

Principal Author

Name of Principal Author (Candidate)	Shashi N. Goonetilleke		
Contribution to the Paper	Prepared the sequencing library, critically analysed all sequence data, conducted phylogenetic analysis, interpreted results and wrote the first draft of the manuscript and took primary responsibility for the manuscript revision.		
Overall percentage (%)	70%		
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.		
Signature		Date	

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

i. the candidate's stated contribution to the publication is accurate (as detailed above);
ii. permission is granted for the candidate in include the publication in the thesis; and
iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Adam E. Croxford		
Contribution to the Paper	Helped make the sequencing library, conducted the sequencing, contributed to the analysis and interpretation of the results. Read the manuscript and suggested revisions.		
Signature		Date	

Name of Co-Author	Michelle G. Wirthensohn		
Contribution to the Paper	Suggested the research topic. Provided the plant materials and information on their *S* genotypes. Read the manuscript and suggested revisions.		
Signature		Date	

Name of Co-Author	Timothy J. March		
Contribution to the Paper	Contributed to the analysis of the sequencing data.		
Signature		Date	

Name of Co-Author	Diane E. Mather		
Contribution to the Paper	Provided overall supervision of the research. Contributed to the interpretation of the results. Supervised the revision of the manuscript.		
Signature		Date	

3.2 Abstract

3.2.1 Background

The self-incompatibility system in almond [*Prunus dulcis* (Mill.) D.A. Webb] is under the genetic control of a complex multi-allelic locus, *S*. The almond *S* locus, which is about 70,000 bp long, harbours the *S* locus F-box gene (*SLF*), two specificity determination genes (*S-RNase* and *SFB*) and long terminal repeat (LTR) retrotransposons. Prior to this research, the entire *S* locus had been sequenced from only one haplotype (S_7). In this research, the *S* locus was sequenced from 48 cultivars and breeding selections, which carry 15 *S* haplotypes, using multiplexed samples in one instrument run.

3.2.2 Results

The S_7 haplotype sequence was refined and complete haplotype sequences were obtained for three other *S* haplotypes (S_1, S_8 and S_f). Partial sequences were obtained for eleven *S* haplotypes (S_3, S_5, S_6, S_9, S_{13}, S_{14}, S_{22}, S_{19}, S_{23}, S_{25} and S_{27}). Comparisons of the *S* locus sequences from these haplotypes indicated high sequence variability among them. Sequences were less variable at the beginning and the end of the *S* locus. High sequence variability was observed among the alleles from the *S-RNase* and *SFB* genes. Considerable sequence variation was observed in region between *S-RNase* and *SFB* genes. The number of LTRs in the *S* locus varied among these haplotypes.

3.2.3 Conclusions

The almond *S* locus is highly variable in some regions, especially where it harbours the *S-RNase* and *SFB* genes. The *SLF* gene was less polymorphic among these haplotypes. LTRs and intergenic regions contribute further variation to the almond *S* locus.

3.2.4 Keywords

Almond, *S* locus, *S-RNase* gene, *SFB* gene, *SLF* gene, *S* haplotype-specific region

3.3 Introduction

Almond [*Prunus dulcis* (Mill.) D.A. Webb] is a predominantly out-crossing species with most cultivars being self-incompatible. Its gametophytic self-incompatibility (SI) system prevents self-fertilisation by discriminating between pollen tubes from self and non-self pollen grains. Pollen tubes that are genetically similar to pollinated plants fail to reach the ovule and effect fertilisation (De Nettancourt 1997). In almond, SI is under the genetic control of a complex multi-allelic locus, *S*, on linkage group 6 (Ballester 1998). This locus is about 70,000 bp long. The S_7 haplotype of the almond cultivar Nonpareil has been nearly fully sequenced (Ushijima et al. 2003). Based on the S_7 haplotype sequence, the main structural features of the almond *S* locus are three genes and four pairs of long terminal repeat retrotransposons (LTRs) (Ushijima et al. 2003). The *S* locus F-box (*SLF*) gene, which marks one boundary of the *S* locus, is expressed in both pollen and pistil, but is not known to play any significant role in determining *S* allele specificity. The *S-RNase* and *SFB* genes are specifically expressed in pistil (Ushijima et al. 1998) and in pollen (Ushijima et al. 2003), respectively. They are highly polymorphic and act as *S* allele specificity determinants in almond. Within the *S* locus, these two genes are oriented inversely relative to each other, and this may facilitate the haplotype-specific interaction between these genes (Ushijima et al. 2003). This is mediated by the *S* haplotype specific domain in the *SFB* gene product (SFB protein), which interacts differently with self and non-self S-RNases (McCubbin and Kao 2000). Length variations have been observed in the regions between the *S-RNase* and *SFB* genes in the *S* locus sequences of S_5, S_7 and S_8 haplotypes, and it has been suggested that these sequence heteromorphisms may prevent recombination, keeping the alleles of these two determinants of SI in linkage disequilibrium (Ushijima et al. 2003).

The S_7 sequence was obtained using Sanger sequencing technology (Sanger et al. 1977) after PCR amplification of overlapping segments (100 to 1,000 bp) from a single individual (Ushijima et al.

2003). With next-generation sequencing (NGS) procedures, it has become possible to simultaneously sequence large numbers of DNA segments from many individuals (Buermans and den Dunnen 2014; Hodkinson and Grice 2015; van Dijk et al. 2014; Voelkerding et al. 2009).

In this research, NGS was used to verify and complete the S_7 sequence and to obtain complete sequences for three other *S* haplotypes (S_1, S_8 and S_f) and partial sequences for eleven *S* haplotypes (S_3, S_5, S_6, S_9, S_{13}, S_{14}, S_{22}, S_{19}, S_{23}, S_{25} and S_{27}). The structure of the *S* locus was refined for the S_7 haplotype and was determined for 14 other haplotypes by examining the distances between genes and the number and distribution of LTRs. Phylogenetic relationships among 15 *S-RNase* alleles from this work, previously sequenced *S* alleles of the *S-RNase* gene from other *Prunus* species (*P. avium*, *P. armeniaca*, *P. cerasus*, *P. pseudocerasus* and *P. spinosa*), from *Malus* × *domestica*, from *Pyrus pyrifolia,* from three species belong to the Solanaceae (*Physalis crassifolia*, *Solanum lycopersicum* and *Nicotiana alata*) and from *Antirrhinum hispanicum* were evaluated using deduced amino acid sequences.

3.4 Materials and methods

3.4.1 Plant materials and DNA extraction

Genomic DNA was extracted from young leaves of 43 almond cultivars and five breeding selections from the University of Adelaide almond breeding program (Table 3.1) using a Isolate II plant DNA Extraction Kit (Bioline, NSW, Australia). DNA quality and quantity of each sample (1 µL) were assessed by comparison to a DNA ladder (HyperLadder I (Bioline, NSW, Australia)) after electrophoresed on 1% agarose at 100 V for 30 min.

Table 3.1. Almond cultivars and breeding selections used in this analysis.

Cultivar/Breeding selection	*S* genotype	Origin
Antoñeta	S_1S_f	Spain
Atkinson's Hardshell	$S_{14}S_{27}$	Australia
Baxendale	S_5S_7	Australia
Biggs Hardshell	S_6S_{14}	Australia
Brown Nonpareil	S_1S_7	Australia
Brown Brandis	$S_{23}S_{25}$	Australia
Bruce	$S_{22}S_{23}$	Australia
Capella	S_7S_f	Australia
Carina	S_7S_f	Australia
Carmel	S_5S_8	USA
Chellaston	Ambiguous	Australia
Clements	S_6S_{14}	Australia
Constantí	S_3S_f	Spain
Federation	S_3S_{22}	Australia
Francolí	S_1S_f	Spain
Frenzy	S_5S_8	Australia
Johnston	Unknown	Australia
Jordan	S_8S_{23}	Spain
Keanes	S_7S_{27}	Australia
Kapareil	S_8S_{13}	USA
Lauranne	S_3S_f	France
LeGrand	S_1S_8	USA
Mandaline	S_1S_f	France
Marion Sturt Creek	S_1S_{22}	Australia
Maxima	S_3S_8	Australia

Table 3.1., continued.

Cultivar/Breeding selection	*S* genotype	Origin
McKinlays	S_7S_8	Australia
Milow	S_1S_8	USA
Mira	S_7S_f	Australia
Monarto 2	$S_{25}S_8$	Australia
Monarto 3	Unknown	Australia
Nonpareil	S_7S_8	USA
Parkinson	$S_{22}S_{23}$	Australia
Pethick Wonder	$S_{23}S_{27}$	Australia
Pearce	Unknown	Australia
Peerless	S_1S_6	USA
Ramillete	S_6S_{23}	Spain
Softshell Jordan	S_8S_{14}	Australia
Somerton	S_1S_{23}	Australia
Steliette	Ambiguous	France
Tom Strout	Unknown	Australia
Strout's Papershell	$S_{22}S_{25}$	Australia
Vairo	S_9S_f	Spain
White Brandis	S_6S_{23}	Australia
12-350	S_1S_f	Spain
T5	S_7S_f	Australia
T6	S_8S_f	Australia
T7	S_7S_f	Australia
T8	unknown	Australia

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

To identify the regions of the *S* locus that might be suitable for primer design, the S_7 reference sequence (AB081587) was aligned with the available *S-RNase* allele sequences (AB433984 for S_f, AB011469 for S_1, AF490505 for S_3, DQ150569 for S_5, AM231657 for S_6, AY291118 for S_7, AB481108 for S_8, AF454001 for S_9, AM231662 for S_{13}, AM231663 for S_{14}, AM231668 for S_{19}, AM231671 for S_{22}, AB488496 for S_{23}, AM231673 for S_{25} and AM231675 for S_{27}), *SFB* gene sequences (AB361036 for S_f, AB376968 for S_3, AB101660 for S_8, DQ677588 for S_9, AM746960 for S_{13} and EU310403 for S_{23}) and with the sequence of scaffold 6 of the peach (*Prunus persica*) whole genome assembly v1.0 (www.rosaceae.org) using the Map to Reference alignment algorithm (iteration = 10) as implemented in Geneious software version 9.0.2 (Kearse et al. 2012). Primers were designed to obtain overlapping amplicons of between 5 kb and 10 kb in length that would cover the entire *S* locus. These primers were named with the prefix WriPdSL, Wri referring to the Waite Research Institute, Pd referring to *Prunus dulcis* and SL referring to the *S* locus. To identify a suitable enzyme for PCR amplification, five commercially available DNA polymerases (Biomix (Bioline, NSW, Australia), Q5 (New England Biolabs, Ipswitch, MA, USA), Q5 Hot Start High-Fidelity (New England Biolabs, Ipswitch, MA, USA), SensiFAST (Bioline, NSW, Australia) and Phusion High-Fidelity (New England Biolabs, Ipswitch, MA, USA)) were used in combination with one primer pair (WriPdSL5) to amplify products from three almond cultivars (Nonpareil, Carmel and Constantí) using DNA samples of 20 ng (2 µL of 10 ng/µL) with 4 µL of buffer, 3.2 µL of 1.25 mM dNTPs, 4 µL of 1 µM forward and reverse primer mix and 0.2 U of DNA polymerase in each 20 µL of reaction. The PCR conditions used were 98˚C for 30 sec, 34 cycles of 98˚C for 10 sec, 60°C for 30 sec, 72°C for 10 min followed by a final extension at 72˚C for 15 min. The amplified products (5 µL) were run on 1% agarose gel at 100 V for 30 min to quantify and to evaluate the amplified products. The most suitable buffer to use with Phusion High-Fidelity DNA polymerase was determined by testing two standard buffers (Phusion®GC and Phusion®HF) with the same DNA samples using the same primer pair and PCR conditions. The amplified products were evaluated as described above.

3.4.3 Primer testing and amplification of the *S* locus

Once a suitable polymerase and a suitable buffer were identified, PCR amplification was conducted for each primer pair from two DNA samples (Nonpareil and 12-350). Amplified products were separated and quantified on 1% agarose as described above. Primer pairs that produced clear bands were selected to amplify products from all the samples.

3.4.4 Library preparation and sequencing

The intensity of the amplicons obtained was not uniform across the entire *S* locus. Hence, regions were classified as strongly or weakly amplified based on the approximate intensity of DNA bands as visualised on 1% agarose. Amplicons that were strongly amplified were bulked into one tube while those that were weakly amplified were bulked into another tube. Each of the bulked samples was purified using AMpure® XP beads (Agencourt Bioscience Co., Massachusetts, USA) as per the manufacturer's instructions. For each sample, strong and weak bulks were mixed together at an appropriate ratio to have approximately uniform coverage of each amplified region across the *S* locus. A total of 50 ng (20 µL of 2.5 ng/µL) of DNA from each sample was used to prepare a sequencing library, using an Illumina Nextera DNA Library Prep Kit (V3) (Illumina Inc., Australia). The tagmentation, purification, indexing and PCR amplification of each DNA sample were carried out as per the manufacturer's instructions. The Tn5 transposase in the Nextera DNA Library Prep Kit was used to digest the DNA samples to generate segments of about 300 bp containing read 1 (5'-TCGTCGGCAGCGT-3') and read 2 (5'-GTCTCGTGGGCTCGG-3') sequences. Index primers i5, i7 and paired-end (PE) primers P5 and P7 were then annealed to each sample of the library using a reduced-cycle PCR amplification. The amplified products in the each sample were purified using AMpure® XP beads (Agencourt Bioscience Co., Massachusetts, USA) as described above, quantified by qPCR using Kapa SYBR FAST Master Mix (Kapa Biosystems, Massachusetts, USA) on a Rotor-Gene instrument (QIAGEN, UK) and assayed for quality using a TapeStation 2002 instrument (Agilent Technologies, Australia). Each sample in the *S* locus library was normalised to 4 nM and pooled. The quality of the library was assessed in a Bioanalyser 2001 instrument (Agilent

Technologies, Australia). The paired end library was mixed with 1% PhiX (Illumina Inc., Australia) as a control before sequencing on a MiSeq instrument (Illumina Inc., Australia) using paired end 300 bp reads.

3.4.5 Sequence data analysis

All raw sequence reads were assessed for quality, adapter sequences and barcode contamination using FASTQC v0.11.5 (Andrews 2010). Adapter sequences were removed using the IllUMINACLIP option in Trimmomatic V0.32 (Bolger et al. 2014) and followed by another run of FASTQC. Trimmed reads from Baxendale (S_5S_7), Brown Nonpareil (S_1S_7), Keanes (S_7S_{27}), McKinlays (S_7S_8), Nonpareil (S_7S_8), Capella (S_7S_f), Mira (S_7S_f), Carina (S_7S_f), T5 (S_7S_f) and T7 (S_7S_f) were aligned to the S_7 reference sequence (AB081587) (Ushijima et al. 2003) using the BWA-mem algorithm as implemented in the Burrows-Wheeler alignment (BWA) tool (Li and Durbin 2009). Binary Alignment/Map (BAM) files were visualised using the Tablet graphical viewer (Milne et al. 2010).

In addition, cleaned reads from each individual were assembled using Mimicking Intelligent Read Assembler (MIRA) version 4.0.2 (Chevreux et al. 2004). The resulting contigs were mapped to the S_7 reference sequence (AB081587). These contigs were visualised using CONTIGuator software (Galardini et al. 2011). Further, all large contigs (N ≥ 50 and contig size ≥ 500 bp) were aligned to the verified S_7 sequence using the Map to Reference function in Geneious software version 9.0.2 (Kearse et al. 2012). Polymorphisms in the *S* locus between two haplotypes in each sample and among samples were graphically visualised using the Integrated Genomic Viewer (IGV) version 2.3 (Robinson et al. 2011). The consensus sequence of the *S* locus of each cultivar and selection was generated using SAM (Sequence Alignment/Map) tools (Li 2011; Li et al. 2009).

The resulting *S* locus sequences for each individual were evaluated for completeness. Based on sequences from individuals with the genotypes S_7S_f, S_7S_8 and S_1S_7, it was possible to obtain the full sequences of haplotypes S_1, S_7, S_8 and S_f by using the S_7 haplotype sequence as a reference and

using the variant call format (VCF) in VCFtools (Danecek et al. 2011). Further sequence analysis was conducted as follows:

1. Sequences from 10 individuals known to carry the S_7 haplotype (Baxendale (S_5S_7), Brown Nonpareil (S_1S_7), Keanes (S_7S_{27}), McKinlays (S_7S_8), Nonpareil (S_7S_8), Capella (S_7S_f), Mira (S_7S_f), Carina (S_7S_f), T5 (S_7S_f) and T7 (S_7S_f)) were aligned with the reference (AB081587) using the Clustal W multiple sequence alignment algorithm (Thompson et al. 1994) as implemented in Geneious software version 9.0.2 (Kearse et al. 2012). Firstly, each individual S_7 haplotype was compared with the published sequence. Secondly, a consensus S_7 haplotype sequence was established by comparing among S_7 haplotype sequence obtained for each individual. The resulting verified S_7 haplotype sequence was used as a reference to obtain the other haplotype sequences.

2. Based on the individuals with genotypes S_1S_7 (Brown Nonpareil), S_7S_8 (Nonpareil and McKinlays) and S_7S_f (Carina, Mira, Capella, T5 and T7) full sequences of haplotypes S_1, S_8 and S_f were obtained using the verified S_7 haplotype sequence as a reference.

3. Sequences of other haplotypes (S_3, S_5, S_6, S_9, S_{13}, S_{14}, S_{19}, S_{22}, S_{23}, S_{25} and S_{27}) were obtained in a similar manner using the newly derived haplotype sequences (S_7, S_8, S_1 and S_f) as references.

4. All completed haplotype sequences were annotated with reference to the S_7 haplotype sequence (AB081587) and the physical distances between the genes (*SLF*, *S-RNase* and *SFB*) were determined.

5. Conserved blocks in the *S* locus (using complete haplotype sequences) and in individual genes (*SLF*, *S-RNase* and *SFB)* were identified using the Gblocks version 0.91b (Castresana 2000) tool in the Phylogeny.fr online server (www.phylogeny.fr) using the 'less stringent' data selection setting. Conserved blocks in the complete *S* locus and the *SLF* gene were identified using DNA sequences. Conserved blocks in the *S-RNase* and *SFB* genes were identified using both DNA and predicted amino acid sequences.

6. Pair-wise sequence differences among complete *S* locus sequences of S_1, S_7, S_8 and S_f haplotypes and among the complete allele sequences for the *SLF*, *S-RNase* and *SFB* genes, were determined using the Clustal W multiple sequence alignment algorithm (Thompson et al. 1994) as implemented in Geneious software version 9.0.2 (Kearse et al. 2012).

3.4.6 Intron–exon structure of the *S-RNase* gene

The *S-RNase* gene sequences from all the *S* alleles obtained here were annotated using published annotations for relevant *S* alleles (AB433984 for S_f, AB011469 for S_1, AF490505 for S_3, DQ150569 for S_5, AM231657 for S_6, AY291118 for S_7, AB481108 for S_8, AF454001 for S_9, AM231662 for S_{13}, AM231663 for S_{14}, AM231668 for S_{19}, AM231671 for S_{22}, AB488496 for S_{23}, AM231673 for S_{25} and AM231675 for S_{27}). Length polymorphisms in the two introns of the *S-RNase* gene were determined.

3.4.7 Distribution of LTR retrotransposons in the *S* locus

Fifteen haplotypes (S_1, S_3, S_5, S_6, S_7, S_8, S_9, S_{14}, S_{22}, S_{23}, S_{25}, S_{27} and S_f) were analysed using LTR_Finder (Xu and Wang 2007) (http://tlife.fudan.edu.cn/ltr_finder) to detect LTR retrotransposons, using *Arabidopsis thaliana* tRNAdb to predict protein binding sites.

3.4.8 Phylogenetic relationships among the *S-RNase* and *SFB* alleles

A multiple sequence alignment of the deduced amino acid sequences for conserved region 1 through conserved region 5 of the *S-RNase* gene of fifteen haplotypes (S_1, S_3, S_5, S_6, S_7, S_8, S_9, S_{14}, S_{22}, S_{23}, S_{25}, S_{27} and S_f) was generated using the MUSCLE sequence alignment algorithm (Edgar 2004) as implemented in Geneious software version 9.0.2 (Kearse et al. 2012). Poorly aligned positions within the deduced amino acid alignment were removed using the Gblocks tool (Castresana 2000) at the Phylogeny.fr:Gblocks online server (www.phylogeny.fr). The resulting refined alignment data set and ModelGenerator v0.85 (Keane et al. 2006) were employed to find a best substitution model to implement in the phylogenetic construction. Two phylogenetic trees were constructed using MEGA7 (Kumar et al. 2016) software. One considered deduced amino acid sequences from almond haplotypes only. The other also considered sequences from the other *Prunus* species (*P. avium* (CAC27784), *P. armeniaca* (AAO33412), *P. cerasus* (ABW74348), *P. pseudocerasus* (ABY65900) and *P. spinosa* (ABG76209), *Physalis crassifolia* (AAB37216), *Solanum lycopersicum* (AEM37151), *Antirrhinum hispanicum* (AJ440731), *Nicotiana alata* (U08860), *Pyrus pyrifolia* (BAB61926) and *Malus* × *domestica* (AAA61820).

Similarly, a phylogenetic tree was constructed using the deduced amino acid sequences from 11 *S* alleles (S_1, S_3, S_5, S_7, S_8, S_9, S_{22}, S_{23}, S_{25}, S_{27} and S_f) of the almond *SFB* gene.

3.5 Results

3.5.1 Enzyme and buffer combination suitable for the *S* locus amplification

Of the five enzymes tested, Phusion High-Fidelity DNA polymerase was selected because it provided very clear and stable amplicons after 35 cycles of PCR. The Phusion®HF buffer was selected because it resulted in more consistent and higher-quality amplicons than the GC buffer.

3.5.2 Primer testing and PCR amplification of the *S* locus

Primer combinations that produced clear bands for the initial two samples tested were selected to run with all DNA samples. The number of primer combinations needed to amplify the entire *S* locus varied among samples, from only seven for Nonpareil up to 25 for Antoñeta (S_1S_f), Mandaline (S_1S_f) and Johnston ($S_{23}S_{25}$). Electrophoretic analysis of the amplified products indicated that some regions (1 to 8,000 bp, 64,000 to 71,000 bp) of the *S* locus were more readily amplified than other regions. Some regions showed length polymorphisms.

3.5.3 Library preparation and sequencing

3.5.3.1 Strong and weak DNA bulks

The DNA concentrations of the strong bulks ranged from 6 ng/µL to 22 ng/µL while that of the weak bulks ranged from 0.8 ng/µL to 10 ng/µL. The strong:weak mixing ratios used ranged from of 1:14 for T7 to 1:2 for Nonpareil.

3.5.3.2 Sequence data analysis

The total number of MiSeq sequence reads generated was 27 million. Twenty-four million reads (QC $\geq$ 30) were selected for further analysis. Visualisation of the resulting BAM files using Tablet indicated that some regions of the *S* locus had amplified well from particular samples, with the presence of single nucleotide polymorphisms (SNPs) indicating the presence of two haplotypes (Figs. 3.1 and 3.2). The sequences assembled with BWA and MIRA were identical. Complete *S* locus sequences were obtained for S_7 (Nonpareil (S_7S_8), Capella (S_7S_f), Mira (S_7S_f), Carina (S_7S_f), T5 (S_7S_f), T8 (S_7S_f)), S_1 (Brown Nonpareil (S_1S_7), S_8 (Nonpareil (S_7S_8)) and S_f (Capella (S_7S_f), T2 (S_7S_f), Carina (S_7S_f), T5 (S_7S_f) and T8 (S_7S_f)). Partial *S* locus sequences were obtained for S_3 (Lauranne (S_3S_f)), S_5 (Carmel (S_5S_8)), S_6 (Ramillete (S_6S_{23})), S_9 (Vairo (S_9S_f)), S_{13} (Kapareil (S_8S_{13})), S_{14} (Biggs Hardshell (S_6S_{14})), S_{19} (Tom Strout ($S_{19}S_{22}$)), S_{22} (Strout's Papershell ($S_{22}S_{25}$)), S_{23} (Chellaston (S_7S_{23})), S_{25} (Johnston ($S_{23}S_{25}$)) and S_{27} (Keanes (S_7S_{27})). The length of the *S* locus of each haplotype ranged from 71,444 to 72,550 bp. For all of the haplotypes evaluated here, sequences

were obtained for the regions corresponding to between 1 and 1,066 bp, 2,500 and 6,700 bp, 7,000 and 17,000 bp and 64,000 and 71,000 bp region of the S_7 haplotype. For the partially sequenced haplotypes, the sequences obtained represented between 47% (S_6) and 99% (S_{23}) of the S_7 haplotype sequence.

Fig. 3.1 Sequence variations observed in 48 almond cultivars and breeding lines used in this research. DNA sequences were assembled using the BWA assembler. The resulting sequences from the BAM files of these cultivars were visualised using the Integrative Genomics Viewer (IGV) software to detect sequence variations. Cultivars are listed in the left panel. Data from each sample are displayed in a horizontal row. The length of the *S* locus reference (AB081587) and sequencing depth of each position of a sample are indicated on the top panel in dark grey. In the lower panel, the genotypes for each base-pair position in an individual are shown. Heterozygous variants are in dark blue, homozygous variants are in cyan, positions with low sequence depth are in white and the reference genotype is in grey.

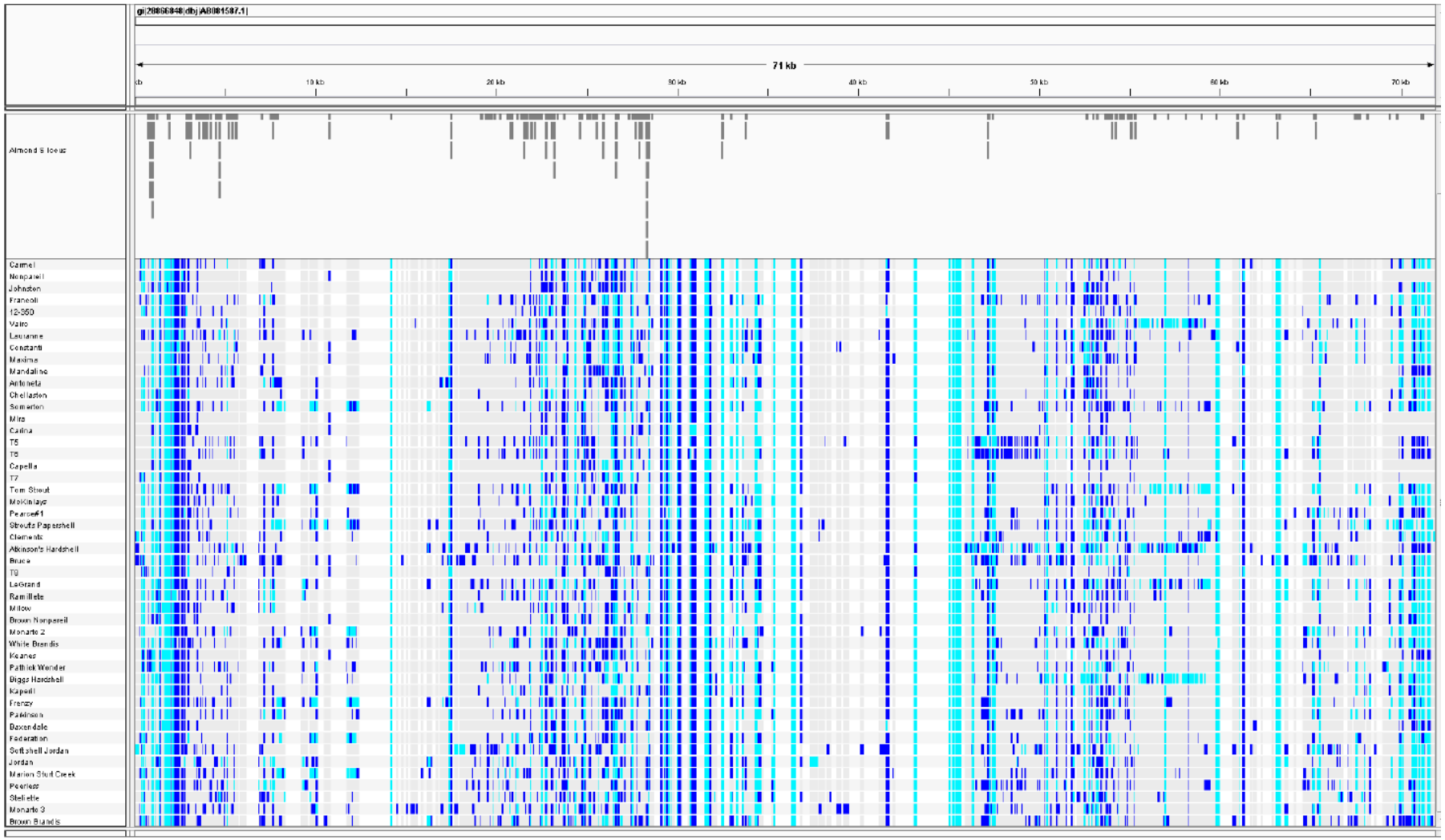
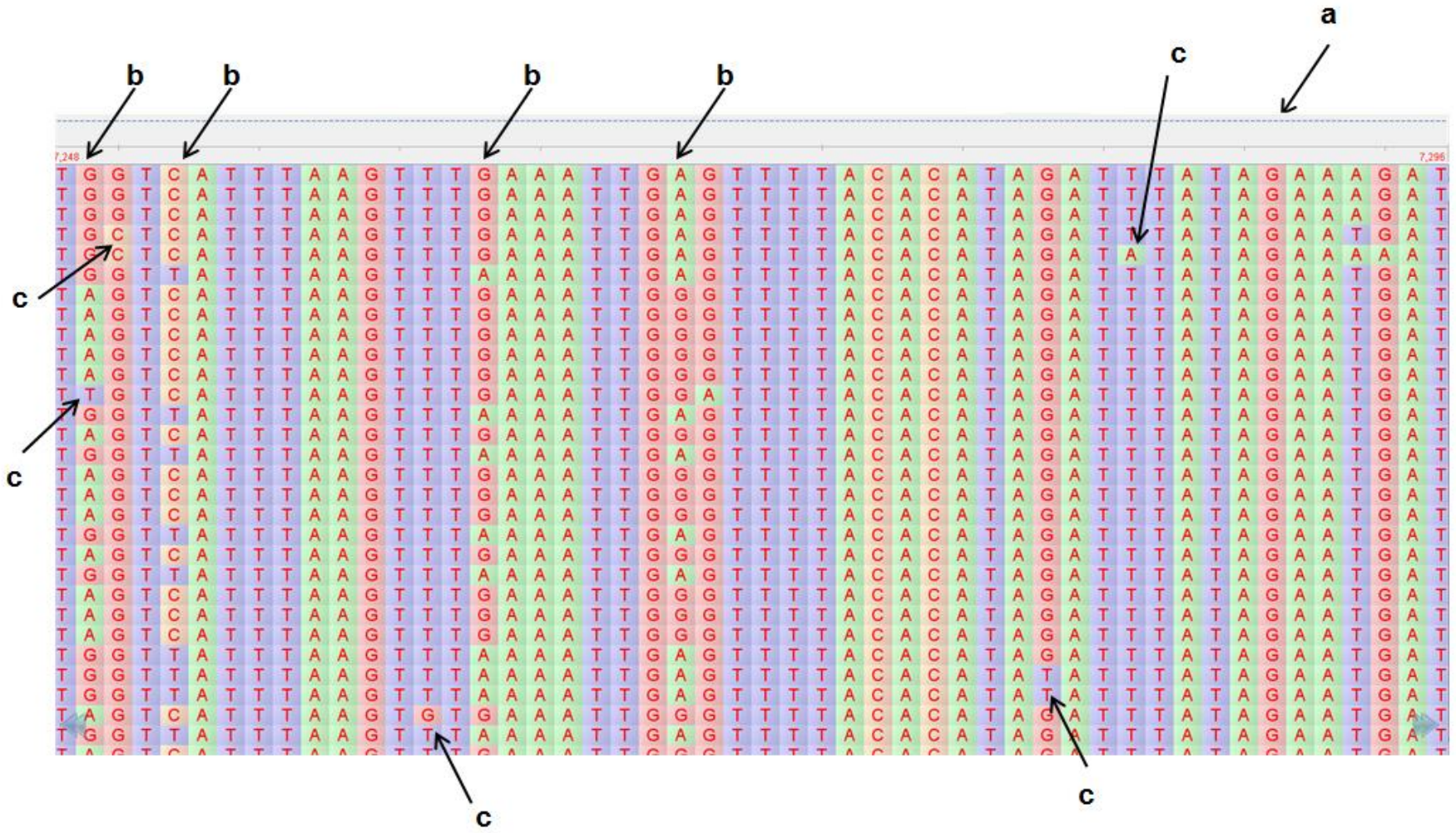

gi|2866848|dbj|AB001587.1|
71 kb
0 kb
10 kb
20 kb
30 kb
40 kb
50 kb
60 kb
70 kb
Almond Blossom
Carmel
Nonpareil
Johnston
Francoli
12-350
Mairo
Lauranne
Constanti
Maxima
Mandalino
Antoneta
Chellaston
Somerton
Mira
Carina
T5
T6
Capella
T7
Tom Sheut
McKinlays
Pearsef1
Sheufs Papershell
Clements
Atkinson's Hardshell
Bruce
T9
LeGrand
Ramillete
Mitow
Brown Nonpareil
Monarte 2
White Brandis
Keanes
Pathick Wonder
Diggs Hardshell
Kapadi
Frenzy
Padinson
Baxendale
Federation
Softshell Jordan
Jordan
Marion Stuft Creek
Peerless
Stellette
Monarte 3
Brown Brandis

Fig. 3.2 Visualisation of a BAM file resulting from assembling the sequence reads from the almond cultivar, Mira (S_7S_f), using the BWA (Burrows-Wheeler) alignment tool. The region from 7,248 bp to 7,296 bp of the *S* locus is shown. The reference sequence (AB081587) is represented by the blue horizontal dotted line on the top (a). Thirty sequence reads are displayed in horizontal rows and each nucleotide position is displayed in a vertical column. Each position indicates the nucleotide obtained by sequencing. The nucleotide positions that are heterozygous are designated (b) and sequencing errors are designated (c).


a
b
b
b
b
c
c
c
c
c

3.5.3.3 Sequence variation and gene organisation in the *S* locus

Comparison of the entire *S* locus nucleotide sequences of the S_1, S_7, S_8 and S_f haplotypes indicated high sequence variation among them, with sequence identity ranging from 51 to 84% (Table 3.2). For each of these haplotypes, the main features in the *S* locus were *SLF, S-RNase* and *SFB* genes and LTRs (Fig. 3.3).

Table 3.2. Percentage of DNA sequence identity among almond *S* haplotypes using the entire *S* locus sequences.

***S* haplotype**	***S*$_1$**	***S*$_7$**	***S*$_8$**
S_7	79	-	
S_8	83	84	-
S_f	60	52	51

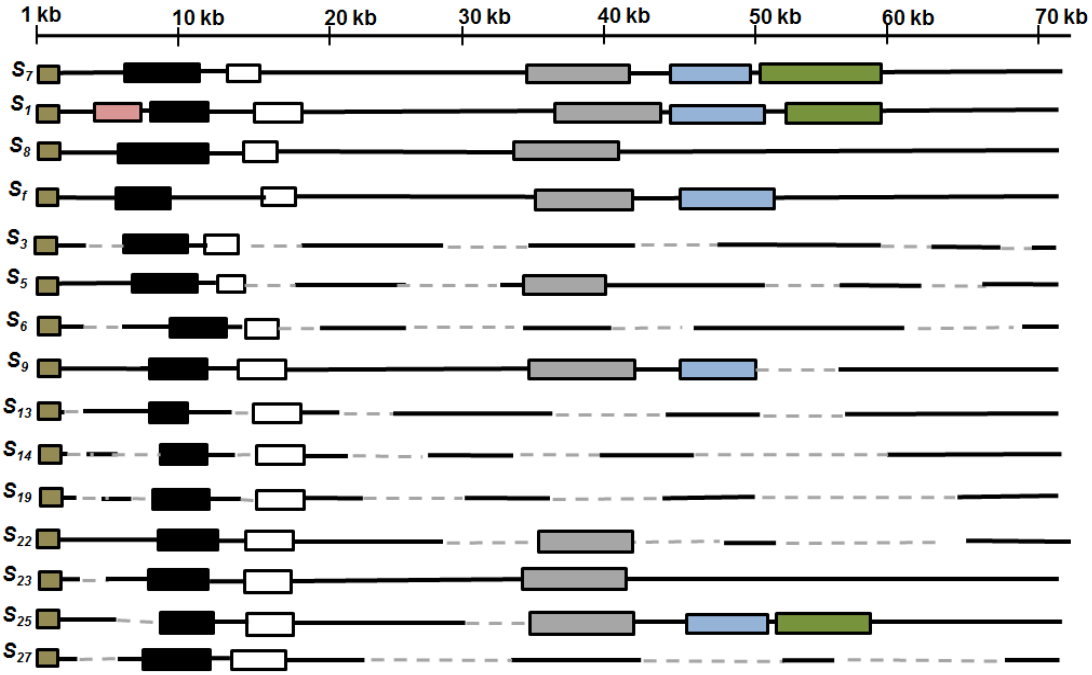


Fig. 3.3 Structure of the almond *S* locus. The *SLF* (yellow box), *S-RNase* (black box), *SFB* (white box) genes and long terminal repeat retrotransposons (LTRs) are shown. LTRs are grouped as described by Ushijima et al. (2003). LTR1 (grey box), LTR2 (blue box), LTR3 (green box) and a new LTR detected for the S_1 haplotype (pink box) are shown. The horizontal bar on the top indicates every 10 kb region in the almond *S* locus. The black lines indicate where sequences are available within the *S* locus and grey dotted lines indicate the regions that did not amplify (gaps) within the *S* locus based on the sequence information generated in this research.

Across all the haplotypes, the region from the start up to 1,500 bp of the *S* locus was less variable than other regions of the *S* locus. This is the region containing the *SLF* gene, which is about 1,200 bp long and has no intron. Sequence identity among the *SLF* alleles ranged from 70 to 98% at the DNA level (Table 3.3) and from 79 to 98% at the amino acid level.

Table 3.3. Percentage of nucleotide identity of the *SLF* alleles in almond.

***S* allele**	***S*$_1$**	***S*$_3$**	***S*$_5$**	***S*$_6$**	***S*$_7$**	***S*$_8$**	***S*$_9$**	***S*$_{13}$**	***S*$_{14}$**	***S*$_{19}$**	***S*$_{22}$**	***S*$_{23}$**	***S*$_{25}$**	***S*$_{27}$**
S_3	96	-												
S_5	90	88	-											
S_6	96	95	88	-										
S_7	94	95	94	95	-									
S_8	92	93	85	93	93	-								
S_9	96	94	88	95	95	92	-							
S_{13}	82	74	81	95	91	71	73	-						
S_{14}	90	89	85	81	93	88	90	72	-					
S_{19}	98	97	98	96	93	93	97	97	91	-				
S_{22}	94	93	92	94	94	93	96	89	92	93	-			
S_{23}	94	93	87	91	90	89	93	70	88	94	94	-		
S_{25}	88	83	87	94	96	83	81	73	81	93	93	73	-	
S_{27}	75	82	85	91	91	79	81	81	78	78	99	79	96	-
S_f	92	91	84	92	93	91	92	70	81	81	92	92	82	78

Among the *SLF* alleles no clear variable regions were observed (Fig. 3.4). Of the tested haplotypes, the greatest sequence variation (amino acid identity of just 70%) was observed between the alleles SLF_{23} *and* SLF_{25} and the least sequence variation (amino acid identity of 98%) was observed between the alleles SLF_{19} *and* SLF_{27}.

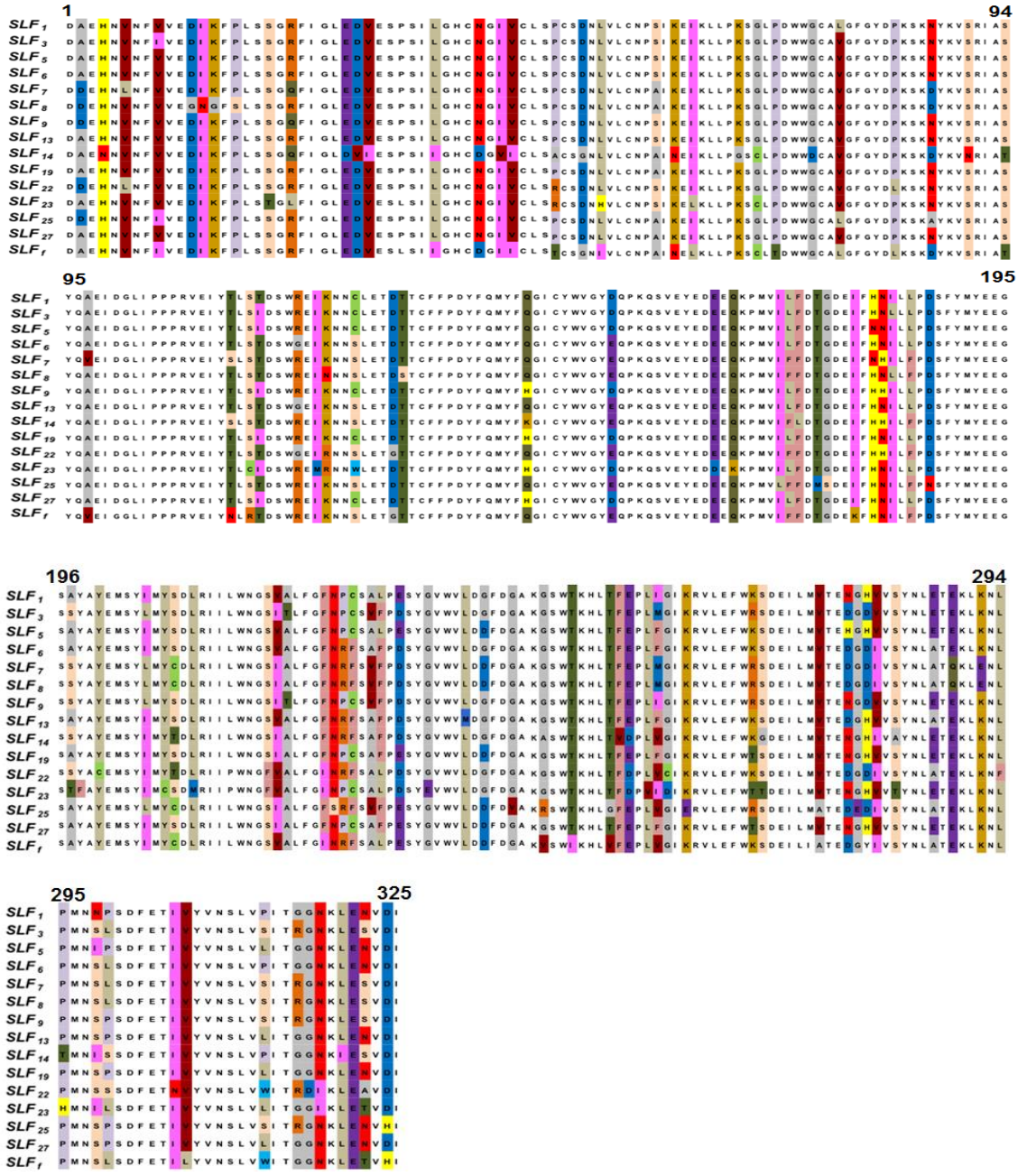


Fig. 3.4 Alignment of deduced amino acid sequences of 15 *S* alleles from the *SLF* gene in almond. Amino acids that are different in each position are shaded by assigning a colour to each amino acid.

The region between 6,500 and 17,000 bp of the *S* locus, in which both the *S-RNase* and *SFB* genes are located, exhibited particularly high sequence variation. Complete *S-RNase* gene sequences

were obtained for the haplotypes *S1*, *S5*, *S7*, *S8*, *S23*, *S27* and *Sf*. For all other haplotypes (*S3*, *S6*, *S9* and *S19*) sequences were obtained for the region between conserved region 1 (C1) and conserved region 5 (C5). Among the complete *S-RNase* gene sequences, the longest was for *S8-RNase* (4,500 bp) and the shortest was for *S1-RNase* (1,010 bp). The pairwise nucleotide sequence identity of the complete *S-RNase* alleles ranged from 19% (between *S1 and S8*) to 51% (between *S7 and S27*) (Table 3.4). Among the completely sequenced alleles, there were eight conserved blocks longer than 10 bp. Together; these comprised only 372 bases (12% of the total). The deduced amino acid sequences of these *S-RNase* alleles contained the five conserved regions (C1, C2, C3, RC4 and C5) and the hypervariable (RHV) region that are considered characteristic of Rosaceae *S-RNases* (Ushijima et al. 1998) (Fig. 3.5). They also contained the two variable regions (VR1 and VR2), which have been identified between the RC4 and C5 regions (Gu et al. 2012; Ortega et al. 2006). Additionally, three highly variable regions were detected between C1 and C2, at the beginning of RHV region and at the beginning of RC4 region (regions A, B and C in Fig. 3.5).

The C1 region was highly conserved among these haplotypes. Within the C2, C3, RC4 and C5 regions there were a few nucleotide polymorphisms. At one position in the C2 region the self-fertile haplotype *Sf* differed from the self-incompatible haplotypes, with the nucleotide A detected for the *Sf* haplotype, the nucleotide C detected for the haplotypes *S1*, *S3*, *S5*, *S6*, *S7*, *S8*, *S9*, *S14*, *S22*, *S23*, *S25* and *S27* and the nucleotide T detected for the haplotypes *S13* and *S19*. In the C3 region, a SNP (C/T) was detected between *S6* and *S7* (C) and the other *S* haplotypes (T). In the same region, an INDEL (TGGAA/-) was observed between the self-fertile haplotype (insertion) and the self-incompatible haplotypes (deletion). Five SNPs were detected in the RC4 region, one of which distinguished the self-fertile haplotype (G) from the self-incompatible haplotypes (C). In the C5 region, two SNPs and two INDELs were observed. One of the INDELs (GTT/-) distinguished the *S23* haplotype from all other haplotypes. The other (AGC/-) distinguished the *S14* haplotype from all other haplotypes.

Table 3.4. Percentage of nucleotide identity of the *S-RNase* alleles in almond. Only *S* alleles that had complete *S-RNase* gene sequences were considered in this analysis.

S allele	S_1	S_5	S_7	S_8	S_{13}	S_{14}	S_{22}	S_{23}	S_{25}	S_{27}
S_5	36	-								
S_7	28	34	-							
S_8	39	29	26	-						
S_{13}	25	34	45	20	-					
S_{14}	29	34	46	21	37	-				
S_{22}	28	27	46	20	41	39	-			
S_{23}	40	39	29	20	31	25	30	-		
S_{25}	40	39	30	29	20	31	25	34	-	
S_{27}	34	24	51	23	49	34	39	33	31	-
S_f	37	32	29	27	27	27	22	29	28	25

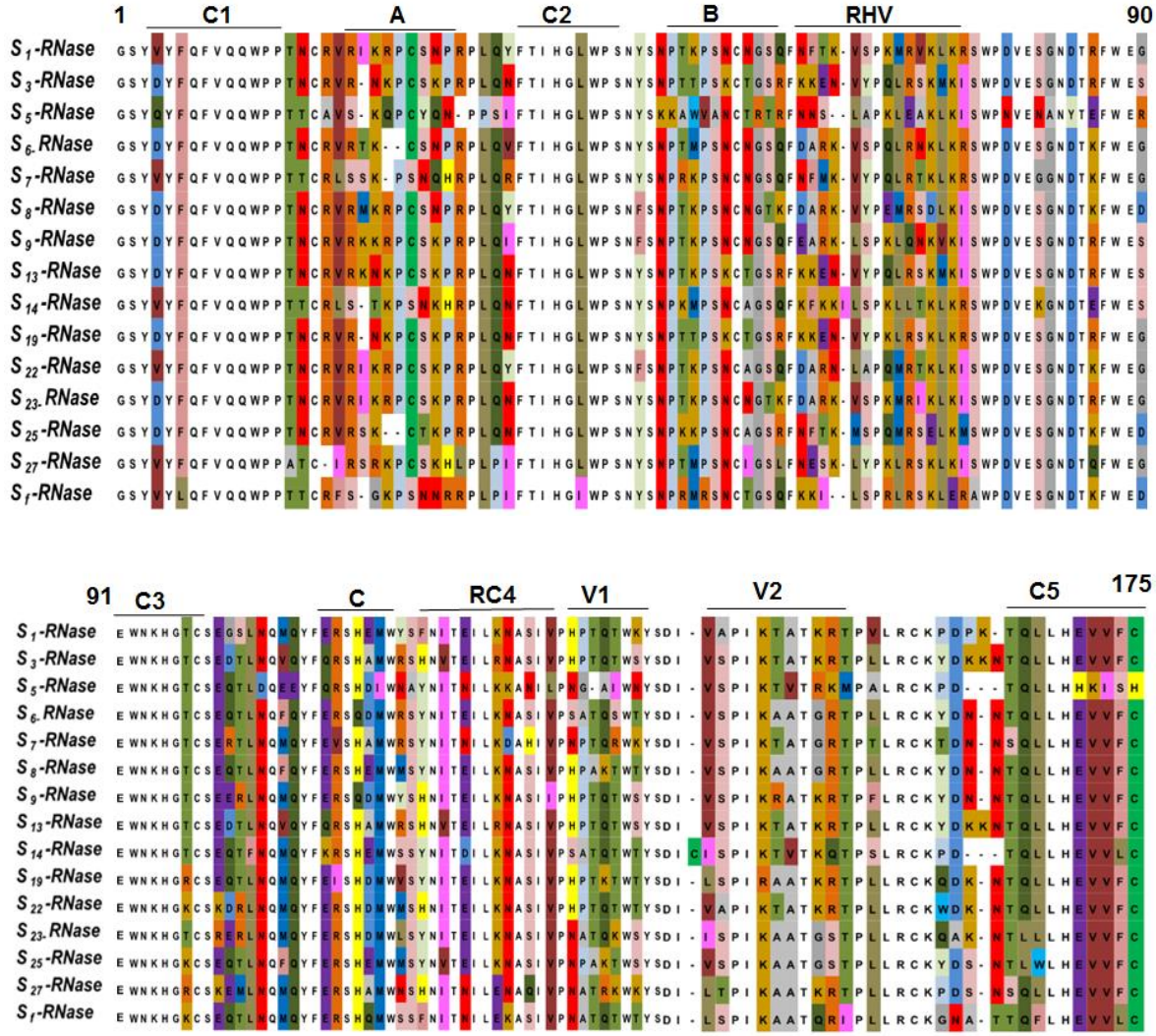


Fig. 3.5 Alignment of deduced amino acid sequences of 15 *S* alleles from the *S-RNase* gene in almond. Conserved regions (C1–C5) and the hypervariable region (RHV) are shown as described by Ushijima et al. (1998), two variable regions (V1 and V2) are shown as described by Ortega et al. (2006) and three additional variable regions observed in this research (A, B and C) are underlined. Amino acids that are different in each position are shaded by assigning a colour to each amino acid.

The length of complete *SFB* allele sequences ranged from 1,100 bp (*SFB*$_9$) to 1,500 bp (*SFB*$_1$) and partial sequences were obtained for *SFB*$_6$, *SFB*$_{13}$ and *SFB*$_{14}$. Sequence identity of complete *SFB* alleles at the DNA level ranged from 35% (between *SFB*$_7$ and *SFB*$_f$) to 86% (between *SFB*$_1$ and *SFB*$_{23}$). Among the completely sequenced 11 alleles (*SFB*$_1$, *SFB*$_3$, *SFB*$_5$, *SFB*$_7$, *SFB*$_8$, *SFB*$_9$, *SFB*$_{14}$, *SFB*$_{22}$, *SFB*$_{23}$, *SFB*$_{27}$ and *SFB*$_f$), there were 25 conserved blocks longer than 10 bp. The deduced

amino acid sequences of these alleles contained a F-box motif, two variable regions (V1 and V2), and two hypervariable regions (HVa and HVb) (Table 3.5 and Fig. 3.6) as previously described for the *Prunus SFB* gene (Ushijima et al. 2003). Within the F-box motif, SFB_6, SFB_{13}, SFB_{14} and SFB_{19} had an insertion of a single arginine (R), while SFB_{23} and SFB_{27} had a deletion of eight amino acids. In some regions of the *SFB* gene, the variability observed here differed from that had been previously described for almond. For example, many residues within variable region 1 (V1) were conserved among these *S* alleles, while two additional short highly variable regions (A and B) were detected between V1 and V2 regions.

Table 3.5. Percentage of nucleotide identity in the almond *SFB* alleles. Only *SFB* alleles that had complete *SFB* gene sequences were considered in this analysis.

S allele	S_1	S_3	S_5	S_7	S_8	S_9	S_{22}	S_{23}	S_{25}	S_{27}
S_3	79	-								
S_5	71	81	-							
S_7	36	39	36	-						
S_8	39	30	29	39	-					
S_9	84	79	80	38	39	-				
S_{22}	85	49	81	38	42	59	-			
S_{23}	86	79	81	37	38	42	51	-		
S_{25}	78	85	76	36	38	54	83	83	-	
S_{27}	85	59	81	37	38	43	81	82	83	-
S_f	39	39	38	35	33	39	38	38	37	40

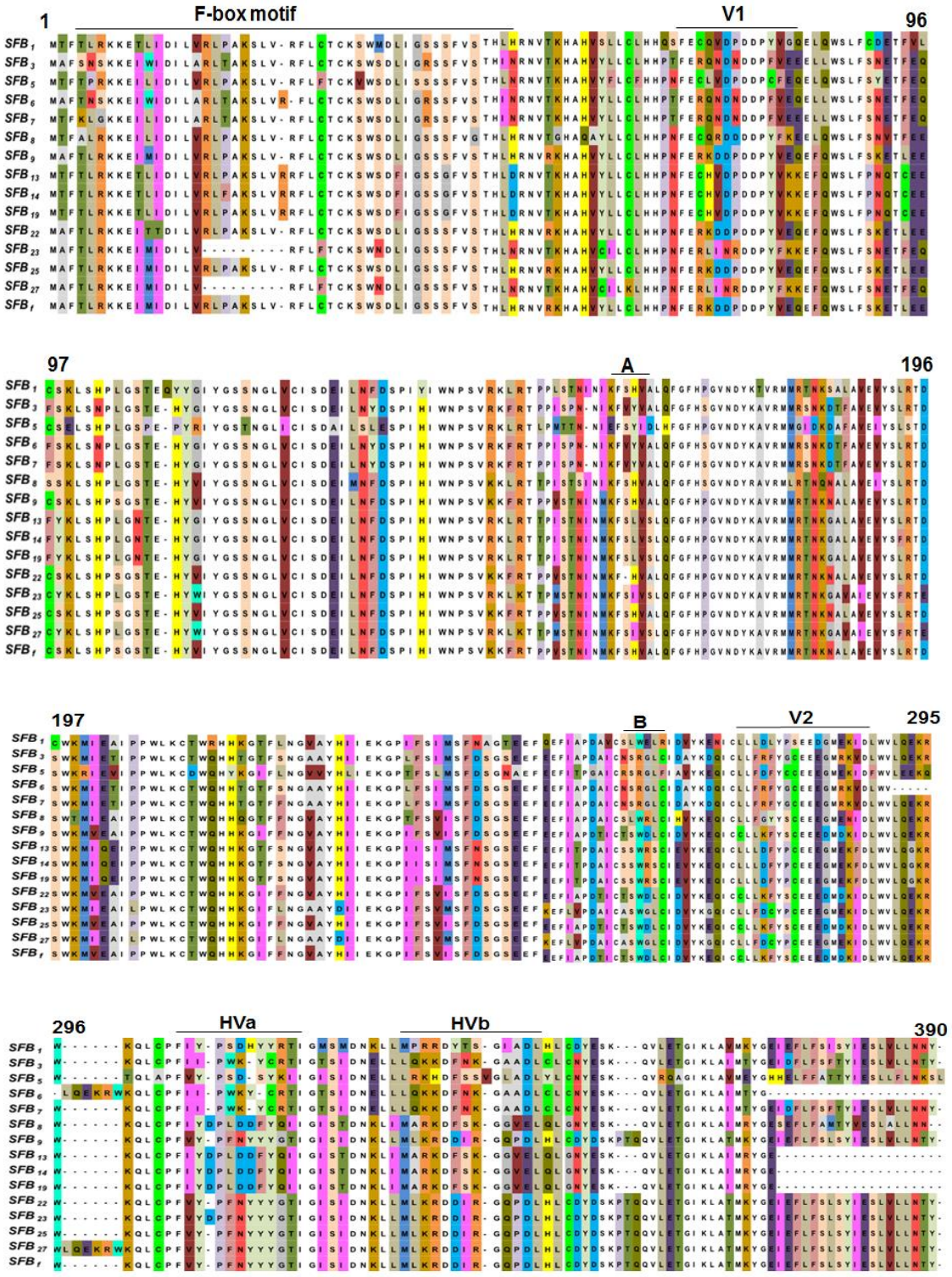


Fig. 3.6 Alignment of deduced amino acid sequences of 15 *S* alleles from the *SFB* gene in almond. For all except the alleles *SFB*$_6$, *SFB*$_{13}$, *SFB*$_{14}$ and *SFB*$_{19}$ complete sequences were used. The F-box motif, two variable regions (V1 and V2), two hypervariable regions (HVa and HVb) and two variable regions identified in this research (A and B) are underlined. Amino acids that are different in each position are shaded by assigning a colour to each amino acid.

Among the ten haplotypes (S_1, S_3, S_5, S_7, S_8, S_9, S_{22}, S_{23}, S_{25}, S_{27} and S_f) for which complete sequences were generated in the region between the *S-RNase* and *SFB* genes, the length of the region between these genes ranged from 1,200 (S_9) to 6,600 bp (S_f). These intergenic regions were rich in adenine (A) and thymine (T) (65 to 70% A/T). DNA sequence identity in this region ranged from 21% (between S_f and S_1) to 98% (between S_7 and S_{23}) (Table 3.6).

The region between 17,000 bp and 64,000 bp did not sequence uniformly and gaps were observed in the S_3, S_5, S_6, S_9, S_{13}, S_{14}, S_{22}, S_{19}, S_{23}, S_{25} and S_{27} haplotypes. However, the region between 64,000 and 71,000 bp was completely sequenced in almost all haplotypes and had sequence identities above 92%.

3.5.3.4 Intron-exon structure of the *SLF*, *S-RNase* and *SFB* genes

No introns were detected in the *SLF* gene. The *S-RNase* gene contained two introns while the *SFB* gene contained one intron. Both *S-RNase* introns exhibited length polymorphisms among alleles, with intron 2 exhibiting more polymorphism than intron 1 (Fig. 3.7). Sequences flanking the exon/intron junctions were highly conserved. The intron of the *SFB* gene of these haplotypes exhibited little polymorphism.

3.5.3.5 Distribution of LTR retrotransposons in the *S* locus

Long terminal repeat retrotransposons were detected in all haplotypes except S_3, S_{13}, S_{14}, S_{19} and S_{27} haplotypes. All of the LTRs are *Ty1/copia*-like retrotransposons, with TG/CA boxes at the 5′ and 3′ LTR regions. The protein binding sites in these LTRs are TyrGTA, IleAAT, MetCAT and AlaTGC. The number of LTRs detected ranged from one (S_5, S_8, S_{22} and S_{25}) to four (S_1). LTRs were scattered across the *S* locus and the size of LTRs ranged from 1,600 bp (the shortest LTR in S_1) to 7,800 bp (the longest in S_{25}) (Fig. 3.3).

Table 3.6. Percentage of nucleotide identity in the region between the *S-RNase* and *SFB* genes in almond *S* haplotypes. *S* haplotypes with the complete intergenic regions were considered in this analysis.

S allele	S_1	S_3	S_5	S_7	S_8	S_9	S_{22}	S_{23}	S_{25}	S_{27}
S_3	52	-								
S_5	60	61	-							
S_7	68	62	57	-						
S_8	51	65	58	65	-					
S_9	55	67	59	64	39	-				
S_{22}	52	69	62	65	42	51	-			
S_{23}	56	59	65	98	38	45	51	-		
S_{25}	59	58	64	67	38	54	78	83	-	
S_{27}	71	59	63	27	38	43	81	82	59	-
S_f	21	39	45	35	33	39	38	38	37	52

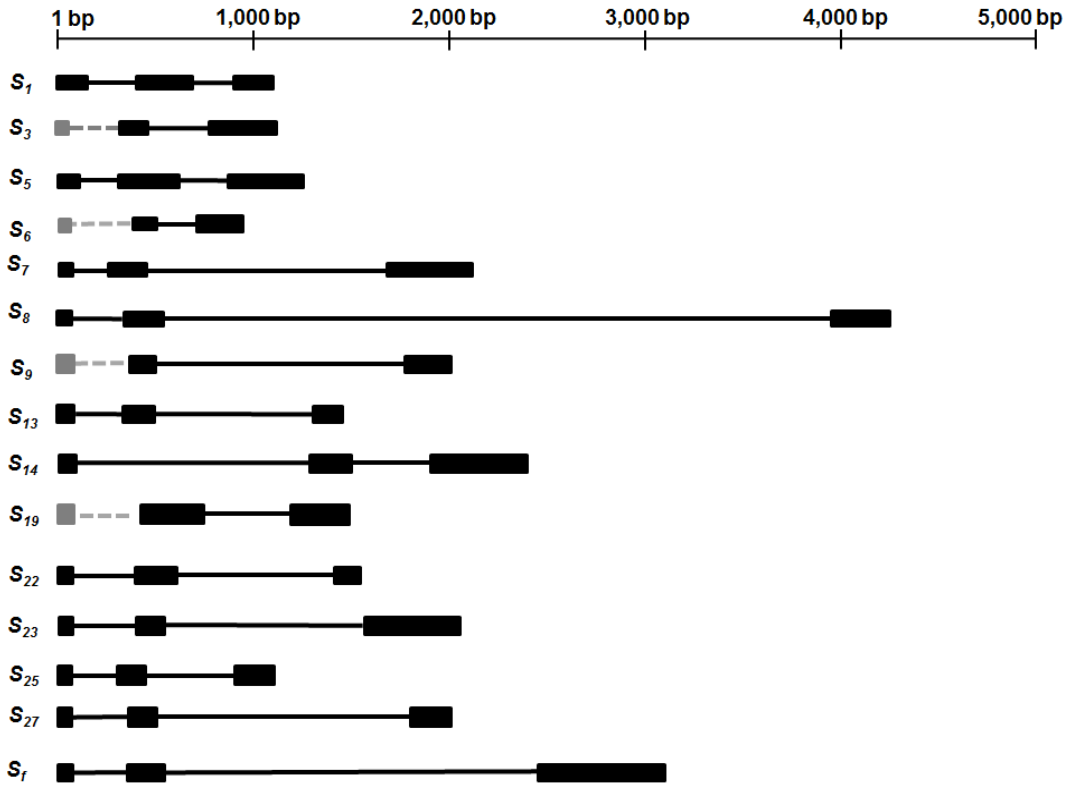


Fig. 3.7 Intron–exon structure of 15 almond *S-RNase* alleles from the sequences generated in this analysis. Length of the introns and exons are represented by the bar on the top. Introns are shown in black lines and exons are ahown in black boxes. In the sequences, where partial sequences are available, the missing regions are indicated in grey.

3.5.3.6 New sequence information

The S_7 haplotype sequence obtained from Nonpareil in this analysis showed 99% nucleotide similarity with the published S_7 haplotype sequence (AB081587) with only five nucleotide mismatches. The S_7 haplotype sequence obtained from other cultivars (Brown Nonpareil (S_1S_7), Keanes (S_7S_{27}), McKinlays (S_7S_8), Capella (S_7S_f), Mira (S_7S_f), Carina (S_7S_f), T5 (S_7S_f) and T7 (S_7S_f)) showed 97% (Keanes) to 99% (Capella, Mira, Carina, T5 and T7) DNA similarity with the published S_7 haplotype sequence .

Comparison of the S_7 haplotype sequences obtained from eight cultivars and breeding selections enabled the determination of 21 nucleotides that had not been determined in the published S_7

haplotype sequence (AB081587). Similarly, eight nucleotides that had not been identified in the published *S_8-RNase* sequence (AB481108) were identified here, making it possible to generate a complete sequence for this allele.

Some published *S-RNase* sequences (*S_3* (AF490505), *S_6* (AF510419), *S_9* (AF454001), *S_{13}* (AM231662), *S_{19}* (AM231668), *S_{22}* (EF690370) and *S_{25}* (EF690372)) did not have complete sequences even for the region from C1 to the C5. In this analysis, complete sequences were generated for that region.

3.5.3.7 Phylogenetic analysis of the almond *S* locus

The most suitable substitution model predicted by the ModelGenerator for the deduced amino acid sequences from the *S-RNase* datasets (a data set containing solely almond and a dataset containing almond and the other *Prunus*, *Malus*, *Pyrus*, *Antirrhinum* species and Solanaceae) and the *SFB* dataset was the Jones-Taylor-Thronton matrix-based model (Jones et al. 1992) with Gamma distribution parameter (JTT+G). One thousand bootstrap replications were used to generate consensus trees. The phylogenetic tree resulting from almond *S-RNase* gene sequences indicated that the *S_{13}* and *S_3* alleles are closely related to each other, while the *S_5*, *S_7*, *S_{14}* and *S_f* alleles are distantly related to other *S* alleles (Fig. 3.8, Fig. S1.1). Comparisons of the *S-RNase* sequences from these 15 *S* alleles with *S-RNase* sequences from other species indicated that the almond *S-RNases* are similar to those from other *Prunus* species but diverge from *S-RNases* from *Malus, Pyrus, Antirrhinum* species and the Solanaceae (Fig. 3.9, Fig. S1.2).

The phylogenetic tree resulting from the almond *SFB* allele sequences also indicated that the *S_5* and *S_f* alleles are distantly related to other *S* alleles (Fig. 3.10, Fig. S1.3).

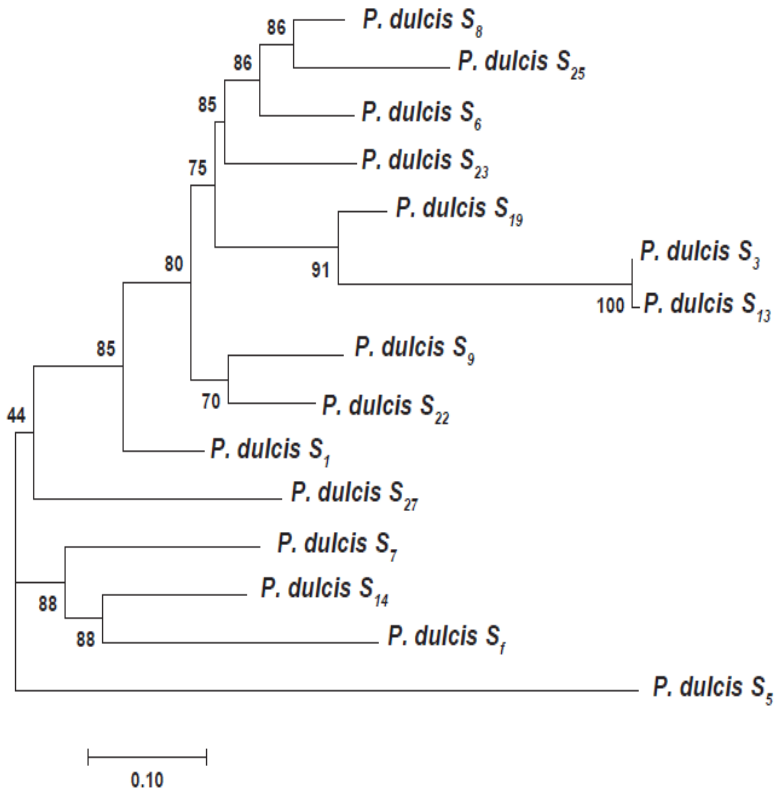


Fig. 3.8 Phylogenetic relationships based on 15 *S* alleles from the *S-RNase* gene in almond, using the deduced amino acid sequences from conserved region 1 (C1) to conserved region 5 (C5) of the *S-RNase* gene. A total of 162 amino acid positions were in the final dataset. Phylogenetic relationships were inferred by using the Maximum Likelihood method based on the JTT matrix based model (Jones et al. 1992). The tree with the highest log likelihood (-2173.44) is shown. The branch support values are shown next to the branches. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+*G*, parameter = 0.6285)). The tree is drawn to scale, with a scale bar 0.1 amino acid substitutions per site.

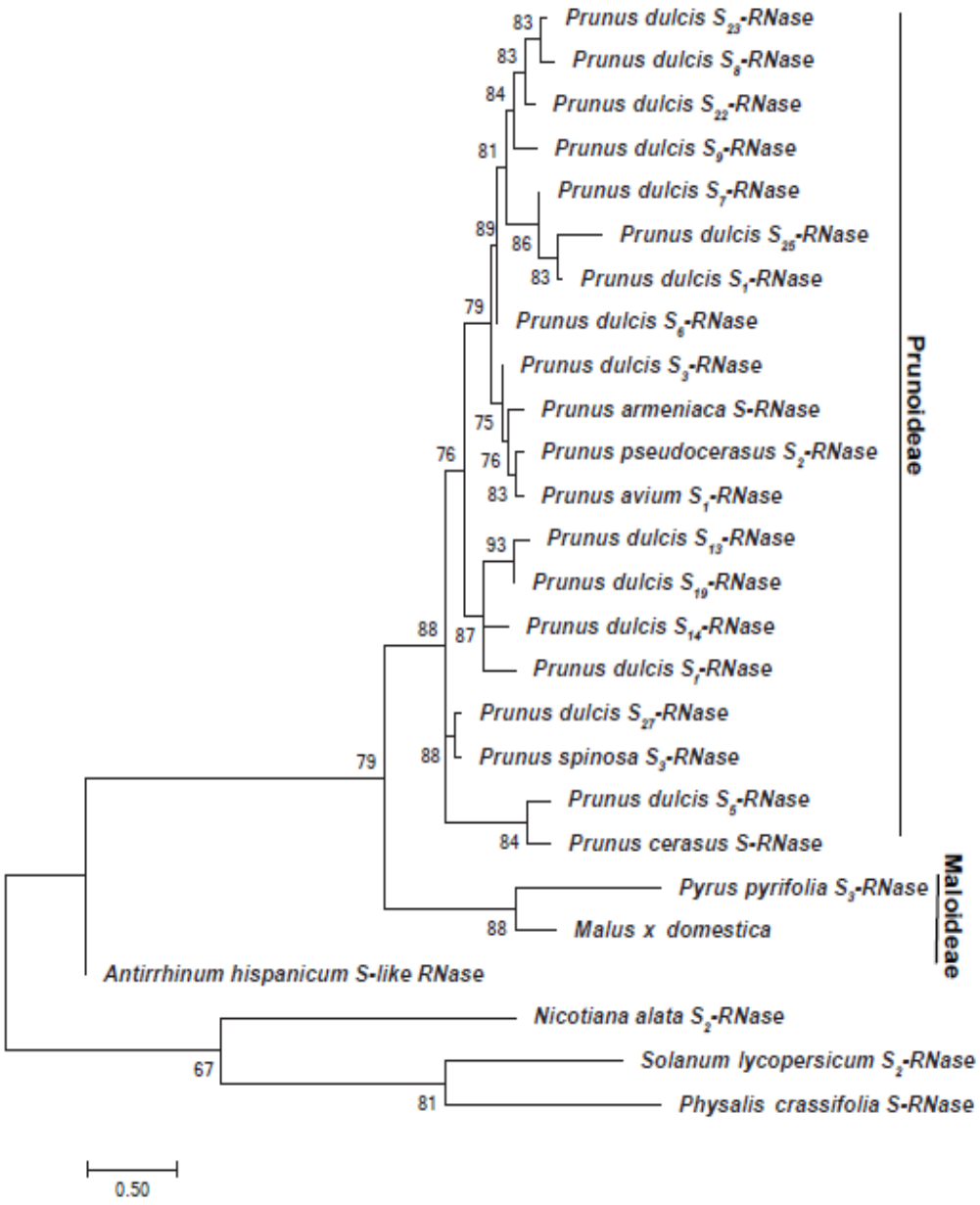


Fig. 3.9 Phylogenetic relationships based on 15 almond *S-RNase* alleles, other *Prunus*, *Malus, Pyrus*, *Antirrhinum* and Solanaceae *S-RNases*, using deduced amino acid sequences from conserved region 1 (C1) to conserved region 5 (C5) of the *S-RNase* gene. There were a total of 47 amino acid positions in the final dataset. Phylogenetic relationships were inferred by using Maximum Likelihood method based on the JTT matrix based model (Jones et al. 1992). The tree with the highest log likelihood (-1573.96) is shown. The branch support values are shown next to the branches. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+*G*, parameter = 2.833)). The tree is drawn to scale, with a scale bar 0.5 amino acid substitutions per site.

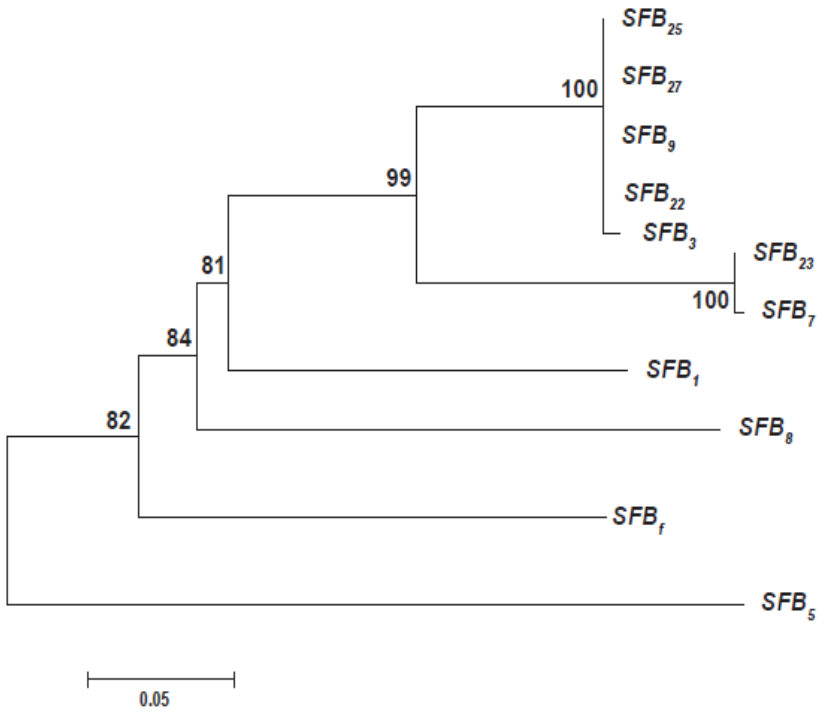


Fig. 3.10 Phylogenetic relationships based on 11 *SFB* alleles in almond. There were a total of 64 amino acid positions in the final dataset. Phylogenetic relationships were inferred by using Maximum Likelihood method based on the JTT matrix based model (Jones et al. 1992). The tree with the highest log likelihood (-2717.17) is shown. The branch support values are shown next to the branches. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+*G*, parameter = 1.0702)). The tree is drawn to scale, with a scale bar 0.05 amino acid substitutions per site

.

3.6 Discussion

In this research, NGS technology was used to obtain sequences from several *S* haplotypes simultaneously using one instrumental run. The S_7 haplotype sequence obtained from eight cultivars had only five nucleotide mismatches to the published sequence (AB081587). Sequencing several individuals that carry the same haplotype enabled the determination of nucleotides that have not been detected previously. In cases where sequences were obtained for the same haplotype from

several individuals, there was little sequence variation except that in at some positions, only one haplotype rather than two was amplified from some cultivars.

While length polymorphisms were observed on the electrophoretic gels for many regions in the S locus, some cultivars only had one band. The appearance of a single band, rather than two could be due to the lack of length polymorphisms of amplicons, or to preferential amplification of short amplicons or to effects of one allele on the amplification of the other -in other words, PCR competition effects (either preferential or exclusive in nature) (Channuntapipat et al. 2003).

The sequence obtained for each haplotype was highly variable across the *S* locus, particularly in the region from 6,500 to 30,000 bp, which contains the highly polymorphic *S-RNase* and *SFB* genes. The complete *S* locus haplotype sequences for S_1, S_7, S_8 and S_f generated here provide new resources for almond research and breeding. This is the first report of sequencing the *S* locus of the self-fertile (S_f) haplotype from almond and from any *Prunus* species. This could be of particular interest to almond breeding and research. Comprehensive analysis of polymorphic regions could provide new information on *S* haplotype-specificity determination. For example, the number of LTRs detected in some of the self-incompatible haplotypes was higher than that in the self-fertile haplotype. This structural difference may be associated with functional difference/s in the *S* haplotypes. While in this research no LTRs were detected within the *S-RNase* and *SFB* genes, LTRs have been detected as an insertion to the *SFB* gene that led to the breakdown of SI in *P. mume* (Ushijima et al. 2004).

Although the sequences generated here for the S_3, S_5, S_6, S_9, S_{13}, S_{14}, S_{19}, S_{22}, S_{23}, S_{25} and S_{27} haplotypes are not complete, they could be used in future to design primers to amplify products to complete the sequences for these haplotypes.

Variation in the length and sequence of the region between the *S-RNase* and *SFB* genes may help prevent recombination between these two genes, helping to maintain *S* locus specificity. The AT-rich

nature of this region may also contribute and consistent with evidence from maize (Sundararajan et al. 2016), rice (Muyle et al. 2011; Serres-Giardi et al. 2012) and eucalyptus (Gion et al. 2016) that recombination-enriched sites are often in regions with high GC content.

Nucleotide variation within the *S-RNase* gene, was due to differences in coding regions and to length polymorphisms in introns, particularly in intron 2, as previously reported (Ushijima et al. 1998). Among the SNPs observed in conserved regions of the *S-RNase* gene, a SNP in conserved region 2 (C2) is of particular interest. This SNP, which causes an amino acid difference between the SI alleles (leucine) and the S_f allele (isoleucine) can be exploited to design assays to differentiate the S_f allele from the SI alleles. Similarly, other SNPs and INDELs detected in the *S-RNase* gene can be exploited for marker design. Sequence variations observed among the alleles of the *SFB* gene are similar to what Ushijima et al. (2003) reported for SFB_1, SFB_5 and SFB_7. These could also be useful for marker assay design.

While all the characteristic regions described for the *S-RNase* (Gu et al. 2012; Ortega et al. 2006; Ushijima et al. 1998) and *SFB* (Ushijima et al. 2003) genes were detected here, additional variable regions (A, B and C) and two more variable regions (A and B) detected in this research provide further evidence for structural variations in the proteins encoded by the alleles from the *S-RNase* and *SFB* genes. Previously, 3D structures constructed for the S_f, S_8 and S_{23} alleles indicated the S_f protein has a long loop between conserved regions RC4 and C5 (Fernández i Marti et al. 2012) that was not present in either the S_8 or S_{23} proteins. Here, comparison of the deduced amino acid sequences obtained for the S_f and 14 SI alleles revealed that five amino acid residues (isoleucine, asparagine, glycine, phenylalanine and alanine) between RC4 and C5 regions are specific to the S_f allele. Of these amino acids, isoleucine is a branched-chain amino acids (C-beta branched), which is involved in substrate recognition and binding (Betts and Russell 2007). Of the previously identified variable regions of Rosaceae *S-RNase* genes, only V1, V2 and RHV are considered to be exposed on the surface of the S-RNase protein, where they could be accessible to mediate the *S* allele

specificity reaction (Matsuura et al. 2001). Here, additional variable regions were detected among almond *S-RNase* alleles. These regions might also affect the *S* allele specificity determination in almond. Construction of protein structures for these SI alleles could provide information how these regions are exposed on the S-RNase protein and how they act in the *S* allele specificity reaction.

The F-box motif of the *SFB* gene is believed to be relatively conserved among *S* alleles (Ikeda et al. 2004; Ushijima et al. 2003; Yamane et al. 2003a; Yamane et al. 2003b), However, according to the results of this analysis, some positions in the F-box motif are quite variable. This motif is involved in forming the E3 ubiquitin ligase complex (Deshaies 1999), which can specifically bind to non-self S-RNases for ubiquitination. The amino acid variations observed here could be due to the specificity in the E3 ubiquitin ligase complex encoded by the *S* alleles and hence substrate selection for degradation.

Long terminal repeat retrotransposons were detected in S_1, S_5, S_7, S_8, S_9 and S_{22}, S_{23} and S_{25} haplotypes. Among haplotypes, the locations of these LTRs were relatively conserved. All these LTRs except one detected for S_1 haplotype, were located towards the end of the *S* locus. The variability in number and length of the LTRs in these haplotypes could contribute to the sequence heteromorphism in the *S* locus to maintain the tight association between the *S-RNase* and *SFB* genes.

In this analysis, three LTRs detected for the S_7 haplotype were similar to LTR1, LTR2 and LTR3 that Ushijima et al. (2003) detected for S_7 haplotype using AutoPredLTR tool in RiceGAAS. However, the LTR0, which was reported by Ushijima et al. (2003) could not be detected. To resolve this issue, LTR analysis was conducted for the published S_7 sequence (AB081587) using *Arabidopsis thaliana* as the reference to predict protein binding sites and only three LTRs were detected. The variation in the number of predicted LTRs could be caused by the references used in the different LTR prediction databases.

After consideration of 96 models with Akaike information criterion (AIC) and Bayesian information criterion (BIC) and hierarchical Likelihood-ratio tests (hLTRs), the JTT model with Gamma substitution factor was selected as the most appropriate substitution model for the data sets used in this analysis. For the almond *S-RNase* gene, the many clades that were identified here reflect the considerable variation observed in the region from C1 to C5. The S_5 allele, which had a very short CDS and an amino acid sequence that differed from those of the other *S* alleles analysed here, grouped separately from other self-incompatibility alleles. The S_7 and S_{14} alleles, which have similar CDS lengths and deduced amino acid sequences also clustered separately from the other *S* alleles. The S_f allele, which has a long exon 3, clustered separately from the SI alleles. Consistent with this finding, the S_5 and S_f alleles of the *SFB* gene were also distant from other alleles.

The phylogenetic relationship analysis using other species showed that almond *S-RNases* are similar to those of other *Prunus* species and different from those of *Malus, Pyrus, Antirrhinum* species and the Solanaceae. These results are similar to what Vieira et al. (2008) has reported for the phylogenetic analysis of *Prunus spinosa SLFL1* gene using a few *SFB* gene sequences of *P. dulcis*, *P. mume* and *P. avium* based on minimum evolution tree construction.

Although resequencing the *S* locus using multiple samples in a single reaction seemed successful and produced a large amount of information, it was difficult to obtain uniform amplification across the entire *S* locus from all haplotypes and samples. Some regions in the *S* locus tended to amplify more readily than some other regions. Furthermore, in some cases it was not possible to amplify both haplotypes using common primers, presumably due to sequence polymorphism in the primer binding sites or PCR competition effects as discussed previously. In most cases, length polymorphisms were observed indicating amplification of both haplotypes, but in a few cases no length polymorphism was observed. This created some gaps and uncertainties in the assembled sequences. It complicated the sequence analysis and necessitated the uses of several bioinformatics tools to establish the final sequences of these haplotypes.

In conclusion, the process by which the sequencing library was generated and the other procedures followed here provide a high-throughput strategy for sequencing haplotypes from multi-allelic loci in heterozygous combination. The data analysis strategy followed here could provide a model for complex multi-allelic sequence data analysis, while the sequence information generated here provides new genomic resources for almond breeding and research.

CHAPTER 4

Marker design for the multi-allelic gametophytic self-incompatibility locus of almond

Shashi N. Goonetilleke[1], Michelle G. Wirthensohn[1], Adam E. Croxford[1], Timothy J. March[1] and Diane E. Mather[1]

[1]School of Agriculture, Food and Wine, Waite Research Institute, The University of Adelaide, Australia.

Corresponding author: diane.mather@adelaide.edu.au

Tel: +61 8 8313 7156

Fax: +61 8 8312 7102

SG: shashi.goonetilleke@adelaide.edu.au

MW: michelle.wirthensohn@adelaide.edu.au

AC: adam.croxford@adelaide.edu.au

TM: timothy.march@adelaide.edu.au

4.1 Statement of Authorship

Title of Paper	Marker design for the multi-allelic gametophytic self-incompatibility locus of almond
Publication Status	☐ Published ☐ Accepted for Publication ☐ Submitted for Publication ☑ Unpublished and Unsubmitted work written in manuscript style
Publication Details	This chapter will be prepared as a manuscript for submission to a refereed journal.

Principal Author

Name of Principal Author (Candidate)	Shashi N. Goonetilleke		
Contribution to the Paper	Prepared the sequencing library, critically analysed all sequence data, designed assays, carried out genotyping, critically examined all genotypic and phenotypic data, interpreted results and wrote the first draft of the manuscript and took primary responsibility for the manuscript revision.		
Overall percentage (%)	75%		
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.		
Signature		Date	

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

i. the candidate's stated contribution to the publication is accurate (as detailed above);
ii. permission is granted for the candidate in include the publication in the thesis; and
iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Michelle G. Wirthensohn		
Contribution to the Paper	Suggested the research topic. Made crosses and provided the plant materials. Read the manuscript and suggested revisions.		
Signature		Date	

Name of Co-Author	Adam E. Croxford		
Contribution to the Paper	Helped make the sequencing library and conducted the sequencing.		
Signature		Date	

Name of Co-Author	Timothy J. March		
Contribution to the Paper	Contributed to the analysis of the sequencing data. Read the manuscript and suggested revisions.		
Signature		Date	

Name of Co-Author	Diane E. Mather		
Contribution to the Paper	Provided overall supervision of the research. Contributed to the interpretation of the results. Supervised the revision of the manuscript.		
Signature		Date	

4.2 Abstract

Self-incompatibility (SI) systems in higher plants are mainly controlled by multi-allelic *S* loci. In almond, in which SI is gametophytically controlled, there are at least 50 *S* alleles. The *S* locus of almond harbours at least two self-incompatibility determining genes: the *S-RNase* gene and the *S* haplotype-specific F-box (*SFB*) gene. Molecular markers have been designed to differentiate among some of these alleles based on length polymorphisms in introns, signal peptide and conserved regions (C1 and C5) of the *S-RNase* gene. In this research, sequence polymorphisms in the *S-RNase* and the *SFB* genes were used to design fluorescence-based marker assays for almond *S* allele detection. A total of 17 marker assays were developed to differentiate among *S* alleles. Some of these assays not only identified the target *S* alleles, but also the companion allele.

4.3 Introduction

Self-incompatibility (SI) prevents self-fertilisation by discriminating between self and non-self pollen grains. In a gametophytic SI system, elongation of pollen tubes that are genetically similar to pollinated plants is retarded in the style (De Nettancourt 1997). The gametophytic SI system of almond [*Prunus dulcis* (Mill.) D.A. Webb] is genetically controlled by a complex, multi-allelic locus, *S,* on chromosome 6 (Ballester et al. 1998). This locus harbours three genes (*SLF*, *S-RNase* and *SFB*) and long terminal repeat retrotransposons (LTRs) (Ushijima et al. 2003; Ushijima et al. 1998). The *S-RNase* gene, encodes a cytotoxic style ribonuclease (S-RNase). It is expressed only in the pistil and is considered to determine pistil-part specificity in almond. It has five conserved regions (C1, C2, C3, RC4 and C5) (Ushijima et al. 1998). The *SFB* (*S* haplotype-specific F-box) gene is expressed only in pollen. This gene is considered to contribute to pollen-part specificity, possibly *via* discrimination of self pollen from non-self pollen by specific binding to the S-RNases and ubiquitination of non-self S-RNases (Ushijima et al. 2003). Each *S-RNase* allele is co-inherited with a corresponding *SFB* allele. These two genes are in the opposite orientation in the *S* locus (Ushijima et al. 2003). The other F-box gene, *S* locus F-box (*SLF*) is expressed in both the pollen and the pistil. Heteromorphism of LTR

retrotransposons within the *S* locus may help to keep the *S-RNase* and *SFB* genes in linkage disequilibrium.

Most almond cultivars are self-incompatible, but a few are self-fertile. There are about 50 known *S* alleles in almond (Bošković et al. 2007; Channuntapipat et al. 2001; Hafizi et al. 2013; Halász et al. 2010; Rahemi et al. 2010). One of these (S_f), which is dominant to the others, confers self-fertility (Grasselly and Olivier 1976; Socias i Company 1990). Cultivars carrying S_f have a little or no *S-RNase* activity in the pistil (Alonso and Socias i Company 2005; Bošković et al. 2003), this could be due to low transcription (Fernández i Martí et al. 2010; Hanada et al. 2009). This allele is sometimes known as S_{fi} (S_f-inactive) to differentiate it from an epigenetic variant S_{fa} (S_f-active) that has the same *S-RNase* genomic sequence but a different methylation pattern (Fernández i Martí et al. 2014). There are also a few cases of self-fertility due to presence of an allele (S_n) that encodes an inactive S-RNase (Bošković et al. 2007) or due to the interactions of other proteins that are involved in determining compatibility reactions (Gómez et al. 2015; Martínez-García et al. 2015).

In almond breeding, SI limits the parental combinations that can be used for crossing. If parental *S* genotypes are not known, the success rate of the resulting crosses can be very low. In almond production, SI necessitates the inclusion of pollinisers with the main cultivar in plantations. Therefore, detection of *S* genotypes would be useful in almond breeding. Cross compatibility can be assessed by evaluating fruit set after controlled pollination (Socias i Company et al. 2005; Tufts and Philip 1922), by examining pollen tube growth after pollination using fluorescence microscopy (Socias i Company 1979), by analysing S-RNase isozymes (Bošković et al. 1997) or by using molecular markers to detect *S* genotypes in almond cultivars (Channuntapipat et al. 2005; Channuntapipat et al. 2001; Channuntapipat et al. 2003; Gu et al. 2015; Ma and Oliveira 2002; Ortega et al. 2005; Sánchez-Pérez et al. 2004; Tamura et al. 2000). Most previously reported molecular markers for *S* genotypes rely upon gel electrophoresis to detect length polymorphisms within the *S-RNase* gene. These markers have some limitations, including masking of the S_8 allele by the S_7 allele and

masking of the S_3 allele by the S_1 allele (Channuntapipat et al. 2001), similarity in size between S_3 and S_f amplicons obtained with primer pairs designed based on *S-RNase* intron sequences (Sánchez-Pérez et al. 2004), presence of multiple-band artefacts (Hanada et al. 2009), and the time and labour required for gel electrophoresis and visual scoring. Some of these issues were partially addressed by the development of a multiplex PCR approach to assay 10 *S* alleles in a single reaction (Sánchez-Pérez et al. 2004) or by the use of degenerate primers (Ma and Oliveira 2002). Limitations could be further overcome by developing assays based on single nucleotide polymorphisms (SNPs), which are the most abundant polymorphisms in plants (Mammadov et al. 2012; Semagn et al. 2014).

In this research, nine *S* alleles were resequenced, SNPs and allele-specific sequences were identified within both the *S-RNase* and *SFB* genes and were used to develop simple uniplex PCR assays to distinguish among *S* alleles in almond. Each of the resulting assays was applied to progeny that segregate for relevant *S* alleles.

4.4 Materials and methods

4.4.1 Plant materials and DNA extraction

The plant materials used here included eight self-fertile cultivars (Antoñeta (S_1S_f), Capella (S_7S_f), Carina (S_7S_f), Francolí (S_1S_f), Lauranne (S_3S_f), Mandaline (S_1S_f), Mira (S_7S_f), Vairo (S_9S_f)), one breeding selection (12-350 (S_1S_f)) and five self-incompatible cultivars (Carmel (S_5S_8), Johnston ($S_{23}S_{25}$), Maxima (S_3S_8), Nonpareil (S_7S_8) and Somerton (S_1S_{23})) and a total of 3,221 progeny that segregate for the relevant *S* alleles (Table 4.1). These populations were maintained in Dareton, NSW, Australia, using standard management practices. Young leaves were harvested and stored at -80^{o}C. Genomic DNA was extracted using an Oktopure™ DNA extraction protocol that had been optimised for almond (LGC Ltd, Teddington, UK).

Table 4.1. Populations used in this analysis.

Female parent and *S* genotype	Male parent and *S* genotype	Number of F_1 progeny
Maxima (S_3S_8)	12-350 (S_1S_f)	232
Maxima (S_3S_8)	Vairo (S_9S_f)	144
Maxima (S_3S_8)	Mira (S_7S_f)	32
Antoñeta (S_1S_f)	Nonpareil (S_7S_8)	183
Carmel (S_5S_8)	Francolí (S_1S_f)	198
Carmel (S_5S_8)	Mandaline (S_1S_f)	67
Carmel (S_5S_8)	Capella (S_7S_f)	71
Carmel (S_5S_8)	Vairo (S_9S_f)	100
Carmel (S_5S_8)	Antoñeta (S_1S_f)	76
Carmel (S_5S_8)	12-350 (S_1S_f)	82
Johnston ($S_{23}S_{25}$)	12-350 (S_1S_f)	127
Johnston ($S_{23}S_{25}$)	Mandaline (S_1S_f)	12
Johnston ($S_{23}S_{25}$)	Lauranne (S_3S_f)	103
Johnston ($S_{23}S_{25}$)	Vairo (S_9S_f)	109
Johnston ($S_{23}S_{25}$)	Capella (S_7S_f)	185
Johnston ($S_{23}S_{25}$)	Constantí (S_3S_f)	69
Nonpareil (S_7S_8)	Lauranne (S_3S_f)	231
Nonpareil (S_7S_8)	Marta (S_1S_f)	49
Nonpareil (S_7S_8)	Vairo (S_9S_f)	199
Nonpareil (S_7S_8)	12-350 (S_1S_f)	135
Nonpareil (S_7S_8)	Constantí (S_3S_f)	350
Nonpareil (S_7S_8)	Mira (S_7S_f)	47
Somerton (S_1S_{23})	Capella (S_7S_f)	94
Somerton (S_1S_{23})	Mira (S_7S_f)	133

4.4.2 *S* allele sequencing and sequence data analysis

Sequence data obtained for the entire *S* locus (Chapter 3) of the above-mentioned cultivars and a breeding selection were used to obtain the sequence of each haplotype used in this research. Allele-specific forward primers were designed for S_f, S_1, S_3, S_5, S_7, S_8, S_9, S_{23} and S_{25} based on sequences upstream from the signal peptide region of the *S-RNase* gene. Allele-specific reverse primers were designed based on sequences downstream from exon 1 region of the *SFB* gene. DNA samples from Carmel (S_5S_8), Francolí (S_1S_f), Johnston ($S_{23}S_{25}$), Lauranne (S_3S_f), Nonpareil (S_7S_8) and Vairo (S_9S_f) were used to separately amplify the complete *S-RNase* and the *SFB* genes from each of the nine *S* alleles separately. A sequencing library was prepared using an Illumina Nextera Library Preparation Kit (Illumina Inc., Australia). For this library, amplicons from each *S* allele (Carmel (S_5), Francolí (S_1), Johnston (S_{23} and S_{25}) Lauranne (S_3 and S_f), Nonpareil (S_7 and S_8) and Vairo (S_9)) were tagged with a barcode adapter. Paired-end sequencing (150 bp reads) was performed on a MiSeq sequencing system (Illumina Inc., Australia).

All raw sequence reads were assessed for quality, adapter sequences and barcode contamination using FASTQC v0.11.5 (Andrews 2010). Adapter sequences were removed using the IllUMINACLIP option in Trimmomatic V0.32 (Bolger et al. 2014) and followed by another run of FASTQC. Trimmed reads obtained for S_f, S_1, S_5, S_7, S_8 and S_{23} were aligned with reference *S-RNase* gene sequences for those alleles (AB433984 for S_f, AB011469 for S_1, DQ150569 for S_5, AY291118 for S_7, AB481108 for S_8, and AB488496 for S_{23}) using the Burrows-Wheeler alignment tool version 0.7 (Li and Durbin 2009). For alleles for which complete *S-RNase* gene sequences were not available (S_3, S_9 and S_{25}), *de novo* assembly was performed using the Velvet assembler version 1.1 (Zerbino and Birney 2008) and Mimicking Intelligent Read Assembler (MIRA) version 4.02 (Chevreux et al. 2004). Similarly, for the *SFB* gene, reads obtained for SFB_f, SFB_1, SFB_5, SFB_7 and SFB_8 were aligned with reference *SFB* gene sequences for those alleles (AB361036 for S_f, AB092967 for S_1, AB092966 for S_5, AB081587 for S_7 and AB081648 for S_8) using the Burrows-Wheeler alignment tool (Li and Durbin 2009) and *de novo* assembly was performed using the Velvet assembler version 1.1 (Zerbino and

Birney 2008) and MIRA version 4.02 (Chevreux et al. 2004) for alleles for which complete *SFB* gene sequences are not available (S_3 and S_{23}) and for alleles which *SFB* sequences could not be found in the nucleotide databases (S_9 and S_{25}). Consensus sequences were obtained for each *S-RNase* and *SFB* allele using SAM (Sequence Alignment/Map) tools (Li et al. 2009). Complete sequences for *S-RNase* alleles were aligned using the Clustal W multiple sequence alignment algorithm (Thompson et al. 1994) as implemented in Geneious software version 9.0.2 (Kearse et al. 2012).

4.4.3 Primer design for *S* allele detection

Primers were designed following KASP™ (LGC Ltd, Teddington, UK) primer design guidelines provided by He et al. (2014) and using Primer 3 software version 4.0 (Untergasser et al. 2012). For detection of the S_3 allele, the same procedure was applied to design primers on the *SFB* sequences. Some primer sets consisted of two primers (an allele-specific primer for a target allele and a second primer) and others consisted of three primers (two allele-specific primers designed based to query a SNP and a common primer). Tail sequences complementary to the FRET cassettes in the KASP Master Mix were added to the 5' ends of the allele-specific primers. The resulting primer sets were named with the prefix WriPdS_x with Wri referring to the Waite Research Institute, Pd referring to *Prunus dulcis* and S_x referring to the relevant *S* allele. In cases where more than one primer is available to detect the target allele, primers are labelled in chronological order.

4.4.4 SNP genotyping

Two DNA samples of Nonpareil, one DNA sample of each of Antoñeta, Carmel, Capella, Carina, Francolí, Johnston, Lauranne, Mandaline, Mira, Maxima, Somerton, Vairo and 12-350 and six water samples (negative controls) were assayed with all primer sets. DNA samples of 10 ng (5 µL of 2 ng/µL) were dried at 55^{o}C for 1 h. A mixture of 0.028 µL (containing 12 µM of each allele specific forward primer and 30 µM of the common primer) and 1.972 µL of 1 × KASP Master Mix was added to each reaction. Amplification was conducted using the standard KASP PCR protocol in a Hydrocycler-16 thermocycler. Fluorescence detection was performed in a Pherastar Plus plate reader

(BMG LABTECH, Germany). Data were analysed using Kraken™ software (LGC Ltd, Teddington, UK). Each primer set that was informative based on results from this panel was assayed on progeny of at least five crosses for which it was expected to be informative. The results were used to assign progeny to genotypic classes. Observed genotypic ratios were compared to expected ratios using a chi-square test (χ^2, $\alpha = 0.05$). In cases where two or more markers were used to screen the same population, results from different markers were compared with each other.

4.4.5 Assessment of self-fertility

For each of six populations (Johnston × Lauranne, Maxima × Mira, Nonpareil × Lauranne, Nonpareil × Marta, Carmel × Capella and Maxima × Vairo), 40 trees (20 that had been genotyped as having inherited S_f from their self-fertile parent and 20 that had been genotyped as having inherited the alternative allele from their self-fertile parent) were selected for evaluation of fruit set. On trees, one branch with about 100 unopened flower buds was covered with a mesh bag (30 cm × 60 cm, with 1 mm mesh size) in early July 2015. In November 2015, the numbers of fruits were counted and the fruit set percentage was calculated.

4.5 Results

4.5.1 The *S-RNase* and *SFB* allele sequences

The raw sequence data for haplotypes yielded 14.6 million reads (QC ≥ 30). The sequences generated here were deposited in the GenBank (KY059853 for S_3, KY059852 for S_9 and KY059854 for S_{25}). Multiple alignments of the nine complete *S-RNase* gene sequences revealed a high level of sequence variation. The greatest difference (88%) was found between the S_5 and the S_9 alleles and the least (53%) was found between the S_{23} and the S_{25} alleles (Fig. S2.1). The interval between conserved regions (C1 and C5) of the *S-RNase* gene of these nine alleles had a total length of 2,960 bp and only 299 bp (10%) were identical across all those alleles. Even the conserved regions (C1, C2, C3, RC4 and C5) each harboured at least two SNPs (Figs. 4.1 and 4.2). In the *SFB* gene,

sequence differences among six *S* alleles ranged from 15% (between S_3 and S_{25}) to 71% (between S_5 and S_8) (Fig. S2.2).

4.5.2 Allele-specific primers to detect the S_1, S_3, S_5, S_7, S_8, S_9, S_{23} and S_{25} alleles of the *S-RNase* gene

Based on multiple alignment of allelic sequences for the region between C1 and C5 of each *S* allele of the *S-RNase* gene, allele-specific primer annealing sites were identified in intron 2 (for the alleles S_5, S_7 and S_f) (Fig. 4.1), in C2 (for the allele S_7), in the region between C2 and C3 (for the allele S_{23}), in C3 (for the alleles S_1 and S_8), in the region in between C3 and RC4 (for the allele S_{25}) and in conserved region 4 (for the alleles S_3, S_9 and S_{25}) (Fig. 4.2). In each case, a suitable annealing site for a second primer was identified. These sites were in C1 (for the allele S_7), in C3 (for the alleles S_3 and S_9), in between region of C3 and RC4 (for alleles S_{25} and S_1), in C4 (for the allele S_8) and in intron 2 (for the alleles S_5, S_7 and S_f).

4.5.3 Allele-specific primers to detect the S_f allele of the *S-RNase* gene

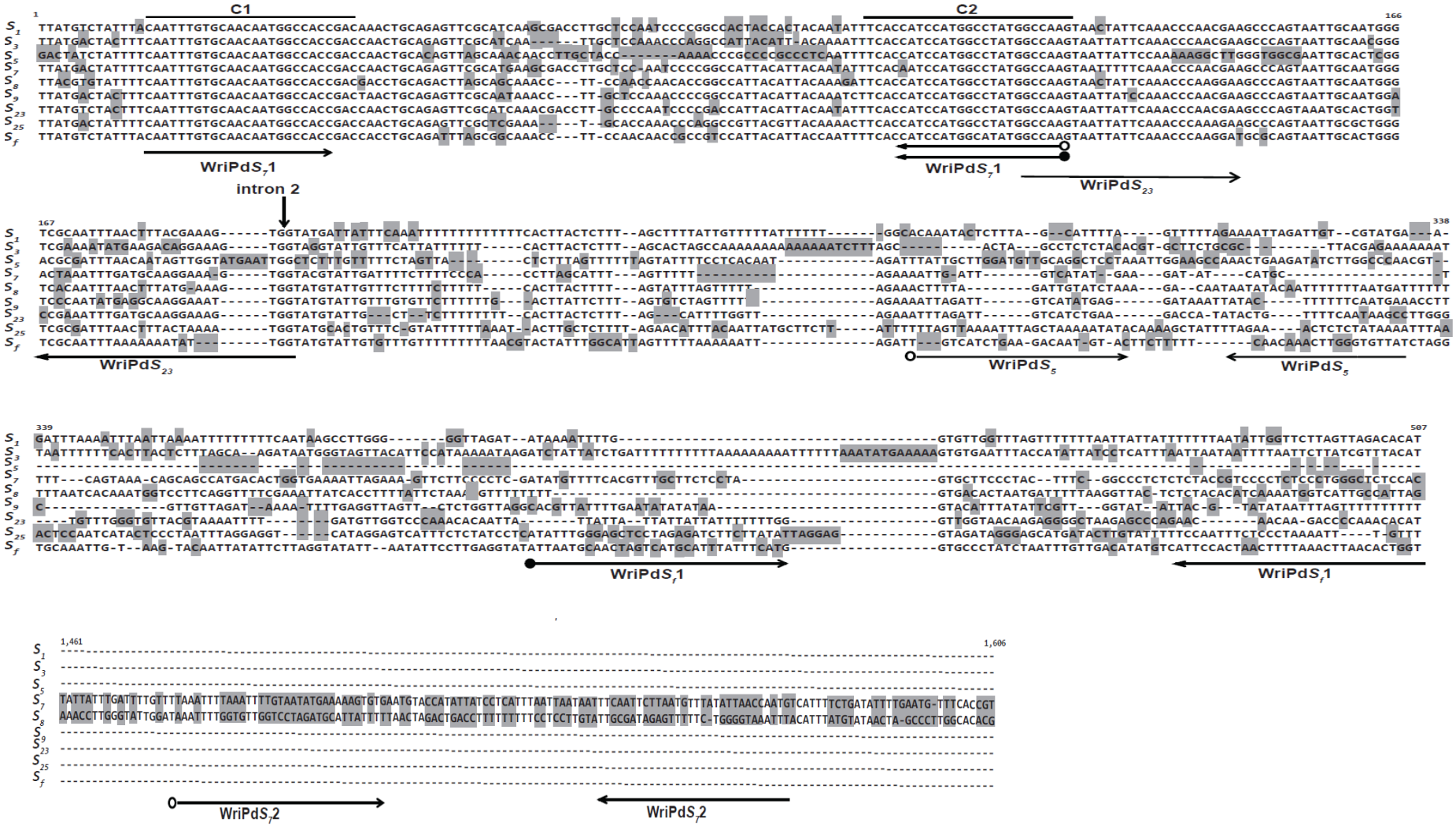
An allele-specific S_f primer was designed based on an annealing site within intron 2 of the *S-RNase* gene (Fig. 4.1).

In addition, a series of primers was designed to distinguish the S_f allele from the SI alleles. For this, several allele-specific primers were designed to query an A/C SNP (A in the S_f allele, C in the each of the SI alleles considered here) (Fig. 4.3) and a common primer was designed based on a conserved segment in the C1 region of the *S-RNase* gene. To distinguish the S_f allele from the S_3, S_9, S_{23} and S_{25} alleles, two allele-specific primers were designed. To distinguish the S_f allele from S_7, S_8 and S_5 alleles, three pairs of degenerate allele-specific primers were designed (Fig. 4.3).

4.5.4 Allele-specific primers to detect the S_3 allele based on the *SFB* gene

New sequences that were generated in this research were deposited in the GenBank (KY059855 for S_3 and KY059857 for S_{25}). Based on multiple alignment of the *SFB* gene sequence of the S_3 allele with those of the S_5, S_7, S_8, S_{23}, S_{25} and S_f alleles, exon 2 was selected to design primer pairs (Fig. 4.4).

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, showing the positions at which primers were designed. Arrows indicate annealing sites for primers. For each primer set, the allele-specific primer to which the FAM fluorescence tail was added is labelled with a light circle and the allele-specific primer with the HEX fluorescence tail is labelled with a dark circle. Nucleotides that are least common in each positon are shaded.


C1
C2
intron 2
WriPdS7_1
WriPdS7_1
WriPdS23
WriPdS23
WriPdS5
WriPdS5
WriPdS7_1
WriPdS7_1
WriPdS7_2
WriPdS7_2

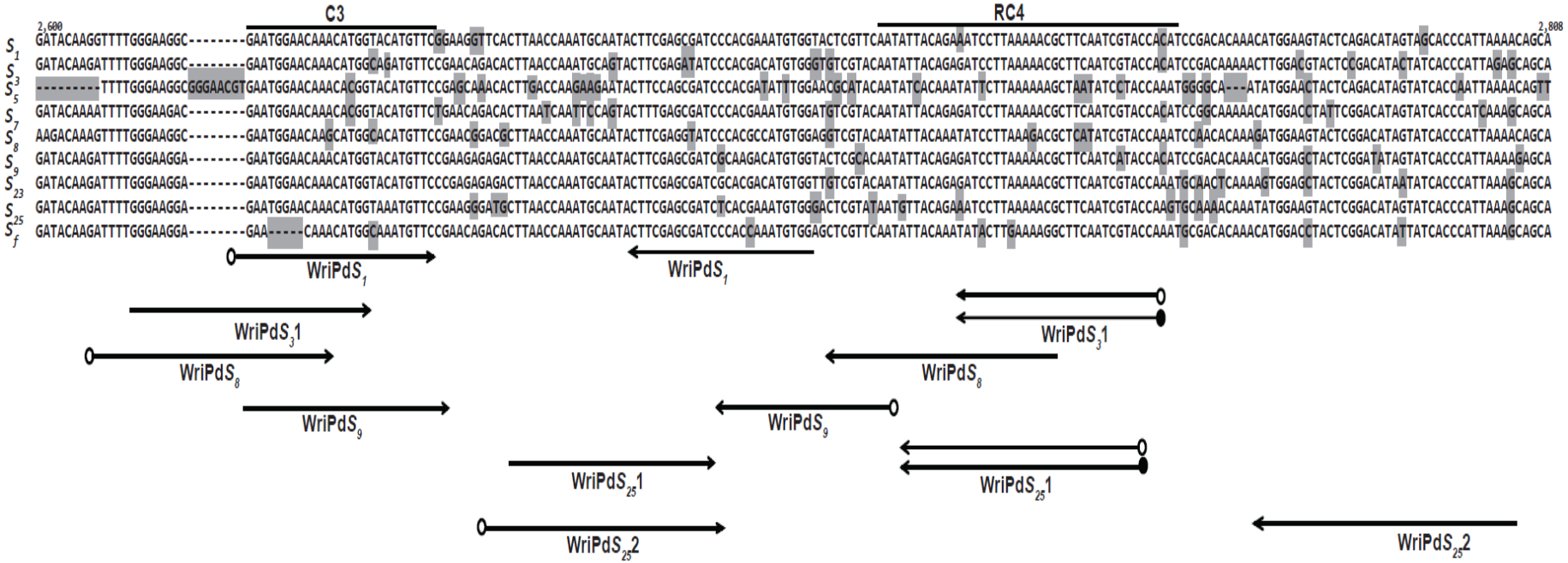


Fig. 4.2 Sequence alignment between conserved region 3 (C3) and conserved region 4 (RC4) of nine *S* alleles of the *S-RNase* gene showing the positions at which primers were designed. Arrows indicate annealing sites for primers. For each primer set, the allele-specific primer to which the FAM fluorescence tail was added is labelled with a light circle and the allele-specific primer with the HEX fluorescence tail is labelled with a dark circle. Nucleotides that are least common in each positon are shaded.

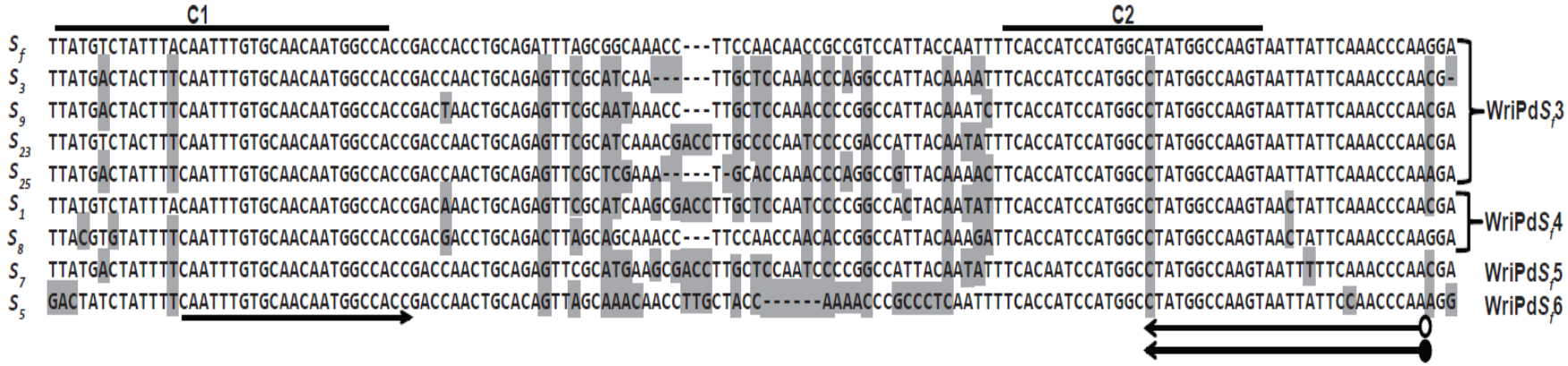


Fig. 4.3 Sequence alignment of nine *S* alleles between conserved region1 (C1) and conserved region 2 (C2) of the *S-RNase* gene. Arrows indicate annealing sites for a common primer and for allele-specific primers. The allele-specific primer with the FAM fluorescence tail is labelled with a light circle and the allele-specific primer with the HEX fluorescence tail is labelled with a dark circle. Allele-specific primer annealing sites are grouped based on the sequence similarity for primer sets WriPd*S*$_f$3, WriPd*S*$_f$4, WriPd*S*$_f$5 and WriPd*S*$_f$6. Nucleotides that are different to the *S*$_f$ allele in each positon are shaded.

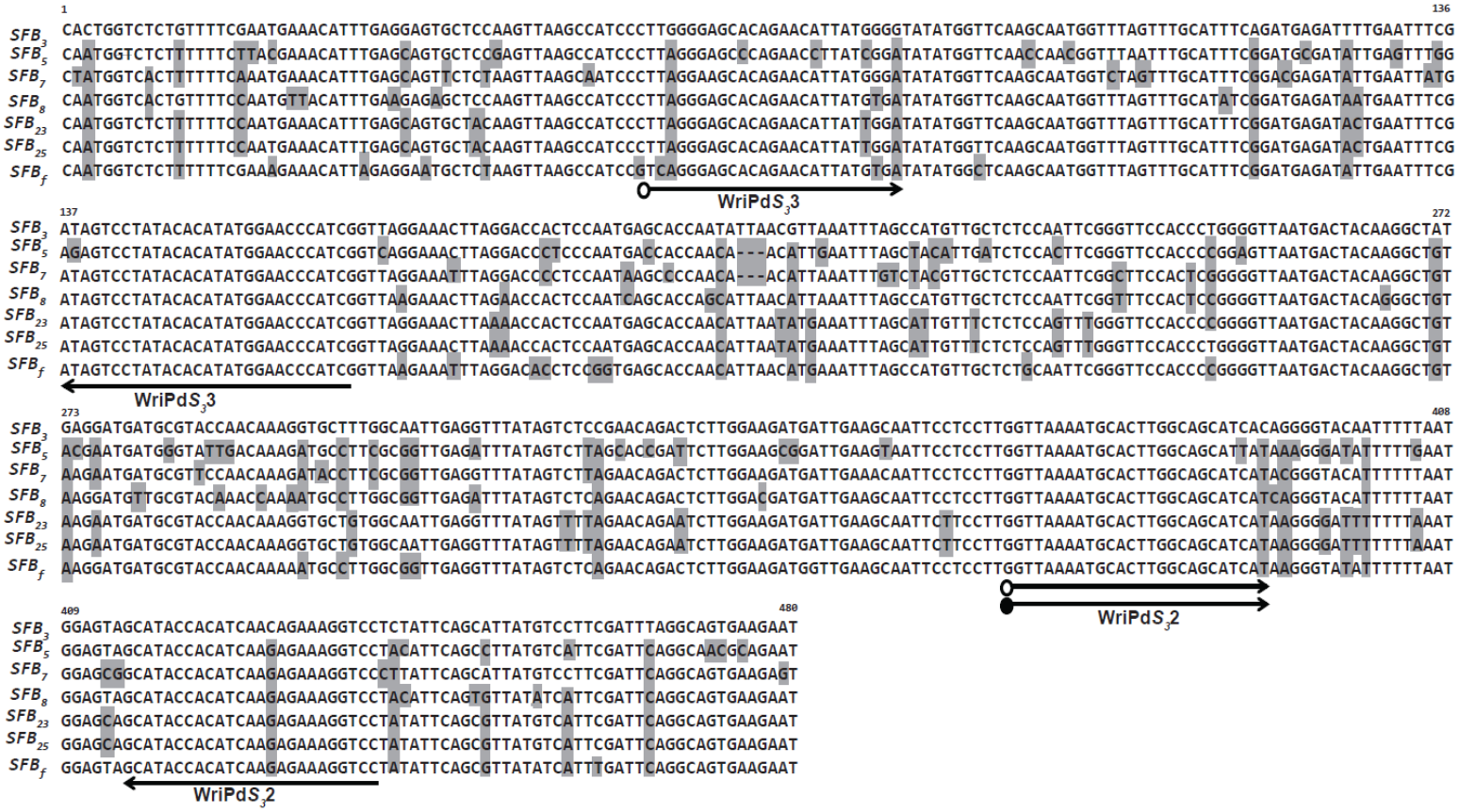


Fig. 4.4 Sequence alignment of seven alleles of the *SFB* gene, showing the positions at which primers were designed. Arrows indicate annealing sites for primers. For each primer set, the allele-specific primer to which the FAM fluorescence tail was added is labelled with a light circle and the allele-specific primer with the HEX fluorescence tail is labelled with a dark circle. Nucleotides that are different to the S_3 allele in each positon are shaded.

4.5.5 Primer validation and population screen

Each of the markers designed in this research differentiated its target *S* allele from each of the other *S* alleles considered here. In some cases markers could also detect which other *S* allele is present in combination with the target *S* allele. Depending on which allele-specific primer was amplified, the PCR product would emit FAM fluorescence, HEX fluorescence, or both (Fig. 4.5). With each of the eight primer sets (WriPdS_1, WriPd$S_3$3, WriPdS_5, WriPd$S_7$2, WriPdS_8, WriPdS_9, WriPdS_{23} and WriPdS_{25}2) FAM fluorescence was detected for each genotype that carries the corresponding target *S* allele, indicating the amplification of the target *S* allele of S_1, S_3, S_5, S_7, S_8, S_9, S_{23} and S_{25}, respectively. Some primer sets amplified products from both target and alternative *S* alleles, resulting in both HEX and FAM fluorescence. For example, primer set WriPd$S_3$1 amplified products from S_3, S_1 and S_{25}. Similarly, WriPd$S_7$1, WriPd$S_3$2 and WriPdS_{25}1 produced both HEX and FAM signal for some allele combinations (Table 4.2).

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. Primer set, genes that primers were designed, primer sequences and fluorescence emitted for each of nine alleles are shown.

Primer set	*S* gene	Primer sequences (5' - 3')[1]	*S* alleles								
			S_1	S_3	S_5	S_7	S_8	S_9	S_{23}	S_{25}	S_f
WriPdS$_1$	*S-RNase*	GAAGGTGACCAAGTTCATGCTGAATGGAACAAACATGGTACATGTTCG CCACATTTCGTGGGATCGCTCGAAG	FAM	-	-	-	-	-	-	-	-
WriPdS$_3$1	*S-RNase*	GAAGGTGACCAAGTTCATGCTTGGTACGATTGAAGCGTTTTTAAGGATC GAAGGTCGGAGTCAACGGATTTGGTACGATTGAAGCGTTTTTAAGGATT GGGAAGGCGAATGGAACAAACATGG	HEX	FAM	-	-	-	-	-	HEX	-
WriPdS$_3$2	*SFB*	GAAGGTGACCAAGTTCATGCTGTTAAAATGCACTTGGCAGCATCAC GAAGGTCGGAGTCAACGGATTGTTAAAATGCACTTGGCAGCATCAT GGACCTTTCTGASTGATGTGGTATGC	-	FAM	-	HEX	HEX	-	-	HEX	HEX
WriPdS$_3$3	*SFB*	GAAGGTGACCAAGTTCATGCTTTGGGGAGCACAGAACATTATGGGG CGATGGGTTCCATATGTGTATAGGACTAT	-	FAM	-	-	-	-	-	-	-
WriPdS$_5$	*S-RNase*	GAAGGTGACCAAGTTCA**TGCT**TGGATGTTGCAGGCTCCTAAAT ACGTTGGGCCAAGATATCTTCA	-	-	FAM	-	-	-	-	-	-

Table 4.2., continued.

Primer set	*S* gene	Primer sequences (5' - 3')[1]	*S* alleles								
			S_1	S_3	S_5	S_7	S_8	S_9	S_{23}	S_{25}	S_f
WriPdS$_7$1	*S-RNase*	GAAGGTGACCAAGTTCATGCTCTTGGCCATAKGCCATGGATT GAAGGTCGGAGTCAACGGATTCTTGGCCATAKGCCATGGATG CAATTTGTGCAACAATGGCCACC	HEX	HEX	HEX	FAM	HEX	HEX	HEX	HEX	HEX
WriPdS$_7$2	*S-RNase*	GAAGGTGACCAAGTTCATGCTAAATTTTAAATTTTGTAATATGAAAAAGTGTG CATTGGTTAATATAAACATTAAGAATTGAA	-	-	-	FAM	-	-	-	-	-
WriPdS$_8$	*S-RNase*	GAAGGTGACCAAGTTCATGCTGTTTTGGGAAGGCGAATGGAACAAG CGTCTTTAAGGATATTTGTAATATTGTACGACC	-	-	-	-	FAM	-	-	-	-
WriPdS$_9$	*S-RNase*	GAAGGTGACCAAGTTCATGCTTTGTGCGAGTACCACATGTCTTGC GAATGGAACAAACATGGTACATGTTCCG	-	-	-	-	-	FAM	-	-	-
WriPdS$_{23}$	*S-RNase*	GAAGGTGACCAAGTTCATGCTCCACTTTCCTTGCATCAAATTTCGG GCCAAGTAATTATTCAAACCCAACGAA	-	-	-	-	-	-	FAM	-	-
WriPdS$_{25}$1	*S-RNase*	GAAGGTGACCAAGTTCATGCTTACGATTGAAGCGTTTTTAAGGATYTCTGTAAC GAAGGTCGGAGTCAACGGATTTACGATTGAAGCGTTTTTAAGGATYTCTGTAAT CTTAACCAAATGCAATACTTCGAGCGATC	HEX	-	-	-	-	-	HEX	FAM	-

Table 4.2., continued.

Primer set	*S* gene	Primer sequences (5' - 3')[1]	*S* alleles								
			S_1	S_3	S_5	S_7	S_8	S_9	S_{23}	S_{25}	S_f
WriPdS_{25}2	*S-RNase*	GAAGGTGACCAAGTTCATGCTATGCTTAACCAAATGCAATACTTCGAGCGATCT CTTTAATGGGTGATACTATGTCCGAGTACTTCCATA	-	-	-	-	-	-	-	FAM	-
WriPdS_f1	*S-RNase*	GAAGGTCGGAGTCAACGG**ATT**AATGCAACTAGTCATGCATTTATTTCATG ACCAGTGTTAAGTTTAAAAGTTAGTGGAAT	-	-	-	-	-	-	-	-	HEX
WriPdS_f3	*S-RNase*	GAAGGTGACCAAGTTCATGC**T**TGGGTTTGAATAATTACTTGGCCATAT GAAGGTCGGAGTCAACGGAT**T**TGGGTTTGAATAATTACTTGGCCATAG CAATTTGTGCAACAATGGCCACC	-	HEX	-	-	-	HEX	HEX	HEX	FAM
WriPdS_f4	*S-RNase*	GAAGGTGACCAAGTTCATGC**T**TGGGTTTGAATARTTACTTGGCCATAT GAAGGTCGGAGTCAACGGA**TT**GGGTTTGAATARTTACTTGGCCATAG CAATTTGTGCAACAATGGCCACC	HEX	HEX	-	-	HEX	HEX	HEX	HEX	FAM
WriPdS_f5	*S-RNase*	GAAGGTGACCAAGTTCATGC**T**TGGGTTTGAAWAATTACTTGGCCATAT GAAGGTCGGAGTCAACGGA**TT**GGGTTTGAAWAATTACTTGGCCATAG CAATTTGTGCAACAATGGCCACC	-	HEX	HEX	HEX	-	HEX	HEX	HEX	FAM
WriPdS_f6	*S-RNase*	GAAGGTGACCAAGTTCATGC**T**TGGGTTKGAATAATTACTTGGCCATAT GAAGGTCGGAGTCAACGGA**TT**GGGTTKGAATAATTACTTGGCCATAG CAATTTGTGCAACAATGGCCACC	-	HEX	HEX	-	-	HEX	HEX	HEX	FAM

[1] Allele-specific primers include tails (underlined) that are complementary to FRET cassettes in the KASP™ Master Mix. Overlapping nucleotides between the tail and allele-specific sequences are in bold text.

Of the S_f primer pairs, WriPdS_f3 amplified the products from S_f, S_9, S_{23} and S_{25}, WriPdS_f4 amplified products from S_f, S_1, S_3, S_8, S_9, S_{23} and S_{25}, WriPdS_f5 amplified products from S_f, S_3, S_5, S_7, S_9, S_{23} and S_{25} and WriPdS_f6 amplified products from S_f, S_3, S_5, S_9, S_{23} and S_{25}. With the WriPdS_f1 primer set, HEX fluorescence was emitted when the target S_f allele was present (Fig. 4.5). With each of the other primer sets HEX fluorescence was detected for genotypes that do not carry target S alleles (Fig. 4.5).

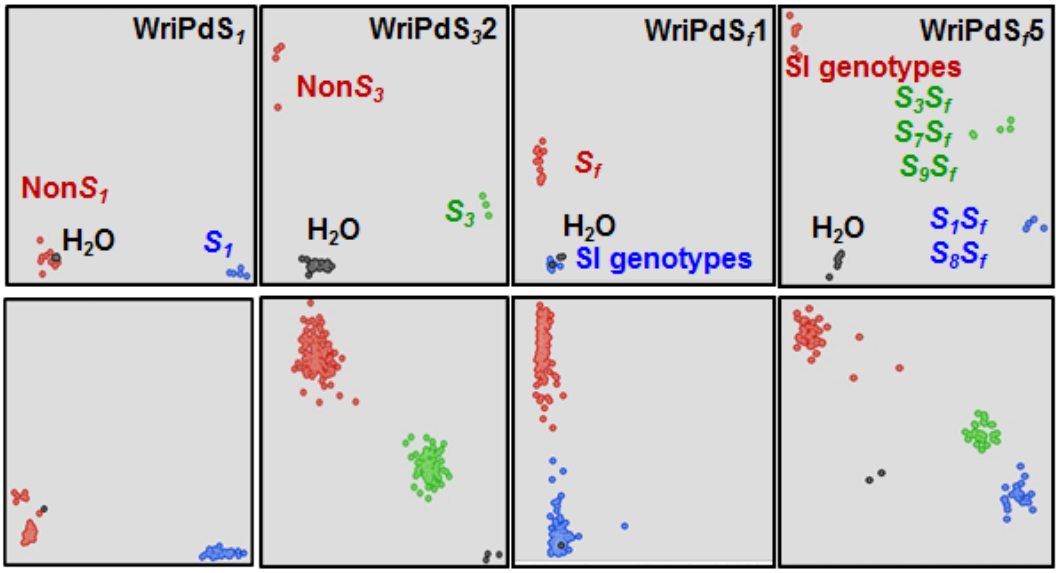


Fig. 4.5 Results obtained with fluorescence-based S allele markers. The upper panel shows results obtained with the WriPdS_1, WriPd$S_3$2, WriPdS_f1 and WriPdS_f5 primer sets on a validation panel of almond cultivars and breeding selections. The lower panel shows results obtained with the same primer sets on F_1 progeny from the crosses (from left to right) Carmel (S_5S_8) × 12-350 (S_1S_f), Nonpareil (S_7S_8) × Lauranne (S_3S_f), Nonpareil (S_7S_8) × Constantí (S_3S_f) and Carmel (S_5S_8) × Francolí (S_1S_f). The horizontal and vertical axes represent intensities of FAM and HEX fluorescence, respectively.

When the primer sets designed here were used to assay F_1 progeny from relevant crosses, no significant deviations were detected from the expected 1:1 ratio. In populations that were assayed with more than one primer set, no discrepancies were detected in the results (Table 4.3).

Table 4.3. Summary of results obtained from assessment of each of the 17 KASP assays on F_1 progeny from two crosses, each showing the numbers of progeny for which HEX fluorescence, FAM fluorescence or both (HEX:FAM) were emitted.

Assay	Population screened	Number of progeny		
		HEX/Null allele	HEX:FAM	FAM
WriPd*S*$_1$	Carmel (*S*$_5$*S*$_8$) × 12-350 (*S*$_1$*S*$_f$)	52	-	40
	Antoñeta (*S*$_1$*S*$_f$) × Nonpareil (*S*$_7$*S*$_8$)	97	-	86
WriPd*S*$_3$1	Johnston (*S*$_{23}$*S*$_{25}$) × Lauranne (*S*$_3$*S*$_f$)	53	28	32
WriPd*S*$_3$2	Nonpareil (*S*$_7$*S*$_8$) × Lauranne (*S*$_3$*S*$_f$)	139	92	-
	Johnston (*S*$_{23}$*S*$_{25}$) × Lauranne (*S*$_3$*S*$_f$)	53	28	32
WriPd*S*$_3$3	Maxima (*S*$_3$*S*$_8$) × Mira (*S*$_7$*S*$_f$)	16	-	16
	Nonpareil (*S*$_7$*S*$_8$) × Lauranne (*S*$_3$*S*$_f$)	139	-	92
WriPd*S*$_5$	Carmel (*S*$_5$*S*$_8$) × Francolí (*S*$_1$*S*$_f$)	102	-	96
	Carmel (*S*$_5$*S*$_8$) × Mandaline (*S*$_1$*S*$_f$)	36	-	31
WriPd*S*$_7$1	Nonpareil (*S*$_7$*S*$_8$) × Mira (*S*$_7$*S*$_f$)	23	24	-
	Maxima (*S*$_3$*S*$_8$) × Mira (*S*$_7$*S*$_f$)	16	16	-
WriPd*S*$_7$2	Nonpareil (*S*$_7$*S*$_8$) × Mira (*S*$_7$*S*$_f$)	23	-	24
	Nonpareil (*S*$_7$*S*$_8$) × Lauranne (*S*$_3$*S*$_f$)	102	-	129
WriPd*S*$_8$	Nonpareil (*S*$_7$*S*$_8$) × Lauranne (*S*$_3$*S*$_f$)	129	-	102
	Carmel (*S*$_5$*S*$_8$) × Mandaline (*S*$_1$*S*$_f$)	31	-	36

Table 4.3., continued.

Assay	Population screened	Number of progeny		
		HEX/Null allele	HEX:FAM	FAM
WriPdS_9	Nonpareil (S_7S_8) × Vairo (S_9S_f)	100	-	99
	Johnston ($S_{23}S_{25}$) × Vairo (S_9S_f)	58	-	51
WriPdS_{23}	Johnston ($S_{23}S_{25}$) × 12-350 (S_1S_f)	66	-	61
	Johnston ($S_{23}S_{25}$) × Vairo (S_9S_f)	55	-	54
WriPdS_{25}1	Johnston ($S_{23}S_{25}$) × 12-350 (S_1S_f)	63	35	29
	Johnston ($S_{23}S_{25}$) × Mandaline (S_1S_f)	4	4	4
WriPdS_{25}2	Johnston ($S_{23}S_{25}$) × Vairo (S_9S_f)	62	-	47
	Johnston ($S_{23}S_{25}$) × Constantí (S_3S_f)	34	-	35
WriPdS_f1	Johnston ($S_{23}S_{25}$) × Lauranne (S_3S_f)	54	-	49
	Nonpareil (S_7S_8) × Lauranne (S_3S_f)	103	-	128
WriPdS_f3	Johnston ($S_{23}S_{25}$) × Lauranne (S_3S_f)	49	54	-
	Maxima (S_3S_8) × Mira (S_7S_f)	15	8	9
WriPdS_f4	Nonpareil (S_7S_8) × Lauranne (S_3S_f)	128	52	51
	Nonpareil (S_7S_8) × Marta (S_1S_f)	24	13	12
WriPdS_f5	Carmel (S_5S_8) × Capella (S_7S_f)	32	20	19
	Nonpareil (S_7S_8) × Lauranne (S_3S_f)	128	52	51
WriPdS_f6	Johnston ($S_{23}S_{25}$) × Vairo (S_9S_f)	57	52	-
	Maxima (S_3S_8) × Vairo (S_9S_f)	75	28	41

4.5.6 Fruit set evaluation

Among the trees that were genotyped as carrying the S_f allele, the mean fruit set percentage on bagged branches ranged from 30% in Maxima × Vairo to 43% in Nonpareil × Lauranne (Table 4.4). In contrast, only three fruits were set on bagged branches of trees that had been determined not to carry the S_f allele, one fruit on a Nonpareil × Marta tree and two fruits on a Carmel × Capella tree.

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.

Assay	Population screened	Fruit set %	Mean fruit set %
WriPdS$_f$1	Johnston ($S_{23}S_{25}$) × Lauranne (S_3S_f)	8 – 62	35
	Nonpareil (S_7S_8) × Lauranne (S_3S_f)	13 – 72	43
WriPdS$_f$3	Johnston ($S_{23}S_{25}$) × Lauranne (S_3S_f)	8 – 62	35
WriPdS$_f$4	Nonpareil (S_7S_8) × Lauranne (S_3S_f)	13 – 72	43
	Nonpareil (S_7S_8) × Marta (S_1S_f)	8 – 67	38
WriPdS$_f$5	Carmel (S_5S_8) × Capella (S_7S_f)	8 – 65	37
	Nonpareil (S_7S_8) × Lauranne (S_3S_f)	13 – 72	43
WriPdS$_f$6	Maxima (S_3S_8) × Vairo (S_9S_f)	8 – 52	30

4.6 Discussion

Although there are two SI specificity determination genes in the *S* locus (*S-RNase* and *SFB*), previous marker design has used only the allele sequences from the *S-RNase* gene (Channuntapipat et al. 2005; Channuntapipat et al. 2003; Ortega et al. 2006; Ortega et al. 2005; Sánchez-Pérez et al. 2004), possibly because the *S-RNase* gene was the first to be associated with

SI specificity (Ushijima et al. 1998). Here, the *S-RNase* gene was confirmed to be highly polymorphic and suitable for assay design, and *SFB* was also found to be polymorphic and useful for distinguishing among *S* alleles.

Within the *S-RNase* gene, the sequences of intron 2 and the conserved regions C1 and C5 have previously been used to design primers to detect length polymorphisms among alleles (Channuntapipat et al. 2001; Channuntapipat et al. 2003; Ma and Oliveira 2002; Ortega and Dicenta 2003; Sánchez-Pérez et al. 2004). Here, sequences from intron 2 were used to design primer sets (WriPdS_5, WriPd$S_7$2 and WriPdS_f1) that could detect alleles S_5, S_7 and S_f, respectively. Among the conserved regions, C1 was completely conserved across the nine *S* alleles evaluated here, but polymorphisms were detected within C2, C3 and RC4. These enabled the development of allele-specific primers for S_7 and S_f (C2), S_1, S_8 and S_9 (C3) and S_{25} (RC4).

KASP™ assays for genotyping of SNPs usually involve sets of two allele-specific primers and one common primer. Ideally, the allele-specific primer for a particular allele is designed based on a sequence for which that allele differs from all others just by a SNP at the 3' end. Here, examination of the *S-RNase* and *SFB* sequences for nine *S* alleles revealed few sequences of this type due to high sequence variability among these *S* alleles. Nevertheless, it was possible to design a three-primer assay (WriPd$S_7$1) that distinguished the S_7 allele from eight other alleles based on a SNP. Several other sequences were found that distinguished a particular target allele from one or two of the other alleles, permitting the design of the three-primer assays WriPd$S_3$1, WriPdS_{25}1, WriPdS_f3, WriPdS_f4, WriPdS_f5 and WriPdS_f6.

Two additional marker development strategies were implemented here. In one of these, an allele-specific primer pair was designed for each *S* allele. This provided presence/absence markers for each of the nine *S* alleles analysed here. Eight of these were based on *S-RNase* allele sequences and one (WriPd$S_3$3) on an *SFB* allele sequence. In the other strategy, the use of degenerate primers

(in the WriPd*S*$_3$2, WriPd*S*$_7$1, WriPd*S*$_{25}$1, WriPd*S*$_f$4, WriPd*S*$_f$5 and WriPd*S*$_f$6) made it possible to discriminate target alleles from several other *S* alleles. In some cases, the same third primer (CAATTTGTGCAACAATGGCCACC) was implemented to amplify the products from several *S* alleles (WriPd*S*$_7$1, WriPd*S*$_f$3, WriPd*S*$_f$4, WriPd*S*$_f$5 and WriPd*S*$_f$6).

In this analysis, only nine of 50 known almond *S* alleles were analysed. However, the presence/absence markers designed here (WriPd*S*$_1$, WriPd*S*$_3$3, WriPd*S*$_5$, WriPd*S*$_7$2, WriPd*S*$_8$, WriPd*S*$_9$, WriPd*S*$_{23}$, WriPd*S*$_{25}$2 and WriPd*S*$_f$1) could provide a broader applicability as the choice of marker does not depend on the other *S* allele that an individual carries.

Primer pairs (sets involving an allele-specific primer and a second primer) amplified only the target *S* allele. In most of these assays, FAM fluorescence was used to detect the target allele (WriPd*S*$_1$, WriPd*S*$_3$3, WriPd*S*$_5$, WriPd*S*$_7$2, WriPd*S*$_8$, WriPd*S*$_9$, WriPd*S*$_{23}$ and WriPd*S*$_{25}$2). The exception was the WriPd*S*$_f$1 primer pair for which HEX fluorescence was used for the target allele *S*$_f$ and FAM fluorescence was used for the alternative alleles. This is because the annealing site for the allele-specific primers shared three nucleotides with the HEX tail, making it possible to design shorter allele-specific primers than if FAM fluorescence had been used. With all of these primer pairs, genotypes that do not carry any of the targeted *S* alleles gave results similar to those of the water sample (negative control), with neither FAM nor HEX fluorescence detected, indicating no amplification.

Assays involving allele-specific primers that were designed based on annealing sites that differ by SNPs at the 3' end (WriPd*S*$_3$1, WriPd*S*$_7$1 and WriPd*S*$_f$3) or allele-specific degenerate primers (WriPd*S*$_7$1, WriPd*S*$_{25}$1, WriPd*S*$_f$4, WriPd*S*$_f$5 and WriPd*S*$_f$6) could amplify the products from both target and alternative alleles, resulting in both HEX and FAM fluorescence. In some cases, it was possible to detect which *S* allele was present in combination with the target *S* allele. In such cases, genotypes that carry target alleles resulted in FAM fluorescence or both FAM and HEX fluorescence,

while genotypes that do not carry the target *S* allele generated only HEX fluorescence. Although there are different allele-specific primers that emit HEX fluorescence for each of the alternative alleles, they all target the same nucleotide in the target site. So, despite the mismatches elsewhere, they are all a better match than the target allele-specific primer which emits FAM fluorescence and detect HEX signal.

For a multi-allelic locus with a high level of sequence variation, one set of primers is not always sufficient to differentiate a target allele from all other alleles. In some cases, it is required to have a series of markers that could differentiate the target from the other *S* alleles in the locus. In the S_f series of primer sets (WriPdS_f3, WriPdS_f4, WriPdS_f5 and WriPdS_f6), the primers with the HEX fluorescence tail were designed for the SI alleles. All of these HEX primers target the same nucleotide at the target SNP site. Despite mismatches elsewhere, they are all a better match than the FAM primer so they outcompete it and result in a HEX signal.

The genotyping method used here is more cost effective than many other currently available genotyping methods. It requires little turnaround time to generate results and is suitable for high-throughput genotyping. These assays could also be deployed using a real-time PCR device such as LightCycler® 480 System.

It was expected that among the F_1 progeny evaluated with these markers, half would carry the target *S* allele. With one exception (WriPdS_3 markers on Nonpareil × Lauranne) the results obtained were consistent with that expectation. No significant deviations from the expected 1:1 ratio (allele of interest : null allele) were observed. Further, whenever a population was analysed with more than one marker, exactly the same results were obtained for the presence or absence of the allele of interest. All the trees that were genotyped as self-fertile set self-fruits. For each of these trees, the fruit set was above the minimum of 6% that was used by Grasselly et al. (1981) as the criterion for considering a population to be self-fertile. Only a few fruits were observed on the trees that were

genotyped as self-infertile, these may have been set due to pollen transfer by small insects such as thrips.

4.7 Conclusions

It is possible to use both the *S-RNase* and *SFB* genes for the *S* allele detection in almond. For a multi-allelic locus such as the almond *S* locus, markers designed based on allele-specific SNPs provided robust markers that can readily distinguish the allele of interest from the rest of the alleles. These markers could facilitate high-throughput detection of *S* alleles in almond. The process by which these markers were designed could serve as a model for the development of high-throughput markers for other crops with self-incompatibility systems and for other multi-allelic loci.

CHAPTER 5

Linkage and quantitative trait locus maps for almond

This chapter consists of three sections reporting on genetic mapping in almond. Section 5.1 reports the use of genotyping-by-sequencing to discover and assay single nucleotide polymorphism (SNP) markers in a Nonpareil × Lauranne F_1 population, on the use of SNP data to construct linkage maps and on the development of simple marker assays for repeated genotyping based on the polymorphisms detected in Nonpareil and Lauranne. Section 5.2 reports on the construction of a composite linkage map for Nonpareil based on genotypic data from four crosses. Section 5.3 reports on phenotypic evaluation of several almond mapping populations and the mapping of quantitative trait loci (QTLs) for physical and chemical traits.

Section 5.1

Application of genotyping-by-sequencing to construct linkage maps for almond

5.1.1 Introduction

Almond [*Prunus dulcis* (Mill.) D.A. Webb] is an important nut crop with an annual global production of 1.2 million tonnes (FAOSTAT 2015). It has a small genome of 270 Mb (Arumuganathan and Earle 1991; Bennett and Leitch 1997). Almond breeding relies mostly on phenotypic assessment of parents, crossing between selected parents, vegetative propagation of progeny and phenotypic selection among progeny, with only limited use of molecular information (Gradziel 2009; Koepke et al. 2013; Sánchez-Pérez et al. 2007; Scorza 2001). Development and implementation of modern molecular tools could increase the speed and the precision of the almond breeding process.

Due to the long juvenile period of almond, progeny cannot be phenotypically assessed for kernel and nut traits until 3 years after planting. This limitation could be partially overcome by using trait-associated molecular markers to select among juvenile progeny. Identification of suitable markers for this purpose requires construction of linkage maps and mapping of trait loci relative to marker loci. As a result of the out-crossing nature of almond, parents and progeny are highly heterozygous. This necessitates mapping approaches that differ from those that have been widely used for self-pollinated plants such as, pseudo-testcross mapping strategy and genotypic data analysis tools based on the inbred backcross design. These approaches have been applied for almond (Arús et al. 1994; Fernández i Martí et al. 2013; Font i Forcada et al. 2012; Font i Forcada et al. 2015; Tavassolian et al. 2010), peach (*Prunus persica*) (Dirlewanger et al. 2006; Dirlewanger et al. 2002; Ogundiwin et al. 2009; Zeballos et al. 2016) and other clonally propagated cross-pollinated perennial

plants (e.g., eucalyptus, grape, oil palm and poplar) (Grattapaglia and Resende 2010; Murphy 2007; Neale and Kremer 2011).

One of the linkage maps that has been published for almond is a reference map that was derived from analysis of F_2 progeny from an inter-specific cross between the peach cultivar Earlygold and the almond cultivar Texas (Joobeur et al. 1998). It was initially constructed using isozymes and restriction fragment length polymorphism (RFLP) markers. Simple sequence repeat (SSR) markers were later added to the map (Aranzana et al. 2003; Dirlewanger et al. 2004). Other marker types that have been mapped in almond include random amplified polymorphic DNA (RAPD) markers, inter-simple sequence repeat (ISSR) markers, sequence characterised amplified region (SCAR) markers (Joobeur et al. 2000; Tavassolian et al. 2010) and single nucleotide polymorphism (SNP) markers (Wu et al. 2010; Wu et al. 2009). Single nucleotide polymorphism is the most abundant type of sequence variation in plants (Mammadov et al. 2012; Michael 2014). They are widely used for genetic mapping and molecular breeding in other species, but have not been extensively exploited in almond.

Next-generation sequencing approaches can make it possible to directly assay large numbers of sequence polymorphisms (Mammadov et al. 2012; Michael 2014; Reuter et al. 2015). One key feature of these approaches is that they do not require prior knowledge about the polymorphisms or their genomic positions. Given the size and complexity of plant genomes, these approaches require the preparation of reduced representation libraries. Of several available library preparation protocols, one proposed for genotyping-by-sequencing (GBS) (Elshire et al. 2011) is a simple method that has been shown to be effective for plant and animal species (Bayer et al. 2015; Bielenberg et al. 2015; Elshire et al. 2011; Etter et al. 2011; Guajardo et al. 2015; Hyma et al. 2013; Kujur et al. 2015; Lu et al. 2013). Using software such as TASSEL (Bradbury et al. 2007; Lu et al. 2013) it is possible to identify many thousands of SNPs from GBS data.

In this research, the GBS protocol was adapted for almond and applied to F_1 progeny from a cross between the almond cultivars Nonpareil and Lauranne. Linkage maps were constructed for each parent and quantitative trait loci were mapped for kernel and nut traits. Simple uniplex assays were developed for selected SNPs, providing tools that could be used to detect polymorphisms in almond breeding materials.

5.1.2 Materials and methods

5.1.2.1 Plant materials

The initial mapping population used here consisted of 89 F_1 progeny from a cross between Nonpareil and Lauranne that was made in 1997. This population is a subset of a Nonpareil × Lauranne population used by Tavassolian et al. (2010). Nonpareil, the female parent of the cross, is a cultivar from the U.S.A. It has paper shell nuts and is self-incompatible. Lauranne, the pollen donor, is a cultivar from France. It has hard shell nuts and is self-compatible (Alonso et al. 2012). In addition, 231 other Nonpareil/Lauranne F_1 progeny derived from crosses made in 2007 and 2009 were used in this research. All of these materials were maintained using standard management practices, at Lindsay Point, Victoria, Australia (initial 89 progeny) and Dareton, NSW, Australia (the other 231 progeny).

5.1.2.2 Selection of a restriction enzyme(s)

To select a restriction enzyme that might be suitable for digestion of the almond genome, a custom python script (www.python.org) was developed and applied to the peach whole genome sequence assembly v1.0 (www.rosaceae.org) to obtain the distribution of fragment lengths that might be obtained from digestion with each of three methylation sensitive restriction enzymes (*Ape*KI, *Pst*I, *Hpa*II) and from combinations among them. A treatment that was predicted to generate a large number of fragments between 150 and 500 bp was selected.

5.1.2.3 DNA extraction

The DNA samples used for the first set of 89 progeny were aliquots from the original samples that had been used for mapping by Tavassolian et al. (2010). These samples had been extracted from young leaves of Nonpareil × Lauranne F_1 plants using the Lamboy and Alpha DNA extraction method (Lamboy 1998). The DNA was quality checked on 1% agarose gel, quantified using PicoGreen® intercalating dye (Invitrogen, Carlsbad, CA, USA) and normalized to a working concentration of 20 ng/µL. DNA from each of the additional 231 progeny was extracted from young leaves using an Oktopure™ DNA extraction protocol that had been optimised for almond (LGC Ltd, Teddington, UK).

5.1.2.4 Library construction and sequencing

A set of 96 barcode sequences and a pair of common adapter sequences for the selected enzyme (*Ape*KI) were downloaded from the Cornell University GBS website (www.maizegenetics.edu). The GBS protocol proposed by Elshire et al. (2011) was slightly modified for almond. The original protocol used 100 ng of DNA for each sample. Here, 200 ng samples were used. In the original protocol, DNA samples and adapters were dried before proceeding to the restriction digestion. In this research, DNA was used in liquid form and restriction digestion of DNA was carried out using *Ape*KI before adding adapters.

The complementary top and bottom strands of each barcode and common adapter were diluted to 10 µM with 10 × adapter buffer and annealed using following PCR conditions: 95°C for 1 min (ramp down to 30°C, by decreasing 1°C per cycle). The resulting double-stranded barcode and common adapters were diluted separately in 1 × TE to 0.6 ng/µL. Diluted adapters were quantified using PicoGreen® intercalating dye, normalised to 0.1 µM with 1 × TE and mixed together in a 1:1 ratio in a 96 well plate.

A titration experiment was carried out to investigate the optimal adapter to DNA ratio. A DNA pool was prepared by mixing equal amounts of DNA from 10 Nonpareil/Lauranne F_1 progeny. Eight 200 ng samples were drawn from the DNA pool. Each of these samples was incubated for 2 h at 75°C with *Ape*KI (New England Biolabs, Ipswitch, MA, USA) in 20 µL of volumes containing 2 µL of 10 × NEB buffer 3 and 3.2 U *Ape*KI. Water was added as required to make a total volume of 20 µL from the eight volumes of 0.1 µM adapters (2, 5, 8, 10, 12, 15, 18, 20 µL) and were ligated to the digested DNA by adding a total volume of 10 µL of a solution containing 200 U of T4 DNA ligase (New England Biolabs, Ipswitch, MA, USA), 5 µL of 10 × ligation buffer with the following thermocyclic conditions: 2 h at 22°C followed by a 20 min ligation denaturation at 65°C. Ligation products were purified using a Purelink PCR Purification Kit (Invitrogen, Carlsbad, CA, USA) as per the manufacturer's instructions. Each purified ligation product was resuspended in a final volume 50 µL. For the final library, 10 µL of each of purified ligation product was used in a 25 µL PCR reaction with 2 µL of 10 µM paired end (PE) primers (Table S3.1) and 12.5 µL of 2 × Taq Master Mix (New England Biolabs, Ipswitch, MA, USA). The PCR conditions used were: 30 s at 95°C, 15 cycles of 30 s at 95°C, 20 s at 65°C, 30 s at 68°C followed by a final extension at 72°C for 5 min. Each amplified library was purified using a Purelink PCR Purification Kit as described above and eluted in a final volume of 30 µL. Each of the libraries (2 µL) was run on 2% agarose at 90 V for 30 min to evaluate the library and the adapter dimer peaks.

After selecting a suitable adapter amount (4.5 ng in a volume of 15 µL, which provided a satisfactory library with no adapter dimer peak), library preparation was carried out for the first set of 89 Nonpareil × Lauranne progeny, the parents (in triplicate) using 200 ng (10 µL of 20 ng/µL) of DNA template and a water sample following the method described above. Initial reactions were carried out in a 96-well plate using a separate well for each individual in the population. After adapter ligation, samples were pooled for purification, PCR amplification, evaluation and sequencing. The pooled library was sequenced using single end sequencing (100 bp reads) on one flow-cell lane of an

Illumina HiSeq 2000 instrument at the Australian Genome Research Facility Ltd (Melbourne, Australia).

In addition, a second GBS library was prepared for sequencing. This library consisted of 87 progeny from the first GBS library described above, with the samples taken from the same digestion and adapter ligation reactions. Further, two pooled DNA samples were included with one consisting of DNA from 25 progeny that had been evaluated as having high tocopherol concentration and the other consisting of DNA from 25 progeny having low tocopherol concentrations, these 50 lines were selected from the second set of 231 Nonpareil × Lauranne progeny. This library was sequenced using paired end sequencing (74 bp reads) on one flow-cell with four lanes of an Illumina NextGen 500 instrument at the Australian Genome Research Facility Ltd (Melbourne, Australia).

5.1.2.5 SNP discovery

The raw GBS data were analysed using the universal network enabled analysis kit (UNEAK) pipeline in the TASSEL 3.0 software (Bradbury et al. 2007; Lu et al. 2013). The level of almond genome coverage obtained was calculated using the Lander–Waterman equation (Lander and Waterman 1988). Raw SNP calls were filtered to select tags with 80% coverage across samples. Tag pairs with minor allele frequencies (MAF) between 0.2 and 0.3 ($0.2 < MAF < 0.3$) were selected for use in linkage map construction. This was based on the expectation that the most informative SNPs (those that are heterozygous in only one parent) would have a minor allele frequency of 0.25.

The GBS data from the second library were analysed as described above. The MasterTags file, which contains all the unique tags, resulting from the initial GBS data was used as a reference to identify the common tag pairs between the first and second libraries.

5.1.2.6 Construction of linkage maps

For the construction of initial framework linkage maps for Nonpareil and Lauranne, the data set of SNP-bearing tags with MAF of about 0.25 (0.2 < MAF < 0.3) was filtered to exclude tags with low coverage across the genotypes using a read depth cut off of 5 and a heterozygosity cut off of 0.01. The resulting data set was separated into two parental datasets based on whether the tags were homozygous in Nonpareil and heterozygous in Lauranne, or *vice versa*. Each parental dataset was further filtered to retain only the SNPs missing no more than 20 data points per marker and with segregation ratios not deviating significantly from 3:1. A double pseudo-test cross (two-way pseudo-testcross) strategy was used to construct separate parental maps for Nonpareil and Lauranne using the backcross (BC) format in the ASMap package in R (www.CranR.org), according to the following map construction strategy:

1. An initial framework linkage map was constructed using data from progeny for which there were no missing data for the selected SNPs. A few SSR, RAPD and ISSR markers that Tavassolian et al. (2010) had reported to be homozygous in one parent and heterozygous in the other were included in addition to the selected SNPs. The linkage mapping was carried out using the minimum spanning tree map algorithm (MSTmap) (Wu et al. 2008) as implemented in ASMap to assign markers to linkage groups and to order them within linkage groups. A p-value of 0.0001 was used to declare whether markers belong to the same linkage group. The Kosambi mapping function was used to calculate genetic distances in centiMorgans (cM) (Kosambi 1944).

2. For each linkage group (LG), ASMap was used to generate heat maps (rf/LOD plots) to evaluate pairwise associations between markers. In cases where markers or a group of markers appeared to have had their alleles assigned to the incorrect parents, genotype designations were reassigned using the 'switchAlleles' function of R/qtl. The maps were re-estimated using the mstmap.cross function.

3. To further improve the quality of the Nonpareil and Lauranne linkage maps, markers were checked for segregation distortion and for the numbers of double crossover events involving adjacent marker intervals. Markers that deviated significantly from the expected 1:1 ratio and those that associated with high numbers of double crossover events were removed manually. The maps were re-estimated using the mstmap.cross function.

4. The orientation of each linkage group of the Nonpareil and Lauranne maps was established by comparing the maps constructed using the SNP data with the published maps of Tavassolian et al. (2010) using the AlignCross function of ASMap.

From the resulting framework maps for Nonpareil and Lauranne, sets of evenly distributed GBS markers from the eight linkage groups were selected for the design of fluorescence-based allele-specific assays.

5.1.2.7 Primer design

Primer sets were designed based on Nonpareil heterozygosity (using GBS tags heterozygous in Nonpareil and homozygous in Lauranne) and Lauranne heterozygosity (using GBS tags heterozygous in Lauranne and homozygous in Nonpareil). Each GBS tag (64 bp long) was aligned to the Nonpareil genomic sequence (unpublished data provided by Pere Arús, IRTA, Spain) using the Blast tool in Geneious software version 9.1.3 (Kearse et al. 2012), to obtain a sequence of about 100 bp long with the SNP located near its midpoint. Each of these SNP-bearing sequences was used to design a set of three primers (two allele-specific primers and one common primer) using the Kraken™ software (LGC Ltd, Teddington, UK). The primer sets were named using the prefix WriPdK, with Wri referring to the Waite Research Institute, Pd referring to almond (*Prunus dulcis*) and K referring to the competitive allele-specific primer (KASP) technology (LGC Ltd, Teddington, UK).

5.1.2.8 Primer validation

A total of 308 primer sets (146 designed from GBS tags that were heterozygous in Nonpareil and 162 designed from GBS tags that were heterozygous in Lauranne) were assayed on a validation panel consisting of DNA samples of Nonpareil, Lauranne and seven Nonpareil × Lauranne F_1 progeny and a water sample (as a negative control). Samples of 10 ng of DNA (5 µL of 2 ng/µL) were dried at 55°C for 1 h. A mixture of 0.028 µL containing 12 µM of the allele-specific forward primers and 30 µM of the common reverse primer and 1.972 µL of 1 × KASP Master Mix (LGC Ltd, Teddington, UK) was added to each reaction sample. PCR amplification was conducted using the standard KASP PCR protocol in a Hydrocycler-16 PCR system (LGC Ltd, Teddington, UK). Fluorescence detection was performed in a Pherastar Plus plate reader (BMG LABTECH, Germany). Data were analysed using Kraken™ software (LGC Ltd, Teddington, UK). Primer sets that detected polymorphism in the validation panel were selected and used to screen 311 Nonpareil × Lauranne progeny: 80 of the 89 progeny that had been used to prepare the GBS library, plus the additional 231 Nonpareil × Lauranne progeny.

5.1.2.9 Reconstruction of linkage maps for Nonpareil and Lauranne using newly designed KASP markers

Linkage maps were constructed for each parent using KASP marker data from the 231 Nonpareil × Lauranne progeny. The mapping procedure was same as described in the initial framework linkage map construction. The maps were drawn using the MapChart software (Voorrips 2002).

5.1.2.10 Comparative mapping

All unique sequence reads of 64 bp long (sequence coverage ≥ 10) obtained for almond were aligned against the peach (*Prunus persica*) whole genome sequence assembly v2.0.a1 (www.rosaceae.org) using BLAST+ tool version 2.2.27 (http://www.ncbi.nlm.nih.gov/blast) and plotted against the physical positions on the first eight pseudomolecules and scaffolds in the peach genome assembly. Each of the sequences (ranging in length from 64 bp to 200 bp) for which an

assay had been developed was aligned against the peach whole genome assembly as described above. Sequences were considered to have been anchored to the peach genome if they mapped to a unique site with greater than 90% sequence similarity and an E-value below 1e-15. The relative positions of the markers were visualised by plotting the almond genetic maps against the *peach* pseudomolecules and scaffolds (the first eight pseudomolecules and in the peach genome assembly) using the Circlize package (Gu et al. 2014) in the R statistical environment (www.CranR.org).

Relationships between the physical and genetic distances were examined by plotting the genetic positions of markers in the Nonpareil and Lauranne linkage maps to the peach against the physical positions at which the markers had been anchored.

5.1.3 Results

5.1.3.1 Selection of a suitable restriction enzyme for almond

The total number of fragments between 150 and 500 bp (Fig. 5.1.1) that resulted from *in-silico* digestion analysis with individual restriction enzymes was higher for *Ape*KI (77,153) than for *Hpa*II (40,873) or *Pst*I (4,325). The combination of *Ape*KI with *Hpa*II gave an even higher number of these fragments: 144,065. The numbers and the size distributions of fragments indicated that either *Ape*KI or the combination of *Ape*KI with *Hpa*II could digest at least 50% of the peach genome into fragments between 150 and 500 bp. Based on this result and the fact that *Ape*KI had already been tested with other crops (Bielenberg et al. 2015; Elshire et al. 2011; Guajardo et al. 2015; Kujur et al. 2015; Lu et al. 2013), *Ape*KI was selected here.

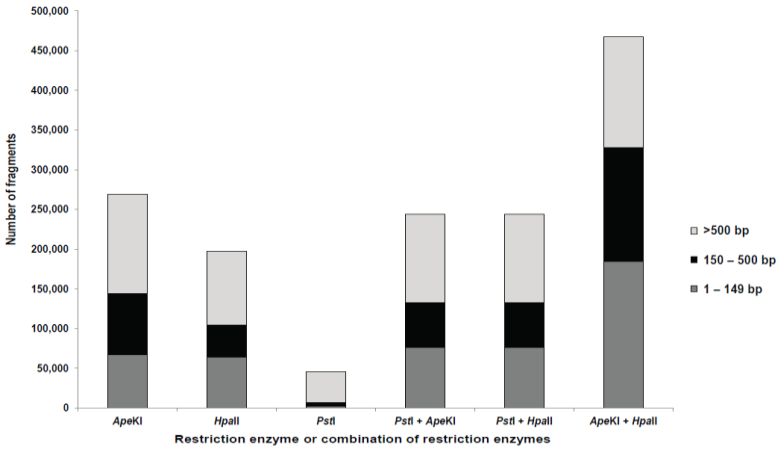


Fig. 5.1.1 Fragment size distributions for *in-silico* digestion of the peach genomic sequence with the restriction enzymes *Ape*KI, *Hpa*II, *Pst*I and combinations of these enzymes. Fragments ranging between 150 and 500 bp are considered to be useful for genotyping-by-sequencing.

5.1.3.2 Nonpareil × Lauranne GBS library preparation and sequencing data analysis

Based on the results of the DNA adapter titration experiment, 4.5 ng of adapter was selected for use with 200 ng of DNA. The library obtained with this amount of adapter contained fragments ranging from 150 bp to 400 bp, with a higher proportion of fragments between 150 bp and 300 bp, and did not exhibit any evidence of adapter dimers (≤ 120 bp) (Fig. 5.1.2). The initial GBS library generated 21.6 Gb of sequence data, with a total of 186 million good sequence reads (call rate > 0.8, QC ≥ 30). The mean number of sequence reads per sample was 2,129,827 and the number of unique tags obtained increased with the increasing number of reads (Fig. 5.1.3). Regression analysis indicated strong linear relationship between the number of sequence reads and the number of unique tags (r^2 = 0.92).

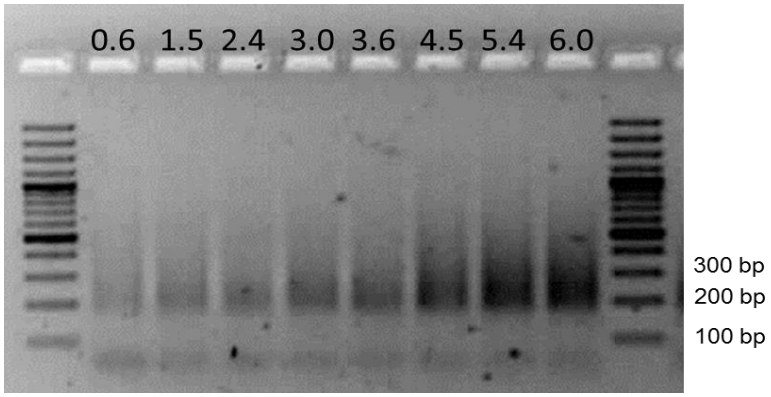


Fig. 5.1.2 Electrophoresis of GBS libraries resulting from different adapter concentrations ligated with 200 ng of DNA. Lanes are marked with the amounts (in ng) of adapter used. The first and last lanes contained a 100 bp DNA ladder.

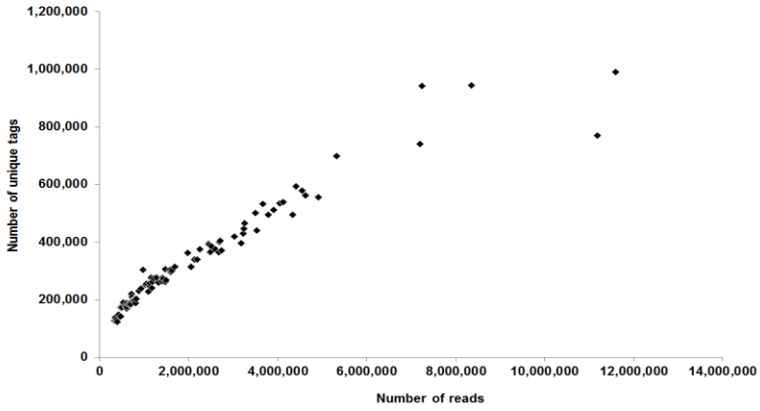


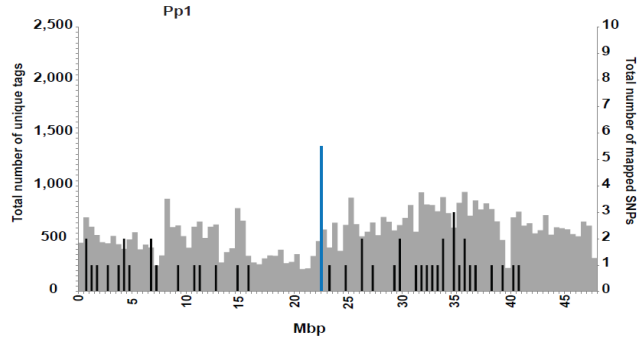
Fig. 5.1.3 The relationship between the number of sequence reads and the number of unique tags obtained for each individual in the Nonpareil × Lauranne GBS library. This was drawn using the data obtained for the initial GBS library.

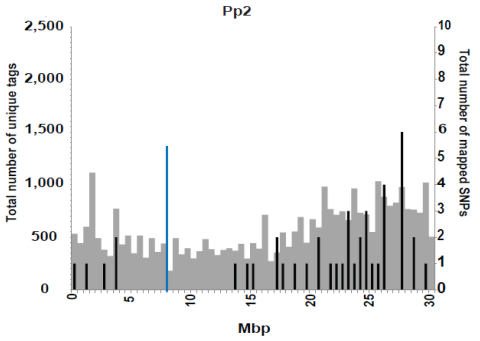
A total of 453,648 unique tags were obtained across all samples. Only 68% tags were mapped to the peach genome. The numbers mapped to the peach scaffolds (Pp) ranged from 33,045 (Pp5) to 53,923 (Pp1) (Table 5.1.1). Tags were mapped throughout the entire length of each scaffold, but with some variation in read depth (Fig. 5.1.4). There were a few regions (e.g. on Pp5 and Pp7) with very high read depth.

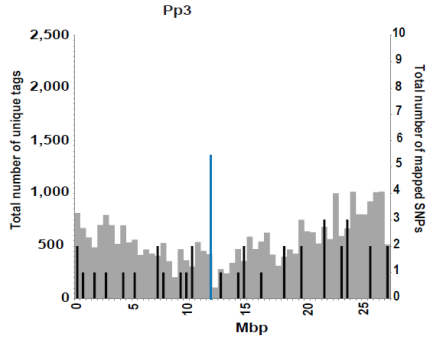
Table 5.1.1. Total number of unique tags mapped to the peach genome sequence assembly using the data from the initial GBS library.

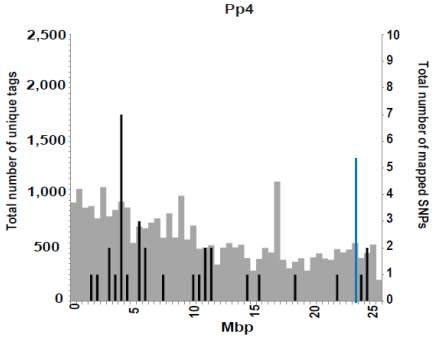
Peach scaffold (Pp)	Total number of unique tags
Pp1	53,923
Pp2	34,341
Pp3	30,594
Pp4	30,881
Pp5	33,045
Pp6	41,743
Pp7	37,331
Pp8	40,334
Other scaffolds	6,608
Unmapped	1,44,677

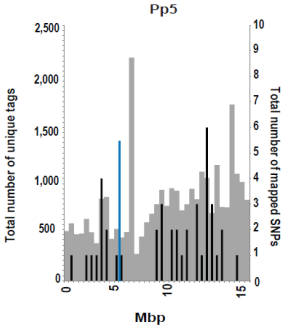
Fig. 5.1.4 Comparison of unique tags and positions of SNPs (KASP markers) mapped to the peach genome sequence assembly (Pp). The unique tags and positions of SNPs that were mapped to the first eight scaffolds (Pp1 to Pp8) of the peach genome sequence assembly. In each case, physical distances of the peach genome scaffolds are shown on the horizontal axis (bp), the number unique tags mapped to each 500-kb region is represented by a grey bar, the number of SNPs mapped to each 500-kb region is represented by a black bar, and the estimated position of the centromeric region based on information from the peach whole genome sequence assembly v2.0.a1 is represented by a vertical blue line.

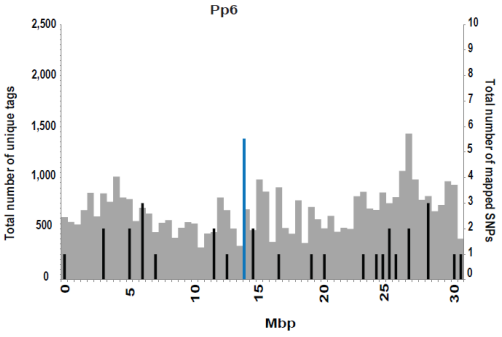

Pp1
Total number of unique tags
Total number of mapped SNPs
Mbp

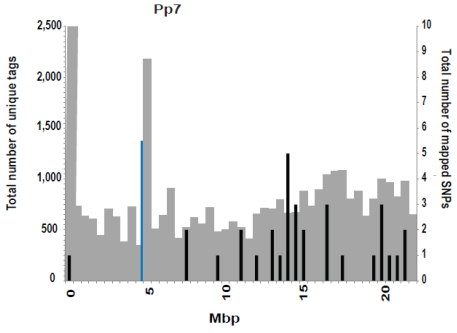

Pp2
Total number of unique tags
Total number of mapped SNPs
Mbp

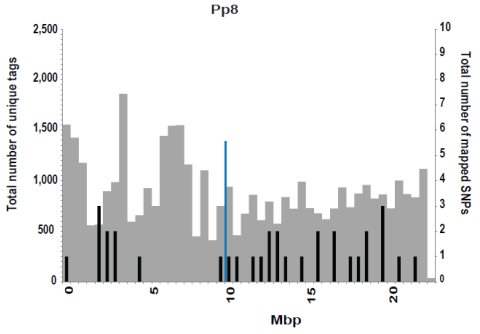

Pp3
Total number of unique tags
Total number of mapped SNPs
Mbp


Pp4
Total number of unique tags
Total number of mapped SNPs
Mbp


Pp5
Total number of unique tags
Total number of mapped SNPs
Mbp


Pp6
Total number of unique tags
Total number of mapped SNPs
Mbp


Pp7
Total number of unique tags
Total number of mapped SNPs
Mbp


Pp8
Total number of unique tags
Total number of mapped SNPs
Mbp

From the unique tags, a total of 11,936 SNP containing tag pairs were identified. With the application of a series of filters, over 300 tag pairs that were considered suitable for mapping were obtained for each parent (Table 5.1.2).

Table 5.1.2. Sorted tag pairs for the almond genome using data from the initial GBS library.

Filtered tag pair groups	**Total number of tag pairs**
Tag pairs with minor allele frequency (MAF ≥ 0.05, missing data ≤ 20%)	11,936
Tag pairs with minor allele frequency (0.2 ≤ MAF ≥ 0.3)	4,567
Tag pairs with read depth (≥ 4), heterozygote cut off (0.1)	666
Tag pairs in Nonpareil parental data set (heterozygous in Nonpareil and homozygous in Lauranne, after removing distorted markers)	333
Tag pairs in Lauranne parental data set (heterozygous in Lauranne and homozygous in Nonpareil, after removing distorted markers)	302

The second GBS library generated 52.82 Gb of sequence data with a total of 356 million good paired-end sequence reads (call rate > 0.8, QC ≥ 30). The mean number of sequence reads per sample was 4,254,567. A total of 4,600 tag pairs were obtained for linkage map construction. Of these, only 33 were new.

5.1.3.3 Linkage maps for Nonpareil and Lauranne

The initial framework map that was constructed for Nonpareil based on complete data for 52 progeny had 321 markers (302 GBS markers, 9 SSR markers, 5 ISSR markers, 3 RAPD markers and 2 markers for the self-incompatibility locus, *S*) on eight linkage groups with a total length of 1,152.1 cM (Figs. 5.1.5 and 5.1.6). The longest interval on the map was a 20.3 cM interval on LG4.

Similarly for Lauranne, an initial map was constructed using 55 progeny. It had eight linkage groups, 283 markers (265 GBS markers, 5 SSR markers, 3 RAPD markers, 8 ISSR markers and 2 markers for the *S* locus) and was 1,371.3 cM long (Figs. 5.1.7 and 5.1.8). On this map, the maximum distance between markers was 25.7 cM for an interval on LG1.

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 with maps constructed for the same linkage groups using data from KASP primer sets designed to assay the same polymorphisms. Framework maps are based on data from 52 progeny (left) and the KASP maps are based on data from 231 progeny (right). Markers that are common to both maps are underlined in framework maps and KASP markers that were designed from SNP-bearing sequence tags resulting from the second GBS data are marked with asterisks (right).

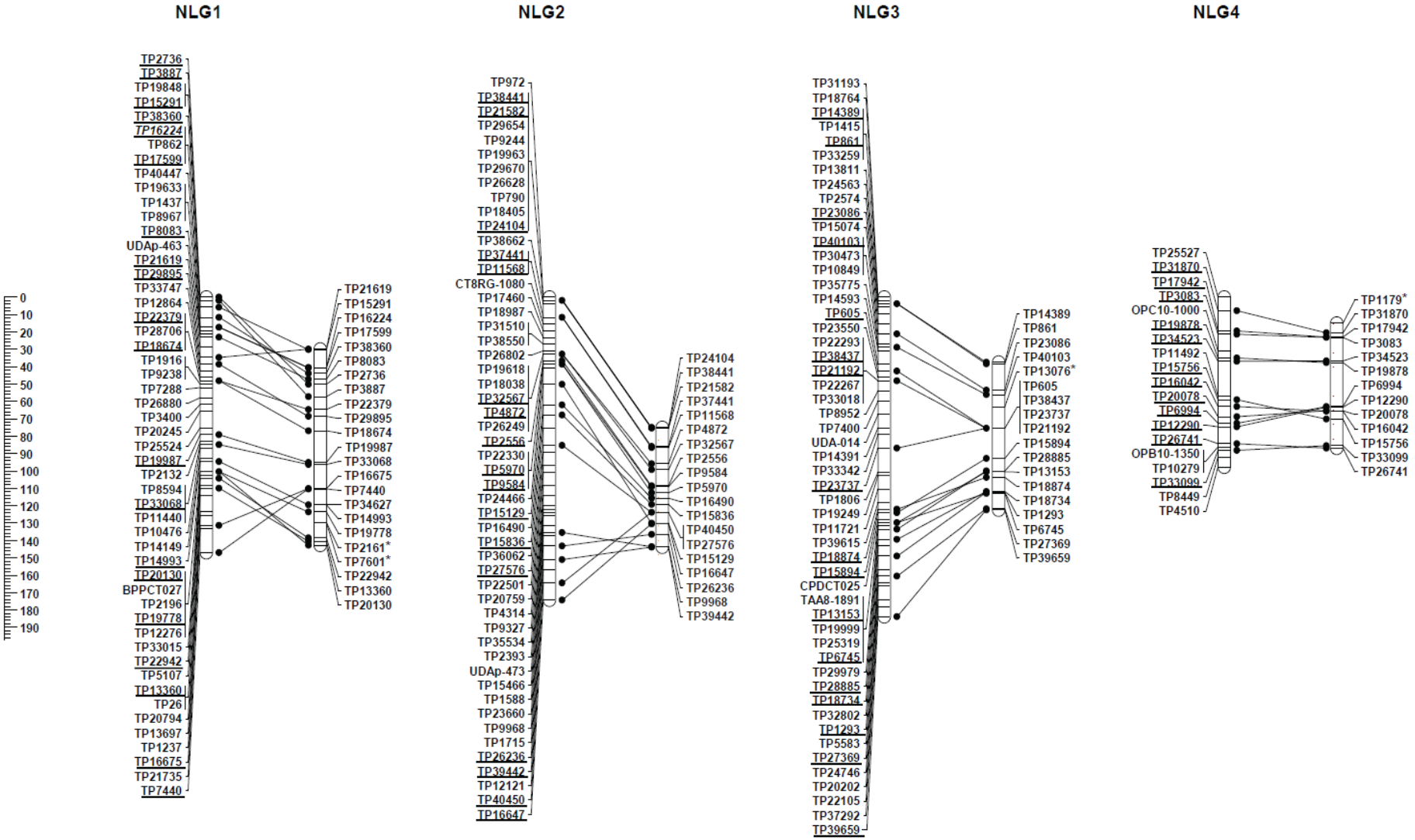

NLG1
NLG2
NLG3
NLG4
0
10
20
30
40
50
60
70
80
90
100
110
120
130
140
150
160
170
180
190
TP2736
TP3887
TP19848
TP15291
TP38360
TP16224
TP862
TP17599
TP40447
TP19633
TP1437
TP8967
TP8083
UDAp-463
TP21619
TP29895
TP33747
TP12864
TP22379
TP28706
TP18674
TP1916
TP9238
TP7288
TP26880
TP3400
TP20245
TP25524
TP19987
TP2132
TP8594
TP33068
TP11440
TP10476
TP14149
TP14993
TP20130
BPPCT027
TP2196
TP19778
TP12276
TP33015
TP22942
TP5107
TP13360
TP26
TP20794
TP13697
TP1237
TP16675
TP21735
TP7440
TP21619
TP15291
TP16224
TP17599
TP38360
TP8083
TP2736
TP3887
TP22379
TP29895
TP18674
TP19987
TP33068
TP16675
TP7440
TP34627
TP14993
TP19778
TP2161*
TP7601*
TP22942
TP13360
TP20130
TP972
TP38441
TP21582
TP29654
TP9244
TP19963
TP29670
TP26628
TP790
TP18405
TP24104
TP38662
TP37441
TP11568
CT8RG-1080
TP17460
TP18987
TP31510
TP38550
TP26802
TP19618
TP18038
TP32567
TP4872
TP26249
TP2556
TP22330
TP5970
TP9584
TP24466
TP15129
TP16490
TP15836
TP36062
TP27576
TP22501
TP20759
TP4314
TP9327
TP35534
TP2393
UDAp-473
TP15466
TP1588
TP23660
TP9968
TP1715
TP26236
TP39442
TP12121
TP40450
TP16647
TP24104
TP38441
TP21582
TP37441
TP11568
TP4872
TP32567
TP2556
TP9584
TP5970
TP16490
TP15836
TP40450
TP27576
TP15129
TP16647
TP26236
TP9968
TP39442
TP31193
TP18764
TP14389
TP1415
TP861
TP33259
TP13811
TP24563
TP2574
TP23086
TP15074
TP40103
TP30473
TP10849
TP35775
TP14593
TP605
TP23550
TP22293
TP38437
TP21192
TP22267
TP33018
TP8952
TP7400
UDA-014
TP14391
TP33342
TP23737
TP1806
TP19249
TP11721
TP39615
TP18874
TP15894
CPDCT025
TAA8-1891
TP13153
TP19999
TP25319
TP6745
TP29979
TP28885
TP18734
TP32802
TP1293
TP5583
TP27369
TP24746
TP20202
TP22105
TP37292
TP39659
TP14389
TP861
TP23086
TP40103
TP13076*
TP605
TP38437
TP23737
TP21192
TP15894
TP28885
TP13153
TP18874
TP18734
TP1293
TP6745
TP27369
TP39659
TP25527
TP31870
TP17942
TP3083
OPC10-1000
TP19878
TP34523
TP11492
TP15756
TP16042
TP20078
TP6994
TP12290
TP26741
OPB10-1350
TP10279
TP33099
TP8449
TP4510
TP1179*
TP31870
TP17942
TP3083
TP34523
TP19878
TP6994
TP12290
TP20078
TP16042
TP15756
TP33099
TP26741

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 with maps constructed for the same linkage groups using data from KASP primer sets designed to assay the same polymorphisms. Framework maps are based on data from 52 progeny (left) and the KASP maps are based on data from 231 progeny (right). Markers that are common to both maps are underlined in framework maps and KASP markers that were designed from SNP-bearing sequence tags resulting from the second GBS data marked with asterisks (right).

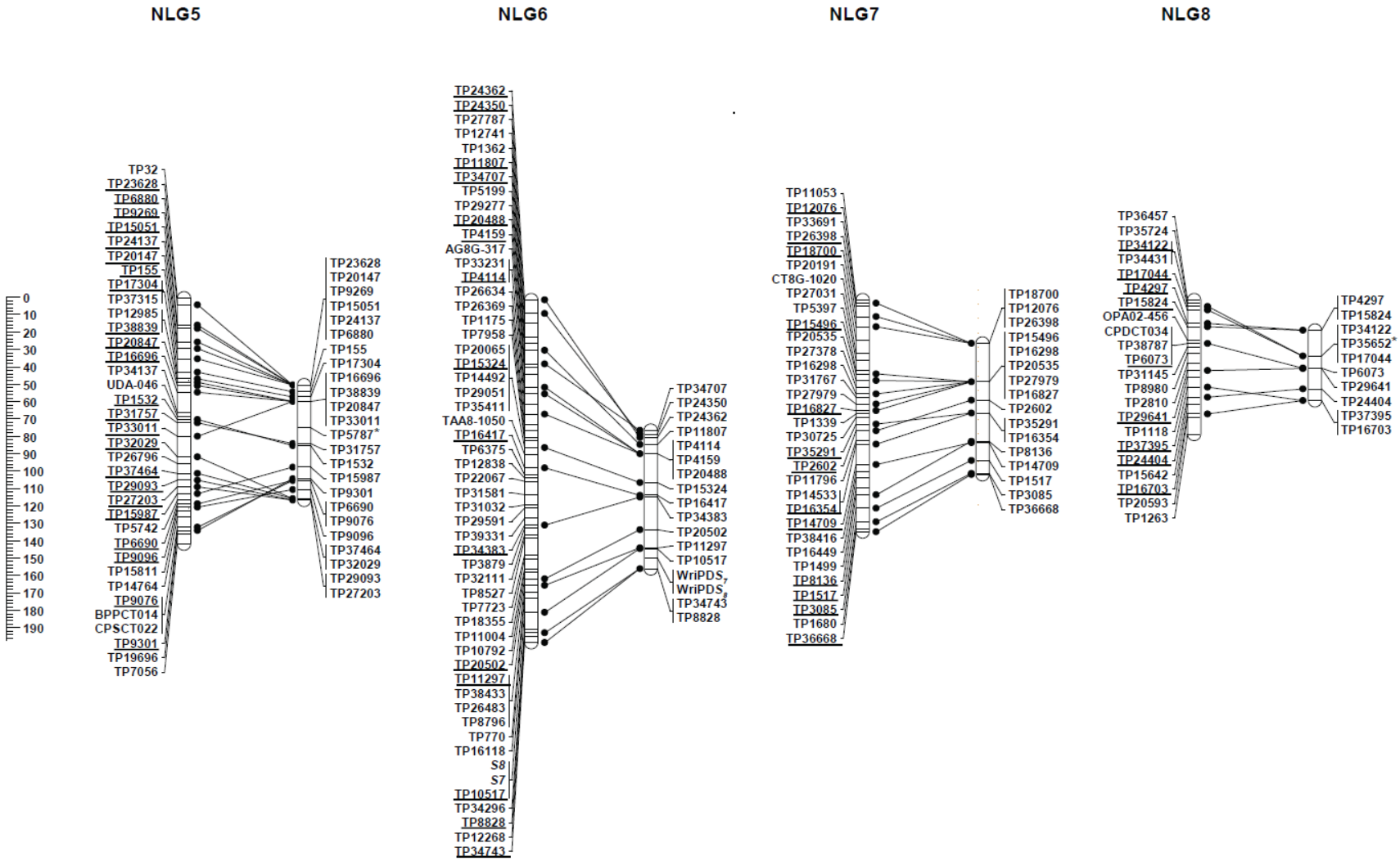

NLG5
NLG6
NLG7
NLG8
0
10
20
30
40
50
60
70
80
90
100
110
120
130
140
150
160
170
180
190
TP32
TP23628
TP6880
TP9269
TP15051
TP24137
TP20147
TP155
TP17304
TP37315
TP12985
TP38839
TP20847
TP16696
TP34137
UDA-046
TP1532
TP31757
TP33011
TP32029
TP26796
TP37464
TP29093
TP27203
TP15987
TP5742
TP6690
TP9096
TP15811
TP14764
TP9076
BPPCT014
CPSCT022
TP9301
TP19696
TP7056
TP23628
TP20147
TP9269
TP15051
TP24137
TP6880
TP155
TP17304
TP16696
TP38839
TP20847
TP33011
TP5787*
TP31757
TP1532
TP15987
TP9301
TP6690
TP9076
TP9096
TP37464
TP32029
TP29093
TP27203
TP24362
TP24350
TP27787
TP12741
TP1362
TP11807
TP34707
TP5199
TP29277
TP20488
TP4159
AG8G-317
TP33231
TP4114
TP26634
TP26369
TP1175
TP7958
TP20065
TP15324
TP14492
TP29051
TP35411
TAA8-1050
TP16417
TP6375
TP12838
TP22067
TP31581
TP31032
TP29591
TP39331
TP34383
TP3879
TP32111
TP8527
TP7723
TP18355
TP11004
TP10792
TP20502
TP11297
TP38433
TP26483
TP8796
TP770
TP16118
S8
S7
TP10517
TP34296
TP8828
TP12268
TP34743
TP34707
TP24350
TP24362
TP11807
TP4114
TP4159
TP20488
TP15324
TP16417
TP34383
TP20502
TP11297
TP10517
WriPDS7
WriPDS3
TP34743
TP8828
TP11053
TP12076
TP33691
TP26398
TP18700
TP20191
CT8G-1020
TP27031
TP5397
TP15496
TP20535
TP27378
TP16298
TP31767
TP27979
TP16827
TP1339
TP30725
TP35291
TP2602
TP11796
TP14533
TP16354
TP14709
TP38416
TP16449
TP1499
TP8136
TP1517
TP3085
TP1680
TP36668
TP18700
TP12076
TP26398
TP15496
TP16298
TP20535
TP27979
TP16827
TP2602
TP35291
TP16354
TP8136
TP14709
TP1517
TP3085
TP36668
TP36457
TP35724
TP34122
TP34431
TP17044
TP4297
TP15824
OPA02-456
CPDCT034
TP38787
TP6073
TP31145
TP8980
TP2810
TP29641
TP1118
TP37395
TP24404
TP15642
TP16703
TP20593
TP1263
TP4297
TP15824
TP34122
TP35652*
TP17044
TP6073
TP29641
TP24404
TP37395
TP16703

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 with maps constructed for the same linkage groups using data from KASP assays designed to assay the same polymorphisms. Framework maps are based on data from 55 progeny (left) and the KASP maps are based on data from 231 progeny (right). Markers that are common to both maps are underlined in framework maps and KASP markers that were designed from SNP-bearing sequence tags resulting from the second GBS data are marked with asterisks (right).

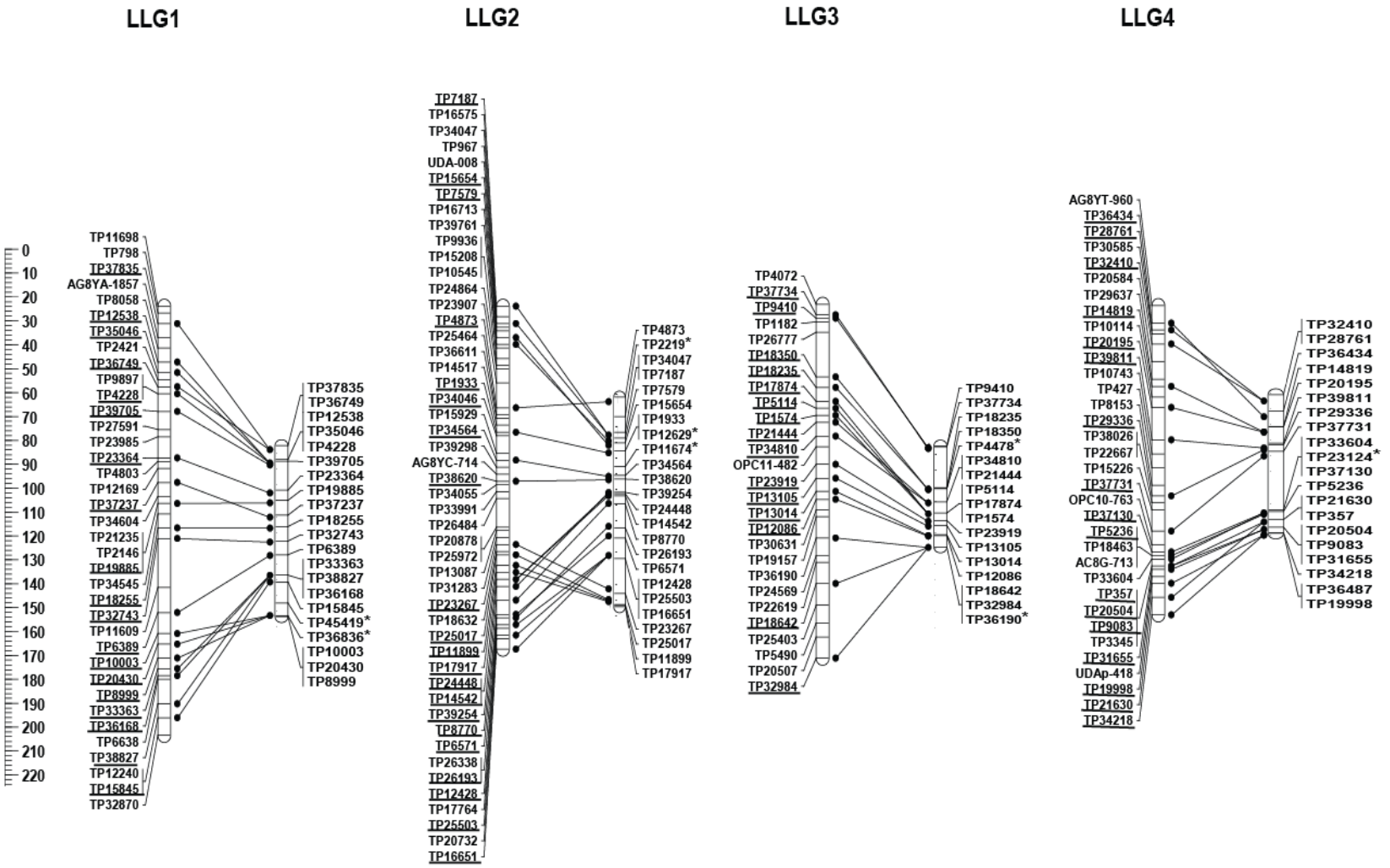
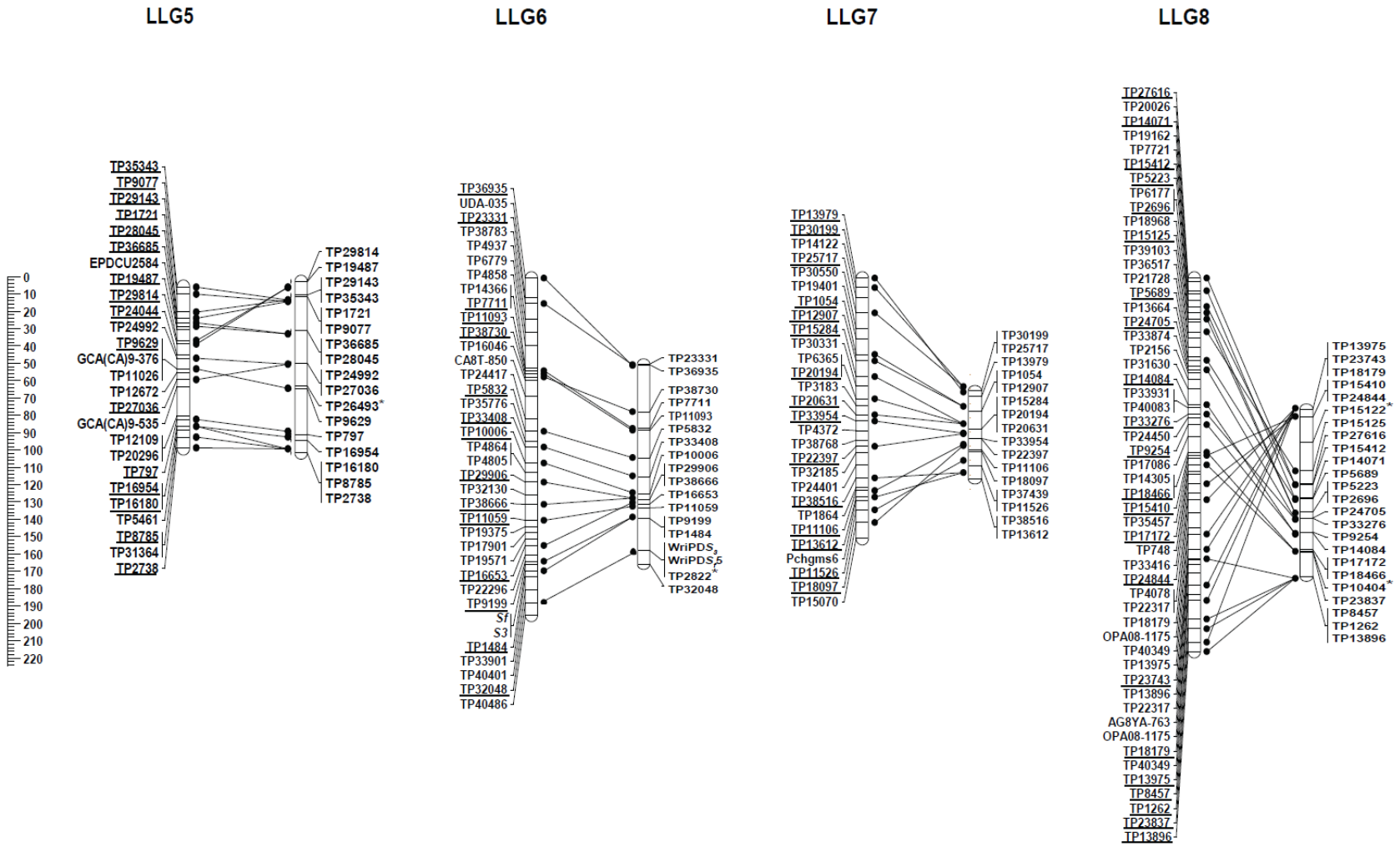

LLG1
LLG2
LLG3
LLG4
0
10
20
30
40
50
60
70
80
90
100
110
120
130
140
150
160
170
180
190
200
210
220
TP11698
TP798
TP37835
AG8YA-1857
TP8058
TP12538
TP35046
TP2421
TP36749
TP9897
TP4228
TP39705
TP27591
TP23985
TP23364
TP4803
TP12169
TP37237
TP34604
TP21235
TP2146
TP19885
TP34545
TP18255
TP32743
TP11609
TP6389
TP10003
TP20430
TP8999
TP33363
TP36168
TP6638
TP38827
TP12240
TP15845
TP32870
TP37835
TP36749
TP12538
TP35046
TP4228
TP39705
TP23364
TP19885
TP37237
TP18255
TP32743
TP6389
TP33363
TP38827
TP36168
TP15845
TP45419*
TP36836*
TP10003
TP20430
TP8999
TP7187
TP16575
TP34047
TP967
UDA-008
TP15654
TP7579
TP16713
TP39761
TP9936
TP15208
TP10545
TP24864
TP23907
TP4873
TP25464
TP36611
TP14517
TP1933
TP34046
TP15929
TP34564
TP39298
AG8YC-714
TP38620
TP34055
TP33991
TP26484
TP20878
TP25972
TP13087
TP31283
TP23267
TP18632
TP25017
TP11899
TP17917
TP24448
TP14542
TP39254
TP8770
TP6571
TP26338
TP26193
TP12428
TP17764
TP25503
TP20732
TP16651
TP4873
TP2219*
TP34047
TP7187
TP7579
TP15654
TP1933
TP12629*
TP11674*
TP34564
TP38620
TP39254
TP24448
TP14542
TP8770
TP26193
TP6571
TP12428
TP25503
TP16651
TP23267
TP25017
TP11899
TP17917
TP4072
TP37734
TP9410
TP1182
TP26777
TP18350
TP18235
TP17874
TP5114
TP1574
TP21444
TP34810
OPC11-482
TP23919
TP13105
TP13014
TP12086
TP30631
TP19157
TP36190
TP24569
TP22619
TP18642
TP25403
TP5490
TP20507
TP32984
TP9410
TP37734
TP18235
TP18350
TP4478*
TP34810
TP21444
TP5114
TP17874
TP1574
TP23919
TP13105
TP13014
TP12086
TP18642
TP32984
TP36190*
AG8YT-960
TP36434
TP28761
TP30585
TP32410
TP20584
TP29637
TP14819
TP10114
TP20195
TP39811
TP10743
TP427
TP8153
TP29336
TP38026
TP22667
TP15226
TP37731
OPC10-763
TP37130
TP5236
TP18463
AC8G-713
TP33604
TP357
TP20504
TP9083
TP3345
TP31655
UDAp-418
TP19998
TP21630
TP34218
TP32410
TP28761
TP36434
TP14819
TP20195
TP39811
TP29336
TP37731
TP33604
TP23124*
TP37130
TP5236
TP21630
TP357
TP20504
TP9083
TP31655
TP34218
TP36487
TP19998

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 with maps constructed for the same linkage groups using data from KASP assays designed to assay the same polymorphisms. Framework maps are based on data from 55 progeny (left) and the KASP maps are based on data from 231 progeny (right). Markers that are common to both maps are underlined in framework maps and KASP markers that were designed from SNP-bearing sequence tags resulting from the second GBS data are marked with asterisks (right).


LLG5
LLG6
LLG7
LLG8
0
10
20
30
40
50
60
70
80
90
100
110
120
130
140
150
160
170
180
190
200
210
220
TP35343
TP9077
TP29143
TP1721
TP28045
TP36685
EPDCU2584
TP19487
TP29814
TP24044
TP24992
TP9629
GCA(CA)9-376
TP11026
TP12672
TP27036
GCA(CA)9-535
TP12109
TP20296
TP797
TP16954
TP16180
TP5461
TP8785
TP31364
TP2738
TP29814
TP19487
TP29143
TP35343
TP1721
TP9077
TP36685
TP28045
TP24992
TP27036
TP26493*
TP9629
TP797
TP16954
TP16180
TP8785
TP2738
TP36935
UDA-035
TP23331
TP38783
TP4937
TP6779
TP4858
TP14366
TP7711
TP11093
TP38730
TP16046
CA8T-850
TP24417
TP5832
TP35776
TP33408
TP10006
TP4864
TP4805
TP29906
TP32130
TP38666
TP11059
TP19375
TP17901
TP19571
TP16653
TP22296
TP9199
S1
S3
TP1484
TP33901
TP40401
TP32048
TP40486
TP23331
TP36935
TP38730
TP7711
TP11093
TP5832
TP33408
TP10006
TP29906
TP38666
TP16653
TP11059
TP9199
TP1484
WriPDS_3
WriPDS_5
TP2822*
TP32048
TP13979
TP30199
TP14122
TP25717
TP30550
TP19401
TP1054
TP12907
TP15284
TP30331
TP6365
TP20194
TP3183
TP20631
TP33954
TP4372
TP38768
TP22397
TP32185
TP24401
TP38516
TP1864
TP11106
TP13612
Pchgms6
TP11526
TP18097
TP15070
TP30199
TP25717
TP13979
TP1054
TP12907
TP15284
TP20194
TP20631
TP33954
TP22397
TP11106
TP18097
TP37439
TP11526
TP38516
TP13612
TP27616
TP20026
TP14071
TP19162
TP7721
TP15412
TP5223
TP6177
TP2696
TP18968
TP15125
TP39103
TP36517
TP21728
TP5689
TP13664
TP24705
TP33874
TP2156
TP31630
TP14084
TP33931
TP40083
TP33276
TP24450
TP9254
TP17086
TP14305
TP18466
TP15410
TP35457
TP17172
TP748
TP33416
TP24844
TP4078
TP22317
TP18179
OPA08-1175
TP40349
TP13975
TP23743
TP13896
TP22317
AG8YA-763
OPA08-1175
TP18179
TP40349
TP13975
TP8457
TP1262
TP23837
TP13896
TP13975
TP23743
TP18179
TP15410
TP24844
TP15122*
TP15125
TP27616
TP15412
TP14071
TP5689
TP5223
TP2696
TP24705
TP33276
TP9254
TP14084
TP17172
TP18466
TP10404*
TP23837
TP8457
TP1262
TP13896

5.1.3.4 Primer validation and population screen

Of 147 primer sets designed based on sequence tags that exhibited heterozygosity in Nonpareil (e.g., Fig. 5.1.9a), 138 detected polymorphism among the progeny (e.g., Fig. 5.1.9b, Table S3.2). Of 161 primer sets designed based on sequence tags that exhibited heterozygosity in Lauranne (e.g., Fig. 5.1.9c), 155 detected polymorphism among the progeny (e.g., Fig. 5.1.9d, Table S3.2). None of the genotypic ratios observed for these polymorphisms deviated significantly from the expected 1:1 ratio.

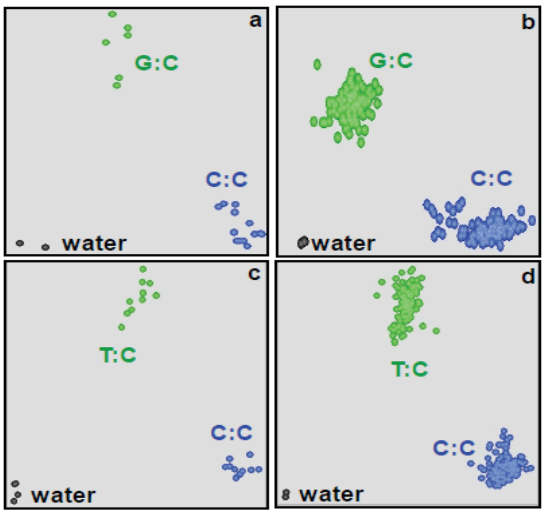


Fig. 5.1.9 Examples of results with primer sets derived from GBS tag sequences: the WriPdK0007 primer set, which assays a SNP within tag TP18674 that is heterozygous (G:C) in Nonpareil and homozygous (C:C) in Lauranne, applied to a validation panel of parents and progeny (a) and to the Nonpareil × Lauranne population (b) and the WriPdK0069 primer set, which assays a SNP within tag TP5689 that is heterozygous (T:C) in Lauranne and homozygous (C:C) in Nonpareil, applied to a validation panel of parents and progeny (c) and to the Nonpareil × Lauranne population (d).

5.1.3.5 Reconstruction of parental linkage maps using KASP markers

Based on screening of markers on 231 progeny, maps were constructed for Nonpareil (138 KASP markers and two *S*-locus markers: total length 608.9 cM) (Fig. 5.1.10) and for Lauranne (155 KASP markers and two *S*-locus markers: total length of 658.7 cM (Fig. 5.1.11). Comparison of these maps with the initial framework maps revealed high conservation of marker order within each linkage group, as illustrated in Fig. 5.1.12 for LG3.

5.1.3.6 Comparison of genetic maps with peach scaffolds

Comparisons of the marker positions on Nonpareil and Lauranne parental maps with positions on peach genome scaffolds confirmed high synteny and collinearity between these two genomes (Fig. 5.1.13). Almost all markers mapped on the expected peach scaffolds. For the Nonpareil map, the exceptions were a few markers that were genetically mapped on LG1, LG4, LG6 and LG8 of almond but that anchored to Pp5, Pp1, Pp1 and Pp4, respectively, in the peach genome (Table S3.3). For the Lauranne map, there were markers that were genetically mapped on LG2, LG3 and LG6 of almond but that anchored to Pp6, Pp6 and Pp4, respectively, in the peach genome (Table S3.4).

Comparisons of marker positions on Nonpareil and Lauranne genetic maps with the positions at which those markers had been anchored on the peach genome scaffolds showed linear relationships between genetic and physical distances in most parts of the genome (Fig. 5.1.14). There are, however, some regions in which one or both parents exhibited no polymorphisms (e.g., parts of Pp2 and Pp7) and some regions in which there was little or no recombination between physically distant markers (e.g., part of Pp2 for Nonpareil and part of Pp8 for Lauranne). There are also some discrepancies in marker order between the genetic maps and the physical scaffolds (e.g., at each end of Pp4 and Pp8 for Nonpareil).

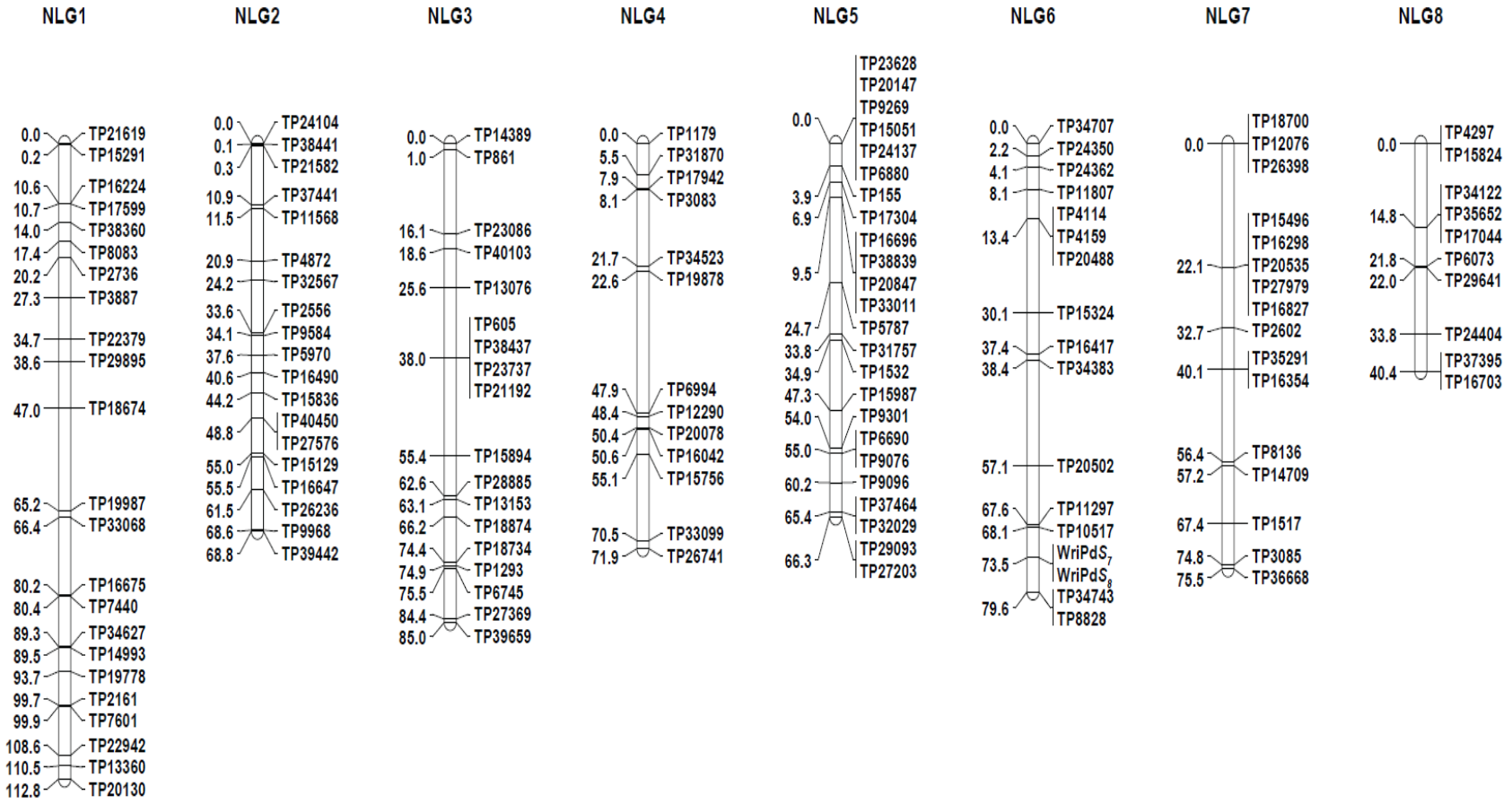


Fig. 5.1.10 A linkage map for Nonpareil, constructed using genotypic data from SNP-based marker assays applied to 231 Nonpareil × Lauranne F_1 progeny, with eight linkage groups labelled as NLG1 to NLG8.

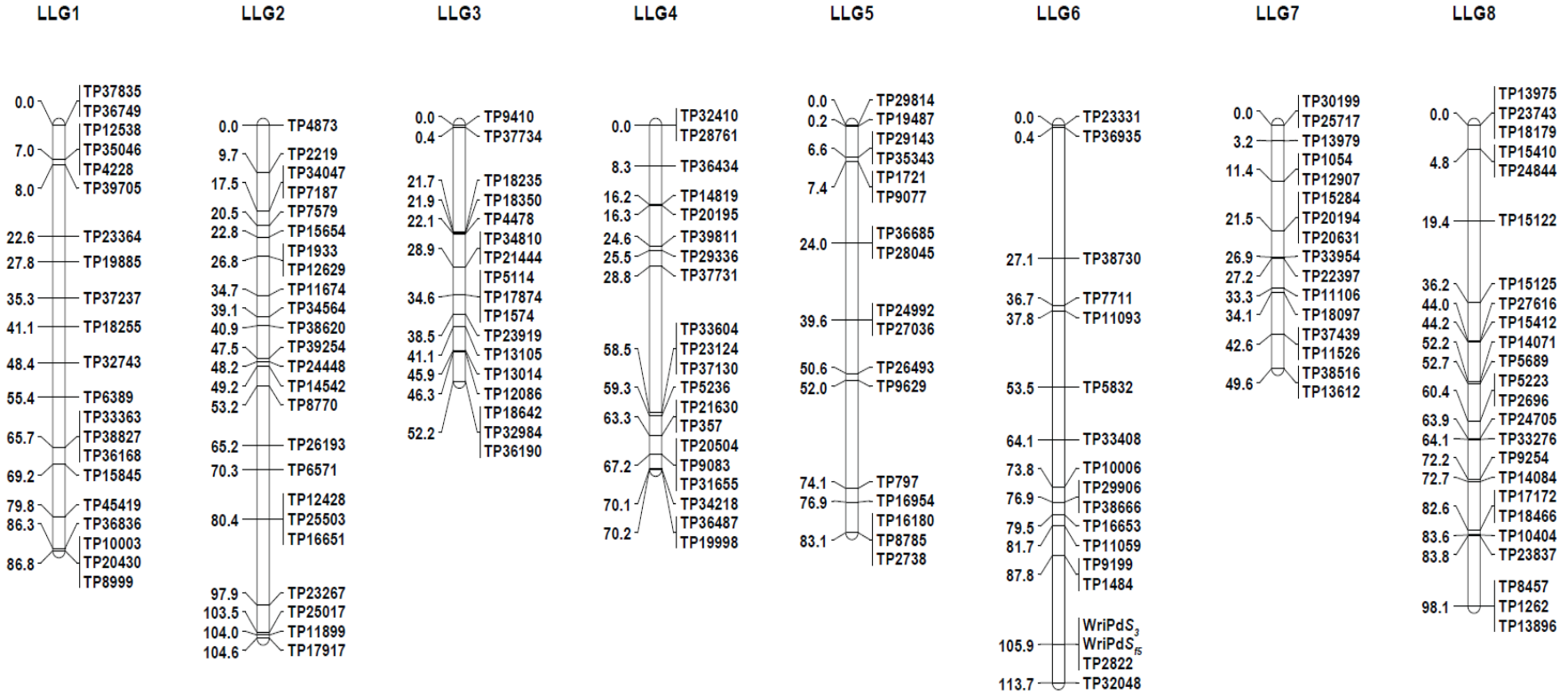


Fig. 5.1.11 A linkage map for Lauranne, constructed using genotypic data from SNP-based marker assays applied to 231 Nonpareil × Lauranne F_1 progeny, with eight linkage groups labelled as LLG1 to LLG8.

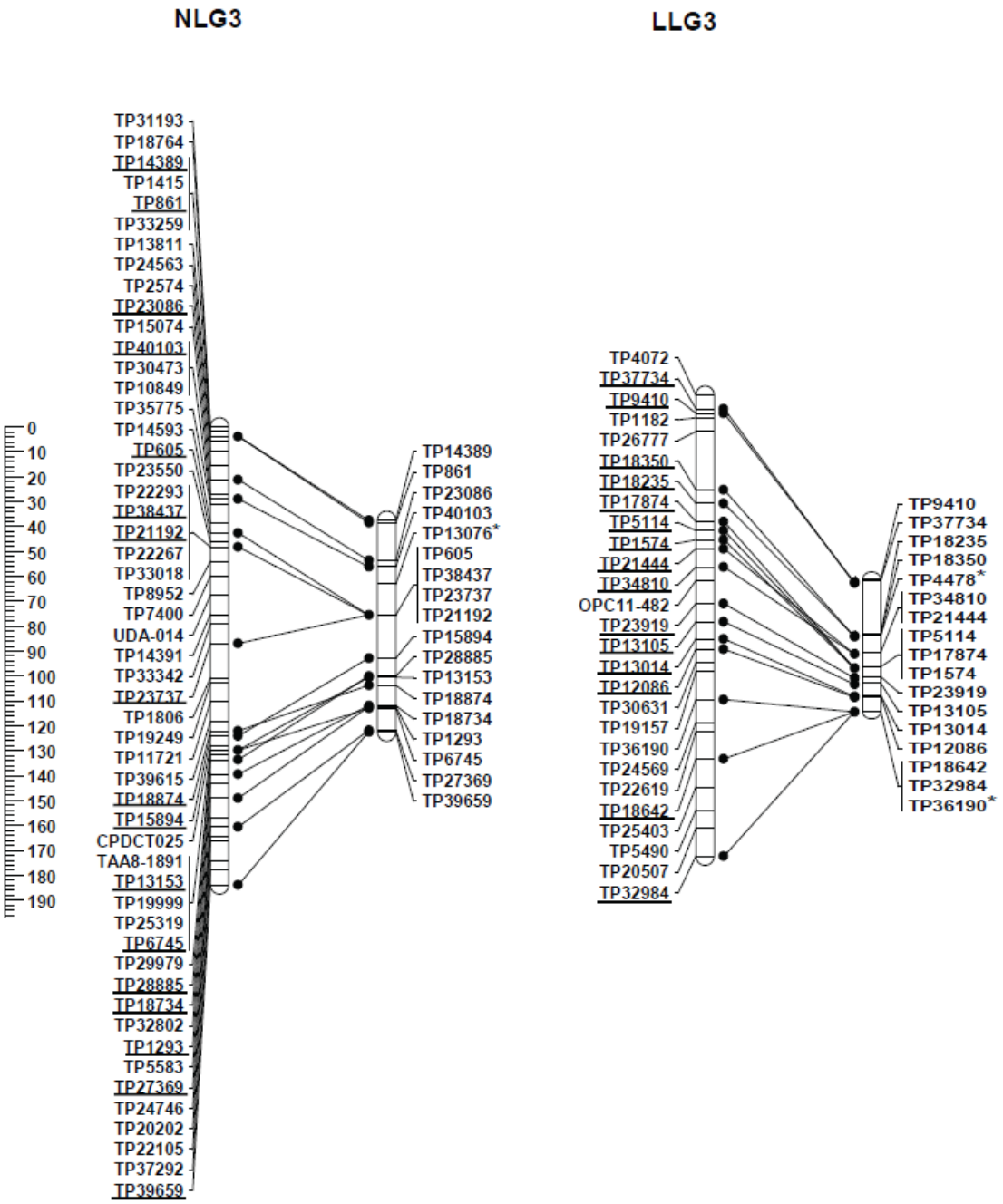


Fig. 5.1.12 Linkage maps constructed for the almond linkage group LG3 for Nonpareil (NLG3) and Lauranne (LLG3). In each case the map on the left is an initial framework map constructed using data from a small number of progeny (52 for Nonpareil and 55 for Lauranne) and the map on the right was constructed using KASP marker data from 231 progeny. Markers that are common to both maps are underlined and the markers that are designed from the SNP-bearing sequence tags from the second GBS data are marked with asterisks.

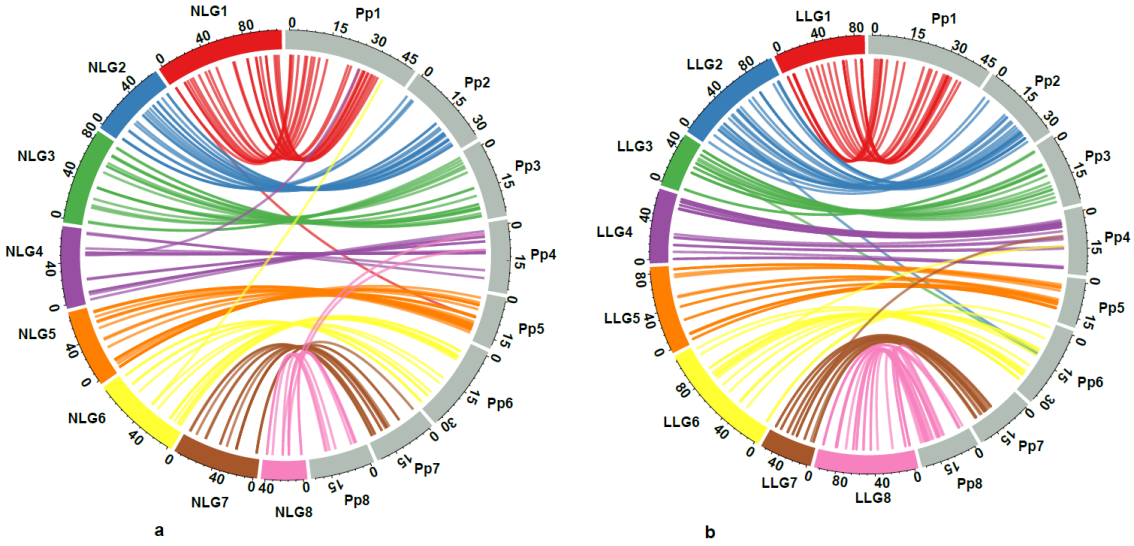
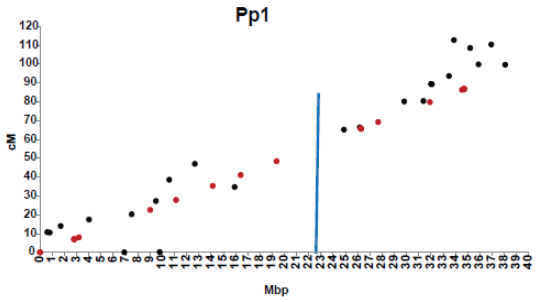
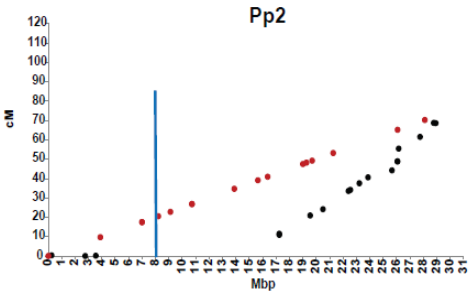
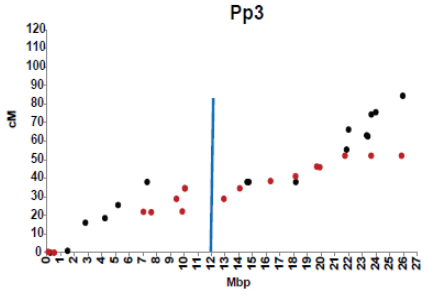


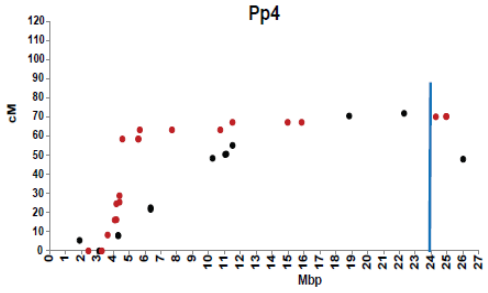
Fig. 5.1.13 Synteny and collinearity between almond genetic maps and the peach genome sequence. Relationship between Nonpareil (a) and Lauranne (b) linkage maps for the eight linkage groups of almond with the first eight scaffolds of the peach genomic sequence assembly. On almond linkage groups (NLG or LLG), genetic distances are given in cM. On peach scaffolds (Pp), physical distances are given in 400,000 bp intervals. Links between linkage groups and scaffolds indicate the positions at which sequences genetically mapped in almond anchor to the genomic sequence of peach.

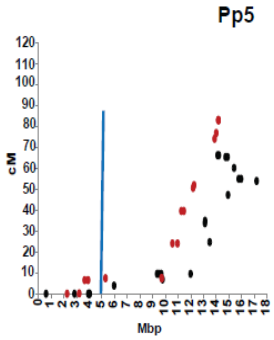
Fig. 5.1.14 Relationships between genetic and physical distances for each linkage group of almond and the peach genome sequence. Peach scaffolds are labelled as Pp1 to Pp8. Positions of the markers that were mapped on each scaffold in the Nonpareil and Lauranne linkage maps are shown in black and red, respectively. In each case, genetic distances are shown on the vertical axis (cM), physical distances are shown on the horizontal axis (Mbp) and the vertical blue lines indicate the estimated positions of the centromeric regions based on information from the peach whole genome sequence assembly v2.0.a1.

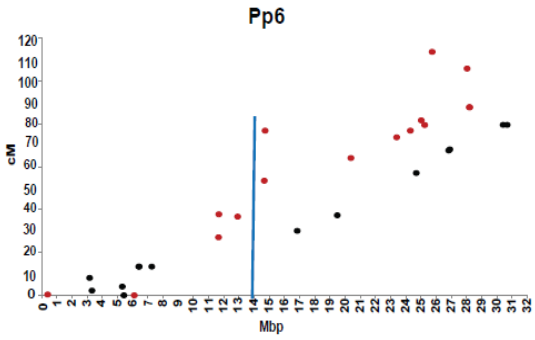

Pp1
cM
Mbp

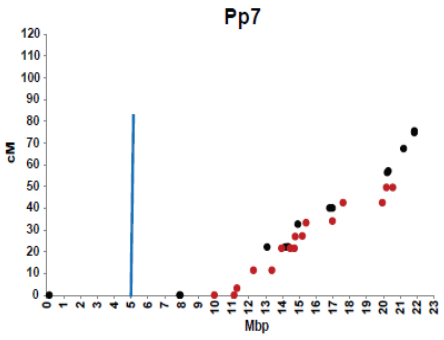

Pp2
cM
Mbp

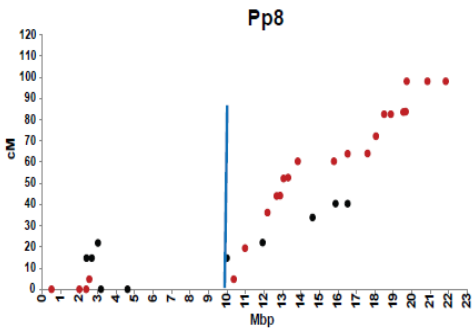

Pp3
cM
Mbp


Pp4
cM
Mbp


Pp5
cM
Mbp


Pp6
cM
Mbp


Pp7
cM
Mbp


Pp8
cM
Mbp

5.1.4 Discussion

Here, implementation of a GBS protocol enabled discovery of hundreds of SNP-bearing GBS tags providing an easy and accurate method to discover and assay SNPs without any knowledge of the almond genome sequence. The restriction enzyme used here, *Ape*KI, is a type II endonuclease that recognises a degenerate 5 bp sequence (GCWGC, where W is A or T). It is useful for removing repetitive sequences, because it has relatively few recognition sites in the major classes of plant retrotransposons and will not cut if the 3´ base of the recognition sequence on the bottom strand is 5´ methyl cytosine (Söllner et al. 2006). It creates a 5´ overhang of 3 bp, providing sites for attachment of adapters to which primers can anneal.

Of the sequences generated, 68% aligned to the peach genome sequence assembly. The SNPs that were used in linkage map construction were distributed across all eight scaffolds of the peach assembly. Some variation was observed in the numbers of sequence tags and SNPs mapped to each scaffold and in the distributions of tags within scaffolds. One reason for this variation could be the distribution of *Ape*KI restriction sites across the almond genome. Another could be methylation differences across the genome. The unusually high numbers of sequence tags obtained in some regions (e.g., on Pp5 (at 6.5 Mbp) and Pp7 (at 0.5 Mbp) may be related to uneven distribution of copy number polymorphisms in the almond genome. Copy number variations are considered as major source of genetic variation and in many cases, they have been discovered close to gene/genes that are associated with disease resistance in plants (Cook et al. 2012) and in human (Sebat et al. 2004; Shaikh et al. 2009). They may change gene structures and modify elements regulating gene expression and may influence gene expression and phenotypic variation. Copy number variation events have been reported for plant species, including maize (Beló et al. 2009; Swanson-Wagner et al. 2010), wheat (Saintenac et al. 2011), rice (Yu et al. 2011), potato (Iovene et al. 2013) and *Arabidopsis thaliana* (Cao et al. 2011). On one arm of almond LG7 (peach Pp7) there were no markers mapped for Lauranne and just two markers mapped for Nonpareil. This could be due to homozygosity in Nonpareil and/or Lauranne throughout this region. This may not be limited to

Nonpareil and Lauranne nor even to almond, as similar observations have been reported for this region for the sweet cherry cultivars Riverdale and Rainer (Guajardo et al. 2015). In addition, analysis of the 32% of unmapped unique tags might also provide structurally and biologically useful information that are specific to the almond genome. Recently, analysis of unmapped reads led to obtain new biological information and insights about structural variations and cross-species contamination in the human (Tae et al. 2014; Ye et al. 2009), cow (Merchant et al. 2014) and pea aphid (Gouin et al. 2015) genomes.

While the number of SNP-bearing tags obtained here was sufficient for genetic mapping and assay design and similar to what has been reported for peach (Bielenberg et al. 2015) and sweet cherry (Guajardo et al. 2015), it is low compared to the tens of thousands of GBS markers that have been reported for some other species (Elshire et al. 2011; Hyma et al. 2013; Li et al. 2015; Lu et al. 2013; Mitchell et al. 2012). A larger number of tags might have been obtained through the use of an enzyme or combination of enzymes that would provide a larger number of digested fragments and increasing the depth of sequencing. For example, based on *in-silico* analysis of the peach genome sequence, combination of *Ape*KI with *Hpa*II would be expected to provide a larger number of restriction fragments of suitable length than *Ape*KI alone. Use of this combination might provide an avenue to generate denser linkage maps for almond.

Based on the strong positive relationship (r^2 = 0.92) between the total read number and the total number of unique tags obtained, it might also be possible to discover additional unique tags and SNP markers by increasing sequencing depth. In this analysis, sequence depth/sample was about 0.9x. With greater sequence depth, it might be possible to discover more tags with sufficient reads from all the individuals. Greater sequence depth could also reduce the effects of technical and biological factors such as PCR amplification bias.

As the TASSEL GBS 3.0 SNP-calling pipeline, which was developed mainly for self-pollinated plants, is considered to be sensitive to low sequence depth in highly heterozygous species (Hyma et al. 2015), a stringent read depth cut-off value (4, compared to the value of 3 that generally used in filtering SNP datasets) (Elshire et al. 2011; Lu et al. 2013) was used here. With a cut-off value of 3, a larger number of GBS tags (over 600) would have been detected, but in maps constructed on this basis (not shown) each linkage group was over 300 cM long, indicating that some of the SNPs may have been spurious.

The GBS approach can generate high numbers of polymorphic markers, but can suffer from incorrect assignment of parental phase, underestimation of heterozygotes and high proportions of missing data (Lu et al. 2013). Here, technical replicates of the parents were included in the genomic library and very stringent filters were applied to select subsets of markers and progeny for initial mapping. During map construction, diagnostic tests were conducted to detect phasing errors and those errors were corrected. Nevertheless, when KASP assays were designed and assayed, one incorrectly phased marker was detected. That marker (GBS tag TP37439), which had originally been assigned to the Nonpareil map was reassigned to the Lauranne map. Three markers (GBS tags TP15642, TP16449 and TP18643) that were originally assigned to the Lauranne map were determined to be heterozygous in both parents and were not used for map construction. Three other markers (GBS tags TP11609, TP12109 and TP25403) that had originally been scored as heterozygous in one parent were determined to be homozygous in both parents and not used for map construction.

The initial genetic maps constructed using GBS data were about twice as long as the Nonpareil × Lauranne map published by Tavassolian et al. (2010), but the final maps constructed using data from KASP assays were similar in length to maps that were published by Tavassolian et al. (2010). This ‘shrinkage’ was due to the correction of genotypes that had been erroneously called in the GBS analysis. In most cases, the corrections were from homozygous to heterozygous, indicating that although two alleles were present, only one of them was sequenced. Of a total of 12,720

heterozygous calls in the final KASP dataset, 1,526 (12%) had been miscalled as homozygous in the GBS analysis. This type of error was evenly distributed among markers. These observations are similar to what has been reported by Lu et al. (2013) for GBS analysis in switchgrass. These are commonly observed in GBS-RAD system and approaches such as increased sequencing depth could mitigate the effects from those that behave as dominant-type markers. To obtain greater sequence read depth, the proportion of the genome that is sequenced can be reduced by restriction enzyme based complexity reduction as mentioned previously or using complexity reduction of polymorphic sequences (CRoPS) (van Orsouw et al. 2007), applying RNAseq (Gore et al. 2007) or sequence capture methods such as SureSelect (Gnirke et al. 2009), Nimblegen (Kiss et al. 2008) and Raindance (Nijman et al. 2010). However, sequence capture methods still cannot be used for almond as these approaches require sequence data to design DNA capture probes. Further, others may not be cost-effective for almond as it has a small genome.

Many of the SNP-bearing GBS tags discovered here did not provide adequate sequence information for primer design, because their SNPs were positioned near one end of the tag. With aligning the tag sequences against Nonpareil genomic sequence, flanking sequences were obtained. Of 308 SNP-bearing GBS sequences that were selected markers that were estimated for assay design, 293 were successfully converted to fluorescence-based SNP marker assays and mapped (138 for Nonpareil and 155 for Lauranne): a success rate of 95%.

Addition of two separate bulks of progeny with the highest and lowest amount of tocopherol in to the second GBS library prepared here, did not lead to discovery of any GBS tags with clear polymorphisms between the two bulks. Nevertheless, it did provide a few new GBS tags that were useful for filling gaps in the linkage maps.

Comparison of the almond genetic maps generated here with the peach genome sequence confirmed the expected high similarity between these two genomes with only a few of the mapped

markers anchoring to unexpected positions. This is the first report of genome-wide anchoring of an almond genetic map to the *Prunus persica* whole genome sequence assembly. Comparisons of marker positions on Nonpareil and Lauranne genetic maps with the positions at which those markers had been anchored on the peach genome scaffolds showed linear relationships between genetic and physical distances in most parts of the genome with some regions with no or less polymorphisms.

The anchoring of the framework maps to the peach genome also indicated that there were no SNPs anchored between 4 Mbp and 11 Mbp for Nonpareil nor between 7 Mbp and 13 Mbp for Lauranne on Pp7. Based on current data, it is not possible to distinguish whether these are simply regions in which one of the almond parents is homozygous, or whether these regions are structurally different between peach and almond.

This is also the first report on the use of GBS in almond to discover SNPs and to generate linkage maps. The processes that were used here to select a restriction enzyme, conduct GBS data analysis and design KASP markers could provide models for GBS data analysis and sequence-based marker design in other species for which a complete genome sequence assembly is not yet available.

Section 5.2

Construction of linkage maps for almond using four populations with a common parent

5.2.1 Introduction

Most of the populations that have been developed for genetic mapping in almond consist of progeny derived from crosses between two unrelated individuals that differ for one or more traits (Arús et al. 1998; Fernández i Martí et al. 2013; Tavassolian et al. 2010). For any cross combination, mapping can only be conducted for genomic regions in which one or both of the parents is heterozygous and map resolution is limited by the number of progeny available from that cross combination. Maintaining large breeding populations can be difficult due to the size of almond trees.

Mapping methods that allow the use of progeny from multiple crosses could make it possible to develop better genetic maps by making use of available breeding materials. One possible approach is to consider all of the progeny plants of an almond breeding program as one large multi-parent population, and to apply genetic analysis approaches such as those that have been developed for nested association mapping (NAM) populations, which consist of large numbers of progeny derived from multiple parents, each crossed with one or more reference parents (Flint-Garcia et al. 2005; Kump et al. 2011; Tian et al. 2011). In multi-parent populations, any individual marker is likely to be informative in only some of the cross combinations. Linkage mapping methods that can handle complex situations associated with multi-parent populations, have been developed and implemented in maize (Yu et al. 2008) and in other plant species for which highly homozygous inbreds can be used as parents. Multi-parent mapping approaches have also been used in peach (*Prunus persica*)

(Fresnedo-Ramírez et al. 2015) and apple (*Malus* × *domestica*) (Allard et al. 2016; Di Pierro et al. 2016).

The University of Adelaide almond breeding program maintains about 8,100 F_1 plants derived from crosses among 39 parents (Fig. 5.2.1). Eleven of the parents have been mostly used as females. Of these, five cultivars (Nonpareil, Johnston, Chellaston, Carmel and Somerton) have been crossed with several other parents, with large numbers of progeny produced (Wirthensohn and Sedgley 2002). For example, Nonpareil has been used as the female parent in crosses with 15 other clones, providing a total of 2,200 F_1 progeny. This set of materials could be regarded as a NAM population. With sufficient genotyping and phenotyping of these materials, it could be possible to develop a composite linkage map for Nonpareil, and investigate how alleles from Nonpareil act in a range of genetic backgrounds.

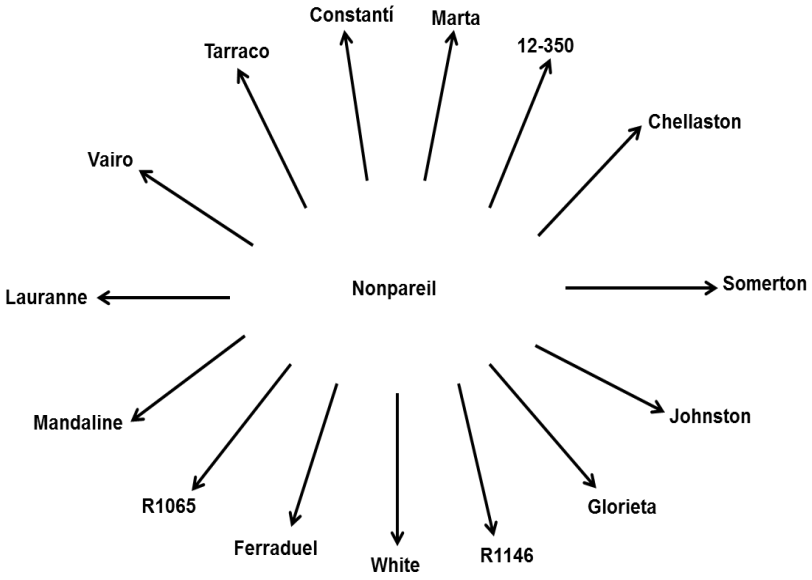


Fig. 5.2.1 A schematic diagram showing the other parents with which Nonpareil cultivar has been crossed in the University of Adelaide almond breeding program.

5.2.2 Materials and methods

5.2.2.1 Plant materials

Four populations derived from crossing Nonpareil with other parents (Lauranne, Constantí, Tarraco and Vairo) were used in this analysis. These populations (Nonpareil × Lauranne (N × L), Nonpareil × Constantí (N × C), Nonpareil × Tarraco (N × T) and Nonpareil × Vairo (N × V)) have 231, 349, 207 and 198 progeny, respectively. All these populations have been maintained in Dareton, NSW, Australia using standard management practices.

5.2.2.2 DNA extraction

DNA from each of the progeny of these populations was extracted using an Oktopure™ DNA extraction protocol that had been optimised for almond (LGC Ltd, Teddington, UK).

5.2.2.3 Polymorphic assay selection and population screen

Marker assays that had been designed based on 138 heterozygous SNP-bearing sequences in Nonpareil and on 155 heterozygous SNP-bearing sequences in Lauranne (Section 5.1) were assayed on Nonpareil, Constantí, Tarraco and Vairo, on four individuals from each population and on water samples as negative controls, using the procedures described in Section 5.1. All assays for which polymorphisms were detected in this validation panel were selected for population screening.

5.2.2.4 Linkage maps for Nonpareil, Constantí, Tarraco and Vairo

A double pseudo-testcross mapping strategy was used to construct linkage maps for each population. The maps for Nonpareil were constructed using markers that were heterozygous in Nonpareil and homozygous in the other parent. The linkage maps for Constantí, Tarraco and Vairo were generated using markers that were homozygous in Nonpareil but heterozygous in the other parent. The ASMap package in the R statistical environment (www.CranR.org) was used to generate linkage maps. Depending on the number of progeny, a p-value of either 1e-8 or 1e-12 was used to

assign markers to linkage groups. The Kosambi mapping function was used to calculate genetic distances in centiMorgans (cM) (Kosambi 1944).

5.2.2.5 A composite map for Nonpareil

A total of 985 progeny from four crosses (N × C, N × L, N × T and N × V) was used to construct a composite map for Nonpareil. This was done with a two-way pseudo-testcross mapping strategy with data coded in the backcross (BC) format. For each marker for which Nonpareil was heterozygous, progeny with the same result as Nonpareil were coded as 'ab'. In populations in which the other parent was homozygous for one of the Nonpareil alleles, progeny with the same result as the other parent were coded as 'aa'. In populations in which the other parent was heterozygous, all progeny were coded as having missing data. The composite map was constructed based on the recombination fractions between adjacent markers that were estimated using the r/qtl (www.CranR.org). Recombination fractions were converted to map distances using Kosambi mapping function (Kosambi 1944). The resulting composite map was compared with each of the Nonpareil maps that had been constructed using data from individual populations. Markers for which there were substantial inconsistencies among maps were removed.

5.2.2.6 Marker order conservation within linkage groups of Nonpareil

The relative positions of markers in the composite Nonpareil map with those of each of the Nonpareil map from four populations were visualized by plotting each of the Nonpareil-derived maps against the composite Nonpareil map using the Circlize package (Gu et al. 2014) in the R statistical environment (www.CranR.org).

5.2.3 Results

5.2.3.1 Polymorphic marker detection and population screen

Of 138 KASP markers that were developed based on Nonpareil heterozygosity observed in the genotyping-by-sequencing analysis of N × L, 85, 92 and 103 markers detected polymorphism in N × T, N × C and N × V, respectively (Fig. 5.2.2a). Of 155 KASP markers that were derived based on Lauranne heterozygosity, 40, 56 and 68 markers detected polymorphism in N × T, N × V and N × C, respectively (Fig. 5.2.2b).

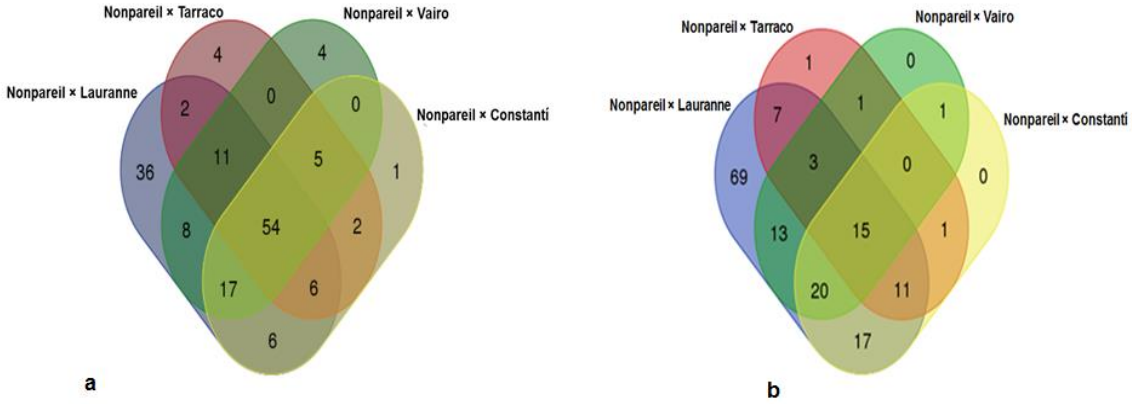


Fig. 5.2.2 Venn diagrams showing the number of KASP markers that detected polymorphisms in the populations used in this analysis, for (a) markers that were designed based on Nonpareil heterozygosity and (b) markers that were designed based on Lauranne heterozygosity.

5.2.3.2 Linkage maps

Of 138 markers that were mapped for Nonpareil based on data from Nonpareil/Lauranne progeny, 92, 85 and 96 were mapped using progeny of Nonpareil/Constantí, Nonpareil/Tarraco, and Nonpareil/Vairo, respectively, resulting in maps with total lengths of 438.9 cM (Fig. 5.2.3), 568.9 cM (Fig. 5.2.4) and 553.4 cM (Fig. 5.2.5). In addition, several markers that had not exhibited polymorphism among Nonpareil × Lauranne progeny were polymorphic in other populations and were mapped. For example, a marker derived from GBS tag TP1263 could not be mapped based on

Nonpareil/Lauranne (Section 5.1) but was mapped based on crosses between Nonpareil and Constantí (Fig. 5.2.3) and between Nonpareil and Tarraco (Fig. 5.2.4).

The Lauranne map presented in Section 5.1 had 155 KASP markers and two *S* locus markers and a total length of 658.7 cM. The maps developed here for Constantí, Tarraco and Vairo had only 65, 39 and 53 markers, respectively, with total lengths of only 381.7 cM (Fig. 5.2.6), 294.8 cM (Fig. 5.2.7) and 148.5 cM (Fig. 5.2.8).

5.2.3.3 Composite linkage map for Nonpareil

The composite linkage map that was constructed by exploiting progeny from four Nonpareil crosses had 96 KASP markers and two *S*-locus markers with a total length of 741.4 cM (Fig. 5.2.9). The marker order in the Nonpareil composite map was highly conserved with what had been observed for the individual Nonpareil maps from Nonpareil × Lauranne, Nonpareil × Constantí, Nonpareil × Tarraco and Nonpareil × Vairo (Fig. 5.2.10), with few exceptions. Some markers that had collocated in Nonpareil genetic maps constructed from individual populations were separated in this map.

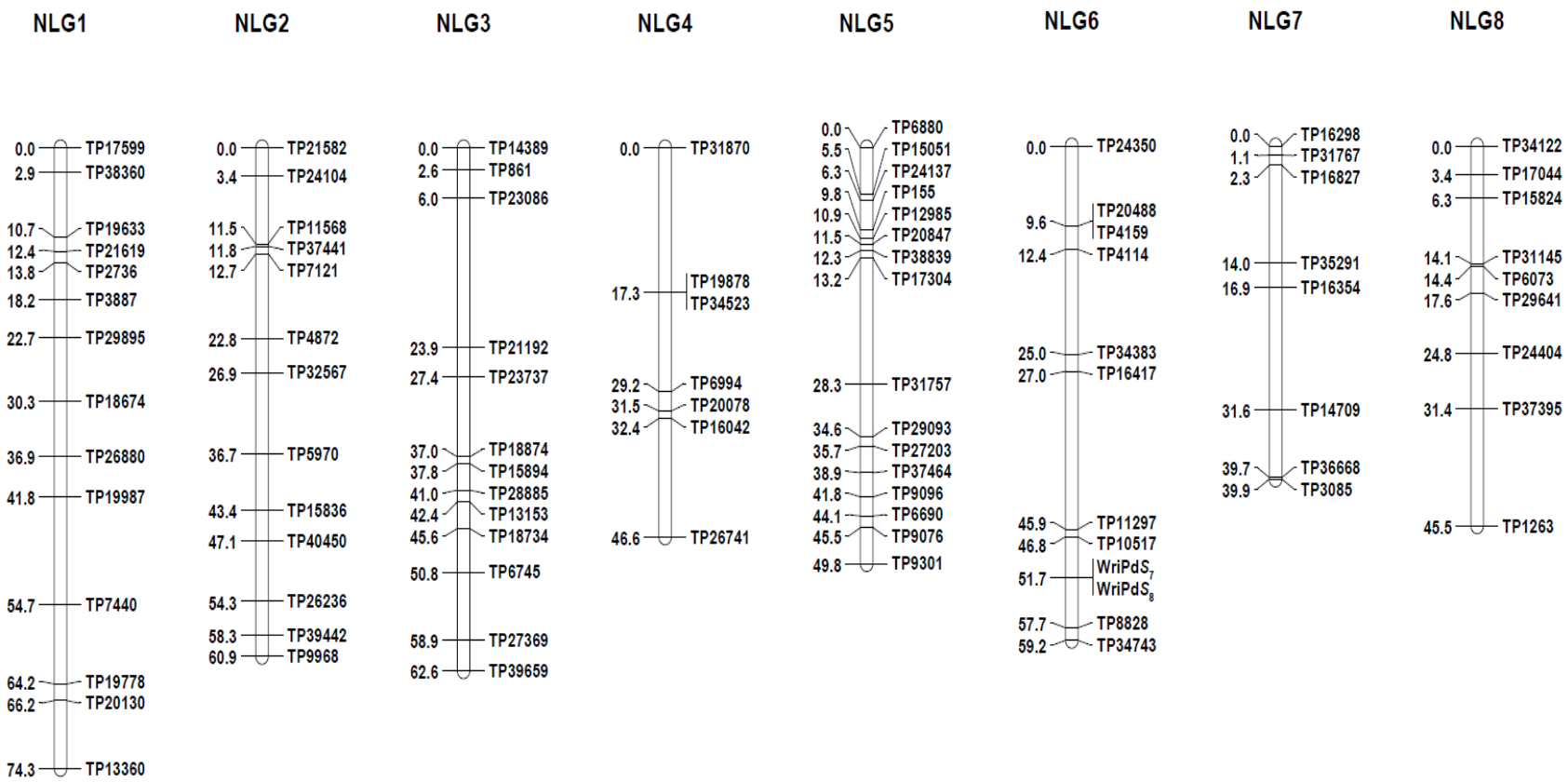


Fig. 5.2.3 A linkage map for Nonpareil, constructed using genotypic data from SNP-based marker assays applied to 349 Nonpareil × Constantí F_1 progeny.

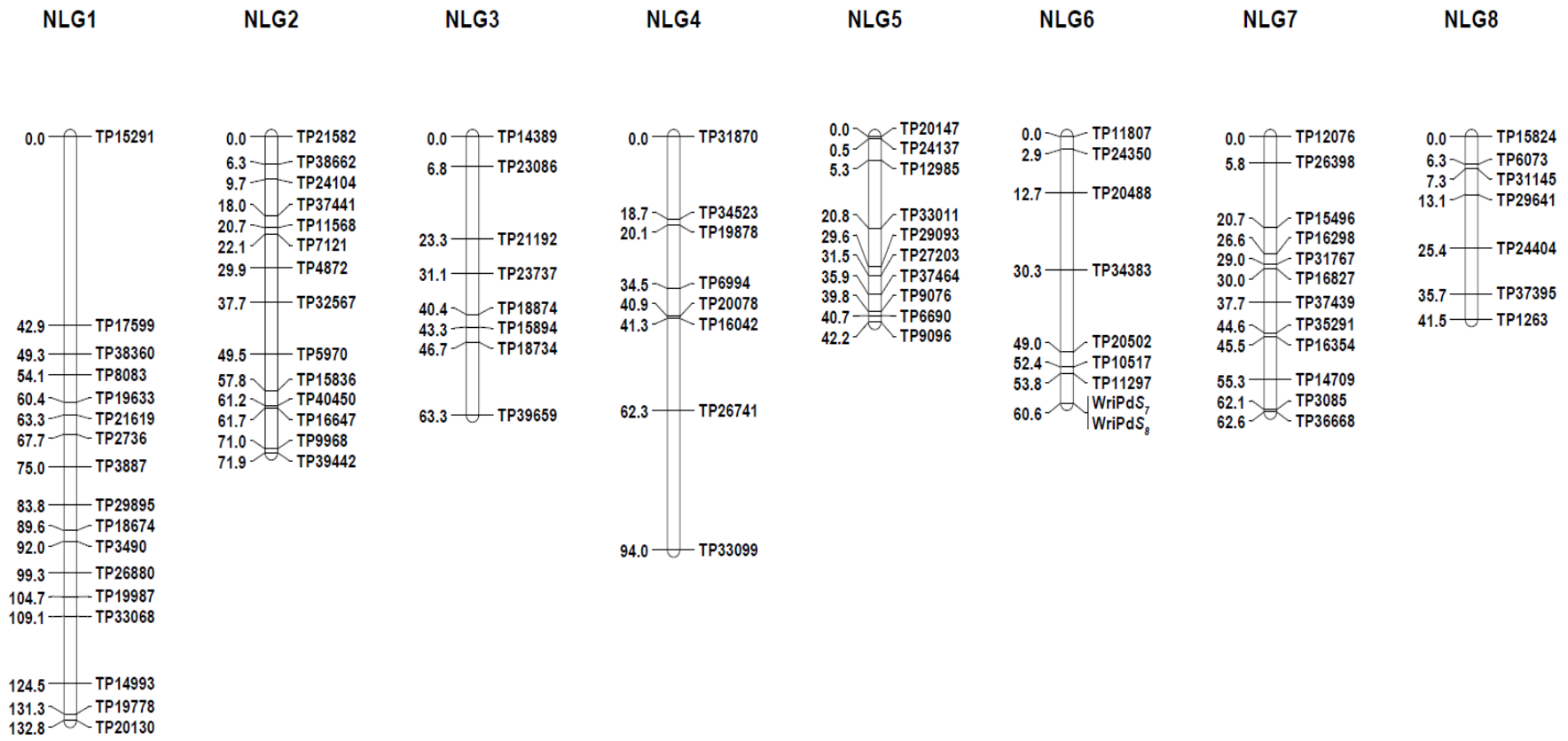


Fig. 5.2.4 A linkage map for Nonpareil, constructed using genotypic data from SNP-based marker assays applied to 207 Nonpareil × Tarraco F_1 progeny.

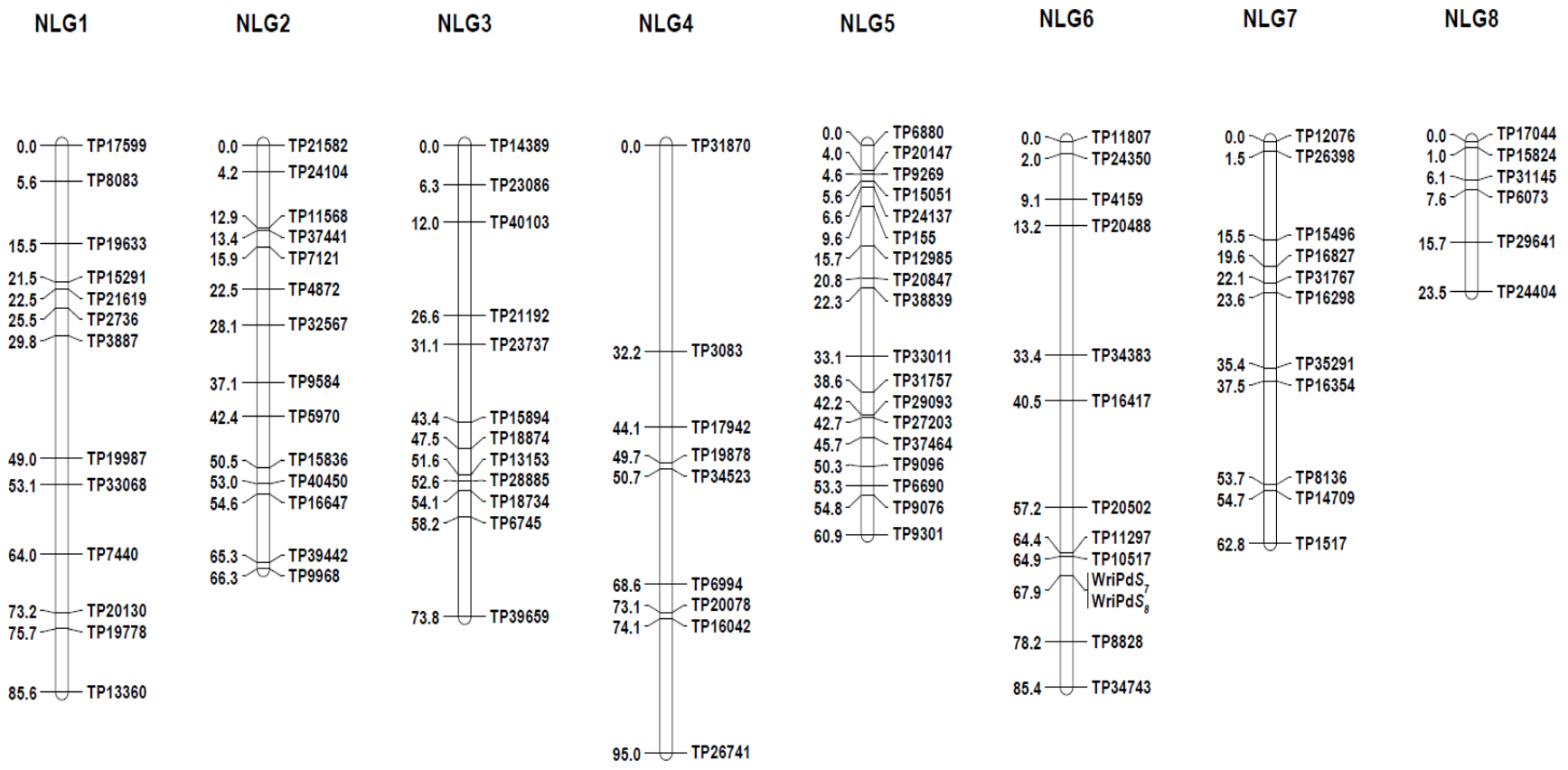


Fig. 5.2.5 A linkage map for Nonpareil, constructed using genotypic data from SNP-based marker assays applied to 198 Nonpareil × Vairo F_1 progeny.

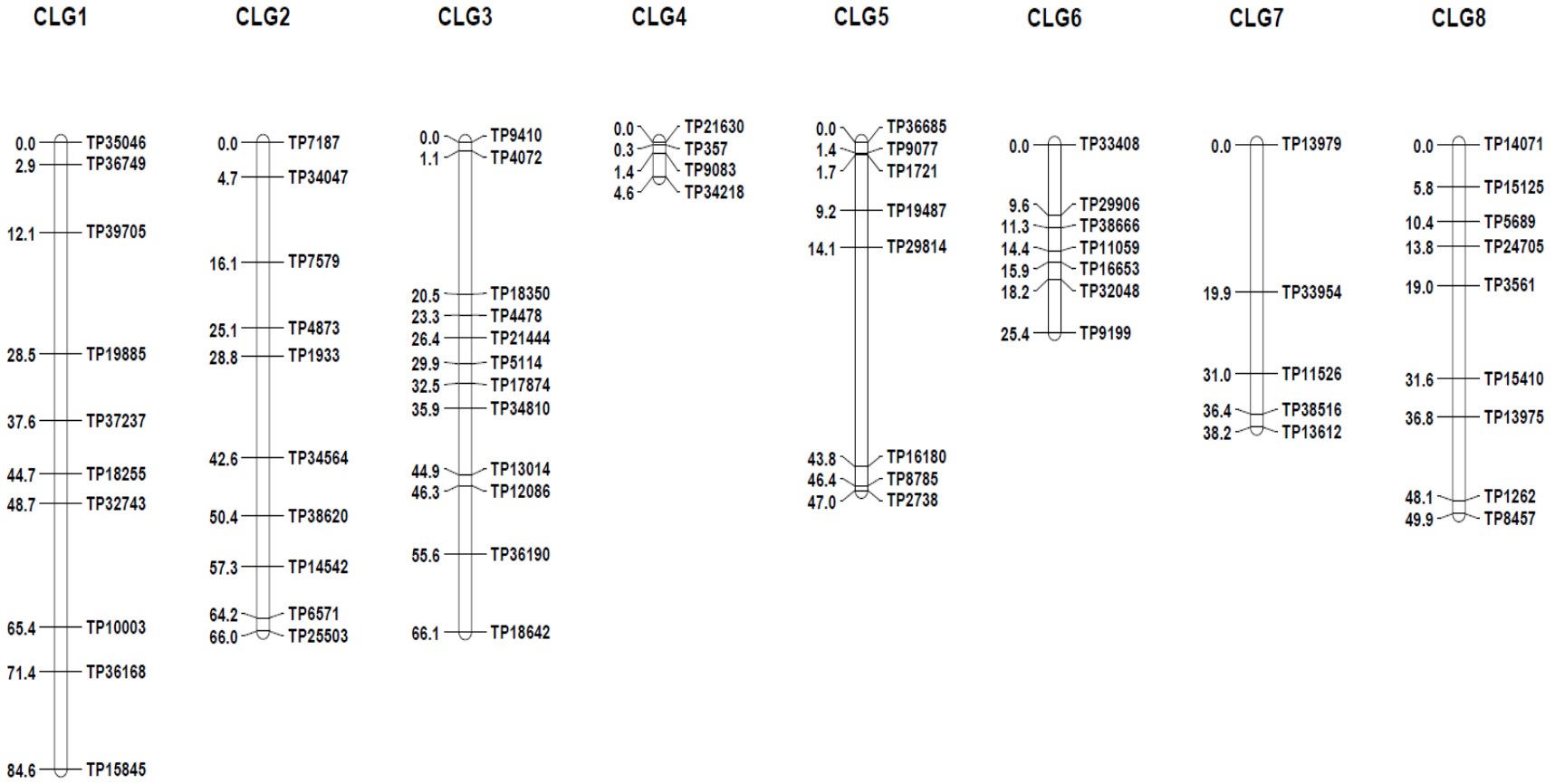


Fig. 5.2.6 A linkage map for Constantí, constructed using genotypic data from SNP-based marker assays applied to 349 Nonpareil × Constantí F_1 progeny.

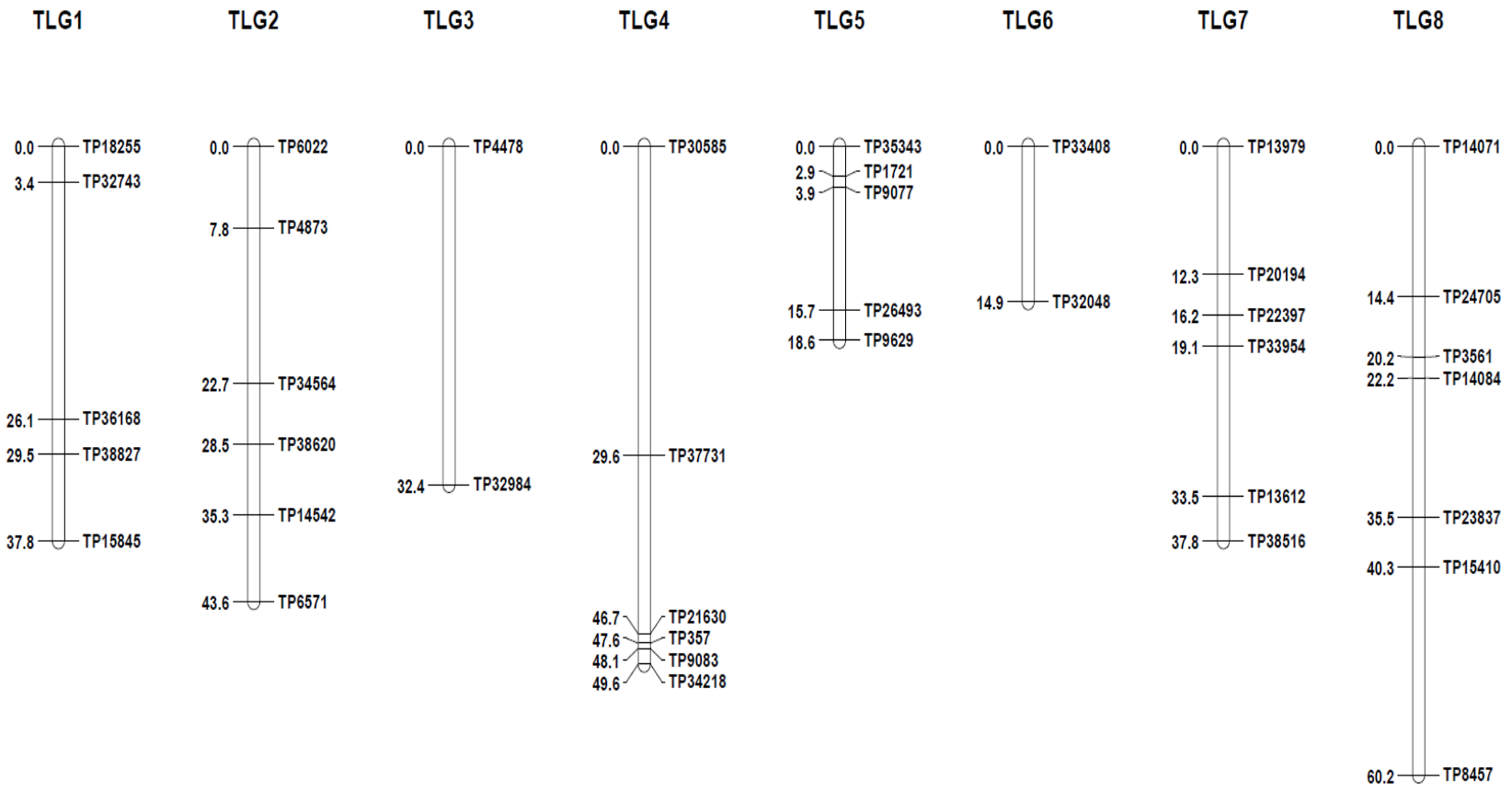


Fig. 5.2.7 A linkage map for Tarraco, constructed using genotypic data from SNP-based marker assays applied to 207 Nonpareil × Tarraco F_1 progeny.

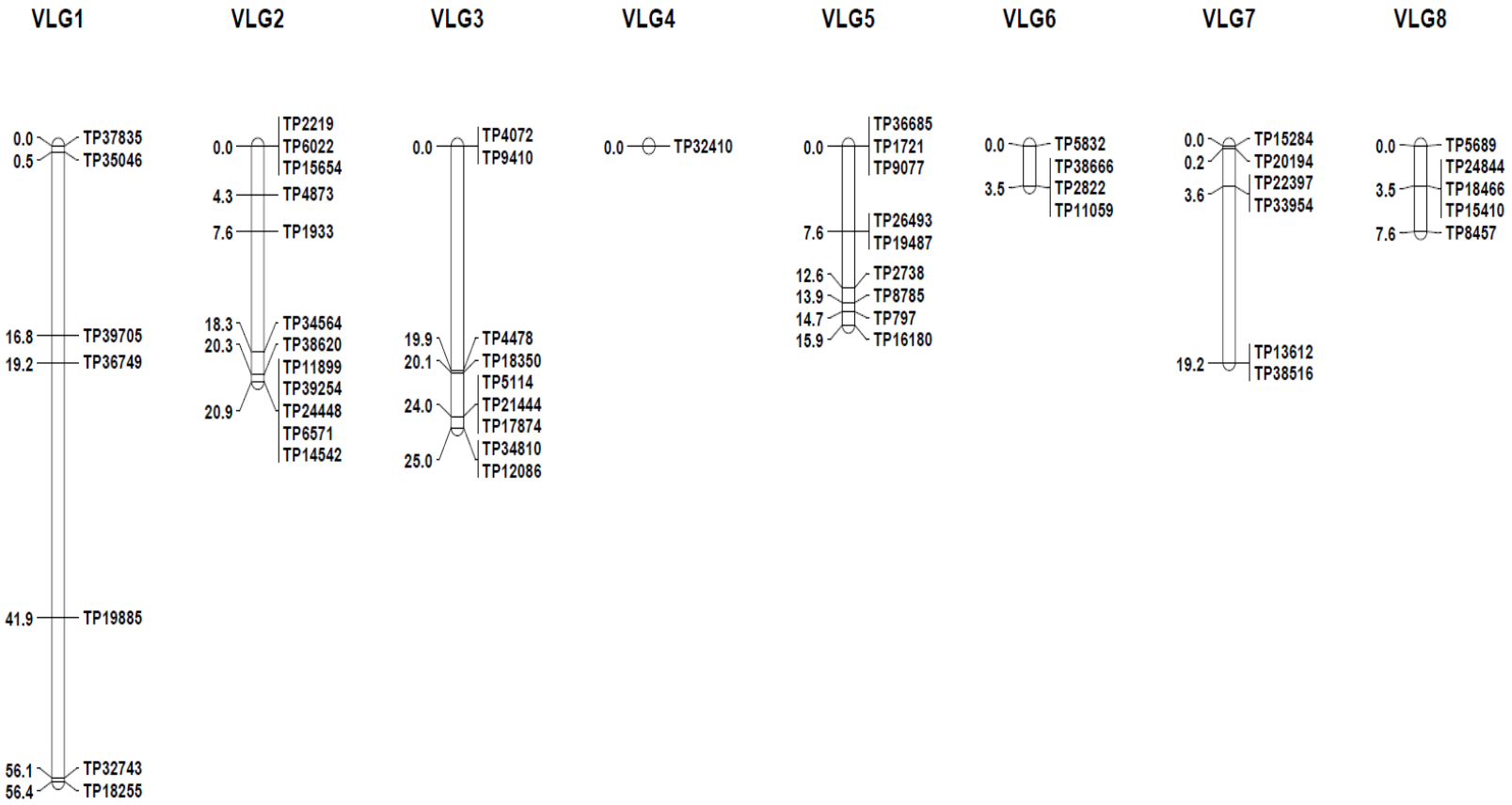


Fig. 5.2.8 A linkage map for Vairo, constructed using genotypic data from SNP-based marker assays applied to 198 Nonpareil × Vairo F_1 progeny.

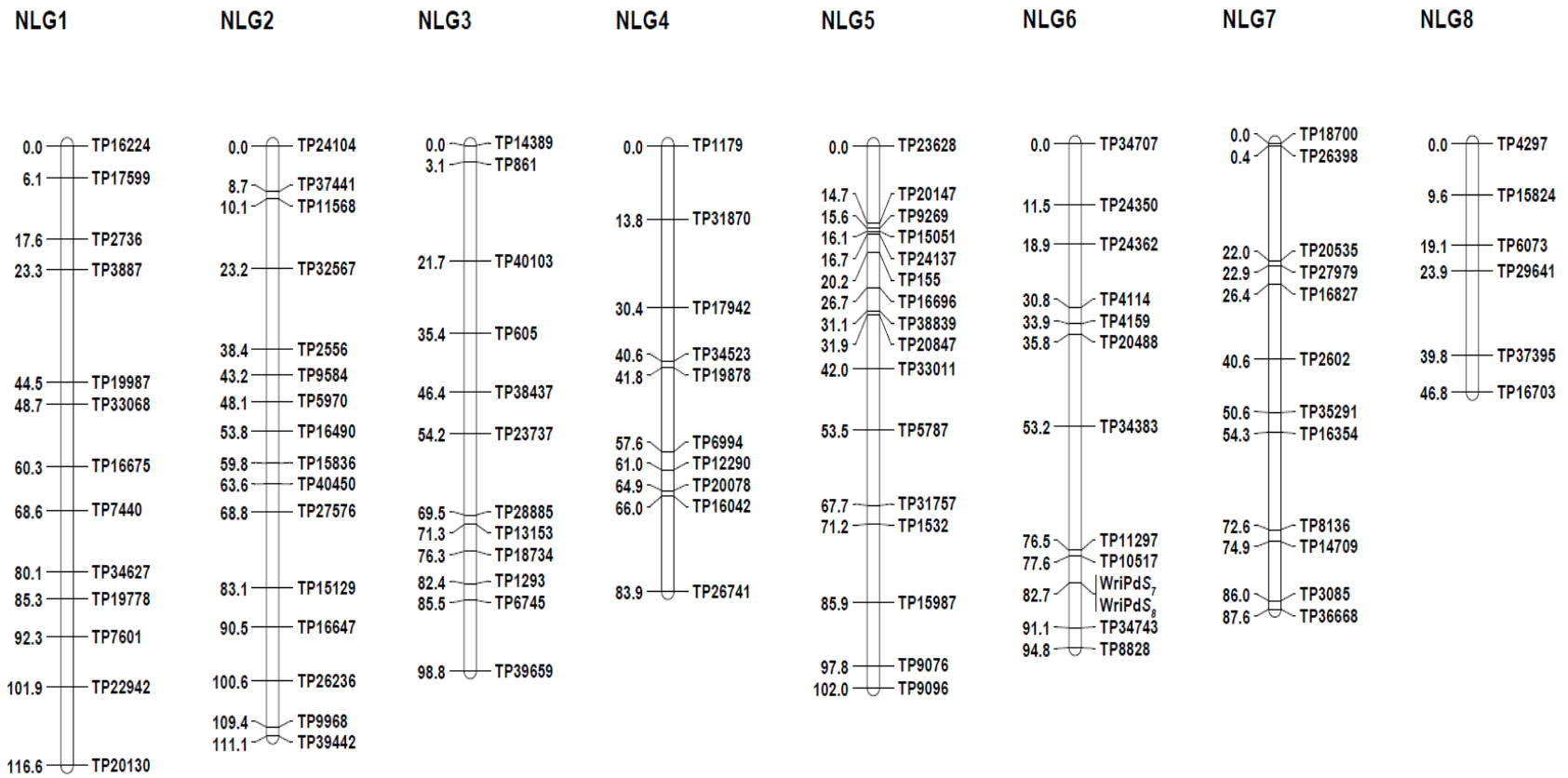


Fig. 5.2.9 A linkage map for Nonpareil, constructed using genotypic data from SNP-based marker assays applied to Nonpareil × Constantí, Nonpareil × Lauranne, Nonpareil × Tarraco and Nonpareil × Vairo F_1 progeny.

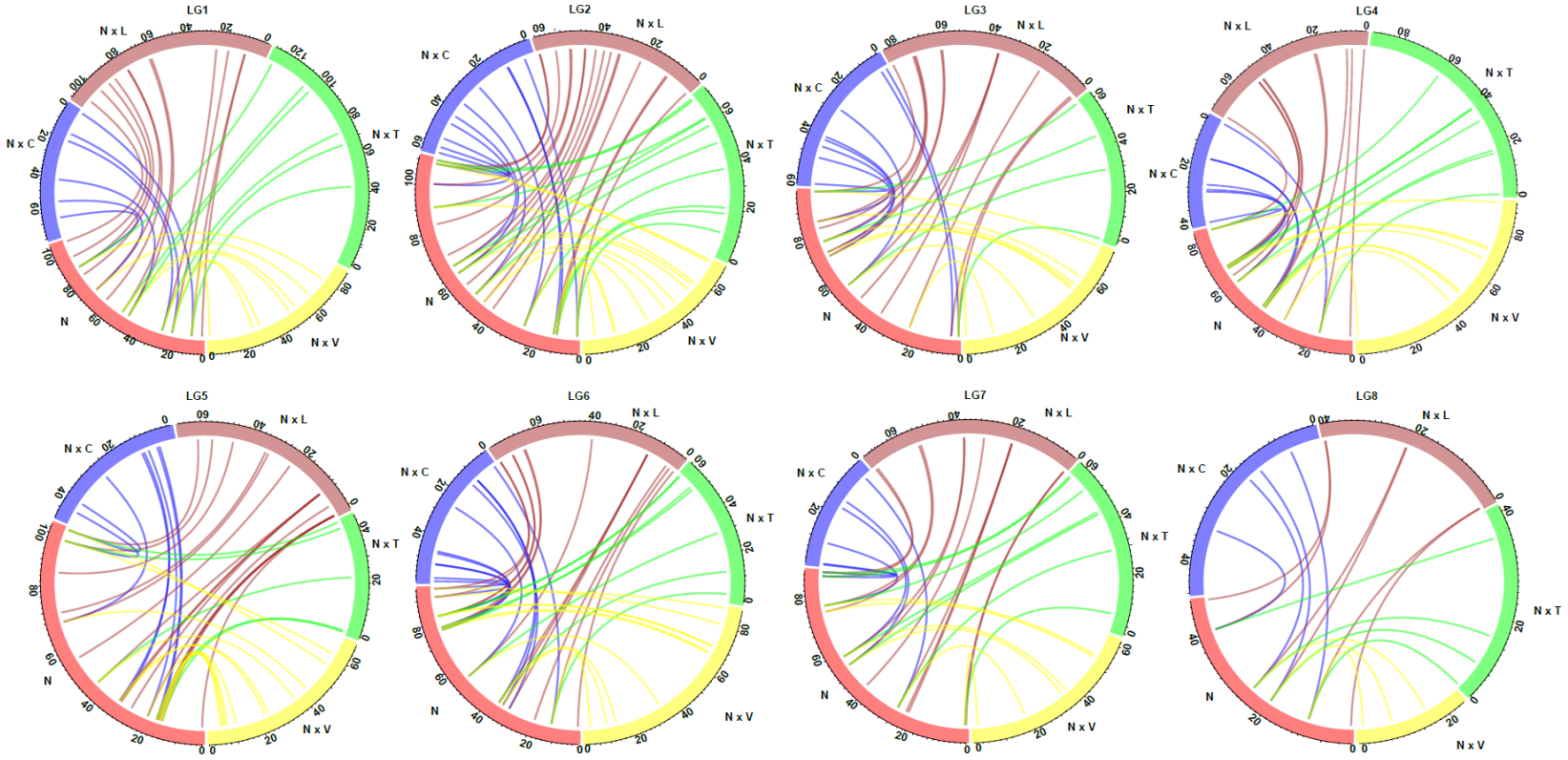


Fig. 5.2.10 Comparison of marker order within linkage groups of a composite Nonpareil map and four individual Nonpareil maps. The upper panel shows linkage groups (LG) 1-4 and the lower panel shows linkage groups (LG) 5-8. In each case, the Nonpareil composite map (N) is in red, the map from Nonpareil × Constantí (N × C) is in blue, the map from Nonpareil × Lauranne (N × L) is in brown, the map from Nonpareil × Tarraco (N × T) is in green and the map from Nonpareil × Vairo (N × V) is in yellow.

5.2.4 Discussion

The maps generated in this analysis provide information about the transferability of single nucleotide polymorphisms discovered in one population to other populations. Of 138 KASP markers that were designed based on heterozygosity in Nonpareil and homozygosity in Lauranne, between 85 (61%) and 96 (70%) were useful in one or more of four other Nonpareil populations. Of 155 KASP markers that were designed based on heterozygosity in Lauranne and homozygosity in Nonpareil, between 39 (25%) and 65 (42%) were useful in one or more of four other Nonpareil populations.

The Nonpareil linkage maps derived from N × V, N × C and N × T were similar in total length (553.4 cM, 438.9 cM and 568.9 cM, respectively) to the Nonpareil map that was obtained from N × L (608.9 cM). The composite Nonpareil map constructed using genotypic data from all four populations was 741.4 cM in total length and provides a high-quality linkage map that resolves the order of some markers that had collocated in maps from individual populations. This map can be used as a consensus map to order markers in other almond populations that originated from Nonpareil.

Although the linkage maps of Tarraco, Vairo and Constantí each have eight linkage groups, the markers are sparsely distributed on most of these groups and the maps of some linkage groups are much shorter than those that were developed for Nonpareil and Lauranne. Improvement of the linkage maps of Tarraco, Vairo and Constantí would require development of additional markers that are heterozygous in these parents but homozygous in Nonpareil. Application of genotyping-by-sequencing to these populations could enable the discovery of additional SNPs that could be exploited to improve the linkage maps.

Nonpareil is the predominant almond cultivar in the USA and in Australia. Until now, linkage mapping for Nonpareil has relied solely on data from N × L progeny (Tavassolian et al. 2010; Wu et al. 2009). Here, the use of progeny from other Nonpareil crosses (N × V, N × C and N × T) provides

opportunities to resolve marker order, to verify QTLs that were detected in N × L (Section 5.1) and to map new QTLs.

The composite linkage map constructed for Nonpareil is the first almond linkage map constructed based on multi-parental genotypic data. The process by which this map was constructed maximises the use of available genetic resources for linkage map construction and provides an initial platform for community-based resources for crop improvement in almond using a unified mapping population. The linkage maps constructed here provide new resources for almond breeding and research. These maps could allow for improved QTL mapping, anchoring to almond genome sequence assemblies and ultimately positional cloning of genes that affect important traits.

Section 5.3

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

5.3.1 Introduction

Improvement of kernel and nut traits is an important objective of almond breeding (Gradziel and Martínez-Gómez 2013). Consumers prefer almonds that are sweet with pleasing colour and high in nutrient content.

Shell hardness, width, length, thickness, geometric diameter and spherical index in the kernel and the nut are important physical parameters in both almond harvesting and processing. Almonds are usually harvested mechanically using shakers and mechanical pickers are used to collect fallen almonds. Almond processing facilities are equipped with various machines such as prehullers, rollers, shell-crackers and sorters. Physical parameters of almond kernels and nuts play important roles in the design and the standardisation of these machines (Fernández i Martí et al. 2013; Kodad and Socias i Company 2008).

Tocopherols and fatty acids are valuable chemical components of almond kernels with potential to reduce risks of cancer, other degenerative disease and cardiovascular disease in humans (Jia et al. 2006; Steinmetz and Potter 1996; Wien et al. 2010). Further, high levels of tocopherols have been reported to enable prolonged storage of almond kernels by conferring resistance against lipid rancidification (Kodad et al. 2010a). In almond, α-tocopherol is the most abundant and biologically active tocopherol (Kodad et al. 2006; Zhu et al. 2015).

Almond kernels have a unique fatty acid composition with very high levels of monounsaturated fatty acids and polyunsaturated fatty acids (Jenkins et al. 2008; Kodad and Socias i Company 2008; Sabate and Hook 1996). Of the various fatty acids in almond kernels, oleic (58 to 71%), linoleic (15 to 29%) and palmitic acids (5 to 7%) are the most abundant while stearic (1 to 3%), palmitoleic acids (3 to 5%) and vaccenic acid (1 to 2%) are also present in minor amounts (Gallier et al. 2012; Zhu et al. 2015). Both oleic (O) and linoleic (L) acids have the ability to prevent cardiovascular diseases in humans (Wien et al. 2010). High ratios of oleic to linoleic acid (O/L ≥ 2.5) reduce oil rancidification, improving stability during storage and transportation (Kodad et al. 2009a; Kodad and Socias i Company 2008).

Tocopherols and fatty acids in almond can be extracted using alkaline assisted, pressurised liquid or supercritical fluid extraction techniques. Generally, determination of tocopherol levels in plants is carried out with high performance liquid chromatography (HPLC) and determination of fatty acid composition is carried out using gas chromatography (GC) (Font i Forcada et al. 2012; Gallier et al. 2012; Zhu et al. 2015). These techniques are time consuming, labour-intensive and require specific instruments and expertise.

Molecular detection tools, linkage mapping and QTL detection could provide avenues to improve understanding of genetic control of these traits. It might be possible to replace some laborious phenotyping methods with molecular marker assays.

5.3.2 Phenotypic evaluation

Phenotypic evaluation was carried out for the 89 Nonpareil × Lauranne F_1 progeny that were used for GBS analysis, 180 of the other 231 Nonpareil × Lauranne F_1 progeny, 95 Nonpareil × Constantí F_1 progeny, 127 Nonpareil × Tarraco F_1 progeny and 90 Nonpareil × Vairo F_1 progeny.

5.3.2.1 Phenotypic evaluation of physical traits

Prior to this thesis research, each of the 89 progeny that were used for GBS analysis had been assessed for kernel and nut traits (shell hardness (SH), in-shell weight (ISWT), shell weight (SWT), kernel weight (KWT), kernel length (KL), kernel width (KW) and kernel thickness (KT) in 2002 and 2003. Forty-eight of these progeny had also been assessed for these traits in each year from 2005 to 2007. The additional progeny of Nonpareil × Lauranne, Nonpareil × Constantí, Nonpareil × Tarraco and Nonpareil × Vairo were assessed for kernel weight, in-shell weight and shell hardness in 2015. After harvest, nuts were stored at room temperature for about a month until completely dry before nut traits were measured. Measurements were taken for a random sample of ten nuts from each tree. An electronic balance with a precision of 0.001 mg was used to obtain the in-shell (almond kernel with the shell) weight and the kernel weight. For each almond nut, the in-shell weight was obtained by weighing the kernel within the shell. The shell was then cracked open using a nutcracker. Weights of the kernel alone and the shell alone were measured for each individual.

A digital calliper with a precision of 0.01 mm was used to measure the length, thickness and width of each kernel. Kernel length was measured as the distance from the basal end to the apex, kernel width was the distance across the kernel at its widest point and thickness was measured as the maximum distance from front to back between the thickest point of the kernel.

The following physical properties were calculated using those measurements.

1. Shell hardness percentage (SH%), calculated according to the formula suggested by Rugini (1986). In this formula kernel weight (KWT) is the weight of the almond kernel and in-shell weight (ISWT) is the weight of the almond kernel together with the shell.

$$SH\% = \frac{KWT}{ISWT} \times 100$$

2. Geometric mean diameter (Dp), calculated based on the formula suggested by Mohsenin (1970).

$$DP = (LWT)^{1/3}$$

3. Spherical index (Φ), calculated by dividing the Dp by the length of the kernel as follows:

$$\Phi = \frac{(LWT)^{1/3}}{L}$$

4. Kernel size (KS), calculated by multiplying length, width and thickness of the kernel as below:

$$KS = L \times W \times T$$

5. Kernel shape (KSH), calculated by dividing the width of the kernel by its length as follows:

$$KSH = \frac{W}{L}$$

Where L is the kernel length, W is the kernel width and T is the kernel thickness.

5.3.2.2 Phenotypic evaluation of almond chemical traits

Tocopherols (α-, β- and γ-) and fatty acids (oleic, linoleic, palmitic, vaccenic and stearic acids) were extracted from almond meal.

5.3.2.2.1 Tocopherol extraction from almond kernel

A total of 180 almond nuts from the 231 Nonpareil × Lauranne F_1 progeny, were harvested when the mesocarps had naturally split, indicating full ripening. These nuts were stored at room temperature for about 1 month until completely dry. The almond shells were removed manually using a nutcracker. Dried almond kernels were ground to a fine powder using an electric grinder, sieved through a 1000 μm mesh, covered with an aluminium foil and stored at -20^{0}C until tocopherol extraction. For tocopherol extraction, two subsamples from each sample were used.

Tocopherol was extracted based on the alkaline saponification and hexane extraction method (Xu, 2002). Approximately 0.25 g of almond powder was mixed with 0.025 g of ascorbic acid (Sigma-Aldrich, USA), 2.5 mL of absolute ethanol (Scharlau Chemie S.A., Sentminant, Spain) and 0.25 mL of 80% aqueous potassium hydroxide (KOH) (Sigma-Aldrich, USA), after thoroughly mixing for 30 s, tubes were incubated in a water bath at 70^{O}C for 30 min with periodically vortexing in every 10 min and were placed in ice for 5 min before adding 2.5 mL of 95% HPLC grade *n*-hexane (Chem Supply, Australia) and 1.5 mL of MilliQ water. Tubes were thoroughly vortexed for 30 s and centrifuged (1000 g at 20^{O}C for 20 min). The supernatant were saved and residue was re-extracted by adding 2.5 mL of hexane to each tube. Extracts were combined and hexane was evaporated using a nitrogen evaporator (N-EVAP 112, Organomation Associates Inc., Berlin, MA, USA) at 45^{O}C. When the hexane had completely evaporated, the residue was dissolved in 1 mL of hexane by vortexing thoroughly, added to 2 mL Crimp vials (Agilent Technologies, Australia) and sealed tightly using a vial capper for HPLC detection.

5.3.2.2.2 Tocopherol determination

In this research, tocopherol analysis was performed using both normal-phase and reversed-phase columns. Of 360 (duplicates of 180) samples, the first 180 samples were analysed using a normal-phase column. Due to decommissioning of the instrument used, the remaining 180 had to be analysed on another instrument, using a reversed-phase column. Each HPLC run contained two samples of blank (mobile phase) one sample of reagent blank, one sample from each of the standards (α-, β- γ- and δ-tocopherol) and 60 almond samples.

The normal-phase column used for this research was a Grace Alltima HP silica column (150 mm, 3 mm, 3 mm; Grace Discovery Sciences, Deerfield, IL, USA) and 98% *n*-hexane and 2% 1,4-dioxane solution was used as the isocratic mobile phase with flow rate of 1.0 mL/min and injection volume 20 mL. The column temperature was 25°C. HPLC analysis was performed using an Agilent 1200

(Agilent Technologies, Deutschland, Germany) that contained a diode array detector (DAD), fluorescence detector (FLD), autosampler, and quaternary pump.

Peaks were recorded using Agilent ChemStation 1100/1200 LC software (Agilent Technologies, Deutschland, Germany). Alpha-tocopherol was detected by DAD at a signal wavelength of 292 nm, while β-, γ- and δ-tocopherol were detected by FLD at signal wavelengths of 292 nm (excitation) and 325 nm (emission). The reversed-phase column used was a Diol LichroCART@ 250-4, Sorbent (dihydroxypropyl bounded) column with a 5 μm particle size (Merck, Darmstadt, Germany). The mobile phase contained a solution of *n*-hexane modified with 1% Isopropanol. The flow rate was 1.2 mL/min, injection volume was 20 mL and column temperature was 25°C. Alpha-tocopherol was detected by DAD at a signal wavelength of 296 nm, while β-, γ- and δ-tocopherol were detected by FLD at signal wavelengths of 296 nm (excitation) and 330 nm (emission). Columns were equilibrated at least 30 min before chromatographic data were collected. Total tocopherol amount of each component (α-, β-, γ- and δ-) was determined using calibration curves prepared from external standards.

5.3.2.2.3 Calibration curve preparation

Calibration curves were prepared for normal-phase and reversed-phase columns separately for α-, β-, γ- and δ-tocopherol. Standard stock solutions were purchased from Merck, Darmstadt. Germany. For standard curve preparation, 6.20 mg of α-, β-, δ- and 4.84 mg of γ-tocopherol standards dissolved in 100 mL of *n*-hexane separately. Calibration curves were prepared using a set of seven data points for each standard by running total volumes from 0.6 μL to 0.005 μL of each diluted stock solutions in HPLC. Wavelengths at which α-, β-, γ- and δ-tocopherol standards were detected, detectors used for each standard and conditions were similar to that mentioned above.

5.3.2.2.4 Fatty acid determination

Lipids were extracted from approximately 0.1 g of milled almond powder from each of 180 samples, using a chloroform-methanol extraction method as described by Folch et al. (1957). Lipid samples were used to prepare methyl esters of the corresponding fatty acids as described by Zhu et al. (2015). These methyl esters were separated in an HP 6890 Gas Chromatograph (Hewlett Packard, Palo Alto, CA. USA) equipped with a flame ionization detector (FID), split/splitless injection, HP 7683 autosampler, HP Chemstation and a capillary GC column (SGE BPX70, 30 m × 0.25mm ID) with 0.25 µm film thickness. The carrier gas was helium with a flow rate of 1 mL/min. The split ratio was 20:1 and injection volume of 1 µL. The temperature of the inlet and the detector was maintained at 250°C and 300°C, respectively. The initial oven temperature was 140°C gradually increasing to 220°C at 5°C/min ramp rate and temperature was maintained at 220°C for 3 min. Methyl esters were identified based on the relative retention time of the internal standard free fatty acid C17:0. Fatty acid determination was conducted by Waite Lipid Analysis Service, The University of Adelaide, Australia.

5.3.3 Statistical analysis

All phenotypic datasets were evaluated for outliers using the linear mixed modelling package ASReml-R (Butler et al. 2009) in the software platform R (www.CranR.org). Statistical analysis was conducted with GenStat16 (Payne et al. 2009). Generalized heritability (h^2_g) was estimated for each trait using ASReml-R (Butler et al. 2009). This uses a residual maximum likelihood approach for estimation of the model parameters of fixed effects, random effects and residual correlations.

Pairwise Pearson phenotypic correlation coefficients were calculated to evaluate the associations among traits. These correlation coefficients were calculated separately for each year. For chemical traits (tocopherols and fatty acids), correlation coefficients were calculated for a subset of 120 individuals for which data were available for all of these traits.

5.3.4 Quantitative trait loci detection

The linkage maps constructed for Nonpareil × Lauranne, Nonpareil × Constantí, Nonpareil × Tarraco and Nonpareil × Vairo (Section 5.1 and Section 5.2) were used for QTLs detection. The R/qtl package available in the statistical software platform R (www.CranR.org) was used for QTLs detection. QTLs were detected using the function Scanone, with testing for putative QTLs at 1 cM intervals throughout the genome. Genome-wide LOD significance threshold for $\alpha = 0.05$ was determined using 10,000 permutations. The QTL regions detected were anchored to the peach genome sequence assembly to enable comparison of QTL regions between the Nonpareil and Lauranne, Constantí and Tarraco maps.

5.3.5 Results

5.3.5.1 Trait means, heritability and correlation between kernel and nut physical traits

Considerable variation was observed among the Nonpareil × Lauranne progeny for mean values of almost all nut and kernel traits (Table 5.3.1). Heritability estimates were high for all nut and kernel traits (Table 5.3.1), ranging from 0.61 for kernel width to 0.79 for in-shell weight. Similar variation of in-shell weight, shell hardness and kernel weight was observed for Nonpareil × Vairo, Nonpareil × Tarraco and Nonpareil × Constantí (Table 5.3.2).

Based on phenotypic data for Nonpareil × Lauranne from 2003, in-shell weight was negatively associated with shell hardness and the ratios L/W, T/L and T/W were positively associated with all other traits (Table 5.3.3). Kernel thickness was positively correlated with all other traits except kernel length and the ratio L/W. Kernel length was positively correlated with kernel width and the ratio L/W, but negatively correlated with the ratios W/L, T/L and T/W. Geometric diameter was positively correlated with kernel size and spherical index.

Table 5.3.1. Means and heritability of physical traits of nuts and kernels assessed on nuts harvested in 2003 from 89 Nonpareil × Lauranne F_1 progeny.

Nut/kernel trait	Progeny			Parents		Heritability
	Minimum	Maximum	Mean ± SD	Nonpareil	Lauranne	
In-shell weight (g)	2.87	4.83	3.18 ± 0.11	3.25	4.12	0.79
Shell weight (g)	2.71	3.19	2.95 ± 0.17	2.95	3.12	0.79
Shell hardness (%)	20	60	33.98 ± 7.36	68	33	0.71
Kernel weight (g)	1.04	1.44	1.09 ± 0.17	0.98	1.25	0.70
Kernel length (mm)	21	24	22.19± 1.03	19.25	23.56	0.70
Kernel width (mm)	12	14.10	13.00 ± 0.61	12.50	13.80	0.61
Kernel thickness (mm)	6.90	8	7.02 ± 2.70	6.20	7.80	0.70
Kernel shape (W/L)	0.50	0.75	0.61 ± 0.05	0.45	0.78	
Length/Width	1.30	2.00	1.60 ± 0.14	1.40	2.20	
Thickness/Length	0.26	0.48	0.37 ± 0.05	0.32	0.43	
Thickness/Width	0.42	0.73	0.61 ± 0.06	0.40	0.68	
Kernel size (mm^3)	2188	4252	3326 ± 441	2557	3975	
Geometric diameter	729	1417	1108 ± 147	1015	1375	

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 F_1 progeny.

Nut/kernel trait	Progeny			Parents	
	Minimum	Maximum	Mean ± SD	Nonpareil	Other parent
Nonpareil × Constantí					
In-shell weight (g)	2.4	6.2	3.8 ± 0.1	3.23	-
Shell weight (g)	0.5	4.9	2.4 ± 1.1	2.93	-
Shell hardness (%)	10.9	32.5	22.2 ± 4.6	67	-
Kernel weight (g)	0.4	1.5	1.1 ± 0.8	0.97	-
Nonpareil × Tarraco					
In-shell weight (g)	1.1	7.9	2.7 ± 1.8	3.23	7.15
Shell weight (g)	0.1	6.2	1.9 ± 0.2	2.93	4.89
Shell hardness (%)	10.5	88.7	36.5 ± 18.1	67	33
Kernel weight (g)	0.6	2.1	1.1 ± 0.9	0.97	2.35
Nonpareil × Vairo					
In-shell weight (g)	1.0	6.3	2.6 ± 0.1	3.23	4.25
Shell weight (g)	0.7	4.4	2.4 ± 0.06	2.93	3.15
Shell hardness (%)	23.1	57.9	38.5 ± 7.1	67	26
Kernel weight (g)	0.3	1.6	1.09 ± 0.2	0.97	1.1

Table 5.3.3. Pair-wise correlation coefficients for almond nut and kernel traits for Nonpareil × Lauranne F_1 progeny in 2003.

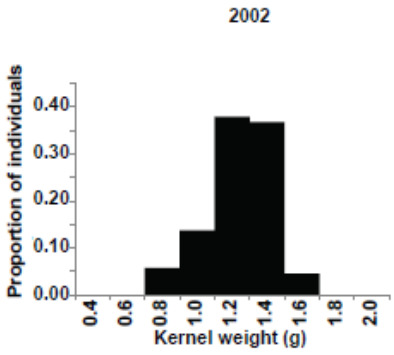
Trait	Shell weight	Shell hardness	Kernel weight	Kernel length	Kernel width	Kernel thickness	Kernel shape	Length/ Width	Thickness/ Length	Thickness/ Width	Kernel size
Shell hardness	-0.42*	-									
Kernel weight	0.61*	0.40*	-								
Kernel length	0.41*	0.28	0.64*	-							
Kernel width	0.72*	-0.19	0.56*	0.31	-						
Kernel thickness	0.14	0.29	0.42	-0.12	0.13	-					
Kernel shape	0.15	-0.40*	-0.17	-0.68*	0.46*	0.20	-				
Length/Width	-0.17	0.39*	0.15	0.67*	-0.48*	-0.21	-0.97*	-			
Thickness/Length	-0.21	-0.21	-0.21	0.79*	-0.15	0.70*	0.63*	-0.62*	-		
Thickness/Width	-0.41*	0.35	-0.13	0.32	-0.64*	0.68*	-0.19	0.19	0.64*	-	
Kernel size	0.65*	0.21	0.85*	0.66*	0.72*	0.51*	-0.07	0.47*	-0.15	-0.14	-
Spherical index	0.55*	0.09	0.64*	0.13	0.72*	0.77*	0.41*	-0.45*	0.39*	0.06	0.80*

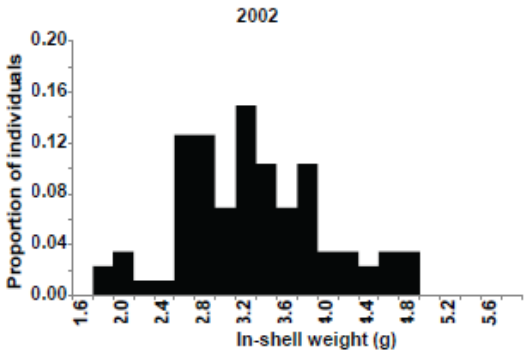
* Significant correlations (probability level, $P \leq 0.01$)

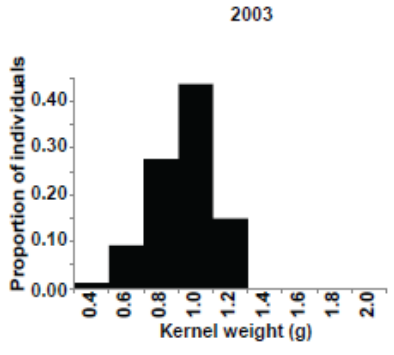
Of major physical traits analysed for 89 individuals from Nonpareil × Lauranne F_1 progeny (Figs. 5.3.1 to 5.3.4) in multiple years, similar phenotypic distributions were observed for all traits across years except for kernel weight (Fig. 5.3.1) and in-shell weight (Fig. 5.3.1). The kernel weight, in-shell weight and shell weight tended to be higher in 2006 than in other years.

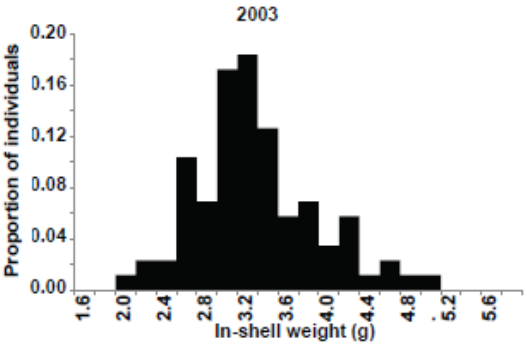
The phenotypic distributions for shell weight (Fig. 5.3.4), shell hardness (Fig. 5.3.4) and in-shell weight (Fig. 5.3.4) in Nonpareil × Tarraco displayed deviations from normality.

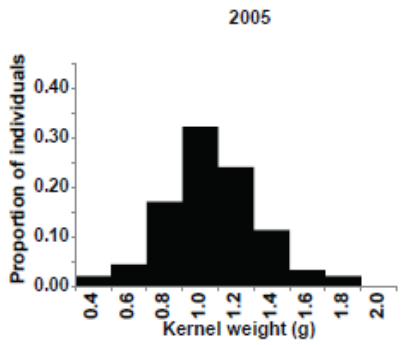
Fig. 5.3.1 Histograms depicting the phenotypic distribution of kernel weight and in-shell weight in the progeny of Nonpareil × Lauranne F_1 population. A total of 89 progeny were assessed from 2002 to 2003 and 48 progeny were assessed from 2005 to 2007. The left panel shows the kernel weight distribution and the right panel shows the in-shell weight distribution in the progeny of Nonpareil × Lauranne. In each case, horizontal axis shows the proportion of individuals and vertical axis indicates kernel and in-shell weight in gram.

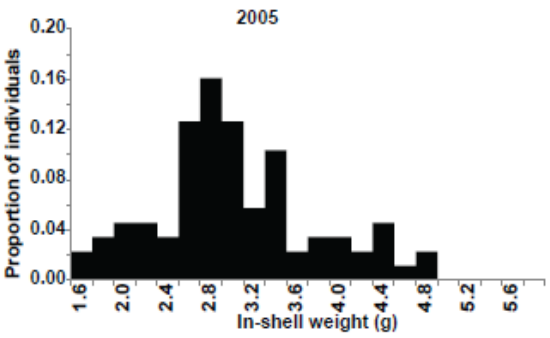

2002
Proportion of individuals
0.00
0.10
0.20
0.30
0.40
0.4
0.6
0.8
1.0
1.2
1.4
1.6
1.8
2.0
Kernel weight (g)

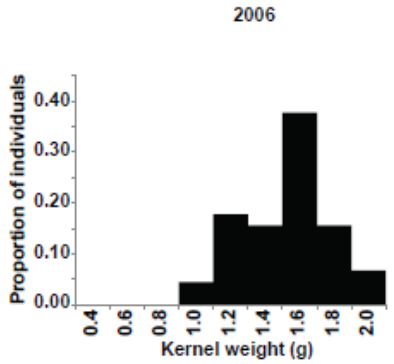

2002
Proportion of individuals
0.00
0.04
0.08
0.12
0.16
0.20
1.6
2.0
2.4
2.8
3.2
3.6
4.0
4.4
4.8
5.2
5.6
In-shell weight (g)

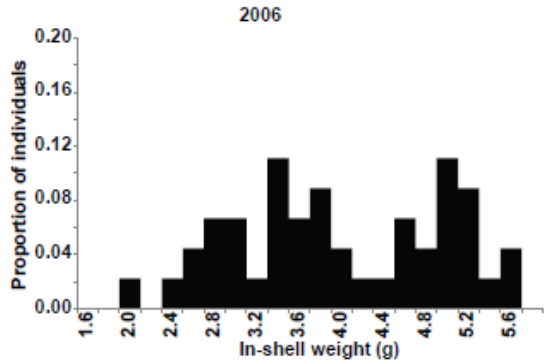

2003
Proportion of individuals
0.00
0.10
0.20
0.30
0.40
0.4
0.6
0.8
1.0
1.2
1.4
1.6
1.8
2.0
Kernel weight (g)

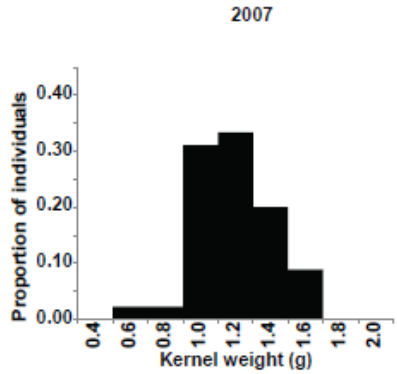

2003
Proportion of individuals
0.00
0.04
0.08
0.12
0.16
0.20
1.6
2.0
2.4
2.8
3.2
3.6
4.0
4.4
4.8
5.2
5.6
In-shell weight (g)


2005
Proportion of individuals
0.00
0.10
0.20
0.30
0.40
0.4
0.6
0.8
1.0
1.2
1.4
1.6
1.8
2.0
Kernel weight (g)


2005
Proportion of individuals
0.00
0.04
0.08
0.12
0.16
0.20
1.6
2.0
2.4
2.8
3.2
3.6
4.0
4.4
4.8
5.2
5.6
In-shell weight (g)


2006
Proportion of individuals
0.00
0.10
0.20
0.30
0.40
0.4
0.6
0.8
1.0
1.2
1.4
1.6
1.8
2.0
Kernel weight (g)


2006
Proportion of individuals
0.00
0.04
0.08
0.12
0.16
0.20
1.6
2.0
2.4
2.8
3.2
3.6
4.0
4.4
4.8
5.2
5.6
In-shell weight (g)


2007
Proportion of individuals
0.00
0.10
0.20
0.30
0.40
0.4
0.6
0.8
1.0
1.2
1.4
1.6
1.8
2.0
Kernel weight (g)

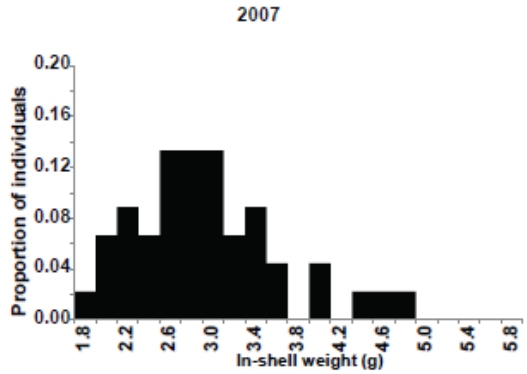
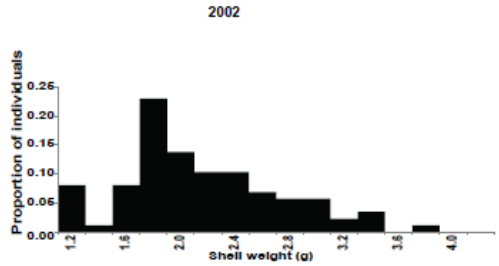
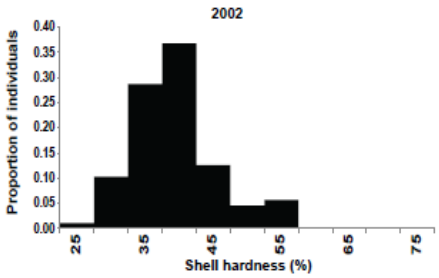
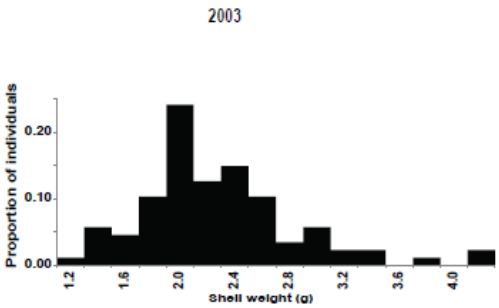
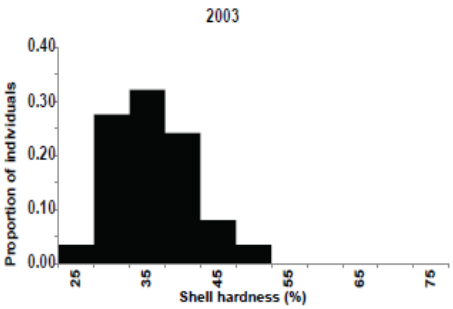

2007
Proportion of individuals
0.00
0.04
0.08
0.12
0.16
0.20
1.8
2.2
2.6
3.0
3.4
3.8
4.2
4.6
5.0
5.4
5.8
In-shell weight (g)

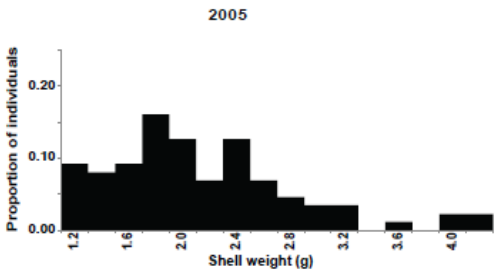
Fig. 5.3.2 Histograms depicting the phenotypic distribution of shell weight and shell hardness in the progeny of Nonpareil × Lauranne F_1 population. A total of 89 progeny were assessed from 2002 to 2003 and 48 progeny were assessed from 2005 to 2007. The left panel shows the shell weight distribution and the right panel shows the shell hardness distribution in the progeny of Nonpareil × Lauranne. In each case, horizontal axis shows the proportion of individuals and vertical axis indicates shell weight in gram and shell hardness as a percentage.

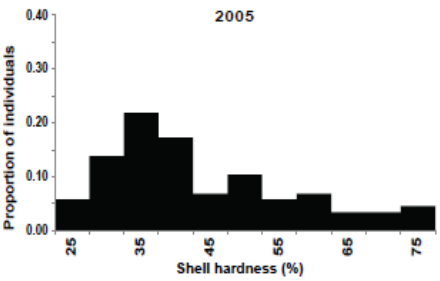

2002
Proportion of individuals
0.26
0.20
0.16
0.10
0.06
0.00
1.2
1.6
2.0
2.4
2.8
3.2
3.6
4.0
Shell weight (g)

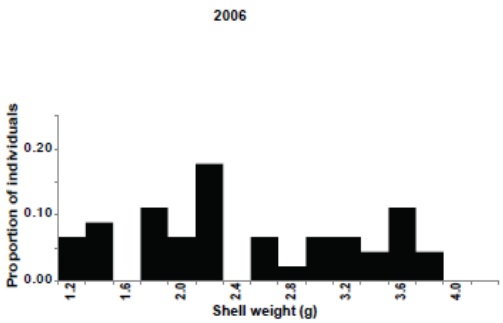

2002
Proportion of individuals
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
25
35
45
55
65
75
Shell hardness (%)

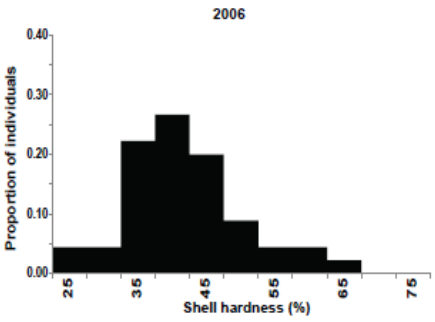

2003
Proportion of individuals
0.20
0.10
0.00
1.2
1.6
2.0
2.4
2.8
3.2
3.6
4.0
Shell weight (g)

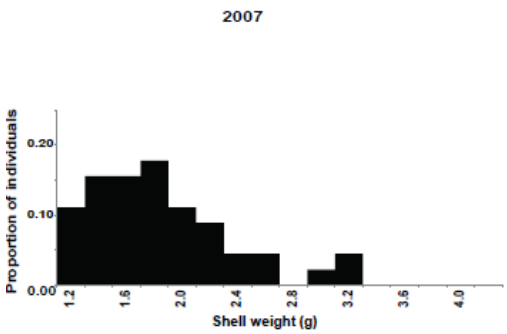

2003
Proportion of individuals
0.40
0.30
0.20
0.10
0.00
25
35
45
55
65
75
Shell hardness (%)


2005
Proportion of individuals
0.20
0.10
0.00
1.2
1.6
2.0
2.4
2.8
3.2
3.6
4.0
Shell weight (g)


2005
Proportion of individuals
0.40
0.30
0.20
0.10
0.00
25
35
45
55
65
75
Shell hardness (%)


2006
Proportion of individuals
0.20
0.10
0.00
1.2
1.6
2.0
2.4
2.8
3.2
3.6
4.0
Shell weight (g)


2006
Proportion of individuals
0.40
0.30
0.20
0.10
0.00
25
35
45
55
65
75
Shell hardness (%)


2007
Proportion of individuals
0.20
0.10
0.00
1.2
1.6
2.0
2.4
2.8
3.2
3.6
4.0
Shell weight (g)

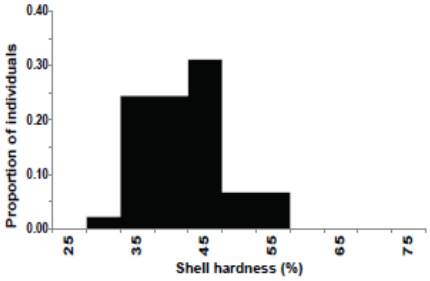
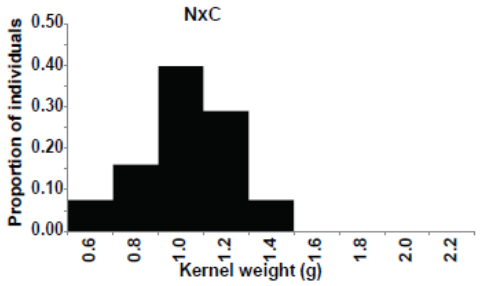
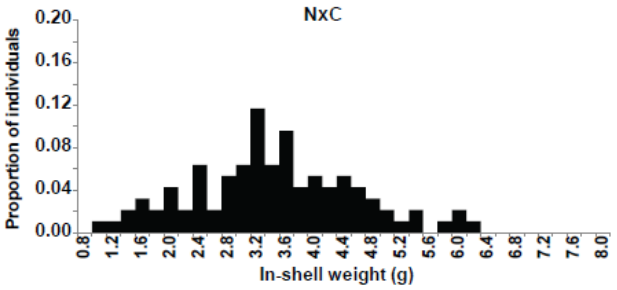
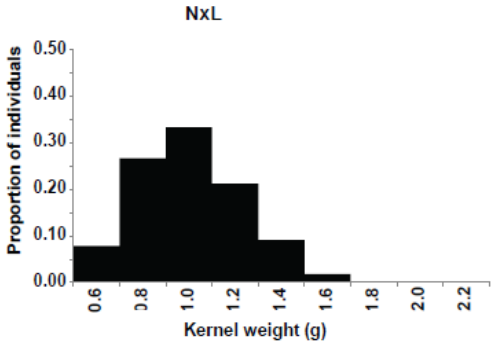
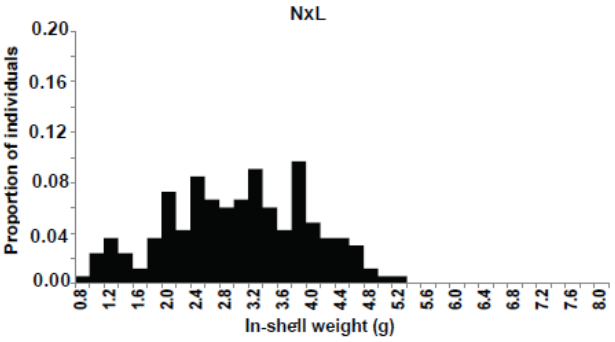

2007
Proportion of individuals
0.40
0.30
0.20
0.10
0.00
25
35
45
55
65
75
Shell hardness (%)

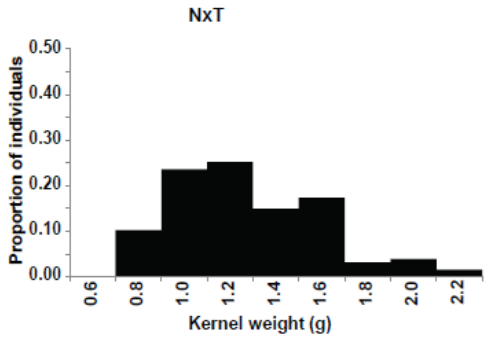
Fig. 5.3.3 Histograms depicting the phenotypic distribution of kernel weight and in-shell weight in the progeny of Nonpareil × Constantí (N × C), Nonpareil × Lauranne (N × L), Nonpareil × Tarraco (N × T) and Nonpareil × Vairo (N × V) F_1 populations in 2015. A total of 95, 180, 127 and 90 individuals were assessed for N × C, N × L, N × T and N × V, respectively. The left shows the kernel weight distribution and the right panel shows the in-shell weight distribution in the each population. In each case, horizontal axis shows the proportion of individuals and vertical axis indicates kernel weight and in-shell weight in gram.

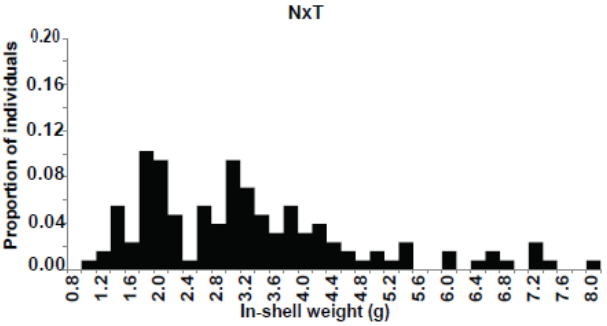

NxC
Proportion of individuals
0.50
0.40
0.30
0.20
0.10
0.00
0.6 0.8 1.0 1.2 1.4 1.6 1.8 2.0 2.2
Kernel weight (g)

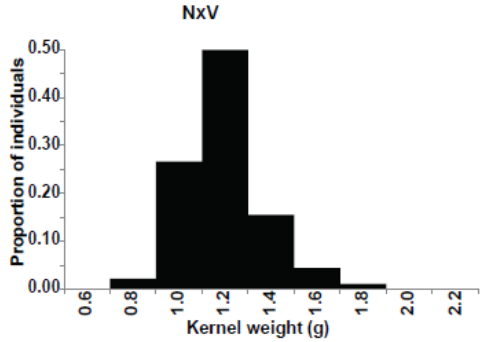

NxC
Proportion of individuals
0.20
0.16
0.12
0.08
0.04
0.00
0.8 1.2 1.6 2.0 2.4 2.8 3.2 3.6 4.0 4.4 4.8 5.2 5.6 6.0 6.4 6.8 7.2 7.6 8.0
In-shell weight (g)


NxL
Proportion of individuals
0.50
0.40
0.30
0.20
0.10
0.00
0.6 0.8 1.0 1.2 1.4 1.6 1.8 2.0 2.2
Kernel weight (g)


NxL
Proportion of individuals
0.20
0.16
0.12
0.08
0.04
0.00
0.8 1.2 1.6 2.0 2.4 2.8 3.2 3.6 4.0 4.4 4.8 5.2 5.6 6.0 6.4 6.8 7.2 7.6 8.0
In-shell weight (g)


NxT
Proportion of individuals
0.50
0.40
0.30
0.20
0.10
0.00
0.6 0.8 1.0 1.2 1.4 1.6 1.8 2.0 2.2
Kernel weight (g)


NxT
Proportion of individuals
0.20
0.16
0.12
0.08
0.04
0.00
0.8 1.2 1.6 2.0 2.4 2.8 3.2 3.6 4.0 4.4 4.8 5.2 5.6 6.0 6.4 6.8 7.2 7.6 8.0
In-shell weight (g)


NxV
Proportion of individuals
0.50
0.40
0.30
0.20
0.10
0.00
0.6 0.8 1.0 1.2 1.4 1.6 1.8 2.0 2.2
Kernel weight (g)

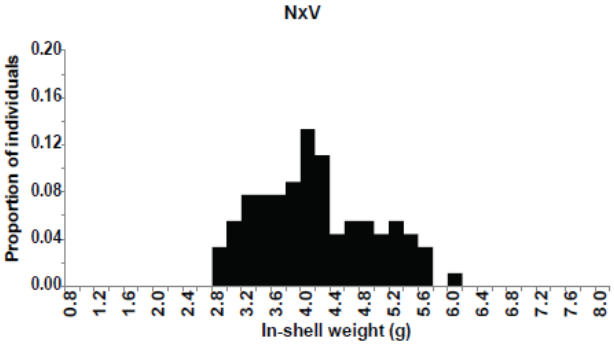
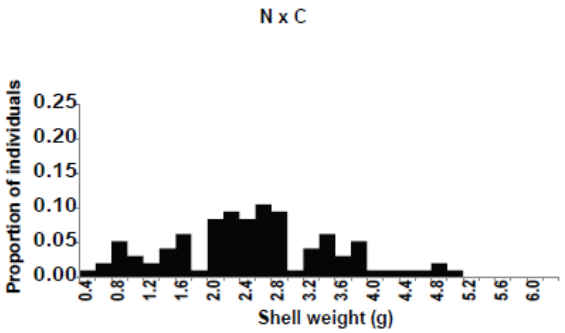
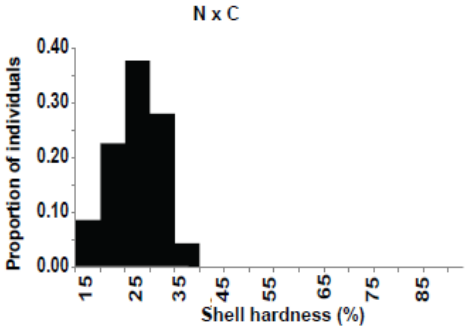
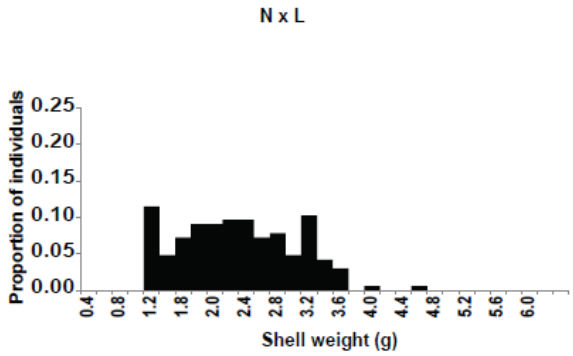
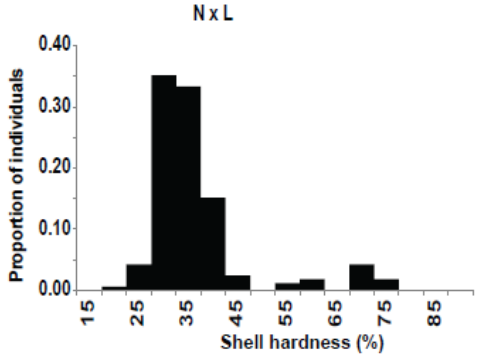

NxV
Proportion of individuals
0.20
0.16
0.12
0.08
0.04
0.00
0.8 1.2 1.6 2.0 2.4 2.8 3.2 3.6 4.0 4.4 4.8 5.2 5.6 6.0 6.4 6.8 7.2 7.6 8.0
In-shell weight (g)

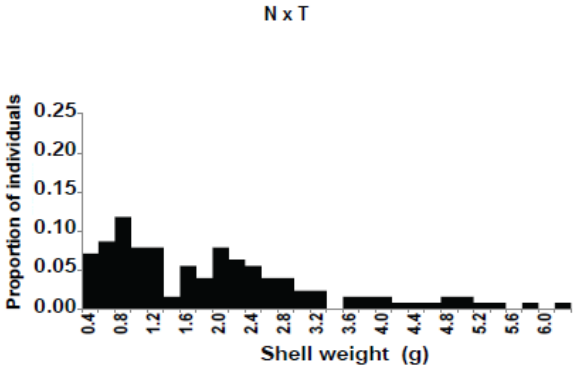
Fig. 5.3.4 Histograms depicting the phenotypic distribution of shell weight and shell hardness in the progeny of Nonpareil × Constantí (N × C), Nonpareil × Lauranne (N × L), Nonpareil × Tarraco (N × T) and Nonpareil × Vairo (N × V) F_1 populations in 2015. A total of 95, 180, 127 and 90 individuals were assessed for N × C, N × L, N × T and N × V, respectively. The upper panel shows the shell weight distribution and the lower panel shows the shell hardness distribution in the each population. In each case, horizontal axis shows the proportion of individuals and vertical axis indicates in-shell weight in gram and shell hardness as a percentage.

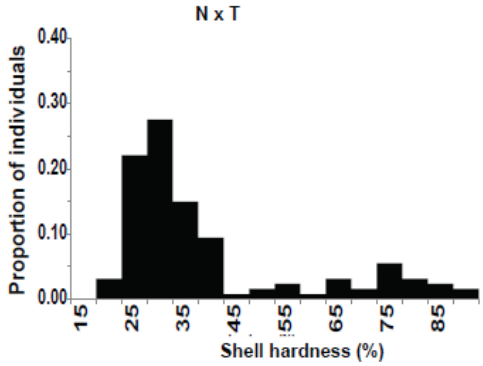

N x C
Proportion of individuals
0.25
0.20
0.15
0.10
0.05
0.00
0.4 0.8 1.2 1.6 2.0 2.4 2.8 3.2 3.6 4.0 4.4 4.8 5.2 5.6 6.0
Shell weight (g)

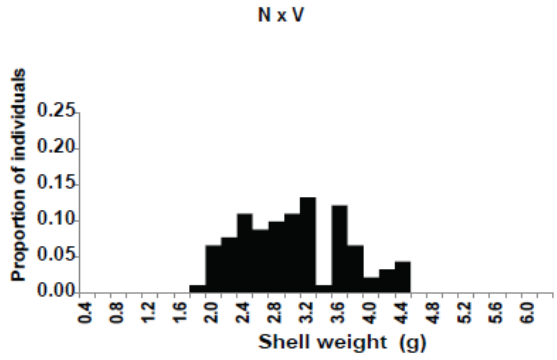

N x C
Proportion of individuals
0.40
0.30
0.20
0.10
0.00
15 25 35 45 55 65 75 85
Shell hardness (%)

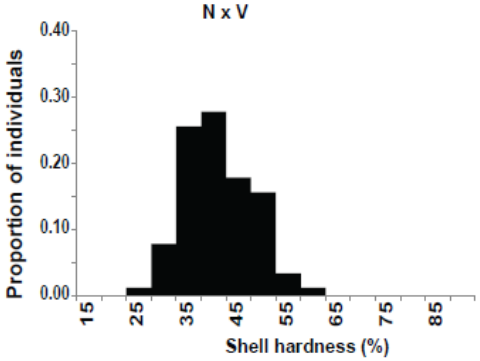

N x L
Proportion of individuals
0.25
0.20
0.15
0.10
0.05
0.00
0.4 0.8 1.2 1.8 2.0 2.4 2.8 3.2 3.6 4.0 4.4 4.8 5.2 5.6 6.0
Shell weight (g)


N x L
Proportion of individuals
0.40
0.30
0.20
0.10
0.00
15 25 35 45 55 65 75 85
Shell hardness (%)


N x T
Proportion of individuals
0.25
0.20
0.15
0.10
0.05
0.00
0.4 0.8 1.2 1.6 2.0 2.4 2.8 3.2 3.6 4.0 4.4 4.8 5.2 5.6 6.0
Shell weight (g)


N x T
Proportion of individuals
0.40
0.30
0.20
0.10
0.00
15 25 35 45 55 65 75 85
Shell hardness (%)


N x V
Proportion of individuals
0.25
0.20
0.15
0.10
0.05
0.00
0.4 0.8 1.2 1.6 2.0 2.4 2.8 3.2 3.6 4.0 4.4 4.8 5.2 5.6 6.0
Shell weight (g)


N x V
Proportion of individuals
0.40
0.30
0.20
0.10
0.00
15 25 35 45 55 65 75 85
Shell hardness (%)

5.3.5.2 Tocopherols and fatty acids in Nonpareil × Lauranne

In almond meal from Nonpareil, the tocopherol concentration was 16.0 mg/100 g. Three types of tocopherols were detected in the Nonpareil × Lauranne: α-tocopherol (detected in all 180 individuals), β-tocopherol (detected in 120 individuals) and γ-tocopherol (detected in 124 individuals). The progeny exhibited a wide range of total tocopherol concentrations, from 1.5 mg/100 g to 40.3 mg/100 g. In all cases, the main component was α-tocopherol, with proportions ranging from 0.60 to 1.00. For all except five progeny, the proportion of α-tocopherol exceeded 0.90. The five exceptions all had low total tocopherol concentration (between 1.9 and 4.4 mg/100g). The other two components (β- and γ-tocopherol) ranged from 0.01 to 0.23 and from 0.005 to 0.17, respectively (Fig. 5.3.5).

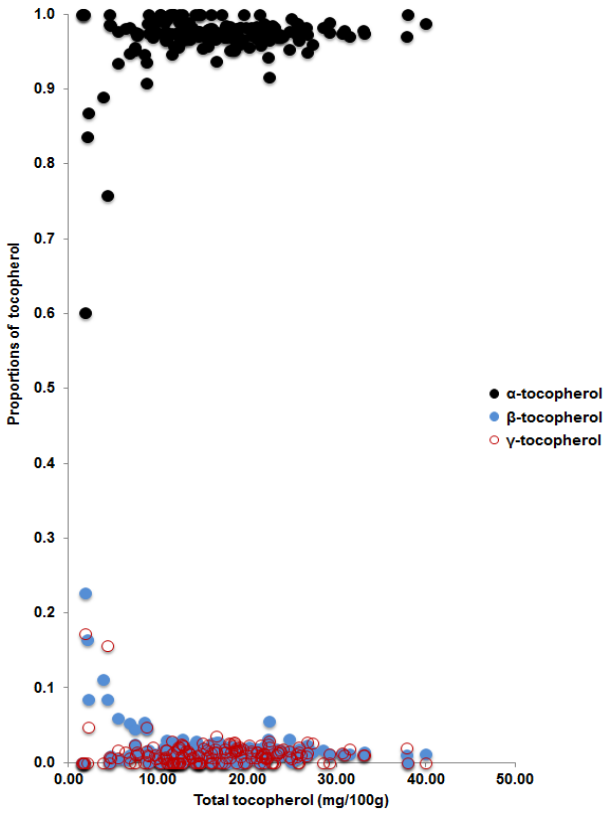


Fig. 5.3.5 Proportions of tocopherols in Nonpareil × Lauranne. Proportions of the tocopherol components (α-, β- and γ-) relative to the total tocopherol concentration in 180 progeny of Nonpareil × Lauranne F_1 population.

The progeny showed wide variation in total fatty acids amounts ranging from 88 mg/100 g to 536 mg/100 g of oil (Fig. 5.3.6). The most abundant fatty acids detected in almond oil extracts were oleic acid, linoleic acid, palmitic acid, and stearic acid, all of which were detected in all 180 individuals. Vaccenic acid was also detected in 130 individuals (Table 5.3.4). Eicosapentaenoic acid (up to 5%) and α-linoleic acid (up to 2%) were detected in 3 and 5 individuals, respectively. Oleic acid was the major component of fatty acid in all, with proportions ranged from 0.41 to 0.66. The proportions of linoleic acid, palmitic acid, stearic acid and vaccenic acid ranged from 0.15 to 0.33, 0.09 to 0.28, 0.02 to 0.14 and 0.01 to 0.03, respectively. Fatty acid composition remained fairly constant regardless of the total amount of fatty acids.

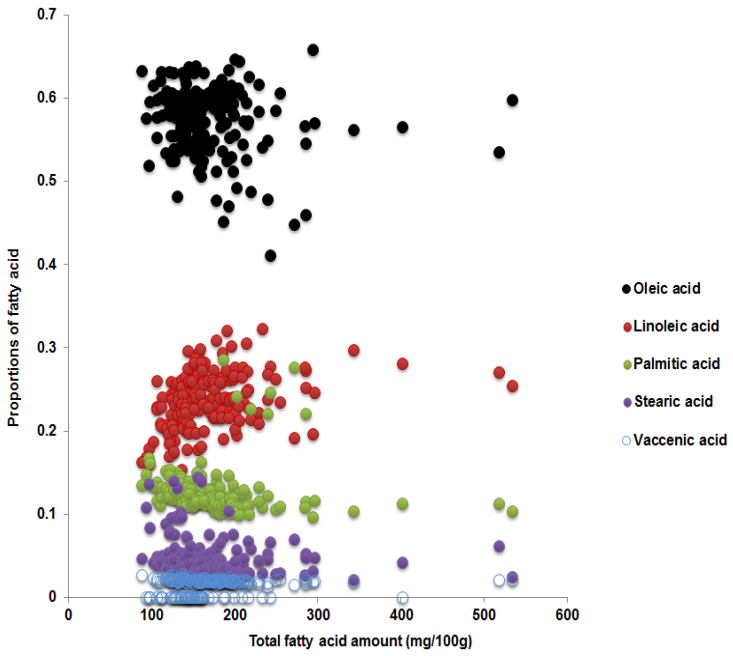


Fig. 5.3.6 Proportions of the major fatty acids in Nonpareil × Lauranne. Proportions of oleic acid, linoleic acid, palmitic acid, stearic acid and vaccenic acid relative to the total fatty acid concentration in 180 progeny of Nonpareil × Lauranne F_1 population.

Table 5.3.4. Means of fatty acids assessed on kernels from nuts harvested in 2015 from 180 Nonpareil × Lauranne F_1 progeny. Fatty acids were measured in mg/100g of total oil.

Trait	Minimum	Maximum	Mean± SD
Oleic acid	56.1	319.1	102 ± 3.2
Linoleic acid	14.3	140.3	42.4 ± 1.6
Palmitic acid	11.9	71.1	21.7 ± 0.8
Stearic acid	3.8	31.7	7.7 ± 0.3
Vaccenic acid	1.9	10.9	3.4 ± 0.1

5.3.6 Quantitative trait loci

Quantitative trait loci were detected on LG1, LG2, LG3, LG4, LG5, LG6 and LG8 of the Nonpareil map (Tables 5.3.5, 5.3.7 and Fig. 5.3.7) and on LG1, LG2, LG3, LG4, LG5 and LG8 of the Lauranne map (Tables 5.3.6, 5.3.7 and Fig. 5.3.8). A total of 23 QTL regions were detected: 9 regions (designated A-I) in the Nonpareil map and 14 (designated J-W) in the Lauranne map. Some QTL regions affected physical traits (7 regions in Nonpareil and 10 regions in Lauranne), some affected chemical traits (2 regions in Nonpareil and 4 regions in Lauranne) and one region affected both physical and chemical traits in Lauranne (genomic region K).

Anchoring of QTL regions detected for Nonpareil and Lauranne to the peach genome sequence assembly indicated that in most cases, QTL regions detected for Nonpareil did not coincide with those detected for Lauranne. The main exception was that region B, which affected physical traits in Nonpareil, may coincide with region O, which affected physical traits in Lauranne. Both of these regions anchored to scaffold 2 (Pp2) of the peach genome assembly (Fig. 5.3.9).

For Nonpareil, two QTLs for shell hardness were detected on LG5 (QTL regions F and G (Table 5.3.5 and Fig. 5.3.7). In these regions, progeny with heterozygous marker genotypes had higher

shell hardness values (means > 35%) than progeny with the homozygous marker genotypes from Lauranne (means < 31%) (Fig. 5.3.10).

For Lauranne, shell hardness QTLs were detected on LG2 (QTL region N), LG5 (QTL regions S and T) and LG8 (QTL region W) (Table 5.3.6 and Fig. 5.3.8). In these regions, progeny with homozygous marker genotypes from Nonpareil had higher shell hardness values (means > 36%) than those with the heterozygous genotype from Lauranne (means < 34%). QTL region on LG2 in Lauranne is close to a region (O) in which in-shell weight and kernel weight QTLs were detected (Table 5.3.6).

Based on individual QTL data, Nonpareil genotypes in seven QTL regions (heterozygosity in regions F and G and homozygosity in regions N, O, S, U and W) increased shell hardness percentages. Of the 180 progeny that were evaluated for shell hardness, 17 had this genotypic combination (Fig. 5.3.11). Selection for heterozygosity at the marker TP23628 (C:T) or at the marker TP15987 (G:A) and for homozygosity at the markers TP38620 (C:C), TP23837 (C:C) and TP29814 (G:G) would be sufficient to retain progeny with high shell hardness values (≥ 53%) and eliminate the progeny with low shell hardness values (≤ 40%). For marker-assisted selection, four markers of the seven evaluated here would be sufficient. The QTL regions, F and G, N and O and S and U are linked to each other.

For Tarraco, a QTL for shell hardness was detected close to the marker TP7579, which maps within the Lauranne QTL region N. For both Tarraco and Constantí, QTLs for shell hardness, in-shell weight and kernel weight traits are close to the marker TP38620 (Fig. 5.3.12), which mapped within Lauranne QTL region O (Fig. 5.3.8, Table 5.3.6). LOD scores for shell hardness, shell weight and in-shell weight were more highly significant for Tarraco than for the other parents.

In the Nonpareil × Constantí population, progeny with homozygous marker genotypes from Nonpareil had somewhat higher shell hardness values (means > 32%) than progeny with

heterozygous marker genotypes from the other parent (means < 23%) (Fig. 5.3.13) but no progeny had the paper shell trait. In the Nonpareil × Tarraco population, progeny with homozygous marker genotypes from Nonpareil had higher shell hardness values (means > 46%) than progeny with heterozygous marker genotypes from the other parent (means < 30%) (Fig. 5.3.13).

For Lauranne, oleic/linoleic (O/L) ratio QTLs were detected on LG1 (QTL regions K and L) and LG6 (QTL region V). The most significant QTL (L) for this trait, with a LOD score 3.9, was mapped close to the marker TP15845 on LG1. Progeny with homozygous genotypes from Nonpareil in these regions had higher O/L ratio values (means > 2.5) than progeny with heterozygous genotype from Lauranne (means < 2.1) (Fig. 5.3.14). Of the 180 progeny that were evaluated for O/L ratio, 67 were homozygous in all three regions (Fig. 5.3.15). For efficient screening of progeny likely to have high O/L ratios, selection could be applied for homozygosity at the markers TP15845 (A:A) and at the marker TP1949 (T:T) or at the marker TP1484 (G:G). Of the four markers evaluated here for high O/L ratio, markers TP15848 and TP37237 and TP9199 and TP1484 are linked to each other. The oleic/linoleic ratio QTL on LG1 (K) was collocated with the kernel length QTL that was detected on LG1 in Lauranne.

Table 5.3.5. Summary of QTLs detected for physical nut and kernel traits in Nonpareil, showing estimated QTL positions (cM), LOD scores and the percentage of phenotypic variation explained (R^2). The letters (A, C, D, F, G, H and I) designating QTL regions correspond with those shown in Figs. 5.3.7 and 5.3.9.

Linkage	QTL	Trait	2002			2003			2005			2006			2007			2015		
group	region		cM	LOD	R^2	cM	LOD	R^2	cM	LOD	R^2	cM	LOD	R^2	cM	LOD	R^2	cM	LOD	R^2
LG1	A	Kernel weight	93	2.8	13	93	3.5	17	93	2.3	11	93	2.3	11	93	2.3	11	93	2.0	6
		Kernel size	95	3.2	15	93	3.2	14	95	2.9	29	93	4.1	35	93	4.6	37	-	-	-
		Geometric diameter	93	3.2	16	92	2.8	14	92	4.6	37	92	4.1	35	92	4.6	37	-	-	-
LG2	C	Kernel thickness	42	2.7	13	42	2.7	13	42	2.3	10	41	3.1	27	41	3.0	26	-	-	-
		Kernel shape	39	3.5	14	40	5.8	27	40	2.4	22	40	2.3	20	41	4.2	35	-	-	-
		Length/Width	42	2.8	14	42	5.9	27	42	2.5	23	42	2.3	20	44	4.1	34	-	-	-
		Thickness/Length	43	3.5	17	44	6.4	29	44	4.0	19	45	3.1	16	45	2.4	10	-	-	-
		Thickness/Width	43	3.2	16	43	3.2	16	44	3.1	27	42	2.2	11	43	2.1	11	-	-	-
		Spherical index	44	4.5	21	45	3.8	18	46	2.4	21	47	3.4	21	44	5.3	41	-	-	-

Table 5.3.5., continued.

Linkage group	QTL region	Trait	2002			2003			2005			2006			2007			2015		
			cM	LOD	R^2	cM	LOD	R^2	cM	LOD	R^2	cM	LOD	R^2	cM	LOD	R^2	cM	LOD	R^2
LG3	D	Kernel length	15	4.7	21	15	3.4	17	15	2.4	21	15	3.6	30	15	2.6	22	-	-	-
		Kernel width	15	2.4	12	15	2.2	11	15	3.4	29	15	2.4	10	15	3.1	27	-	-	-
LG5	F	Shell hardness	0	5.7	26	0	4.5	22	0	2.8	14	0	3.3	17	0	3.1	15	0	2.5	9
	G	Shell hardness	45	2.6	14	45	2.5	12	45	2.6	13	45	2.7	13	45	3.6	18	45	2.5	9
		In-shell weight	45	2.8	13	45	2.7	12	45	2.8	13	45	2.2	11	45	2.2	10	45	2.3	7
		Shell weight	49	2.6	12	49	2.5	13	45	2.6	12	47	2.6	12	45	2.6	12	45	1.9^{ns}	-
LG6	H	Kernel length	2	2.5	12	2	2.5	12	2	2.4	11	2	2.4	11	2	2.4	11	-	-	-
LG8	I	Shell weight	36	2.3	12	38	2.7	13	30	2.5	12	30	2.1^{ns}	-	38	2.5	12	37	1.7^{ns}	-

ns - no significant effect, - not assessed.

Table 5.3.6. Summary of QTLs detected for Lauranne physical nut and kernel traits, showing estimated QTL positions (cM), LOD scores and the percentage of phenotypic variation explained (R^2). The letters (J, K, M, N, O, P, S, T, U and W) designating QTL regions correspond with those shown in Figs. 5.3.8 and 5.3.9.

Linkage	QTL	Trait	2002			2003			2005			2006			2007			2015		
group	region		cM	LOD	R^2	cM	LOD	R^2	cM	LOD	R^2	cM	LOD	R^2	cM	LOD	R^2	cM	LOD	R^2
LG1	J	Kernel thickness	3	3.2	16	3	2.5	12	3	2.6	12	3	2.5	12	3	2.5	12	-	-	-
	K	Kernel length	37	3.2	15	42	3.2	16	45	3.2	16	44	2.8	24	43	2.8	15	-	-	-
	M	Kernel shape	80	5.4	25	81	4.9	22	81	3.9	19	80	3.6	16	82	3.2	15	-	-	-
		Kernel length	80	5.3	25	81	5.1	24	81	3.9	33	79	3.4	30	81	3.8	32	-	-	-
		Length/Width	81	4.9	24	81	4.9	24	84	4.1	34	78	3.6	30	81	3.6	30	-	-	-
		Thickness/Length	80	2.9	14	81	3.5	26	85	3.2	28	77	2.8	24	76	4.2	34	-	-	-
		Spherical index	81	3.9	17	81	2.6	13	86	4.4	36	81	3.1	27	78	4.9	40	-	-	-
LG2	N	Shell hardness	25	2.4	13	20	2.8	15	20	2.7	15	20	2.7	15	24	3.4	26	26	3.4	9

Table 5.3.6., continued.

Linkage group	QTL region	Trait	2002			2003			2005			2006			2007			2015		
			cM	LOD	R^2	cM	LOD	R^2	cM	LOD	R^2	cM	LOD	R^2	cM	LOD	R^2	cM	LOD	R^2
	O	Shell weight	37	2.0^{ns}	-	37	2.6	13	35	2.6	13	27	2.5	13	41	2.1	12	40	2.1^{ns}	-
		Kernel weight	41	3.3	16	41	3.2	16	40	2.8	14	40	2.7	15	41	2.5	12	42	2.0^{ns}	-
		In-shell weight	40	2.5	13	41	4.3	21	40	3.7	17	40	3.6	17	41	2.5	12	42	2.0^{ns}	-
		Kernel width	43	2.5	13	42	3.1	16	43	3.7	28	42	2.8	15	40	2.7	13	-	-	-
		Kernel size	41	2.6	13	42	3.1	16	41	3.7	28	41	2.8	15	41	2.9	22	-	-	-
		Geometric diameter	41	2.6	13	41	3.1	15	41	3.3	26	41	2.2	11	41	2.4	12	-	-	-
LG3	P	Thickness/Width	4	2.2	10	5	2.4	12	4	2.9	15	5	3.0	16	5	2.7	14	-	-	-
LG5	S	Shell hardness	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	4.2	11
		In-shell weight	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	3.1	8
	T	Shell weight	20	3.6	17	20	3.4	16	26	2.8	14	25	2.2	10	25	1.9^{ns}	-	16	2.3	7

Table 5.3.6., continued.

Linkage group	QTL region	Trait	2002			2003			2005			2006			2007			2015		
			cM	LOD	R^2	cM	LOD	R^2	cM	LOD	R^2	cM	LOD	R^2	cM	LOD	R^2	cM	LOD	R^2
	U	Shell hardness	38	2.8	16	36	2.9	17	36	2.8	16	36	2.8	16	36	2.8	16	43	3.2	9
LG8	W	Shell hardness	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	87	3.4	9

ns - no significant effect , - not assessed.

Table 5.3.7. Summary of QTLs detected for chemical traits in Nonpareil and in Lauranne in 2015, showing estimated QTL positions (cM), LOD scores and the percentage of phenotypic variation explained (R^2). The letters (B and E) designating QTL regions correspond with those shown in Figs. 5.3.7 and 5.3.9 in Nonpareil. The letters (K, L, Q, R and V) designating QTL regions correspond with those shown in Figs. 5.3.8 and 5.3.9 in Lauranne.

Linkage group	QTL region	Trait	cM	LOD	R^2
Nonpareil					
LG2	B	Total fatty acid concentration	19	2.6	8
LG4	E	Total tocopherol concentration	34	2.4	6
Lauranne					
LG1	K	Oleic/linoleic ratio	35	3.1	8
LG1	L	Oleic/linoleic ratio	69	3.9	10
LG3	Q	Total fatty acid concentration	52	2.1	5
LG4	R	Total tocopherol concentration	59	2.7	8
LG6	V	Oleic/linoleic ratio	87	2.5	6

Table 5.3.8. Summary of QTLs detected in Constantí and Tarraco maps for the year 2015. Linkage group, QTL region, estimated QTL position (cM), LOD scores and the percentage of phenotypic variation explained (R^2) are shown.

Linkage group	QTL region	Trait	cM	LOD	R^2
Constantí					
LG2	TP7579-TP4873	In-shell weight	11	2.5	6
		Shell weight	11	3.0	8
	TP34564-TP38620	Shell hardness	42	2.5	6
		In-shell weight	46	4.1	10
		Shell weight	56	4.8	11
Tarraco					
	TP6022-TP4873	Shell hardness	11	4.5	11
		Shell weight	13	4.8	11
	TP38620-TP14542	Shell weight	34	10.0	36
		In-shell weight	34	9.7	25
		Shell hardness	35	16.3	45
		Kernel weight	35	3.3	9

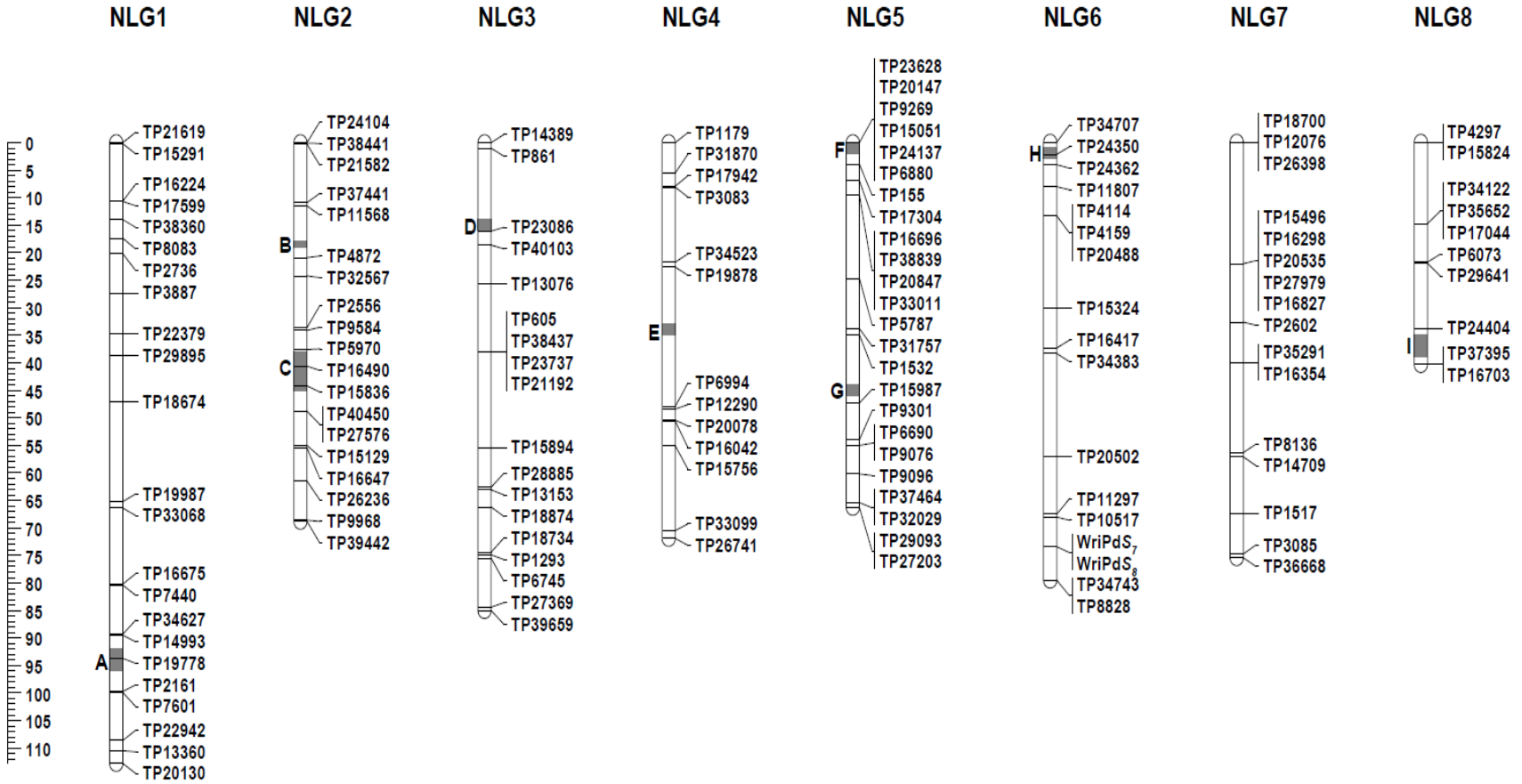


Fig. 5.3.7 A linkage map for Nonpareil, constructed using genotypic data from SNP-based marker assays applied to 231 Nonpareil × Lauranne F_1 progeny, showing QTL regions that affect almond nut and kernel traits as listed in Table 5.3.5. QTL regions detected on the eight Nonpareil linkage groups (NLG1 to NLG8) labelled from A to I and are shaded in dark grey.

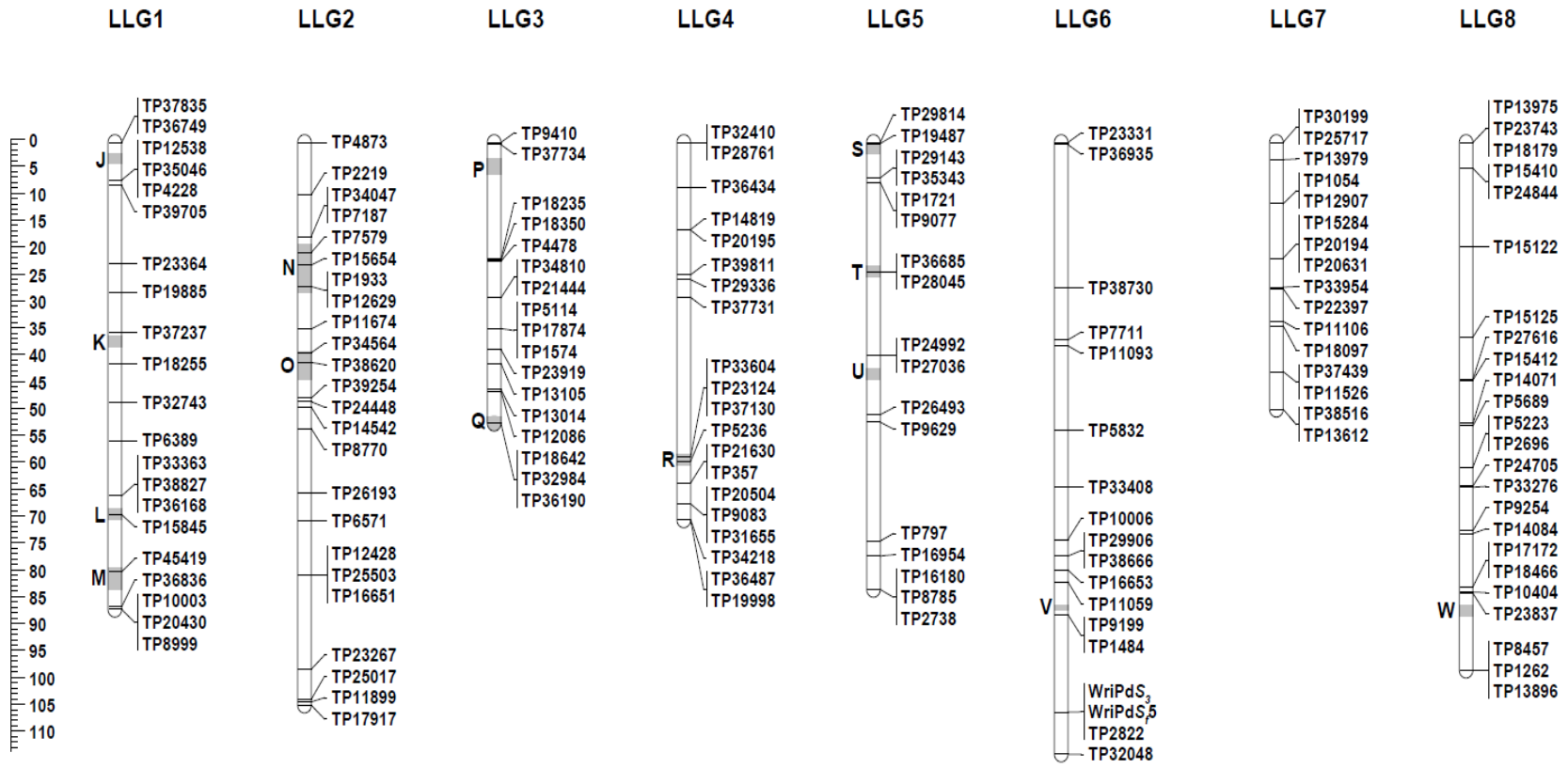


Fig. 5.3.8 A linkage map for Lauranne, constructed using genotypic data from SNP-based marker assays applied to 231 Nonpareil × Lauranne F_1 progeny, showing QTL regions that affect almond nut and kernel traits as listed in Table 5.3.6. QTL regions detected on the eight Lauranne linkage groups (LLG1 to LLG8) labelled from J to W and are shaded in light grey.

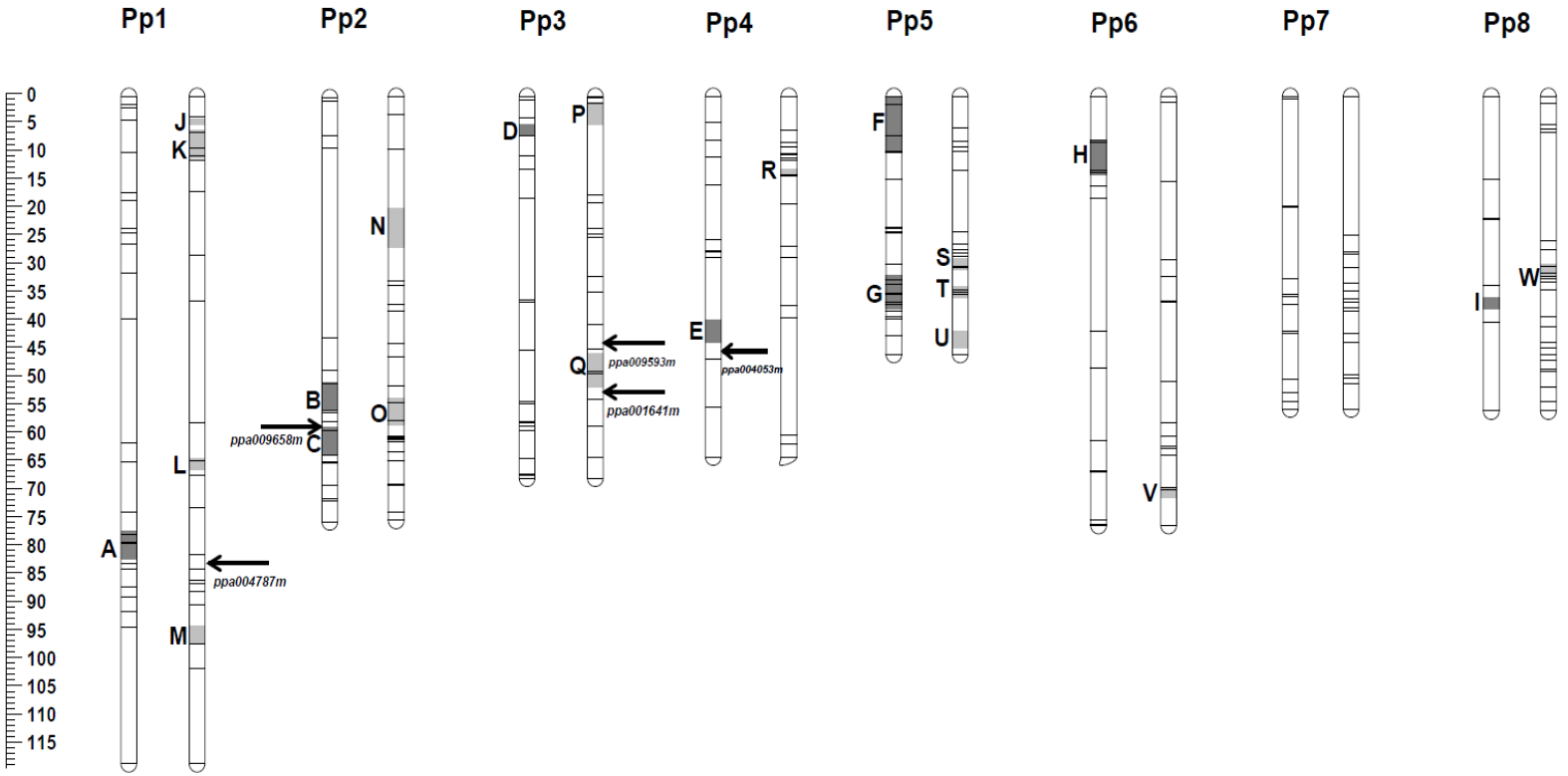


Fig. 5.3.9 Comparisons of the positions of almond quantitative trait loci detected in Nonpareil (A to I, shaded in dark grey), Lauranne (J to W, shaded in light grey) based on the positions at which linked markers were anchored to scaffolds (Pp1 to Pp8) and positions of genes that affect the fatty acid and tocopherol biosynthesis on the peach (*Prunus persica*) whole genome sequence assembly that detected close to the fatty acid and tocopherol QTLs in Nonpareil and Lauranne are marked with arrows.

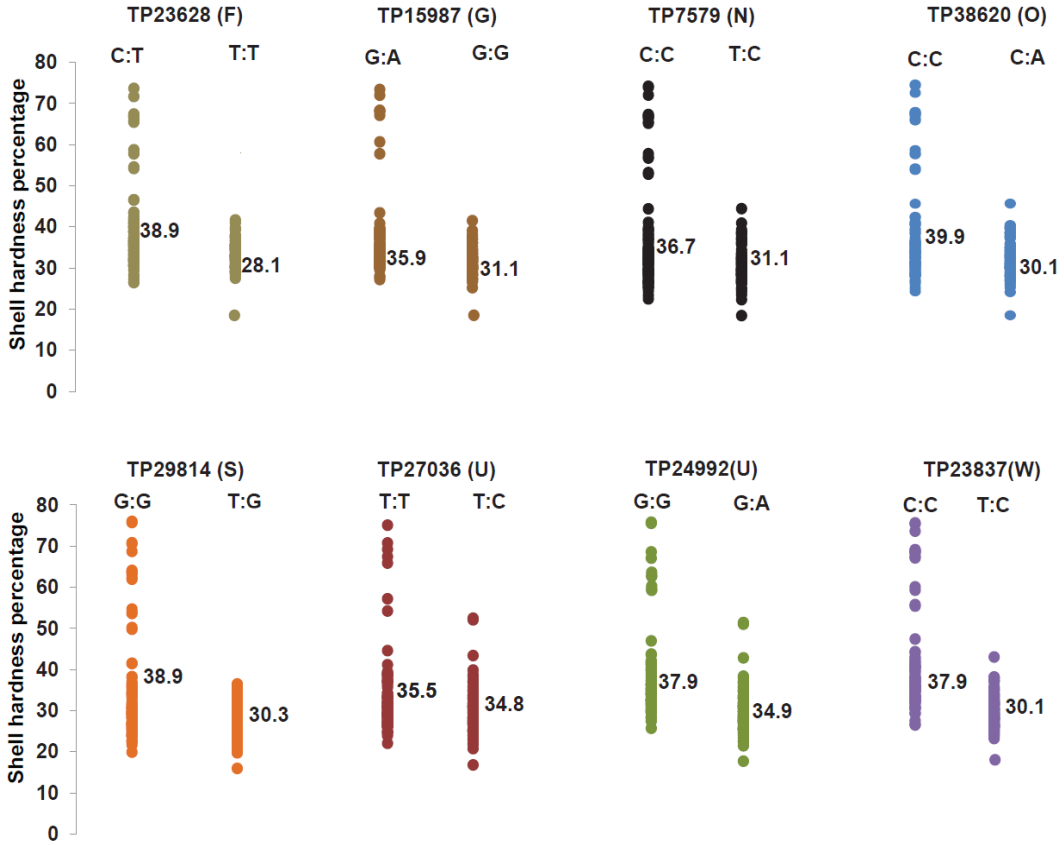


Fig. 5.3.10 Shell hardness percentages and their means for groups of Nonpareil × Lauranne F_1 progeny defined based on their genotypes at markers in QTL regions for shell hardness detected on LG2 (N and O), LG5 (F, G, S and U) and LG8 (W).

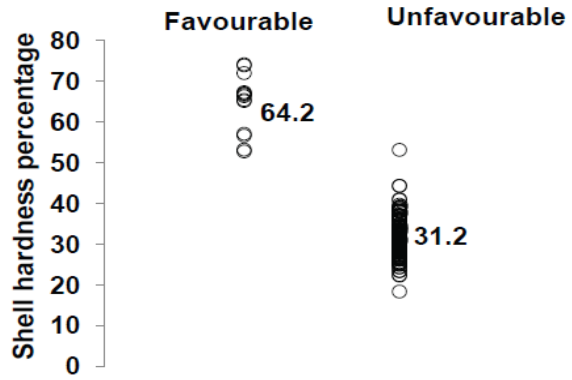


Fig. 5.3.11 Shell hardness percentages and their means for groups of Nonpareil × Lauranne F_1 progeny selected (favourable) to have paper shell traits (> 55%) based on genotypes at TP23628, TP15987, TP7579, TP38620, TP29814 and TP23837 markers in comparison to those not selected (unfavourable).

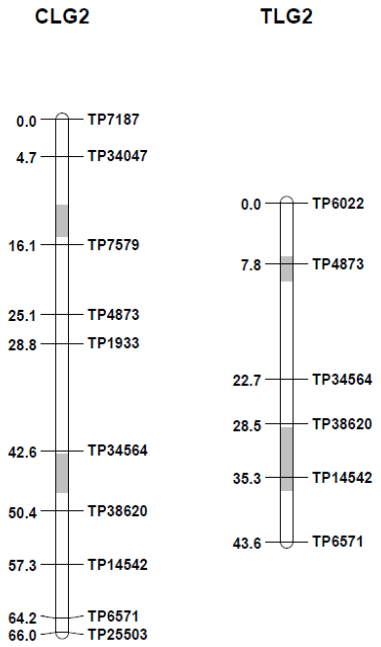


Fig. 5.3.12 Genetic maps of linkage group 2 for Constantí (CLG2) and Tarraco (TLG2). The QTL regions that affect shell hardness, in-shell weight, shell weight and kernel weight are shaded in light grey.

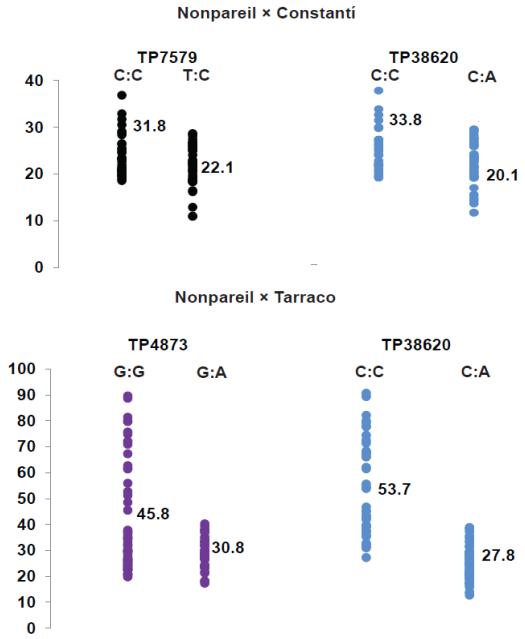


Fig. 5.3.13 Shell hardness percentages and their means for groups of Nonpareil × Constantí F_1 progeny (upper panel) and Nonpareil × Tarraco F_1 progeny (lower panel) defined based on their genotypes at markers in QTL regions for shell hardness detected on LG2.

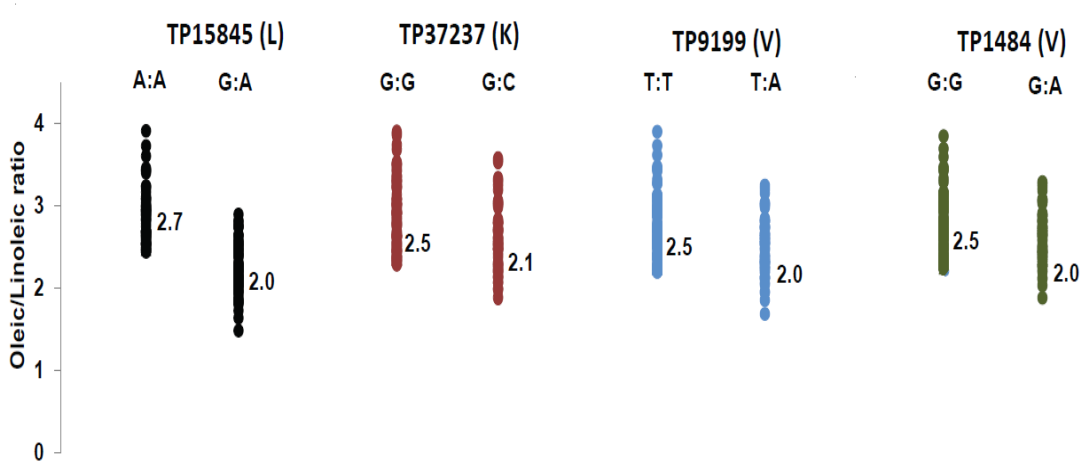


Fig. 5.3.14 Oleic/Linoleic ratios and their means for groups of Nonpareil × Lauranne F_1 progeny defined based on their genotypes at markers in QTL regions O/L ratio detected on LG1 (L and K) and LG6 (V).

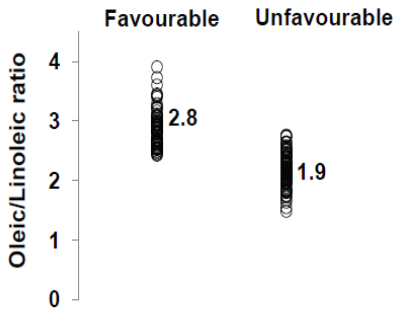


Fig. 5.3.15 Oleic/Linoleic ratios and their means for groups of Nonpareil × Lauranne F_1 progeny selected (favourable) to have high O/L ratio (> 2.5) based on genotypes at TP15845, TP37237, TP9199 and TP1484 markers in comparison to those not selected (unfavourable).

5.3.7 Discussion

A considerable difference detected for kernel weight and in-shell weight in Nonpareil × Lauranne between 2006 and 2007 might be partially explained by the alternate bearing cycle that can occur in almond. Some almond trees adopt an alternate bearing cycle (an on-crop/off-crop cycle) that results in a large crop of small almonds in one year, followed by small crop of large almonds the next year. Adverse climatic conditions, over-pruning, under-fertilisation and water-deficit stress during bloom

and fruit set can induce this phenomenon. For Lindsay Point in 2006/2007, the driest months were reported from December 2006 to March 2007 and this could have influenced fruit development.

Of three types of tocopherol (α-, β- and γ-) detected in the progeny of Nonpareil × Lauranne, the predominance of α-tocopherol (mean 97%) is consistent with what has been reported previously for almond (Kodad et al. 2009a; López-Ortiz et al. 2008; Zhu et al. 2015). The amount of β-tocopherol detected for Nonpareil × Lauranne was very low, (mean 1.8%). and its immediate biosynthetic precursor, δ-tocopherol, was not detected at all. If δ-tocopherol was present, as has been reported for some other almond materials (Font i Forcada et al. 2012), it must have been below the detection limits of the methods used here. Despite considerable variation in tocopherol concentration, only limited variation was observed in tocopherol composition in Nonpareil × Lauranne progeny.

The Nonpareil × Lauranne population showed considerable variation in fatty acid concentration. Although, high percentages of oleic acid and linoleic acid and the presence of palmitic and stearic acids were similar to what has previously been reported for almond (Kodad et al. 2006; Kodad et al. 2011; Zhu et al. 2015), the fatty acid composition detected for progeny of Nonpareil × Lauranne was novel in that substantial amounts of vaccenic (up to 3%), α-linoleic (up to 2%) and eicosapentaenoic acids (up to 5%) were present. The fatty acid composition in Nonpareil × Lauranne progeny remained fairly constant relative to the total fatty acid concentrations.

The oleic/linoleic ratio is considered as an important indicator of almond kernel quality, with high O/L ratios considered desirable. High O/L ratio ($\geq$ 2.5) has the ability to prolong shelf life of almond by preventing oil rancidification during storage and transportation (Kodad and Socias i Company 2008). Of 180 progeny evaluated here, 83 (46%) had O/L ratios above 2.5, and the maximum O/L ratio observed here was 3.9. High O/L selections from Nonpareil × Lauranne might be useful in breeding cultivars with high O/L ratio.

Although QTL mapping has been previously conducted for most of the traits considered here (Arús et al. 1998; Fernández i Martí et al. 2013; Sánchez-Pérez et al. 2007), no QTL analysis has been previously reported for these traits for Nonpareil, which is the major almond cultivar in both the USA and Australia, nor for Lauranne.

The QTLs that were detected for some of the traits in this research had previously been mapped on the same linkage group. For example, a shell hardness QTL that was detected for Lauranne on LG2 is at the same position as one mapped in R1000 × Desmayo Largueta (R × D) population (Sánchez-Pérez et al. 2007). Anchoring of the marker UDP402b, confirmed that the position of the previously mapped QTL corresponds with the genomic region N in Lauranne.

Some QTLs were detected on the same linkage groups as previously mapped QTLs (Fernández i Martí et al. 2013). Those were QTLs for geometric diameter in Nonpareil on LG1, kernel length/width and thickness/length on LG2, kernel width on LG3, kernel length on LG6 and in Lauranne kernel length and kernel thickness/length on LG1.

Of kernel traits mapped here, shell weight, kernel shape, total tocopherol concentration, total fatty acid concentration and oleic/linoleic ratio were mapped for the first time in almond. Shell weight and kernel shape QTLs were mapped very closely with other known QTLs. The shell weight QTL was mapped close to the shell hardness and in-shell weight QTLs on LG2 in Nonpareil, Lauranne, Tarraco and Constantí. The kernel shape QTL was mapped with kernel length and length/width QTLs for Nonpareil and Lauranne on LG2 and LG1, respectively.

For shell hardness, QTLs have previously been reported on LG2 (Sánchez-Pérez et al. 2007) and on LG8 (Arús et al. 1998) using populations that exhibited variation from hard to soft shells. Here, with the Nonpareil x Lauranne population, which exhibited a wider range of variation, from paper shell (shell hardness ≥ 55%) to stone shell (shell hardness ≤ 25%), shell hardness QTLs were detected

on LG2 (Lauranne), LG5 (both Nonpareil and Lauranne) and LG8 (Lauranne). In Lauranne, shell hardness was affected by six QTLs. The most significant of these was located at 40 cM on LG2, at which high shell hardness percentages was associated with homozygosity for marker alleles (i.e., the same genotype as the paper shell parent (Nonpareil). However, to ensure shell hardness percentages above 55, homozygosity for the markers detected on LG2, LG5 and LG8 are important (similar to Nonpareil).

For Tarraco and Constantí, shell hardness was mapped at a similar position (close to the marker TP38620) on LG2 as for Lauranne. In all three populations, shell hardness QTLs were at or near loci affecting in-shell weight, shell weight and kernel weight QTLs, indicating a possible common genetic basis for variation in these traits. In Lauranne, O/L ratio and kernel length QTLs were mapped to the same position suggesting a genetic association between chemical and physical traits. This was similar to the genetic association that has been reported between kernel length and tocopherols for almond (Font i Forcada et al. 2015).

Based on these results, markers TP23628, TP15987, TP38620, TP29814 and TP23837 seem useful for marker-assisted selection for recovery of the paper shell trait of Nonpareil. The utility of these markers could be further validated in other populations in the Nonpareil nested association mapping population (Section 5.2). Suitable populations include Nonpareil × Mandaline, Nonpareil × Marta and Nonpareil × 12-350, as genotyping of the parents and small subsets of the progeny (data not shown) has confirmed that Mandaline, Marta and 12-350 are all heterozygous for TP38620 and 12-350 is heterozygous for TP27036 and TP23837.

Marker-assisted selection for high O/L ratio can be done using the markers TP15845, TP37237, TP9199 and/or TP1484. The applicability of these markers could be further investigated using other

populations in the Nonpareil nested association mapping population. Nonpareil × 12-350 is suitable for this as 12-350 is heterozygous for TP15845.

Of seven genomic regions that were detected in Nonpareil and Lauranne for fatty acid concentration, tocopherol concentration and O/L ratio, some coincide with or close to regions of the peach genome where there are genes that are thought to be involved in fatty acid and lipid biosynthesis. Of nine putative fatty acid biosynthesis initiation genes on scaffold 1 (Pp1) of the peach genome assembly, *ppa006689m*, *ppa004787m* and *ppa006513m* are close to a QTL (L) that affected oleic/linoleic ratio in Nonpareil × Lauranne progeny.

The genes *ppa009593m* and *ppa001641m*, which were anchored to scaffold 3 (Pp3), are very closely located to the fatty acid concentration QTL that was detected for Lauranne on LG3. These genes encode acetyl-CoA synthetase, which initiates the fatty acid initiation from acetyl-CoA and NADPH to produce palmitic acid (16:0) and linoleic acid (18:2). Aligning the sequence of *ppa009593m,* which is anchored between 19,136,301 and 19,128,411 bp of the peach genome with the SNP-bearing GBS tags, showed that there is a GBS tag (TP8464-CAGCCA[A/T]GGAATAATGACAAAAAGTCAAGCATACAATTCCAATAGAAGGTGGAGCTAACAAC CT) that aligned to the respective region of the peach scaffold 3. In addition, mapping this gene sequence with the Nonpareil sequence, which is used to obtain the sequence information when designing KASP primers for Nonpareil and Lauranne, indicated that there is a region in the Nonpareil sequence with a 98% sequence similarity to this gene. This information can use to design primers to resequence the *ppa009593m* gene from Nonpareil and Lauranne that could lead to the identification of sequence polymorphisms for investigation as candidates for the polymorphisms responsible for QTL effects.

The gene *ppa004053m* which were anchored to scaffold 4 (Pp4), was very closely located to the total tocopherol concentration QTL that was detected for Lauranne on LG4. This gene regulates the

synthesis of biotin carboxylase and acetyl -CoA carboxylase, which catalyses the committed step in fatty acid biosynthesis by producing malonyl-CoA. It is important for synthesis of the saturated straight chain fatty acid of stearic acid (18:0) and palmitic acid (16:0). Tocopherols play a major role in protecting the fatty acids from being peroxidised. Aligning the sequence of *ppa004053m* that is anchored between 20,508,795 and 20,515,423 bp of the peach genome, indicated that the GBS tag TP1806 (CAGCAAGCATATCTAGGTGGTAGTGCATTCCAGCAGTCC[A/G]TGGCTGGCATGGATT ACAACTTTC) aligned to the respective region of the peach scaffold 4. Mapping the gene sequence against the Nonpareil genomic sequence provided a segment, which has 98% sequence similarity to the *ppa004053m* gene. It is possible to design primers using this information to resequence this gene from Nonpareil and Lauranne. This could also lead to the identification of sequence polymorphisms for investigation as candidates for the polymorphisms affecting tocopherol concentration QTL effects.

While there are no published reports on genes involved in tocopherol biosynthesis in almond, it is known that in *Arabidopsis thaliana*, the tocopherol biosynthetic pathway is controlled by five genes (*VITE1-5*) (DellaPenna 2005; Gilliland et al. 2006). On the peach genome sequence assembly, *VITE1* (*ppa004789m*) and *VITE2* (*ppa006544m*) were both anchored to scaffold 1 (Pp1) and in *Arabidopsis thaliana*, *VITE4* and *VTE5* were mapped to LG1 and LG5, respectively (Gilliland et al. 2006). However, no QTL that affected tocopherols were detected on LG1 or on LG5 for Nonpareil or Lauranne.

Oleic/linoleic ratio varied substantially among the Nonpareil × Lauranne progeny and three QTLs were detected for this trait. The most significant one located at 69 cM on LG1 in Lauranne. At this locus homozygous allele combination was associated with higher O/L ratios. The other two QTLs (on LG1 and on LG6) had smaller effects but guaranteed opportunities for useful gains.

Most traits (shell hardness, in-shell weight, shell weight, kernel length, oleic/linoleic ratio, total tocopherol concentration and total fatty acid concentration) analysed here were affected by more than one genomic region, and so many genes can be involved. Another year of phenotyping using Nonpareil × Lauranne is recommended to confirm the QTLs identified for chemical traits. Since shell hardness is important in almond breeding, the QTL detected here for shell hardness on LG5 and LG8 are particularly interesting. Further validation of these QTL can easily be done by repeated phenotyping of Nonpareil × Lauranne. Phenotyping complete populations form Nonpareil × Constantí and Nonpareil × Tarraco would be advisable as in this analysis used phenotypic data from a fraction of these populations and/or by analysing QTLs in populations that are segregating from hard shell to paper shell (Nonpareil × 12-350 (hard shell), Nonpareil × Mandaline (hard shell) and Nonpareil × Marta (hard shell)) in the University of Adelaide almond breeding program would be able to confirm these QTLs.

This research was conducted in two regions (NSW and VIC) of Australia, which have similar environmental conditions. Some of these traits may be more affected by environmental conditions than other characters. Chemical traits can be highly affected both by genotype and by environment (Kodad et al. 2006; Zhu et al. 2015). In this research, QTL analysis was not conducted for the components in fatty acids and tocopherols, since Nonpareil × Lauranne progeny did not show much compositional variation in tocopherols and fatty acids. There are experimental evidences that some of the cultivars (Somerton, Johnston and Chellaston) that used as parents in the crosses made by the University of Adelaide almond breeding program had significantly different tocopherol and fatty concentrations to Nonpareil (Zhu et al. 2015). Phenotyping and genotyping progeny of Nonpareil × Somerton, Nonpareil × Johnston and Nonpareil × Chellaston crosses in the University of Adelaide almond breeding program could confirm or validate the previously identified QTLs and possibly lead to the identification of novel QTLs.

Although the QTL mapping for chemical traits reported here was conducted using phenotypic data from only one year, the results seem promising in that a few QTL regions that affect fatty acids and tocopherols coincide with previously published QTL for almond and/or the positions of candidate genes in the peach genome sequence assembly.

Breeding for kernel quality is a challenging task in almond breeding. Some physical traits (shell hardness) and chemical traits (tocopherols and fatty acids) are affected by many QTL regions and some genes may interact with each other. Kernel quality can be improved by combining desirable traits. Some molecular markers identified here could be used in marker-assisted selection. Upon validation these markers can be used for selection at the seedling stage, so that the plants retained have a high probability of carrying the desirable traits. This could result in significant savings of resources in a breeding program.

CHAPTER 6

General discussion

Almond [*Prunus dulcis* (Mill.) D.A. Webb] is highly heterozygous due to its outcrossing nature. Its perennial growth habit, large tree size and long juvenile period hinder both genetic improvement and genetic analysis. Until recently, a limited amount of sequence information had been generated for almond. In addition, the genetic research required to discover marker-trait associations is hampered by the same phenological factors that hamper almond breeding. Therefore, discovery and characterisation of polymorphisms in the almond genome could facilitate genetic analysis and breeding. The research reported in this thesis employed next-generation sequencing (NGS) technology to resequence the self-incompatibility (*S*) locus and to skim sequence the almond genome using genotyping-by-sequencing approach to discover a large number of single nucleotide polymorphisms. Further, QTLs that affect physical and chemical traits were mapped for almond.

In this thesis research, complete *S* locus sequences were generated for a self-compatible haplotype (S_f) and for two additional self-incompatible haplotypes (S_1 and S_8). It verified the published S_7 haplotype sequence (AB081587) and refined it by determining several previously undetected nucleotides. In addition, partial *S* locus sequences were generated for 11 *S* haplotypes (S_3, S_5, S_6, S_9, S_{13}, S_{14}, S_{19}, S_{22}, S_{23}, S_{25} and S_{27}). Comparisons of the *S* locus sequences from all these haplotypes indicated that they are highly polymorphic in some regions and that they have variable numbers of LTRs. The locations of the LTRs in the *S* locus were conserved across most haplotypes. Variable numbers of LTRs were detected for these *S* haplotypes. Consistent with the prediction of Ushijima et al. (2003) that LTRs could be sources of variation in the almond *S* locus, this research demonstrated that LTRs provide considerable variation to the *S* locus. Long terminal repeats were

detected in all haplotypes except S_3, S_{13}, S_{14}, S_{19} and S_{27}. The number of LTRs detected ranged from one (S_5, S_8, S_{22} and S_{25}) to four (S_1). All these LTRs belong to *Ty1/copia*-like retrotransposons. In this research, although a high variability in LTRs was detected between S_f and SI haplotypes and among SI haplotypes, neither functional analysis nor comprehensive sequence analysis was conducted. Given previous reports on the involvement of transposable elements in SI breakdown in other *Prunus* species (Halász et al. 2014; Hauck et al. 2006; Tao et al. 2007; Yamane et al. 2003a), it is possible that comprehensive sequence and/or structural analysis using the sequence information generated here for individual LTRs could provide insights on why the number of LTRs varies among *S* haplotypes and on their functions in *S* locus.

The highly polymorphic regions of the almond *S* locus harbour the *S-RNase* and *SFB* genes. Comparison of 15 *S* alleles indicated presence of three variable regions in the S-RNase gene. These were detected between conserved region 1 and conserved region 2, at the beginning of the RHV and at the beginning of the RC4 region. In the *SFB* gene, high variability was detected in the F-box motif with small conserved regions within it. These variations may facilitate the specific binding of the F-box protein to non-self S-RNases for ubiquitination. In the *SFB* gene, two short variable regions were observed in between the V1 and V2 regions. This information provides some evidence to reconsider the labelling of these characteristics for the *S-RNase* and *SFB* genes.

Highly significant sequence variation was observed in the *S-RNase* and *SFB* genes. The variation in the *S-RNase* gene was mainly due to sequence variation in the coding regions (C2, C3, RC4 and C5) and length polymorphisms in intron regions, particularly in the second intron. Among sequence variations observed in conserved regions, a single nucleotide polymorphism (SNP) detected in conserved region 2 (C2) was particularly interesting. At this position, the amino acid isoleucine was detected in the S_f allele while the amino acid leucine was present in SI alleles. This SNP provided a basis for designing robust markers to differentiate the S_f progeny from SI progeny. Further, sequence information generated here enabled the discovery of useful polymorphisms in C3 (for the

alleles S_1 and S_8) and RC4 (for the alleles S_3, S_9 and S_{25}) regions in the *S-RNase* gene and in the *SFB* gene (for the allele S_3) to differentiate the *S* alleles in almond. The *SLF* gene, which marks one boundary of the *S* locus, is intron-less and is less polymorphic among these haplotypes.

While the relative order and transcriptional orientation of the *S-RNase* and *SFB* genes were conserved among these 15 *S* haplotypes, considerable length polymorphisms were observed in the intergenic region between the *S-RNase* and *SFB* genes, and this region from most of these alleles was highly rich in A and T nucleotides. Not only length polymorphisms, also nucleotide composition in the intergenic region could help in preventing recombination between the *S-RNase* and *SFB* genes and to maintain complete linkage between them. There is not much research reported on intergenic region between the *S-RNase* and *SFB* genes in almond. The sequence and structural variations observed here could lead to interesting discovery related to functional variations and maintenance of specificity in almond *S* alleles.

This research generated new sequences, which could be particularly useful in expanding the sequence availability for almond research and breeding. Among these are *SLF* gene sequences from 14 alleles (S_1, S_3, S_5, S_6, S_8, S_9, S_{13}, S_{14}, S_{19}, S_{22}, S_{23}, S_{25}, S_{27} and S_f), the three *SFB* gene sequences (from the S_9, S_3 and S_{27} alleles) and the *S-RNase* gene sequences that are more complete than previously available from the S_9, S_3 and S_{27} alleles.

A high level of variation within almond *S* alleles was also inferred by the phylogenetic relationship analysis, using deduced amino acid sequences between conserved region 1 (C1) and conserved region 5 (C5) regions in 15 *S-RNase* gene sequences (S_1, S_3, S_5, S_6, S_7, S_8, S_9, S_{13}, S_{14}, S_{19}, S_{22}, S_{23}, S_{25}, S_{27} and S_f) and the S_5 allele, clustered separately from other SI alleles and the S_f allele. Nevertheless, the phylogenetic relationship analysis using deduced amino acid sequences from *S-RNase* alleles of almond, other *Prunus, Malus, Pyrus, Antirrhinum* species and Solanaceae indicated similarity of the *S-RNase* alleles from almond with those from other *Prunus* species, and divergence

of these alleles from orthologous *S-RNases* from the *Malus, Pyrus, Antirrhinum* species and the Solanaceae. The phylogenetic relationship analysis using the almond *SFB* alleles also indicated distant relatedness of the S_5 and S_f allele to other SI alleles.

While resequencing the *S* locus, the main difficulty encountered was failure of amplification in some regions of the *S* locus from some haplotypes. Here, attempts were taken to amplify the two haplotypes in a sample using common primers. In most cases, length polymorphisms were observed indicating amplification of both haplotypes. In a few cases, no size differences were detected, precluding separate barcoding of each haplotype. This led to gaps and uncertainties in the sequence, complicating downstream data analysis and requiring the use of several bioinformatics tools to establish a final sequence. For the regions where the *S-RNase* and *SFB* genes are located, gaps in the sequence were filled by preparing a second sequencing library using amplicons obtained from allele-specific primers. In this research, tools for the analysis of reference sequences (BWA) and *de novo* sequences (MIRA and Velvet) were implemented to obtain sequences for those haplotypes. Each analysis resulted in the same sequences. In occasional cases in which the identity of a nucleotide at a particular position differed among individuals that were expected to carry the same haplotype, the more frequently detected nucleotide was regarded as the correct nucleotide.

The sequencing library preparation strategy used here could be applied to sequence any other *S* haplotype from almond or any other self-incompatible crop species or to resequence complex multi-allelic loci from any species.

The *S* locus sequences generated here provided valuable information that was exploited for *S* allele detection in the almond *S* locus. About 50 *S* alleles have been identified for almond. Most of these confer self- incompatibility (SI), but at least one (S_f) confers self-fertility. In almond orchards, self-incompatibility necessitates inclusion of polliniser trees with the main cultivar and the use of insect pollinators to obtain fruit set. Therefore, some almond breeders attempt to develop self-fertile

cultivars that have no requirement for pollinisers and are less reliant on insect pollinators. Molecular detection of *S* genotypes in almond and *Prunus* species has been widely implemented, using markers that detect polymorphism in *S-RNase* gene sequences (Channuntapipat et al. 2001; Channuntapipat et al. 2003; Ma and Oliveira 2002; Ortega et al. 2005; Sánchez-Pérez et al. 2004). While these markers can distinguish among some SI alleles, they cannot distinguish the S_f allele from all the other SI alleles. Furthermore, some of these markers generate multiple-band artefacts due to improper annealing of PCR fragments (Hanada et al. 2009) and/or, amplicons of similar size from two or more alleles (Sánchez-Pérez et al. 2004). Additionally, as they are gel-based markers, they require significant investments of time from skilled personnel. To overcome above–mentioned limitations, assays were developed based on single nucleotide polymorphisms (SNPs) discovered from the sequences information generated for the *S-RNase* and *SFB* genes.

Additional SNP-based assay development strategies were implemented in this research, as very few sequences were found to be useful for designing standard three-primer SNP-based assays using KASP™ technology. One of these, provided presence/absence primers to detect the relevant *S* alleles using allele-specific primer pairs, and the other strategy involved the use of degenerate primers to enable discrimination of target alleles from several other *S* alleles. In some cases, a common third primer made it possible to amplify the products from several *S* alleles. Each of the markers designed in this research differentiated its target *S* allele from each of the other *S* alleles considered here. In some cases, markers could also detect which other *S* allele is present in combination with the target *S* allele. Depending on which allele-specific product was amplified, the PCR product would emit FAM fluorescence, HEX fluorescence or both. These assays provided high-throughput and efficient *S* allele detection for almond.

A high level of genetic variation in the almond *S* locus called into question the suitability of one set of primers to differentiate a target allele from all other alleles. Therefore, a series of markers that could differentiate the target from other *S* alleles was designed. The S_f series of primer sets made it

possible to differentiate the S_f allele from several SI alleles. This research demonstrated that both the *S-RNase* and *SFB* genes could be used for *S* allele detection in almond. Assays that were designed based on allele-specific SNPs provided robust methods to distinguish the alleles of interest readily from rest of the alleles. In the research conducted for this thesis only nine *S* alleles were considered among the 50 *S* alleles identified for almond. In the future, by using the procedure followed here, high-throughput assays can be designed for other *S* alleles.

In this thesis research, SNPs in the almond genome were used to construct linkage maps. Until now, to generate almond linkage maps restriction fragment length polymorphism (RFLP) markers (Joobeur et al. 1998), simple sequence repeat (SSR) markers (Aranzana et al. 2003; Dirlewanger et al. 2004), RAPD markers (Joobeur et al. 2000) inter-simple sequence repeat (ISSR) markers and sequence characterised amplified region (SCAR) markers (Tavassolian et al. 2010) have been used. Only limited application of SNP markers has been reported for linkage map construction in almond (Wu et al. 2010; Wu et al. 2009). The next-generation sequencing approach using genotyping-by-sequencing (GBS) was applied to Nonpareil × Lauranne F_1 progeny to discover SNPs in the almond genome. A published GBS protocol (Elshire et al. 2011) was slightly modified for almond. Use of the *Ape*KI restriction enzyme to digest the almond genome enabled discovery of hundreds of SNP-bearing GBS tags providing an easy and accurate method to discover and assay SNPs without any knowledge of the almond genome sequence.

Although aligning the unique reads obtained for almond to the peach genome indicated a complete sequence coverage of the almond genome, the number of SNP-bearing tags obtained here was low compared to the tens of thousands of GBS markers that have been reported for some other species (Elshire et al. 2011; Hyma et al. 2013; Li et al. 2015; Lu et al. 2013; Mitchell et al. 2012) and was similar to what has been obtained for peach (Bielenberg et al. 2015) and sweet cherry (Guajardo et al. 2015), using the *Ape*KI restriction enzyme. *In-silico* analysis conducted here based on the peach genome assembly indicated that a larger number of tags might have been obtained using a

combination of *Ape*KI and *Hpa*II enzymes, which could have provided a larger number of digested fragments. An even higher number of unique tags can be obtained through increasing sequencing depth.

In this research, individual genetic maps for Nonpareil and Lauranne were constructed using Nonpareil × Lauranne F_1 progeny genotypic data. Generally, genetic maps in plants have been constructed from crosses between inbred lines that are genetically different or by exploiting the genetic information from two or three filial generations. Such populations are not available and are difficult or impossible to obtain for almond due to perennial growth habit and highly heterozygous nature. Therefore, populations of almond generally involve only two parents and their F_1 progeny. Linkage map construction usually involves separate genetic maps for parents using individual-specific markers and an integrated map using common markers. Here, the linkage map construction was performed for Nonpareil and Lauranne separately, considering DNA polymorphisms present in one parent and absent in the other.

Although a stringent read depth cut-off value (4, compared to the value of 3 that is generally used in filtering SNP datasets) was implemented in this research, the total lengths of the initial linkage maps constructed using GBS data for Nonpareil and Lauranne were 1,152.1 cM and 1,371.3 cM, respectively, about twice as long as previously published Nonpareil × Lauranne maps (Tavassolian et al. 2010). This discrepancy was mainly due to erroneously called genotypes in the GBS analysis. In most cases, these would be heterozygotes that were called as homozygotes in the GBS dataset. To rectify these issues, SNP-bearing sequences from GBS analysis were used to design uniplex marker assays based on 138 tag pairs that were heterozygous in Nonpareil and homozygous in Lauranne and 155 tag pairs that were heterozygous in Lauranne and homozygous in Nonpareil. Application of these assays to an additional 231 Nonpareil × Lauranne progeny improved the linkage maps for Nonpareil and Lauranne reducing their total lengths to 608.9 cM and 658.7 cM, respectively. Although these maps were similar in length to previously published Nonpareil × Lauranne maps

(Tavassolian et al. 2010), both the Nonpareil and Lauranne maps have large gaps of about 20 cM in some linkage groups (e.g., LG3 in Nonpareil and LG6 in Lauranne). In the future, these gaps can be filled by designing uniplex SNP assays using GBS tags that mapped to the relevant scaffolds of the peach genome.

Anchoring of unique reads obtained for almond to the peach genome assembly confirmed the expected high similarity between the almond and peach genomes and hinted where highly structural and genomic differences could be present between the almond and peach genomes. Sixty-seven percent of unique tags from Nonpareil × Lauranne population were distributed across the first eight scaffolds and pseudomolecules of the peach genome sequence assembly with some variation in read depth. There were a few regions with unusually high read depth in Nonpareil and Lauranne that mapped to the peach scaffold 5 at 6.5 Mbp and scaffold 7 at 0.5 Mbp. Comprehensive analysis of these regions could provide information on copy number polymorphisms in the almond genome and hinted at where high levels of genome structure variation could be present. The 32% of GBS tags that were not mapped to the peach genome could be useful in investigating the extent of structural divergence between the almond and peach genomes or the possibility of contamination in the almond genome by the DNA from other species.

The GBS-based genetic maps of Nonpareil and Lauranne provided sources of sequence polymorphisms for the design of uniplex fluorescence-based PCR assays. These assays were useful in detecting polymorphism in not only Nonpareil × Lauranne, but also Nonpareil × Constantí, Nonpareil × Tarraco and Nonpareil × Vairo. This led to the construction of a composite map using the genotypic data from all four populations and was useful in resolving marker order of some markers that had collocated in maps from individual populations. This is the first almond linkage map constructed based on multi-parental genotypic data. In the future, this map could be improved by adding more markers obtained by screening other Nonpareil crosses in the University Adelaide

almond breeding program and it can be used as a reference for ordering markers in other almond populations.

In addition, individual linkage maps were generated for Constantí, Tarraco and Vairo using SNPs that were mappable from the assays that were designed based on Lauranne heterozygosity and Nonpareil homozygosity. Variable numbers of markers were mapped to these populations and these maps maximised the use of available genetic resources for linkage map construction. However, markers were sparsely distributed on most linkage groups within each of these maps and some linkage groups were much shorter than those that were developed for Nonpareil and Lauranne. Improvement of the linkage maps of Constantí, Tarraco and Vairo can be obtained through development of additional markers that are heterozygous in these parents but homozygous in Nonpareil. Application of genotyping-by-sequencing to these populations could enable the discovery of additional SNPs that could be exploited to improve the linkage maps.

Although considerable variation was observed for physical traits and oleic/linoleic ratio in the Nonpareil × Lauranne population, no large-scale variations were observed for content or composition of tocopherol. Of three types of tocopherol (α-, β- and γ-) that were detected in kernels from the Nonpareil × Lauranne F_1 progeny, α-tocopherol was the most abundant (97%). The major fatty acids detected in the progeny of Nonpareil × Lauranne were oleic acid, linoleic acid, palmitic acid and vaccenic acid. Oleic acid concentration in the progeny of Nonpareil × Lauranne ranged from 0.41 to 0.66.

For Nonpareil and Lauranne, QTL analyses were conducted for 13 kernel and nut traits and three chemical traits. A total of 23 genomic regions were detected as affecting nut and/or kernel traits in Nonpareil and Lauranne. Nine and 14 QTLs were detected for Nonpareil and for Lauranne, respectively. In these two maps, some of the regions affected only chemical traits, some affected only physical traits and one affected both physical and chemical traits. Many traits were affected by

several small–effect QTLs and some traits were affected by one large–effect QTL and several small-effect QTLs. For Nonpareil and Lauranne, in most cases, QTLs were detected in the same linkage group but at different positions. Some loci affecting kernel physical traits were co-located with the loci affecting chemical traits suggesting a common genetic basis for these traits. In Lauranne, QTLs for oleic/linoleic ratio and kernel length were detected at the same position on LG, suggesting a common genetic basis for these traits.

The research presented here reports QTLs for nut or kernel traits in Nonpareil, which is the predominant almond cultivar grown both in Australia and the USA for the first time. Except shell weight, kernel shape, total fatty acid concentration, total tocopherol concentration and oleic/linoleic ratio, many of the traits mapped for Nonpareil and for Lauranne have been previously mapped for other parents (Arús et al. 1998; Fernández i Martí et al. 2013; Font i Forcada et al. 2012; Sánchez-Pérez et al. 2007). Some QTLs detected in this research were mapped to the same linkage group as previously published QTLs for the same or similar traits. A shell hardness QTL that was detected for Lauranne on LG2 has previously been mapped on LG2 is at the same position as one mapped in R1000 × Desmayo Largueta (R × D) population (Sánchez-Pérez et al. 2007). Of QTLs mapped, shell hardness and oleic/linoleic ratio QTLs were particularly interesting. For shell hardness, the Nonpareil x Lauranne population exhibited a wider range of variation, from paper shell (shell hardness ≥ 55%) to stone shell (shell hardness $\leq$ 25%) and three QTLs were detected for shell hardness: on LG2 (Lauranne), LG5 (both Nonpareil and Lauranne) and LG8 (Lauranne). Previously, QTLs have been reported on LG2 (Sánchez-Pérez et al. 2007) and on LG8 (Arús et al. 1998) using populations that exhibited variation from hard to soft shells. Shell hardness, shell weight, in-shell weight and kernel weight were mapped at a similar position (close to the marker TP38620) on LG2 for Tarraco, Constantí and Lauranne. Based on the genotypic and phenotypic results obtained here, the paper shell trait seems to require homozygosity at the QTL on LG2, LG5 and LG8. For marker-assisted selection, the markers TP23628, TP15987, TP38620, TP29814 and TP23837 are capable of ensuring recovery of the paper shell trait (shell hardness percentage $\geq$ 55%).

Further validation of the QTLs on LG5 and LG8 for shell hardness, could be done by repeated phenotyping of Nonpareil × Lauranne F_1 progeny. These QTLs can be further confirmed/validated by adding more markers to the respective QTL regions on LG5 and LG8 in the Constantí and Tarraco maps and/or by analysing QTLs in populations that are segregating from hard shell to paper shell (Nonpareil × 12-350 (hard shell), Nonpareil × Mandaline (hard shell) and Nonpareil × Marta (hard shell)) in the University of Adelaide almond breeding program.

Some previous genetic analyses have treated shell hardness as a qualitative character that is affected by a major gene (Grasselly and Crossa-Raynaud 1980; Sánchez-Pérez et al. 2007) while others have treated it as a quantitative character with high heritability (Arús et al. 1998). In the research conducted here, the phenotypic variation for this trait was quantitative, and several QTLs were mapped on LG2, LG5 and LG8, The QTL detected on LG2 seems to have a lager effect than others.

This research provides sequence information that is useful in isolating the genes or components of these genes that are involved in fatty acid and tocopherol concentration in almond. Of seven genomic regions detected in Nonpareil and Lauranne for fatty acid concentration, tocopherol concentration and O/L ratio, some QTLs coincide with or are close to regions of the peach genome where there are genes that are thought to be involved in fatty acid and lipid biosynthesis. The gene *ppa009593m,* which was anchored to scaffold 3 (Pp3), is very close to the fatty acid concentration QTL that was detected for Lauranne on LG3. BLASTN search of *ppa009593m* sequence using the SNP-bearing GBS tags generated in this research, revealed that there is one tag (TP8464-CAGCCA[A/T]GGAATAATGACAAAAAGTCAAGCATACAATTCCAATAGAAGGTGGAGCTAACAACCT) that aligned to a position within the gene. An assay designed for this SNP could be useful for further investigation of the collocation of the gene with the QTL. In addition, aligning the *ppa009593m* sequence to the unpublished Nonpareil genomic sequence indicated that there is a region in the almond genome with 98% sequence similarity to this gene. This information could be

used to design primers to amplify the region from Nonpareil and Lauranne for resequencing. This could lead to the identification of sequence polymorphisms for investigation as candidates for the polymorphisms responsible for QTL effects.

The gene *ppa004053m*, which was anchored to scaffold 4 (Pp4), is very close to the total tocopherol concentration QTL that was detected for Lauranne on LG4. BLASTN analysis of the *ppa004053m* sequence, using the SNP-bearing GBS tags generated in this research, showed that TP1806 (CAGCAAGCATATCTAGGTGGTAGTGCATTCCAGCAGTCC[A/G]TGGCTGGCATGGATTACAACT TTC) tag is aligned close to the gene. An assay designed for this SNP could be useful for further investigation of the collocation of the gene with the QTL. In addition, aligning the *ppa004053m* sequence against the unpublished Nonpareil genomic sequence indicated that there is a region in the almond genome with 98% sequence similarity to this gene. This information could be used to design primers to amplify the region from Nonpareil and Lauranne for resequencing. This could lead to the identification of sequence polymorphisms for investigation as candidates for the polymorphisms responsible for QTL effects.

In Lauranne, oleic/linoleic (O/L) ratio seemed to be affected by three small-effect QTLs. Of three QTLs detected for O/L ratio, two were on LG1 and one was on LG6. Based on the genotypic and phenotypic results obtained here, high O/L ratio seems to require homozygosity at the QTL detected at 69cM on LG1.

Implications for future breeding

The assays developed for *S* allele detection and markers detected close to the shell hardness and O/L ratio QTLs in this thesis research can be implemented in screening juvenile plants and to save resources in breeding programs.

Assays to differentiate the *Sf* allele from the SI alleles

There are five assays (WriPd*Sf*1, WriPd*Sf*3, WriPd*Sf*4, WriPd*Sf*5 and WriPd*Sf*6) that are suitable to identify Sf progeny from SI progeny. Some of these assays can identify which allele combination is present in an individual.

Assays to differentiate among the SI alleles

There are twelve assays (WriPd*S1,* WriPd*S3*1, WriPd*S3*2, WriPd*S3*3, WriPd*S5*, WriPd*S7*1, WriPd*S7*2, WriPd*S8*, WriPd*S9*, WriPd*S23*, WriPd*S25*1 and WriPd*S25*2) can be used for high-throughput detection of *S* alleles in almond.

Markers to apply for marker-assisted selection

There were five markers detected close to shell hardness QTLs that were detected on LG2 (TP7579 and TP38620), LG5 (TP29814and TP24992) and LG8 (TP23837 and TP27036). The markers TP23628, TP15987, TP38620, TP29814 and TP23837 seem beneficial for marker-assisted selection for recovery of the paper shell trait. Hard shells are less prone to insect attack, pathogen infections and damage by birds and they are preferred in European markets. Other markets such as India and the Middle East prefer soft or paper shells which can be easily removed. Depending on the trait that a breeder is looking for, these markers can be used to screen juvenile plants and to make decisions on which ones need to be kept and which ones to discard.

There were four markers detected for O/L ratio, two (TP15845, TP37237) on LG1 and two (TP9199 and TP1484) on LG6. Marker-assisted selection for high O/L ratio can be done using the markers TP15845, TP37237, TP9199 and/or TP1484.

CHAPTER 7

Contributions to knowledge

The significant contributions made by this thesis research to the advancement of scientific knowledge include:

Resequencing of the almond *S* locus and development of high-throughput markers to differentiate almond *S* alleles

1. Complete *S* locus sequences were obtained for self-compatible haplotype (S_f) and two self-incompatible haplotypes (S_1 and S_8). Partial *S* locus sequences were obtained for another 11 *S* haplotypes (S_3, S_5, S_6, S_9, S_{13}, S_{14}, S_{19}, S_{22}, S_{23}, S_{25} and S_{27}).

2. New sequences for the alleles from the *S-RNase* and *SFB* genes. This research generated sequences that are more complete than those that have been previously available for the S_3, S_9, S_{25} and S_{27} alleles from both the *SFB* and *S-RNase* genes.

3. Discovery of high level of sequence variation in the regions other than those that had previously been described for the *Prunus S-RNase* and *SFB* genes.

Development of markers based on different regions of the *S-RNase* gene

1. Novel alternative primer design strategies were successfully employed to design KASP marker assays for *S* alleles. Due to high variability in the *S* alleles, standard KASP primer sets were not adequate to distinguish among all alleles. Therefore, some *S* allele markers were designed by using sequences specific for the target *S* allele and degenerate allele-

specific and/or degenerate third primer to enable discriminating target alleles from several other *S* alleles.

2. Several SNP-based markers were designed for the detection of specific *S* alleles (*S_1*, *S_3*, *S_5*, *S_7*, *S_8*, *S_9*, *S_{23}*, *S_{25}* and *S_f*). These markers are suitable for high-throughput genotyping and they overcome some other limitations associated with previously available gel-based markers (e.g. masking of some alleles by other alleles, similar size amplicons for different *S* alleles and presence of spurious bands).

3. A series of markers were designed to distinguish the *S_f* allele from self-incompatible alleles. Some of these markers not only distinguish *S_f* individuals from non-*S_f* individuals, but can also identify which self-incompatible allele is present in combination with the *S_f* allele.

Application of genotyping-by-sequencing to construct linkage maps for almond and detection of quantitative trait loci for physical and chemical nut traits in almond

1. This is the first report on the use of GBS in almond to discover SNPs and to generate linkage maps.

2. Comparison of almond linkage maps with the peach genome sequence assembly confirmed the high synteny and collinearity between these two genomes.

3. Improved linkage maps were generated for both Nonpareil and Lauranne. The linkage map for Nonpareil was further refined by analysis of data from multiple crosses in which Nonpareil was a common parent.

4. This is the first report of a linkage map of almond generated from multi-parent F_1 populations. Due to the high resolution achieved by exploiting multiple populations, this map provides a reference that could be used for ordering markers in other almond linkage maps.

5. Skeleton linkage maps were developed for Constantí, Tarraco and Vairo.

6. Twenty-three genomic regions affected physical and chemical traits in almond nuts and kernels were detected for Nonpareil and Lauranne. A novel QTL was detected for shell hardness on LG5 in Nonpareil and Lauranne. A few traits (shell weight, kernel shape, total tocopherol concentration, total fatty acid concentration and oleic/linoleic ratio) were mapped for the first time in almond.

7. Confirmation and validation of the shell hardness QTL on LG2 for Lauranne using Constantí and Tarraco linkage maps close to the marker TP38620.

8. The markers TP23628, TP15987, TP38620, TP29814 and TP23837 seem useful for marker-assisted selection for recovery of the paper shell trait.

9. The markers TP15845, TP37237 and TP9199 can be applied for marker-assisted selection to detect progeny likely to have high oleic/linoleic ratio.

10. Useful sequence information was generated for amplification and resequencing genes that may affect fatty acid and tocopherol concentration in almond.

REFERENCES

ABA (2015) ABA Industry statistics. Almond Board of Australia, Australia

ABC (2014) Almond Almanac. Almond Board of California, USA

ABC (2015) Almond Almanac 2015. Almond Board of California, USA

Allard A, Bink MCAM, Martinez S, Kelner J-J, Legave J-M, di Guardo M, Di Pierro EA, Laurens F, van de Weg EW, Costes E (2016) Detecting QTLs and putative candidate genes involved in budbreak and flowering time in an apple multiparental population. J Exp Bot 67:2875-2888

Alonso J, Kodad O, Gradziel T (2012) Almond. In: Badenes ML, Byrne DH (eds) Fruit Breeding. Springer US, pp 697-728

Alonso JM, Socias i Company R (2005) Self-incompatibility expression in self-compatible almond genotypes may be due to inbreeding. J Am Soc Hortic Sci 130:865-869

Andrews S (2010) FASTQC: A quality control tool for high throughput sequence data. http:// www.bioinformatics.babraham.ac.uk/projects/fastqc/

Aranzana MJ, Pineda A, Cosson P, Dirlewanger E, Ascasibar J, Cipriani G, Ryder CD, Testolin R, Abbott A, King GJ, Iezzoni AF, Arús P (2003) A set of simple-sequence repeat (SSR) markers covering the *Prunus* genome. Theor Appl Genet 106:819-825

Arumuganathan K, Earle ED (1991) Nuclear DNA content of some important plant species. Plant Mol Biol Rep 9:208-218

Arús P, Ballester J, Jauregui B, Joobeur T, Truco MJ, de Vicente MC (1998) The European *Prunus* mapping project: Update on marker development in almond. In: Tobutt KR, Alston FH (eds) Eucarpia Symposium on Fruit Breeding and Genetics, pp 331-336

Arús P, Messeguer R, Viruell M, Tobutt K, Dirlewanger E, Santi F, Quartas R, Ritter E (1994) The European *Prunus* mapping project progress in the almond linkage map. Euphytica 77:97-100

Aydin C (2003) Physical properties of almond nut and kernel. J Food Eng 60:315-320

Baird NA, Etter PD, Atwood TS, Currey MC, Shiver AL, Lewis ZA, Selker EU, Cresko WA, Johnson EA (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers. PLoS ONE 3:e3376

Ballester J (1998) Localització i anàlisi de caràcters d'interès agronòmic de l'ametller. Tesis Doctoral, Universitat Autònoma de Barcelona, España

Ballester J, Bokovi R, Batlle I, Arús P, Vargas F, Vicente M (1998) Location of the self-incompatibility gene on the almond linkage map. Plant Breed 117:69 - 72

Ballester J, Company R, Arús P, de Vicente M (2001) Genetic mapping of a major gene delaying blooming time in almond. Plant Breed 120:268 - 270

Bayer PE, Ruperao P, Mason AS, Stiller J, Chan C-KK, Hayashi S, Long Y, Meng J, Sutton T, Visendi P, Varshney RK, Batley J, Edwards D (2015) High-resolution skim genotyping by sequencing reveals the distribution of crossovers and gene conversions in *Cicer arietinum* and *Brassica napus*. Theor Appl Genet 128:1039-1047

Beló A, Beatty MK, Hondred D, Fengler KA, Li B, Rafalski A (2009) Allelic genome structural variations in maize detected by array comparative genome hybridization. Theor Appl Genet 120:355

Bennett M, Leitch IJ (1997) Nuclear DNA amounts in Angiosperms—583 new estimates. Ann Bot 80:169-196

Betts MJ, Russell RB (2007) Amino-acid properties and consequences of substitutions. In: Barnes MR (ed) Bioinformatics for Geneticists: a bioinformatics primer for the analysis of genetic data. John Wiley & Sons, Ltd, Chichester, UK., pp 311-342

Bielenberg DG, Rauh B, Fan S, Gasic K, Abbott AG, Reighard GL, Okie WR, Wells CE (2015) Genotyping by sequencing for SNP-based linkage map construction and QTL analysis of chilling requirement and bloom date in peach *Prunus persica* (L.) Batsch]. PLoS ONE 10:e0139406

Biémont C (2010) A brief history of the status of transposable elements: from junk DNA to major players in evolution. Genetics 186:1085-1093

Bioinnovate-Africa (2014) Resource development for finger millet. Bioinnovate-Africa, Kenya

Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics

Bošković R, Tobutt K, Ortega E, Sutherland B, Godini A (2007) Self-(in)compatibility of the almonds *P. dulcis* and *P. webbii*: detection and cloning of 'wild-type S_f' and new self-compatibility alleles encoding inactive *S-RNases*. Mol Gen Genomics 278:665-676

Bošković R, Tobutt KR, Batlle I, Duval H (1997) Correlation of ribonuclease zymograms and incompatibility genotypes in almond. Euphytica 97:167-176

Bošković R, Tobutt KR, Batlle I, Duval H, Martinez-Gomez P, Gradziel TM (2003) Stylar ribonucleases in almond: correlation with and prediction of incompatibility genotypes. Plant Breed 122:70-76

Bosković R, Tobutt KR, Duval H, Batlle I, Dicenta F, Vargas FJ (1999) A stylar ribonuclease assay to detect self-compatible seedlings in almond progenies. Theor Appl Genet 99:800-810

Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES (2007) TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics 23:2633-2635

Broothaerts W, Janssens GA, Proost P, Broekaert WF (1995) cDNA cloning and molecular analysis of two self-incompatibility alleles from apple. Plant Mol Biol 27:499-511

Buermans HPJ, den Dunnen JT (2014) Next generation sequencing technology: advances and applications. BBA Mol Basis Dis 1842:1932-1941

Butler D, Cullis BR, Gilmour A, Gogel B (2009) ASReml-R reference manual. The State of Queensland, Department of Primary Industries and Fisheries, Brisbane

Cao J, Schneeberger K, Ossowski S, Gunther T, Bender S, Fitz J, Koenig D, Lanz C, Stegle O, Lippert C, Wang X, Ott F, Muller J, Alonso-Blanco C, Borgwardt K, Schmid KJ, Weigel D (2011) Whole-genome sequencing of multiple *Arabidopsis thaliana* populations. Nat Genet 43:956-963

Castresana J (2000) Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol Biol Evol 17:540-552

Channuntapipat C, Collins G, Ramesh SA, Sedgley M, Wirthensohn MG (2005) Determination of almond *S* alleles using PCR primers designed from their introns. In: M.M. O, V. C (eds) XIII GREMPA Meeting on Almonds and Pistachios. CIHEAM, Options Méditerranéennes : Série A. Séminaires Méditerranéens; , Zaragoza pp 333-339

Channuntapipat C, Sedgley M, Collins G (2001) Sequences of the cDNAs and genomic DNAs encoding the S_1, S_7, S_8, and S_f alleles from almond, *Prunus dulcis*. Theor Appl Genet 103:1115-1122

Channuntapipat C, Wirthensohn M, Ramesh S, Batlle I, Arús P, Sedgley M, Collins G (2003) Identification of incompatibility genotypes in almond (*Prunus dulcis* Mill.) using specific primers based on the introns of the *S*-alleles. Plant Breed 122:164-168

Chen T, Wu X, Chen Y, Li X, Huang M, Zheng M, Baluška F, Šamaj J, Lin J (2009) Combined proteomic and cytological analysis of Ca^{2+}-calmodulin regulation in *Picea meyeri* pollen tube growth. Plant Physiol 149:1111-1126

Chevreux B, Pfisterer T, Drescher B, Driesel AJ, Müller WEG, Wetter T, Suhai S (2004) Using the miraEST assembler for reliable and automated mRNA transcript assembly and SNP detection in sequenced ESTs. Genome Res 14:1147-1159

Conn E (1980) Cyanogenic compounds. Ann Rev Plant Physiol 31:433-451

Cook DE, Lee TG, Guo X, Melito S, Wang K, Bayless AM, Wang J, Hughes TJ, Willis DK, Clemente TE, Diers BW, Jiang J, Hudson ME, Bent AF (2012) Copy number variation of multiple genes at *Rhg1* mediates nematode resistance in soybean. Science 338:1206-1209

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, McVean G, Durbin R, Group GPA (2011) The variant call format and VCFtools. Bioinformatics 27:2156-2158

De Donato M, Peters SO, Mitchell SE, Hussain T, Imumorin IG (2013) Genotyping-by-sequencing (GBS): a novel, efficient and cost-effective genotyping method for cattle using next-generation sequencing. PLOS ONE 8:e62137

De Nettancourt D (1997) Incompatibility in angiosperms. Sex Plant Reprod 10:185-199

DellaPenna D (2005) Progress in the dissection and manipulation of vitamin E synthesis. Trends Plant Sci 10:574-579

DellaPenna D, Mène-Saffrané L (2011) Chapter 5 - Vitamin E. In: Fabrice R, Roland D (eds) Advances in botanical research. Academic Press, pp 179-227

Denisov VP (1998) Almond Genetic Resources in the USSR and their use in production and breeding. Acta Hort 224:299-306

Deshaies RJ (1999) SCF and Cullin/RING H2-based ubiquitin ligases. Annu Rev Cell Dev Biol 15:435-467

Di Pierro EA, Gianfranceschi L, Di Guardo M, Koehorst-van Putten HJJ, Kruisselbrink JW, Longhi S, Troggio M, Bianco L, Muranty H, Pagliarani G, Tartarini S, Letschka T, Lozano Luis L, Garkava-Gustavsson L, Micheletti D, Bink MCAM, Voorrips RE, Aziz E, Velasco R, Laurens F, van de Weg WE (2016) A high-density, multi-parental SNP genetic map on apple validates a new mapping approach for outcrossing species. Hortic Res 3:16057

Dicenta F, Garcia JE (1993a) Inheritance of self-compatibility in almond. Heredity 70:313-317

Dicenta F, Garcia JE (1993b) Inheritance of the kernel flavour in almond. Heredity 70:308-312

Dicenta F, Martínez-Gómez P, Ortega E, Duval H (2000) Cultivar pollinizer does not affect almond flavor. HortSci 35:1153-1154

Dicenta F, Ortega E, Martínez-Gómez P (2007) Use of recessive homozygous genotypes to assess genetic control of kernel bitterness in almond. Euphytica 153:221-225

Dirlewanger E, Cosson P, Boudehri K, Renaud C, Capdeville G, Tauzin Y, Laigret F, Moing A (2006) Development of a second-generation genetic linkage map for peach [*Prunus persica* (L.) Batsch] and characterization of morphological traits affecting flower and fruit. Tree Genet Genomes 3:1-13

Dirlewanger E, Cosson P, Tavaud M, Aranzana MJ, Poizat C, Zanetto A, Arús P, Laigret F (2002) Development of microsatellite markers in peach *Prunus persica* (L.) Batsch and their use in genetic diversity analysis in peach and sweet cherry (*Prunus avium* L.). Theor Appl Genet 105

Dirlewanger E, Graziano E, Joobeur T, Garriga-Caldere F, Cosson P, Howad W, Arús P (2004) Comparative mapping and marker-assisted selection in Rosaceae fruit crops. Proc Natl Acad Sci USA 101

Dirlewanger E, Pronier V, Parvery C, Rothan C, Guye A, Monet R (1998) Genetic linkage map of peach [*Prunus persica* (L.) Batsch] using morphological and molecular markers. Theor Appl Genet 97:888-895

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32:1792-1797

Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS ONE 6:e19379

Entani T, Iwano M, Shiba H, Che F-S, Isogai A, Takayama S (2003) Comparative analysis of the self-incompatibility (*S*-) locus region of *Prunus mume*: identification of a pollen-expressed F-box gene with allelic diversity. Genes Cells 8:203-213

Etter PD, Bassham S, Hohenlohe PA, Johnson EA, Cresko WA (2011) SNP discovery and genotyping for evolutionary genetics using RAD sequencing. Methods Mol Biol 772:157-178

FAOSTAT (2015)

Fernández i Martí A, Font i Forcada C, Socias i Company R (2013) Genetic analysis for physical nut traits in almond. Tree Genet Genomes 9:455-465

Fernández i Martí A, Gradziel T, Socias i Company R (2014) Methylation of the S_f locus in almond is associated with *S-RNase* loss of function. Plant Mol Biol 86:681-689

Fernández i Martí A, Hanada T, Alonso J, Yamane H, Tao R, Socias i Company R (2009) A modifier locus affecting the expression of the *S-RNase* gene could be the cause of breakdown of self-incompatibility in almond. Sex Plant Reprod 22:179-186

Fernández i Martí À, Hanada T, Alonso JM, Yamane H, Tao R, Socias i Company R (2010) The almond S_f haplotype shows a double expression despite its comprehensive genetic identity. Sci Hortic 125:685-691

Fernández i Marti Á, Wirthensohn M, Alonso J, Socias i Company R, Hrmova M (2012) Molecular modelling of *S-RNases* involved in almond self-incompatibility. Front Plant Sci 3

Flint-Garcia SA, Thuillet A-C, Yu J, Pressoir G, Romero SM, Mitchell SE, Doebley J, Kresovich S, Goodman MM, Buckler ES (2005) Maize association population: a high-resolution platform for quantitative trait locus dissection. Plant J 44:1054-1064

Folch J, Lees M, Sloane-Stanley G (1957) A simple method for the isolation and purification of total lipids from animal tissues. J biol Chem 226:497-509

Font i Forcada C, i Marti A, i Company R (2012) Mapping quantitative trait loci for kernel composition in almond. BMC Genet 13:47

Font i Forcada C, Oraguzie N, Reyes-Chin-Wo S, Espiau MT, Socias i Company R, Fernández i Martí A (2015) Identification of genetic loci associated with quality traits in almond via association mapping. PLoS ONE 10:e0127656

Franklin-Tong VE (1999) Signaling and the modulation of pollen tube growth. Plant Cell 11:727-738

Franklin-Tong VE, Hackett G, Hepler PK (1997) Ratio-imaging of Ca^{2+} in the self-incompatibility response in pollen tubes of *Papaver rhoeas*. Plant J 12:1375-1386

Franks TK, Yadollahi A, Wirthensohn MG, Guerin JR, Kaiser BN, Sedgley M, Ford CM (2008) A seed coat cyanohydrin glucosyltransferase is associated with bitterness in almond (*Prunus dulcis*) kernels. Funct Plant Biol 35:236-246

Fresnedo-Ramírez J, Bink MAM, van de Weg E, Famula T, Crisosto C, Frett T, Gasic K, Peace C, Gradziel T (2015) QTL mapping of pomological traits in peach and related species breeding germplasm. Mol Breed 35:1-19

Galardini M, Biondi EG, Bazzicalupo M, Mengoni A (2011) CONTIGuator: a bacterial genomes finishing tool for structural insights on draft genomes. Source Code Biol Med 6:1-5

Gallier S, Gordon KC, Singh H (2012) Chemical and structural characterisation of almond oil bodies and bovine milk fat globules. Food Chem 132:1996-2006

Gao Q, Yue G, Li W, Wang J, Xu J, Yin Y (2012) Recent progress using high-throughput sequencing technologies in plant molecular breeding. J Integr Plant Biol 54:215-227

Gilliland LU, Magallanes-Lundback M, Hemming C, Supplee A, Koornneef M, Bentsink L, DellaPenna D (2006) Genetic basis for natural variation in seed vitamin E levels in *Arabidopsis thaliana*. Proc Natl Acad Sci 103:18834-18841

Gion J-M, Hudson CJ, Lesur I, Vaillancourt RE, Potts BM, Freeman JS (2016) Genome-wide variation in recombination rate in Eucalyptus. BMC Genomics 17:590

Gnirke A, Melnikov A, Maguire J, Rogov P, LeProust E, Brockman W, Fennell T, Giannoukos G, Fisher S, Russ C (2009) Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. Nature Biotechnol 27:182-189

Goldraij A, Kondo K, Lee CB, Hancock CN, Sivaguru M, Vazquez-Santana S, Kim S, Phillips TE, Cruz-Garcia F, McClure B (2006) Compartmentalization of S-RNase and HT-B degradation in self-incompatible *Nicotiana*. Nature 439:805-810

Golz JF, Clarke AE, Newbigin E, Anderson M (1998) A relic *S-RNase* is expressed in the styles of self-compatible *Nicotiana sylvestris*. Plant J 16:591-599

Gómez EM, Dicenta F, Martínez-García PJ, Ortega E (2015) iTRAQ-based quantitative proteomic analysis of pistils and anthers from self-incompatible and self-compatible almonds with the S_f haplotype. Mol Breed 35:1-14

Goodier JL (2016) Restricting retrotransposons: a review. Mobile DNA 7:16

Gore M, Bradbury P, Hogers R, Kirst M, Verstege E, van Oeveren J, Peleman J, Buckler E, van Eijk M (2007) Evaluation of target preparation methods for single-feature polymorphism detection in large complex plant genomes. Crop Sci 47:S-135-S-148

Gouin A, Legeai F, Nouhaud P, Whibley A, Simon JC, Lemaitre C (2015) Whole-genome re-sequencing of non-model organisms: lessons from unmapped reads. Heredity 114:494-501

Gradziel TM (2009) Almond (*Prunus dulcis*) breeding. In: Mohan Jain S, Priyadarshan PM (eds) Breeding Plantation Tree Crops: Temperate Species. Springer New York, USA, pp 1-31

Gradziel TM, Martínez-Gómez P (2013) Almond breeding. In: Janick J (ed) Plant breeding reviews. John Wiley & Sons, Inc., NJ, USA, pp 207-258

Grasselly C (1972) L'amandier, caractères morphologiques et physiologiques des varietés, modalité de leurs transmissions chez les hybrides de première generation. Thése de doctorat, Université de Bordeaux, France

Grasselly C, Crossa-Raynaud P (1980) L'amandier. Maisonneuve et Larose, Paris

Grasselly C, Crossa-Raynaud P, Olivier G, Gall H (1981) Transmission du caractère d'autocompatibilité chez l'amandier (*Amygdalus communis*). Options Méditerr CIHEAM/IAMZ 81:71-75

Grasselly C, Olivier G (1976) Mise évidence de quelques types autocompatibles parmi les cultivars d'amandier (*P. amygdalus* Batsch) de la population des Pouilles. Ann Amélio Plantes 26:107-113

Grattapaglia D, Resende M (2010) Genomic selection in forest tree breeding. Tree Genet Genomes

Grattapaglia D, Sederoff R (1994) Genetic linkage maps of *Eucalyptus grandis* and *Eucalyptus urophylla* using a pseudo-testcross: mapping strategy and RAPD markers. Genetics 137:1121-1137

Gregory D, Sedgley M, Wirthensohn M, Arús P, Kaiser B, Collins G (2005) An integrated genetic linkage map for almond based on RAPD, ISSR, SSR and morphological markers. Acta Hort 694:67-72

Grusak MA, DellaPenna D (1999) Improving the nutrient composition of plants to enhance human nutrition and health. Annu Rev Plant Physiol Plant Mol Biol 50:133-161

Gu C, Wang L, Korban SS, Han YP (2015) Identification and characterization of *S-RNase* genes and *S*-genotypes in *Prunus* and *Malus* species. Can J Plant Sci 95:213-225

Gu C, Wu J, Zhang S-J, Yang Y-N, Wu H-Q, Tao S-T, Zhang S-L (2012) Characterization of the *S-RNase* genomic DNA allele sequence in *Prunus speciosa* and *P. pseudocerasus*. Sci Hortic 144:93-101

Gu Z, Gu L, Eils R, Schlesner M, Brors B (2014) circlize implements and enhances circular visualization in R. Bioinformatics 30:2811-2812

Guajardo V, Solís S, Sagredo B, Gainza F, Muñoz C, Gasic K, Hinrichsen P (2015) Construction of high density sweet cherry (*Prunus avium* L.) linkage maps using microsatellite markers and SNPs detected by genotyping-by-sequencing (GBS). PLoS ONE 10:e0127750

Hafizi A, Shiran B, Maleki B, Imani A, Banović B (2013) Identification of new *S-RNase* self-incompatibility alleles and characterization of natural mutations in Iranian almond cultivars. Trees 27:497-510

Halász J, Fodor Á, Pedryc A, Hegedüs A (2010) *S*-genotyping of Eastern European almond cultivars: identification and characterization of new (S_{36}– S_{39}) self-incompatibility ribonuclease alleles. Plant Breed 129:227-232

Halász J, Kodad O, Hegedűs A (2014) Identification of a recently active *Prunus*-specific non-autonomous Mutator element with considerable genome shaping force. Plant J 79:220-231

Hanada T, Fukuta K, Yamane H, Esumi T, Tao R, Gradziel TM, Dandekar AM, Fernández i Martí Á, Alonso JM, Socias i Company R (2009) Cloning and characterization of a self-compatible S_f haplotype in almond [*Prunus dulcis* (Mill.) D.A. Webb. syn. *P. amygdalus* Batsch] to resolve previous confusion in its S_f-RNase sequence. HortSci 44:609-613

Hanada T, Watari A, Kibe T, Yamane H, Wünsch A, Gradziel TM, Sasabe Y, Yaegaki H, Yamaguchi M, Tao R (2014) Two novel self-compatible *S* haplotypes in Peach (*Prunus persica*). J Japan Soc Hort Sci:203-2013

Hauck NR, Ikeda K, Tao R, Iezzoni AF (2006) The mutated S_1-haplotype in sour cherry has an altered *S*-haplotype–specific F-Box protein gene. J Heredity 97:514-520

Havecker ER, Gao X, Voytas DF (2004) The diversity of LTR retrotransposons. Genome Biol 5:1-6

He C, Holme J, Anthony J (2014) SNP genotyping: the KASP assay. In: Fleury D, Whitford R (eds) Crop Breeding: Methods and Protocols. Springer New York, New York, NY, pp 75-86

Heppner MJ (1923) The factors for bitterness in the sweet almond Genetics 8:390-391

Hodkinson BP, Grice EA (2015) Next-generation sequencing: a review of technologies and tools for wound microbiome research. Adv Wound Care 4:50-58

Howad W, Yamamoto T, Dirlewanger E, Testolin R, Cosson P, Cipriani G, Monforte A, Georgi L, Abbott A, Arús P (2005) Mapping with a few plants: using selective mapping for microsatellite saturation of the *Prunus* reference map. Genetics 171:1305-1309

Hua Z-H, Fields A, Kao T-h (2008) Biochemical models for *S-RNase*-based self-incompatibility. Mol Plant 1:575-585

Hyma K, Acharya C, Sun Q, Mitchell S (2013) Genotyping by sequencing applications: outcrossing species and diversity studies. Allele mining workshop Intl PAG XXI, San Diego, CA

Hyma KE, Barba P, Wang M, Londo JP, Acharya CB, Mitchell SE, Sun Q, Reisch B, Cadle-Davidson L (2015) Heterozygous Mapping Strategy (HetMappS) for highresolution genotyping-by-sequencing markers: a case study in grapevine. PLoS ONE 10:e0134880

Ikeda K, Igic B, Ushijima K, Yamane H, Hauck N, Nakano R, Sassa H, Iezzoni A, Kohn J, Tao R (2004) Primary structural features of the *S* haplotype-specific F-box protein, SFB, in *Prunus*. Sex Plant Reprod 16:235-243

Illa E, Sargent DJ, Girona EL, Bushakra J, Cestaro A, Crowhurst R, Pindo M, Cabrera A, Knaap E, Iezzoni A, Gardiner S, Velasco R, Arús P, Chagné D, Troggio M (2011) Comparative analysis of *rosaceous* genomes and the reconstruction of a putative ancestral genome for the family. BMC Evol Biol 11:1-13

Ioerger TR, Clark AG, Kao TH (1990) Polymorphism at the self-incompatibility locus in Solanaceae predates speciation. Proc Natl Acad Sci 87:9732-9735

Iovene M, Zhang T, Lou Q, Buell CR, Jiang J (2013) Copy number variation in potato – an asexually propagated autotetraploid species. Plant J 75:80-89

Ishimizu T, Shinkawa T, Sakiyama F, Norioka S (1998) Primary structural features of rosaceous *S-RNases* associated with gametophytic self-incompatibility. Plant Mol Biol 37:931-941

Jambazian PR, Haddad E, Rajaram S, Tanzman J, Sabaté J (2005) Almonds in the diet simultaneously improve plasma α-tocopherol concentrations and reduce plasma lipids. J Am Diet Assoc 105:449-454

Jenkins DJA, Kendall CWC, Marchie A, Josse AR, Nguyen TH, Faulkner DA, Lapsley KG, Blumberg J (2008) Almonds reduce biomarkers of lipid peroxidation in older hyperlipidemic subjects. J Nutr 138:908-913

Jia X, Li N, Zhang W, Zhang X, Lapsley K, Huang G, Blumberg J, Ma G, Chen J (2006) A pilot study on the effects of almond consumption on DNA damage and oxidative stress in smokers. Nutr cancer 54:179-183

Jiang Y-Q, Ma R-C (2003) Generation and analysis of expressed sequence tags from almond (*Prunus dulcis* Mill.) pistils. Sex Plant Reprod 16:197-207

Jones DT, Taylor WR, Thornton JM (1992) The rapid generation of mutation data matrices from protein sequences. Comput Appl Biosci 8

Joobeur T, Periam N, Vicente M, King G, Arus P (2000) Development of a second generation linkage map for almond using RAPD and SSR markers. Genome 43:649-655

Joobeur T, Viruel M, de Vicente M, Jauregui B, Ballester J, Dettori M, Verde I, Truco M, Messeguer R, Batlle I (1998) Construction of a saturated linkage map for *Prunus* using an almond × peach F_2 progeny. Theor Appl Genet 97:1034-1041

Kamal-Eldin A, Görgen S, Pettersson J, Lampi A-M (2000) Normal-phase high-performance liquid chromatography of tocopherols and tocotrienols: comparison of different chromatographic columns. J Chromatogr A 881:217-227

Keane TM, Creevey CJ, Pentony MM, Naughton TJ, Mclnerney JO (2006) Assessment of methods for amino acid matrix selection and their use on empirical data shows that ad hoc assumptions for choice of matrix are not justified. BMC Evol Biol 6:1-17

Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, Thierer T, Ashton B, Meintjes P, Drummond A (2012) Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28:1647-1649

Kester D, Gradziel T (1996) Almonds. In: Basenes M, Byrne D (eds) Fruit breeding, Nuts Vol 3. Springer New York Dordrecht Heidelberg London, pp 1-97

Kester DE, Gradziel TM, Grasselly C (1991) Almonds (*Prunus*). Acta Hort Sci 290:701-760

Kester DE, Gradziel TM, Micke WC (1994) Identifying pollen incompatibility groups in California almond cultivars. J Am Soc Hortic Sci 119:106-109

Kester DE, Hansche PE, Beres V, Asay RN (1977) Variance components and heritability of nut and kernel traits in almond J Am Soc Hortic Sci 102:264-266

Kiss MM, Ortoleva-Donnelly L, Beer NR, Warner J, Bailey CG, Colston BW, Rothberg JM, Link DR, Leamon JH (2008) High-throughput quantitative polymerase chain reaction in picoliter droplets. Anal Chem 80:8975-8981

Kodad O, Alonso J, Font i Forcada C, R SiC (2010a) Fruit quality in almond: chemical aspects for breeding strategies. Options Mediterr Ser A 94:235-243

Kodad O, Company RSI, Prats MS, Ortiz MCL (2006) Variability in tocopherol concentrations in almond oil and its use as a selection criterion in almond breeding. J Hortic Sci Biotechnol 81:501-507

Kodad O, Estopañán G, Juan T, Socias i Company R (2009a) Xenia effects on oil content and fatty acid and tocopherol concentrations in autogamous almond cultivars. J Agric Food Chem 57:10809-10813

Kodad O, Sánchez A, Saibo N, Oliveira M (2010b) Molecular characterization of five new *S* alleles associated with self-incompatibility in local Spanish almond cultivars. CIHEAM, Zaragoza

Kodad O, Socias i Company R (2008) Variability of oil content and of major fatty acid composition in almond (*Prunus amygdalus* Batsch) and its relationship with kernel quality. J Agric Food Chem 56:4096-4101

Kodad O, Socias i Company R (2009) Review and update of self-incompatibility alleles in almond Acta Hort 814:421-424

Kodad O, Socias i Company R, Estopanan G, Juan T, Mamouni A (2011) Tocopherol concentration in almond oil: genetic variation and environmental effects under warm conditions. J Agric Food Chem 59:6137-6141

Kodad O, Socias i Company R, Sánchez A, Oliveira MM (2009b) The expression of self-compatibility in almond may not only be due to the presence of the S_f allele. J Am Soc Hortic Sci 134:221-227

Koepke T, Schaeffer S, Harper A, Dicenta F, Edwards M, Henry RJ, Møller BL, Meisel L, Oraguzie N, Silva H, Sánchez-Pérez R, Dhingra A (2013) Comparative genomics analysis in Prunoideae to identify biologically relevant polymorphisms. Plant Biotechnol J 11:883-893

Kornsteiner M, Wagner K-H, Elmadfa I (2006) Tocopherols and total phenolics in 10 different nut types. Food Chem 98:381-387

Kosambi D (1944) The estimation of map distances from recombination values. Ann Eugen 12:172-175

Kujur A, Upadhyaya HD, Shree T, Bajaj D, Das S, Saxena MS, Badoni S, Kumar V, Tripathi S, Gowda CLL, Sharma S, Singh S, Tyagi AK, Parida SK (2015) Ultra-high density intra-specific genetic linkage maps accelerate identification of functionally relevant molecular tags governing important agronomic traits in chickpea. Sci Rep 5:9468

Kumar S, Stecher G, Tamura K (2016) MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol Biol Evol 33:1870-1874

Kump KL, Bradbury PJ, Wisser RJ, Buckler ES, Belcher AR, Oropeza-Rosas MA, Zwonitzer JC, Kresovich S, McMullen MD, Ware D, Balint-Kurti PJ, Holland JB (2011) Genome-wide association study of quantitative resistance to southern leaf blight in the maize nested association mapping population. Nat Genet 43:163-168

Ladizinsky G (1999) On the origin of almond. Genet Resour Crop Evol 46:143-147

Lamboy WF (1998) Using simple sequence repeats (SSRs) for DNA fingerprinting germplasm accessions of grape (Vitis L.) species. J Am Soc Hortic Sci 123:182-188

Lander ES, Waterman MS (1988) Genomic mapping by fingerprinting random clones: a mathematical analysis. Genomics 2:231-239

Li H (2011) A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics 27:2987-2993

Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25:1754-1760

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R (2009) The Sequence Alignment/Map format and SAMtools. Bioinformatics 25:2078-2079

Li H, Vikram P, Singh RP, Kilian A, Carling J, Song J, Burgueno-Ferreira JA, Bhavani S, Huerta-Espino J, Payne T, Sehgal D, Wenzl P, Singh S (2015) A high density GBS map of bread wheat and its application for dissecting complex disease resistance traits. BMC Genomics 16:1-15

Li S-C, Liu Y-H, Liu J-F, Chang W-H, Chen C-M, Chen CYO (2011) Almond consumption improved glycemic control and lipid profiles in patients with type 2 diabetes mellitus. Metabolism 60:474-479

López-Ortiz M, Prats-Moya S, Sanahuja A, Maestre-Perez S, Grane-Teruel N, Martin-Carratala M (2008) Comparative study of tocopherol homologue content in four almond oil cultivars during two consecutive years. J Food Comp Anal 21:144-151

López M, Mnejja M, Rovira M, Collins G, Vargas F, Arús P, Batlle I (2004) Self-incompatibility genotypes in almond re-evaluated by PCR, stylar ribonucleases, sequencing analysis and controlled pollinations. Theor Appl Genet 109:954-964

Lu F, Lipka AE, Glaubitz J, Elshire R, Cherney JH, Casler MD, Buckler ES, Costich DE (2013) Switchgrass genomic diversity, ploidy, and evolution: novel insights from a network-based SNP discovery protocol. PLoS Genet 9:e1003215

Luu DT, Qin X, Morse D, Cappadocia M (2000) S-RNase uptake by compatible pollen tubes in gametophytic self-incompatibility. Nature 407:649

Luu DT, Qin X, Laublin G, Yang Q, Morse D, Cappadocia M (2001) Rejection of *S*-heteroallelic pollen by a dual-specific *S-RNase* in *Solanum chacoense* predicts a multimeric SI pollen component. Genetics 159:329-335

Ma RC, Oliveira M (2002) Evolutionary analysis of *S*-RNase genes from Rosaceae species. Mol Gen Genomics 267:71-78

Madawala S, Kochhar S, Dutta P (2012) Lipid components and oxidative status of selected specialty oils. Grasas y Aceites 63:143-151

Maguire LS, O'Sullivan SM, Galvin K, O'Connor TP, O'Brien NM (2004) Fatty acid profile, tocopherol, squalene and phytosterol content of walnuts, almonds, peanuts, hazelnuts and the macadamia nut. Int J Food Sci Nutr 55:171-178

Mammadov J, Aggarwal R, Buyyarapu R, Kumpatla S (2012) SNP markers and their impact on plant breeding. Int J Plant Genomics 2012:11

Marchese A, Bošković RI, Caruso T, Raimondo A, Cutuli M, Tobutt KR (2007) A new self-compatibility haplotype in the sweet cherry 'Kronio, *S5*', attributable to a pollen-part mutation in the *SFB* gene. J Exp Bot 58:4347-4356

Martínez-García PJ, Gãmez EM, Casado-Vela J, Elortza F, Dicenta F, Ortega E (2015) Differential protein expression in compatible and incompatible pollen-pistil interactions in almond [*Prunus dulcis* (Miller) D. A. Webb] by 2D-DIGE and HPLC-MS/MS. J Hortic Sci Biotechnol 90:71-77

Martínez-Gómez P, Arulsekar S, Potter D, Gradziel TM (2003) An extended interspecific gene pool available to peach and almond breeding as characterized using simple sequence repeat (SSR) markers. Euphytica 131:313-322

Martínez-Gómez P, Sánchez-Pérez R, Dicenta F, Howad W, Arús P, Gradziel T (2007) Almond. In: Kole C (ed) Fruits and Nuts. Springer Berlin Heidelberg, pp 229-242

Martinez-Gracía P (2009) Mejora genetica del *amendro* [*Prunus dulcis* (Miller) D. A, webb]: aspectos agronomicos y moleculares de la incompatibilidad floral y su indencia sobre la fructification. Tesis doctoral, Universidad de Murcia, España

Mascher M, Wu S, Amand PS, Stein N, Poland J (2013) Application of genotyping-by-sequencing on semiconductor sequencing platforms: a comparison of genetic and reference-based marker ordering in barley. PLoS One 8:e76925

Matsuura T, Sakai H, Unno M, Ida K, Sato M, Sakiyama F, Norioka S (2001) Crystal structure at 1.5-Å resolution of *Pyrus pyrifolia* pistil ribonuclease responsible for gametophytic self-incompatibility. J Biol Chem 276:45261-45269

Matton DP, Maes O, Laublin G, xe, ve, Xike Q, Bertrand C, Morse D, Cappadocia M (1997) Hypervariable domains of self-incompatibility *RNases* mediate allele-specific pollen recognition. Plant Cell 9:1757-1766

McClure B, Cruz-García F, Romero C (2011) Compatibility and incompatibility in *S-RNase*-based systems. Ann Bot 108:647-658

McCubbin AG, Kao TH (2000) Molecular recognition and response in pollen and pistil interactions. Annu rev cell dev biol 16:333-364.

Merchant S, Wood DE, Salzberg SL (2014) Unexpected cross-species contamination in genome sequencing projects. PeerJ 2:e675

Michael JT (2014) High-throughput SNP genotyping to accelerate crop improvement. Plant Breed Biotechnol 2:195-212

Miller M, Dunham J, Amores A, Cresko W, Johnson E (2007) Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. Genome Res 17:240-248

Milne I, Bayer M, Cardle L, Shaw P, Stephen G, Wright F, Marshall D (2010) Tablet—next generation sequence assembly visualization. Bioinformatics 26:401-402

Mitchell S, Elshire R, Glaubitz J, Lu F, Harriman J, Sun Q, Buckler E (2012) Genotyping-by-sequencing (GBS): optimization and applications for high diversity species PAG XX, San Diego, CA, p 181

Mohsenin N (1970) Physical properties of plant and animal materials. Gordon and Breach Science Publishers, New York

Mousavi A, Fatahi R, Zamani Z, Imani A, Dicenta F, Ortega E (2011) Identification of self-incompatibility genotypes in iranian almond cultivars. Acta Hort 912:303-311

Murphy DJ (2007) Future prospects for oil palm in the 21st century: biological and related challenges. Eur J Lipid Sci Tech 109:296-306

Muyle A, Serres-Giardi L, Ressayre A, Escobar J, Glémin S (2011) GC-biased gene conversion and selection affect GC content in the *Oryza* Genus (rice). Mol Biol Evol 28:2695-2706

Neale D, Kremer A (2011) Forest tree genomics: growing resources and applications. Nat Rev Genet 12:111-122

Nijman IJ, Mokry M, van Boxtel R, Toonen P, de Bruijn E, Cuppen E (2010) Mutation discovery by targeted genomic enrichment of multiplexed barcoded samples. Nat Meth 7:913-915

Norioka N, Norioka S, Ohnishi Y, Ishimizu T, Oneyama C, Nakanishi T, Sakiyama F (1996) Molecular cloning and nucleotide sequences of cDNAs encoding *s*-allele specific *stylar-RNases* in a self-incompatible cultivar and its self-compatible mutant of japanese pear, *Pyrus pyrifolia* Nakai. J Biochem 120:335-345

Nunes MDS, Santos RAM, Ferreira SM, Vieira J, Vieira CP (2006) Variability patterns and positively selected sites at the gametophytic self-incompatibility pollen *SFB* gene in a wild self-incompatible *Prunus spinosa* (Rosaceae) population. New Phytol 172:577-587

O'Neill P, Singh M, Knox R (1988) Cell biology of the stigma of *Brassica campestris* in relation to CO_2 effects on self-pollination. J Cell Sci, pp 541-550

Ogundiwin EA, Peace CP, Gradziel TM, Parfitt DE, Bliss FA, Crisosto CH (2009) A fruit quality gene map of *Prunus*. BMC Genomics 10:1-13

Ortega E, Bošković R, Sargent D, Tobutt K (2006) Analysis of *S-RNase* alleles of almond (*Prunus dulcis*): characterization of new sequences, resolution of synonyms and evidence of intragenic recombination. Mol Genet Genomics 276:413-426

Ortega E, Dicenta F (2003) Inheritance of self-compatibility in almond: breeding strategies to assure self-compatibility in the progeny. Theor App Genet 106:904-911

Ortega E, Sutherland BG, Dicenta F, Boskovic R, Tobutt KR (2005) Determination of incompatibility genotypes in almond using first and second intron consensus primers: detection of new *S* alleles and correction of reported *S* genotypes. Plant Breed 124:188-196

Payne R, Murray D, Harding S, Baird D, Soutar D (2009) GenStat for Windows (16th Edition) Introduction. VSN International, Hemel Hempstead., VSN International, Hemel Hempstead, UK

Peterson B, Weber N, Kay E, Fisher H, Hoekstra H (2012) Double Digest RADseq: an inexpensive method for *de novo* SNP discovery and genotyping in model and non-model species. PLoS ONE 7:e37135

Poland J, Brown P, Sorrells M, Jannink J-L (2012) Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. PLoS ONE 7:e32253

Qiao H, Wang H, Zhao L, Zhou J, Huang J, Zhang Y, Xue Y (2004) The F-box protein AhSLF-S2 physically interacts with *S-RNases* that may be inhibited by the Ubiquitin/26S proteasome pathway of protein degradation during compatible pollination in *Antirrhinum*. Plant Cell 16:582-595

Rahemi A, Fatahi R, Ebadi A, Taghavi T, Hassani D, Gradziel T, Chaparro J (2010) Genetic variation of *S*-alleles in wild almonds and their related *Prunus* species. Aus J Crop Sci 648:648-659

Rahemi A, Gradziel T, Chaparro J, Folta K, Taghavi T, Fatahi R, Ebadi A, Hassani D (2015) Phylogenetic relationships among the first and second introns of selected *Prunus S-RNase* genes. Can J Plant Sci 95:1145-1154

Reuter J, Spacek D, Snyder M (2015) High-throughput sequencing technologies. Mol Cell 58:586-597

Robinson JT, Thorvaldsdottir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP (2011) Integrative genomics viewer. Nat Biotech 29:24-26

Rocher S, Jean M, Castonguay Y, Belzile F (2015) Validation of genotyping-by-sequencing analysis in populations of tetraploid alfalfa by 454 sequencing. PLoS ONE 10:e0131918

Romero C, Vilanova S, Burgos L, Martínez-Calvo J, Vicente M, Llácer G, Badenes ML (2004) Analysis of the *S*-locus structure in *Prunus armeniaca* L. Identification of *S*-haplotype specific *S-RNase* and *F-box* genes. Plant Mol Biol 56:145-157

Romojaro F, Riquelme F, Gimenez JL, Llorente S (1988) Fat content and oil characteristics of some almond varieties. Fruit Sci Rep 15:53-58

Royo J, Kunz C, Kowyama Y, Anderson M, Clarke A, Newbigin E (1994) Loss of a histidine residue at the active site of *S*-locus ribonuclease is associated with self-compatibility in *Lycopersicon peruvianum*. Proc Natl Acad Sci 91:6511-6514

Rugini N (1986) Almond. In: Evans A, Sharp D, Williams R, Phillip A (eds) Hand book of plant cell culture. MacMillan Publishing New York, pp 507-611

Sabate J, Hook D (1996) Almonds, walnuts, and serum lipids. In: GA S (ed) Handbook of lipids in human nutririon. CRC Press, Inc., New York, pp 137–444

Saintenac C, Jiang D, Akhunov ED (2011) Targeted analysis of nucleotide and copy number variation by exon capture in allotetraploid wheat genome. Genome Biol 12:R88

Sánchez-Pérez R, Belmonte F, Borch J, Dicenta F, Møller B, Jørgensen K (2012) Prunasin hydrolases during fruit development in sweet and bitter almonds. Plant Physiol 158:1916-1932

Sánchez-Pérez R, Dicenta F, Martínez-Gómez P (2004) Identification of *S*-alleles in almond using multiplex PCR. Euphytica 138:263-269

Sánchez-Pérez R, Howad W, Dicenta F, Arús P, Martínez-Gómez P (2007) Mapping major genes and quantitative trait loci controlling agronomic traits in almond. Plant Breed 126:310-318

Sánchez-Pérez R, Howad W, García-Mas J, Arús P, Martínez-Gómez P, Dicenta F (2010) Molecular markers for kernel bitterness in almond. Tree Genet Genomes 6:237- 245

Sánchez-Pérez R, Jørgensen K, Olsen C, Dicenta F, Møller B (2008) Bitterness in almonds. Plant Physiol 146:1040-1052

Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci 74:5463-5467

Sassa H, Hirano H (1998) Style-specific and developmentally regulated accumulation of a glycosylated thaumatin/PR5-like protein in Japanese pear (*Pyrus serotina* Rehd.). Planta 205:514-521

Sassa H, Hirano H, Nishio T, Koba T (1997) Style-specific self-compatible mutation caused by deletion of the *S-RNase* gene in Japanese pear (*Pyrus serotina*). Plant J 12:223-227

Sassa H, Nishio T, Kowyama Y, Hirano H, Koba T, Ikehashi H (1996) Self-incompatibility (*S*) alleles of the rosaceae encode members of a distinct class of the T2/S ribonuclease superfamily. Mol Gen Genet 250:547-557

Scorza R (2001) Progress in tree fruit improvement through molecular genetics. Hortsci 36:855-858

Sebat J, Lakshmi B, Troge J, Alexander J, Young J, Lundin P, Månér S, Massa H, Walker M, Chi M, Navin N, Lucito R, Healy J, Hicks J, Ye K, Reiner A, Gilliam TC, Trask B, Patterson N, Zetterberg A, Wigler M (2004) Large-scale copy number polymorphism in the human genome. Science 305:525-528

Semagn K, Babu R, Hearne S, Olsen M (2014) Single nucleotide polymorphism genotyping using Kompetitive Allele Specific PCR (KASP): overview of the technology and its application in crop improvement. Mol Breed 33:1-14

Serres-Giardi L, Belkhir K, David J, Glémin S (2012) Patterns and evolution of nucleotide landscapes in seed plants. Plant Cell 24:1379-1397

Shaikh TH, Gai X, Perin JC, Glessner JT, Xie H, Murphy K, O'Hara R, Casalunovo T, Conlin LK, D'Arcy M, Frackelton EC, Geiger EA, Haldeman-Englert C, Imielinski M, Kim CE, Medne L, Annaiah K, Bradfield JP, Dabaghyan E, Eckert A, Onyiah CC, Ostapenko S, Otieno FG, Santa E, Shaner JL, Skraban R, Smith RM, Elia J, Goldmuntz E, Spinner NB, Zackai EH, Chiavacci RM, Grundmeier R, Rappaport EF, Grant SFA, White PS, Hakonarson H (2009) High-resolution mapping and analysis of copy number variations in the human genome: A data resource for clinical and research applications. Genome Res 19:1682-1690

Sims TL, Ordanic M (2001) Identification of a *S*-ribonuclease-binding protein in *Petunia hybrida*. Plant Mol Biol 47:771-783

Socias i Company R (1979) Aportación a las téchnicas de observación de tubos polinicos. Caso del almendro. An Inst Nac Invest Ser Prod Veg 10:233-236

Socias i Company R (1990) Breeding self-compatible almonds. In: Janick J (ed) Plant Breeding Reviews. John Wiley & Sons, Inc., New York, NYSE, pp 313-338

Socias i Company R, Gómez Aparisi J, Alonso JM (2005) Year and enclosure effects on fruit set in an autogamous almond. Sci Hortic 104:369-377

Socias i Company R, Kodad O, Alonso J, Gradziel T (2008) Almond quality: a breeding perspective. Hortic Rev 34:197-238

Söllner S, Berkner S, Lipps G (2006) Characterisation of the novel restriction endonuclease *SuiI* from *Sulfolobus islandicus*. Extremophiles 10:629-634

Soulard J, Qin X, Boivin N, Morse D, Cappadocia M (2013) A new dual-specific incompatibility allele revealed by absence of glycosylation in the conserved C2 site of a *Solanum chacoense S-RNase*. J Exp Bot 64:1995-2003

Souza LM, Gazaffi R, Mantello CC, Silva CC, Garcia D, Le Guen V, Cardoso SEA, Garcia AAF, Souza AP (2013) QTL mapping of growth-related traits in a full-sib family of rubber tree (*Hevea brasiliensis*) evaluated in a sub-tropical climate. PLoS ONE 8:e61238

Steinmetz K, Potter J (1996) Vegetables, fruit, and cancer prevention: a review. J Am Diet Assoc 96:1027-1039

Suen D, Wu S, Chang H, Dhugga K, Huang A (2003) Cell wall reactive proteins in the coat and wall of maize pollen: potential role in pollen tube growth on the stigma and through the style. J Biol Chem 278:43672-43681

Sundararajan A, Dukowic-Schulze S, Kwicklis M, Engstrom K, Garcia N, Oviedo OJ, Ramaraj T, Gonzales MD, He Y, Wang M, Sun Q, Pillardy J, Kianian SF, Pawlowski WP, Chen C, Mudge J (2016) Gene evolutionary trajectories and GC patterns driven by recombination in zea mays. Front Plant Sci 7

Sutherland B, Robbins T, Tobutt K, Weber W (2004) Primers amplifying a range of *Prunus S*-alleles. Plant Breed 123:582-584

Swanson-Wagner RA, Eichten SR, Kumari S, Tiffin P, Stein JC, Ware D, Springer NM (2010) Pervasive gene content variation and copy number variation in maize and its undomesticated progenitor. Genome Res 20:1689-1699

Tae H, Karunasena E, Bavarva JH, McIver LJ, Garner HR (2014) Large scale comparison of non-human sequences in human sequencing data. Genomics 104:453-458

Tamura M, Ushijima K, Sassa H, Hirano H, Tao R, Gradziel TM, Dandekar AM (2000) Identification of self-incompatibility genotypes of almond by allele-specific PCR analysis. Theor Appl Genet 101:344-349

Tao R, Habu T, Yamane H, Sugiura A (2002) Characterization and cDNA cloning for *S-RNase*, a molecular marker for self-compatibility, in japanese apricot (*Prunus mume*). J Jap Soc Hortic Sci 71:595-600

Tao R, Watari A, Hanada T, Habu T, Yaegaki H, Yamaguchi M, Yamane H (2007) Self-compatible peach (*Prunus persica*) has mutant versions of the *S* haplotypes found in self-incompatible *Prunus* species. Plant Mol Biol 63:109-123

Tavassolian I (2008) Construction of a microsatellite-based genetic map of almond. PhD Thesis, The University of Adelaide, Australia

Tavassolian I, Rabiei G, Gregory D, Mnejja M, Wirthensohn M, Hunt P, Gibson J, Ford C, Sedgley M, Wu S-B (2010) Construction of an almond linkage map in an Australian population Nonpareil × Lauranne. BMC Genomics 11:10

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22:4673-4680

Tian F, Bradbury PJ, Brown PJ, Hung H, Sun Q, Flint-Garcia S, Rocheford TR, McMullen MD, Holland JB, Buckler ES (2011) Genome-wide association study of leaf architecture in the maize nested association mapping population. Nat Genet 43:159-162

Tufts WP, Philip GL (1922) Almond pollination. Cali Agri Bul

Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, Rozen SG (2012) Primer3—new capabilities and interfaces. Nucleic Acids Res 40:e115-e115

Ushijima K, Sassa H, Dandekar AM, Gradziel TM, Tao R, Hirano H (2003) Structural and transcriptional analysis of the *self-incompatibility* locus of almond: identification of a pollen-expressed F-box gene with haplotype-specific polymorphism. Plant Cell 15:771-781

Ushijima K, Sassa H, Tao R, Yamane H, Dandekar AM, Gradziel TM, Hirano H (1998) Cloning and characterization of cDNAs encoding *S-RNases* from almond (*Prunus dulcis*): primary structural features and sequence diversity of the *S-RNases* in Rosaceae. Mol Gen Genet 260:261-268

Ushijima K, Yamane H, Watari A, Kakehi E, Ikeda K, Hauck N, Iezzoni A, Tao R (2004) The *S* haplotype-specific F-box protein gene, *SFB*, is defective in self-compatible haplotypes of *Prunus avium* and *P-mume*. Plant J 39:573-586

van Dijk EL, Auger H, Jaszczyszyn Y, Thermes C (2014) Ten years of next-generation sequencing technology. Trends Genet 30:418-426

van Orsouw N, Hogers R, Janssen A, Yalcin F, Snoeijers S, Verstege E, Schneiders H, van der Poel H, van Oeveren J, Verstegen H (2007) Complexity Reduction of Polymorphic Sequences (CRoPS ™): a novel approach for large-scale polymorphism discovery in complex genomes. PLoS ONE 2

Vezvaei A, Jackson JF (1995) Almond nut analysis. In: Linskens H, Jackson J (eds) Fruit Analysis. Springer Berlin Heidelberg, pp 133-148

Vieira J, Ferreira P, Aguiar B, Fonseca N, Vieira C (2010) Evolutionary patterns at the *RNase* based gametophytic self -incompatibility system in two divergent Rosaceae groups (*Maloideae* and *Prunus*). BMC Evol Biol 10:1-15

Vieira J, Teles E, Santos RAM, Vieira CP (2008) Recombination at *Prunus S*-locus region *SLFL1* Gene. Genetics 180:483-491

Voelkerding KV, Dames SA, Durtschi JD (2009) Next-generation sequencing: from basic research to diagnostics. Clinical Chem 55:641-658

Voorrips R (2002) MapChart: software for the graphical presentation of linkage maps and QTLs. J Hered 93:77-78

Wang L, Peng H, Ge T, Liu T, Hou X, Li Y (2014) Identification of differentially accumulating pistil proteins associated with self-incompatibility of non-heading Chinese cabbage. Plant Biol 16:49-57

Wendler N, Mascher M, Noh C, Himmelbach A, Scholz U, Ruge-Wehling B, Stein N (2014) Unlocking the secondary gene-pool of barley with next-generation sequencing. Plant Biotechnol J 12:1122-1131

Wien M, Bleich D, Raghuwanshi M, Gould-Forgerite S, Gomes J, Monahan-Couch L, Oda K (2010) Almond consumption and cardiovascular risk factors in adults with prediabetes. J Am Coll Nutr 29:189-197

Wirthensohn M, Sedgley M (2002) Almond breeding in Australia. Acta Hort 591:245-248

Wong MML, Gujaria-Verma N, Ramsay L, Yuan HY, Caron C, Diapari M, Vandenberg A, Bett KE (2015) Classification and characterization of species within the genus *Lens* using genotyping-by-sequencing (GBS). PLoS ONE 10:1-16

Wu S-B, Franks T, Hunt P, Wirthensohn M, Gibson J, Sedgley M (2010) Discrimination of SNP genotypes associated with complex haplotypes by high resolution melting analysis in almond: implications for improved marker efficiencies. Mol Breed 25:351-357

Wu S-B, Tavassolian I, Rabiei G, Hunt P, Wirthensohn M, Gibson J, Ford C, Sedgley M (2009) Mapping SNP-anchored genes using high-resolution melting analysis in almond. Mole Genet Genomics 282:273-281

Wu Y, Bhat PR, Close TJ, Lonardi S (2008) Efficient and accurate construction of genetic linkage maps from the minimum spanning tree of a graph. PLoS Genet 4:e1000212

Wünsch A, Hormaza JI (2003) Cloning and characterization of genomic DNA sequences of four self-incompatibility alleles in sweet cherry *(Prunus avium* L.). Theor Appl Genet 108:299-305

Xu Z, Wang H (2007) LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. Nucleic Acids Res 35:W265-W268

Yamane H, Ikeda K, Hauck N, Iezzoni A, Tao R (2003a) Self-incompatibility (*S*) locus region of the mutated S_6-haplotype of sour cherry (*Prunus cerasus*) contains a functional pollen *S* allele and a non-functional pistil *S* allele. J Exp Bot 54:2431-2437

Yamane H, Ikeda K, Ushijima K, Sassa H, Tao R (2003b) A pollen-expressed gene for a novel protein with an F-box motif that is very tightly linked to a gene for *S-RNase* in two species of cherry, *Prunus cerasus* and *P. avium*. Plant Cell Physiol 44:764-769

Ye K, Schulz MH, Long Q, Apweiler R, Ning Z (2009) Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. Bioinformatics 25:2865-2871

Yildirim AN, Akinci-Yildirim F, Polat M, Şan B, Sesli Y (2014) Amygdalin content in kernels of several almond cultivars grown in Turkey. HortSci 49:1268-1270

Yu J, Holland J, McMullen M, Buckler E (2008) Genetic design and statistical power of nested association mapping in maize. Genetics 178:539-551

Yu P, Wang C, Xu Q, Feng Y, Yuan X, Yu H, Wang Y, Tang S, Wei X (2011) Detection of copy number variations in rice using array-based comparative genomic hybridization. BMC Genomics 12:372

Zeballos JL, Abidi W, Giménez R, Monforte AJ, Moreno MÁ, Gogorcena Y (2016) Mapping QTLs associated with fruit quality traits in peach [*Prunus persica* (L.) Batsch] using SNP maps. Tree Genet Genomes 12:1-17

Zerbino DR, Birney E (2008) Velvet: Algorithms for *de novo* short read assembly using de Bruijn graphs. Genome Res 18:821-829

Zhang L, Yan L, Jiang J, Wang Y, Jiang Y, Yan T, Cao Y (2014) The structure and retrotransposition mechanism of LTR-retrotransposons in the asexual yeast *Candida albicans*. Virulence 5:655-664

Zhu Y, Wang X, Zhao P, Ni X (2003) Relationship between glutathione s-transferase activity of restorer anthers and pollen fertility of F_1 hybrid in upland cotton. Acta Agron Sin 29:693-696

Zhu Y, Wilkinson K, Wirthensohn M (2015) Lipophilic antioxidant content of almonds (*Prunus dulcis*): a regional and varietal study. J Food Compst Anal 39:120-127

Zisovich AH, Stern RA, Sapir G, Shafir S, Goldway M (2004) The RHV region of *S-RNase* in the European pear (*Pyrus communis*) is not required for the determination of specific pollen rejection. Sex Plant Reprod 17:151-156

Zonia L, Munnik T (2009) Uncovering hidden treasures in pollen tube growth mechanics. Trends Plant Sci 14:318-327

APPENDIX 1: Supplementary materials of Chapter 3

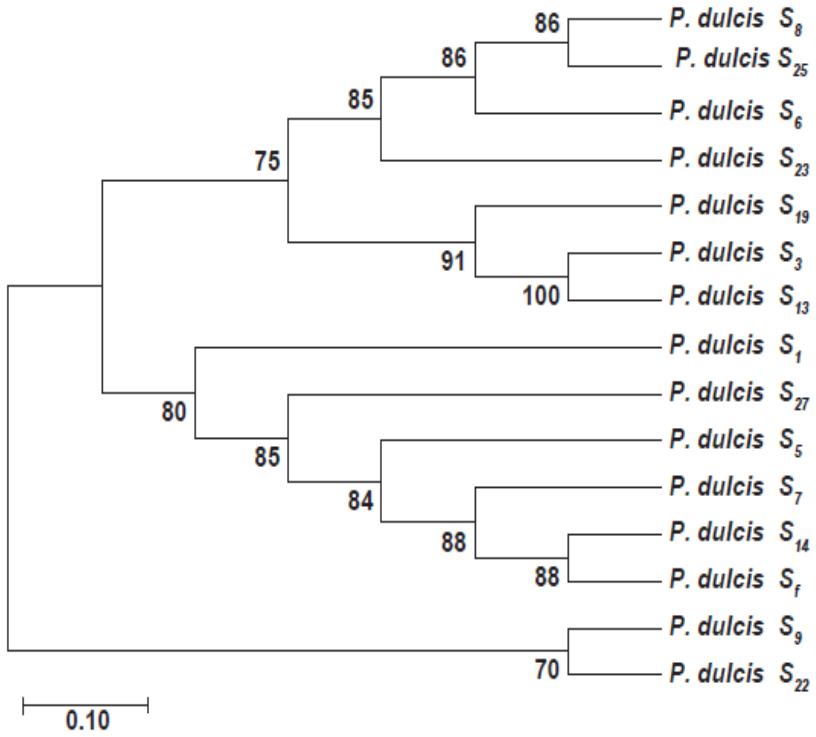


Fig. S1.1 Phylogenetic relationships among 15 almond *S* alleles from the *S-RNase* gene using the bootstrap consensus tree, protein sequences between conserved region 1 and conserved region 5 (C5) of the *S-RNase* gene were used to generate the tree. A total of 162 amino acid positions were in the final dataset. Phylogenetic relationships were inferred by using Maximum Likelihood method based on the JTT matrix based model (Jones et al. 1992). The tree with the highest log likelihood (-2081.69) is shown. The percentage of trees in which the associated branch clustered together in the bootstrap test (1,000 replicates) is shown next to the branches. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+*G*, parameter = 0.6285)). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site.

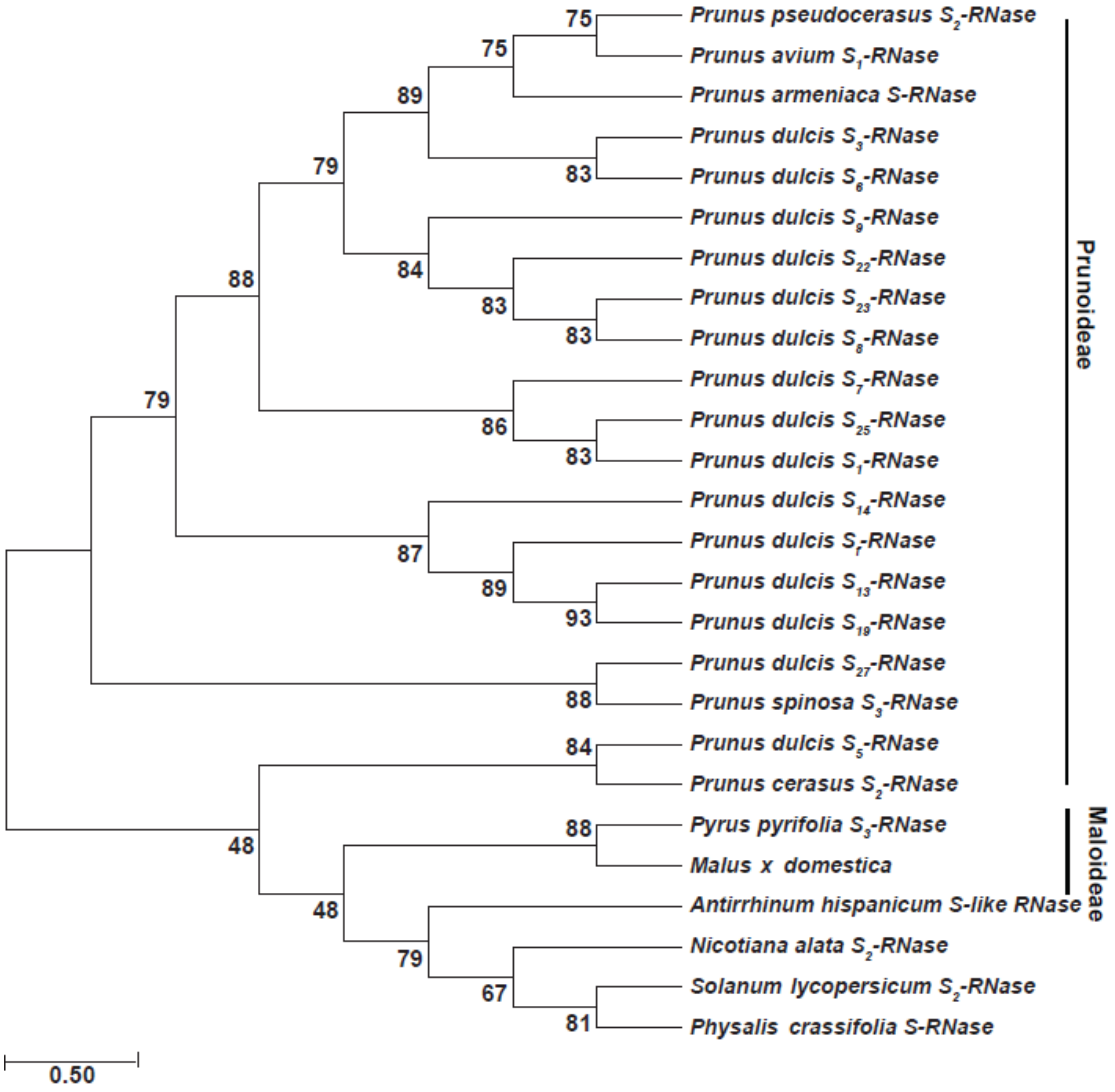


Fig. S1.2 Phylogenetic relationships among the *S-RNase* alleles from *Prunus, Malus*, *Pyrus*, *Antirrhinum* species and species from Solanaceae using the bootstrap consensus tree, protein sequences between conserved region 1 and conserved region 5 (C5) of the *S-RNase* gene were used and there was a total of 53 positions in the final dataset. Phylogenetic relationships were inferred by using Maximum Likelihood method based on the JTT matrix based model (Jones et al. 1992). The tree with the highest log likelihood -980.8521) is shown. The percentage of trees in which the associated branch clustered together in the bootstrap test (1,000 replicates) is shown next to the branches. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+*G*, parameter = 0.7883)). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site.

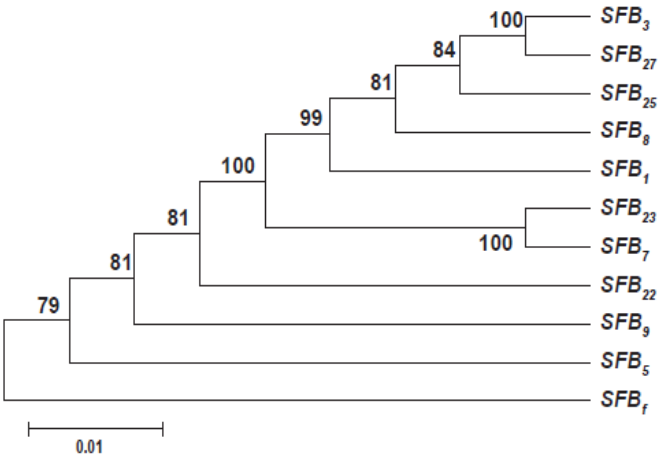


Fig. S1.3 Phylogenetic relationships among the almond *SFB* alleles using the bootstrap consensus tree. A total of 64 amino acid positions were in the final dataset. Phylogenetic relationships were inferred by using Maximum Likelihood method based on the JTT matrix based model (Jones et al. 1992). The tree with the highest log likelihood (-2717.1674) is shown. The percentage of trees in which the associated branch clustered together in the bootstrap test (1,000 replicates) is shown next to the branches. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+*G*, parameter = 1.0702)). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site.

APPENDIX 2: Supplementary materials of Chapter 4

S allele	S_f	S_5	S_8	S_7	S_9	S_3	S_{23}	S_{25}
S_5	28	-						
S_8	25	29	-					
S_7	29	26	35	-				
S_9	28	22	33	35	-			
S_3	32	23	30	36	46	-		
S_{23}	28	29	42	46	32	46	-	
S_{25}	29	29	42	47	32	47	47	-
S_1	37	25	28	28	32	46	47	50

Fig. S2.1 A heat map showing the DNA level sequence identity of nine *S-RNase* alleles. *S* alleles are grouped based on the results from the phylogenetic analysis conducted using the deduced amino acid sequences from conserved region 1 to conserved region 5 of the *S-RNase* gene using the method described in Chapter 3.

S allele	S_8	S_7	S_f	S_{23}	S_3	S_{25}
S_7	39	-				
S_f	33	35	-			
S_{23}	38	37	38	-		
S_3	30	39	39	79	-	
S_{25}	38	36	37	83	85	-
S_5	29	36	38	81	81	76

Fig. S2.2 A heat map showing the DNA level sequence identity of seven *SFB* alleles. *S* alleles are grouped based on the results from the phylogenetic analysis conducted in Chapter 3.

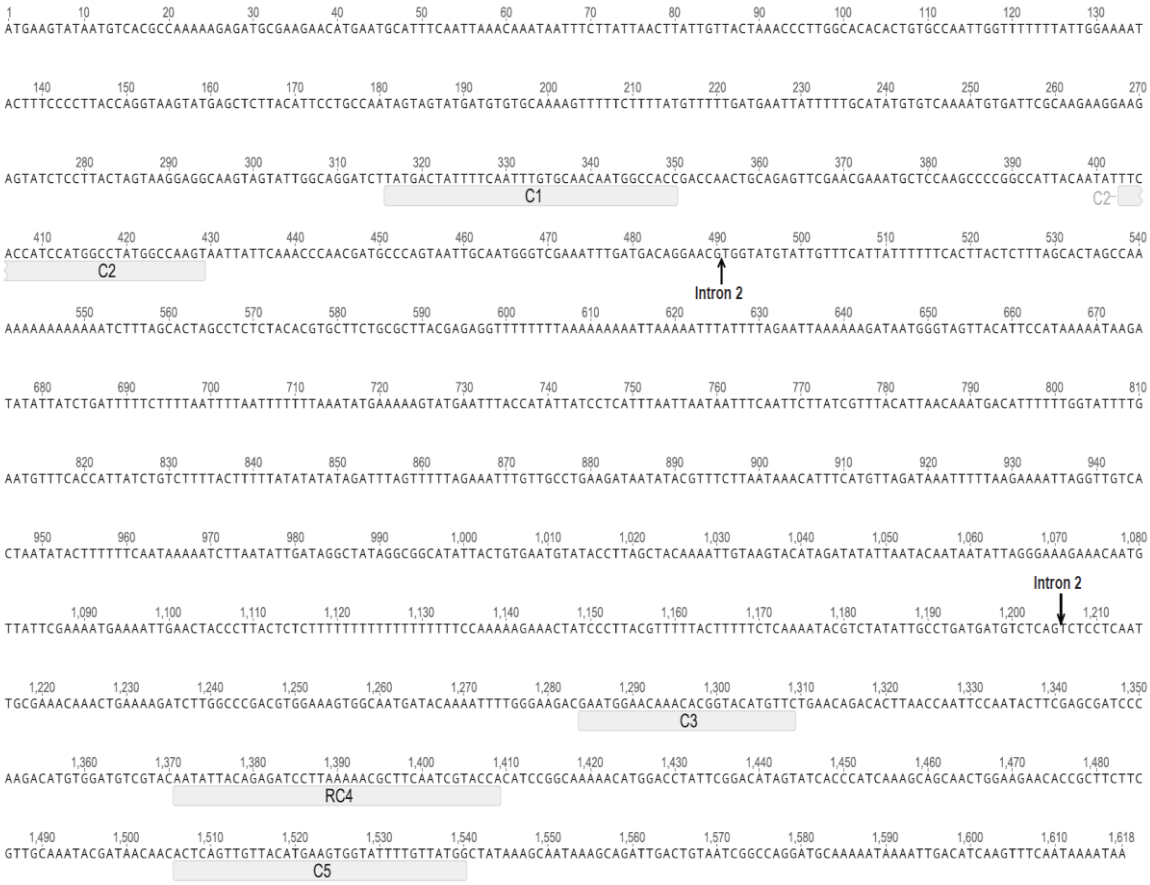


Fig. S2.3 The *S-RNase* gene sequence of the S_3 allele from the almond cultivar, Lauranne. Conserved regions C1, C2, C3, RC4, C5, start and end positions of intron 2 in the *S-RNase* gene are labelled.

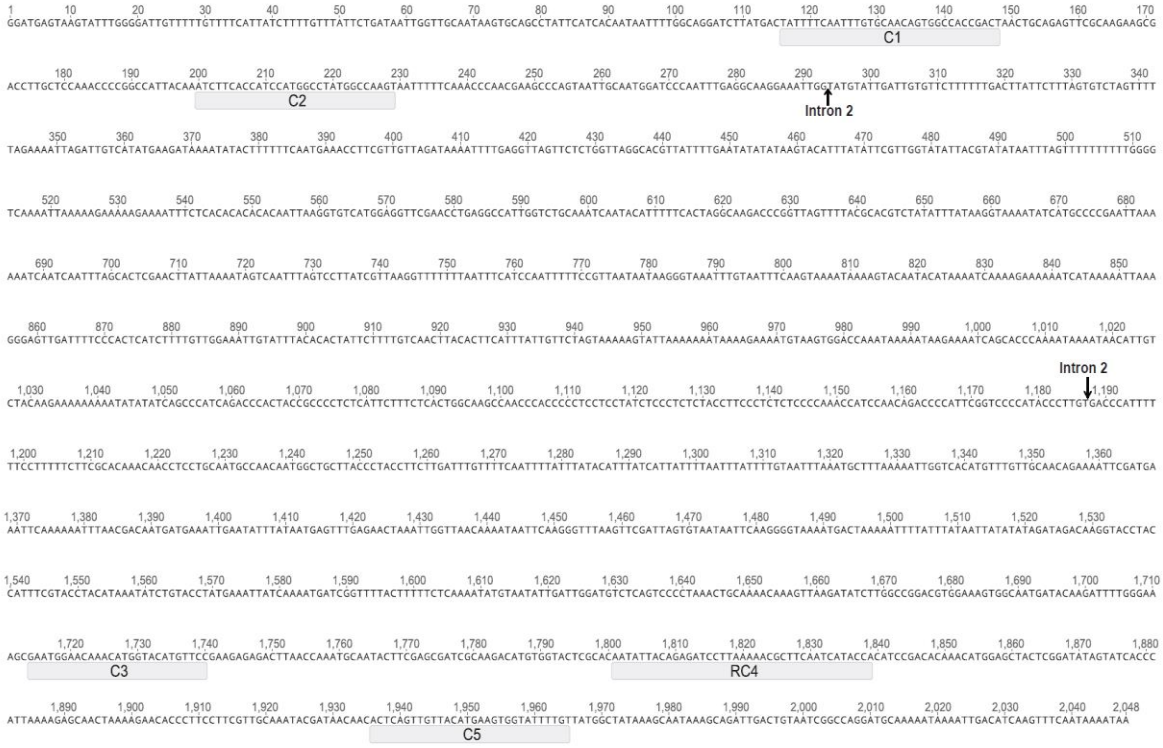


Fig. S2.4 The *S-RNase* gene sequence of the S_9 allele from the almond cultivar, Vairo. Conserved regions C1, C2, C3, RC4, C5, start and end positions of intron 2 in the *S-RNase* gene are labelled.

Fig. S2.5 The *S-RNase* gene sequence of the S_{25} allele from the almond cultivar, Johnston. Conserved regions C1, C2, C3, RC4, C5, start and end positions of intron 2 in the *S-RNase* gene are labelled.

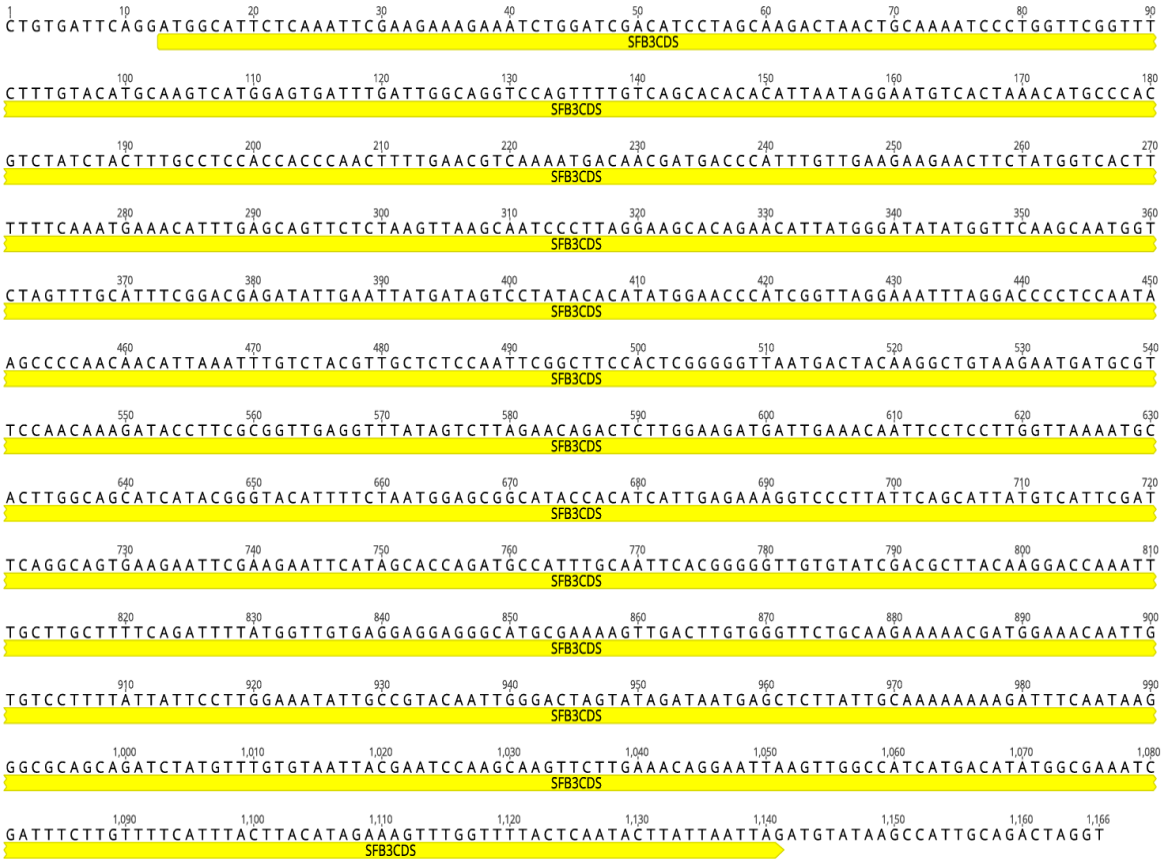


Fig. S2.6 The *SFB* gene sequence of the S_3 allele from the almond cultivar, Lauranne. CDS are in yellow.

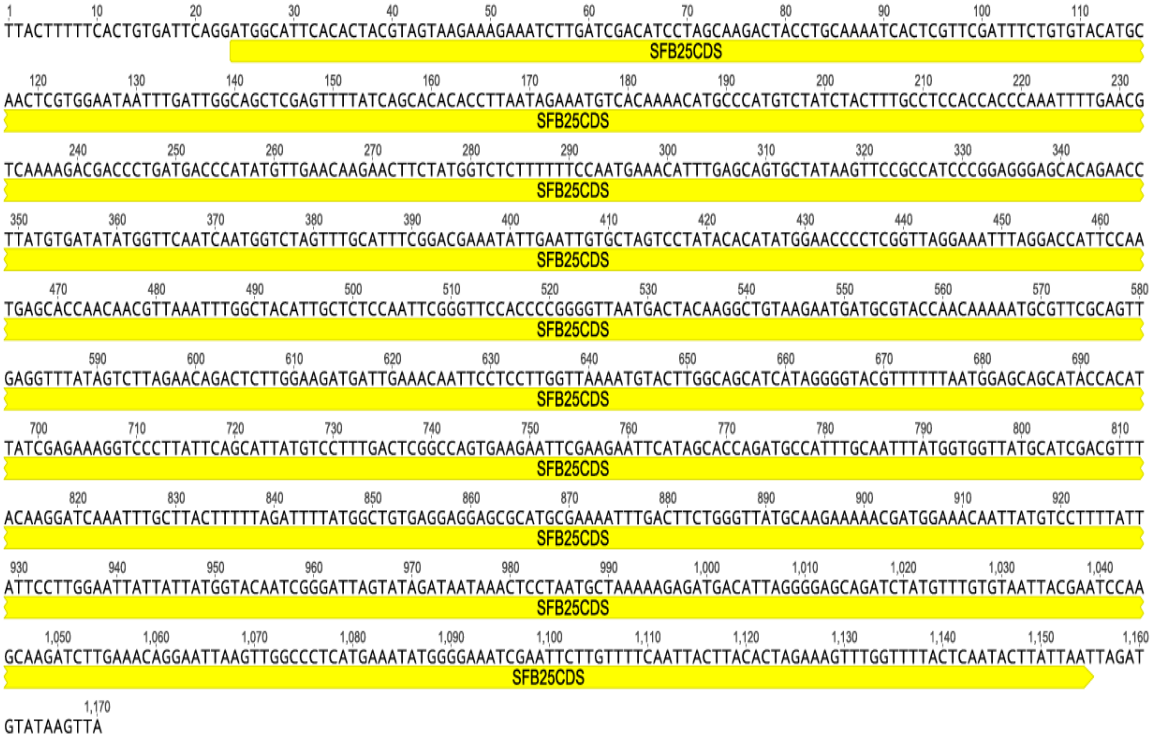


Fig. S2.7 The *SFB* gene sequence of the S_{25} allele from the almond cultivar, Johnston. CDS are in yellow.

APPENDIX 3: Supplementary materials of Chapter 5

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

Primer	Sequence (5'-3')
Forward	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT
Reverse	CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATCT

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

Primer	SNP-bearing GBS Tag	Allele-specific primer sequences (5'-3')[1]	Common primer sequence (5'-3')	Nonpareil genotype	Lauranne genotype
WriPdK0001	TP13153	GAAGGTGACCAAGTTCAT**GCT**CGGCTTCACGGCCCAC GAAGGTCGGAGTCAACGGATTGCTCGGCTTCACGGCCCAT	TCAAACCGGAAACGCCTTAGACGTT	C:T	T:T
WriPdK0002	TP15291	GAAGGTGACCAAGTTCATGCTCCACCACCATAACTGCCACCTTT GAAGGTCGGAGTCAACGGATTCACCACCATAACTGCCACCTTC	GCTATTGGTTCCGGTGGTGGTGTT	A:G	G:G
WriPdK0003	TP24404	GAAGGTGACCAAGTTCATGCTGCCGGTAAAGACCAAGTAATGGTTTT GAAGGTCGGAGTCAACGGATTCCGGTAAAGACCAAGTAATGGTTTC	CTGCAAGGAAGAGAGTTGAAGAGAGA	A:G	G:G
WriPdK0004	TP29093	GAAGGTGACCAAGTTCATGCTCACAGAAGTTCTTTGAGCCCTTTAC GAAGGTCGGAGTCAACGGATTCCACAGAAGTTCTTTGAGCCCTTTAT	TCAAGGAGATCGACTTTAATGCCTTTCAT	T:C	C:C
WriPdK0005	TP9968	GAAGGTGACCAAGTTCAT**GCT**TGTTGGACTTCGAATCCGAGTCT GAAGGTCGGAGTCAACGGATTCTTAGGAGGAAGCTAAATAAGCGATC	CATGGGAGTGTAAGGCCTGGGAT	A:G	G:G
WriPdK0006	TP16417	GAAGGTGACCAAGTTC**ATGCT**TGTTGGACTTCGAATCCGAGTCT GAAGGTCGGAGTCAACGGA**TT**GTTGGACTTCGAATCCGAGTCG	GCGTGAATTGAGTGGGACATTGCAA	A:C	C:C

Table S3.2., continued.

Primer	SNP-bearing GBS Tag	Allele-specific primer sequences (5'-3')[1]	Common primer sequence (5'-3')	Nonpareil genotype	Lauranne genotype
WriPdK0007	TP18674	GAAGGTGACCAAGTTCATGC**T**ACTCCTTGCGGACCTCCTTG GAAGGTCGGAGTCAACGGAT**T**ACTCCTTGCGGACCTCCTTC	AGCTCCGCAGAGACTACGCCAA	C:G	G:G
WriPdK0008	TP33068	GAAGGTGACCAAGTTCAT**CT**AATTTCCAATATAAGTTTGGCTGCGAA GAAGGTCGGAGTCAACGGA**TT**CTAATTTCCAATATAAGTTTGGCTGCGAT	CCAATTAGTAGCTCTCTGGCTGGTT	T:A	T:T
WriPdK0009	TP20078	GAAGGTGACCAAGTTCAT**GCT**GCATGCCTTATGGCTGTCAAGG GAAGGTCGGAGTCAACGGAT**T**ATGCATGCCTTATGGCTGTCAAGT	CTTATTGTGTATGAGAGCTGTGAGGTTAT	G:T	T:T
WriPdK0010	TP16298	GAAGGTGACCAAGTTCATGCTCACATAGCAACTCCTCACCCC GAAGGTCGGAGTCAACGGAT**T**CCACATAGCAACTCCTCACCCT	CTTTGACTGGTCAGAAGTAGCTGGTT	C:T	T:T
WriPdK0011	TP3083	GAAGGTGACCAAGTTCATGCTCAGAAGAATAGCAATTCATGCACGGA GAAGGTCGGAGTCAACGGATTAGAAGAATAGCAATTCATGCACGGG	AAGTTGGTACCTTTCTTGGCCCACT	A:G	G:G
WriPdK0012	TP6073	GAAGGTGACCAAGTTCATGCTGCCCATCACGAAGAGCAACAAC GAAGGTCGGAGTCAACGGATTGCCCATCACGAAGAGCAACAAG	CAATTAGAGAGGAGGAGAGGTTGACAA	G:C	CC
WriPdK0013	TP14389	GAAGGTGACCAAGTTCATGC**T**CAGCTTGCGGTCGGAGGCT GAAGGTCGGAGTCAACGGATTAGCTTGCGGTCGGAGGCC	GGGAGGTCCCCGAGCTGGTT	A:G	G:G

Table S3.2., continued.

Primer	SNP-bearing GBS Tag	Allele-specific primer sequences (5'-3')[1]	Common primer sequence (5'-3')	Nonpareil genotype	Lauranne genotype
WriPdK0014	TP21192	GAAGGTGACCAAGTTCATGCTAACATCGACACCCTCGACGGT GAAGGTCGGAGTCAACGGATTCATCGACACCCTCGACGGC	AGCTTACGCCGGAACTTCCTCAAAA	A:G	G:G
WriPdK0016	TP34383	GAAGGTGACCAAGTTCATGCTGTTGGCCAAACAGAGGACCAC GAAGGTCGGAGTCAACGGATTCGTTGGCCAAACAGAGGACCAT	CTCTTCCACGTCAAGCACACCGAT	C:T	T:T
WriPdK0017	TP13360	GAAGGTGACCAAGTTCATGCTGTGTCGGTAGGTCAATTGTAAACCT GAAGGTCGGAGTCAACGGATTGTCGGTAGGTCAATTGTAAACCG	CAGCCTGCCACAGGTCTTATCCTT	C:A	A:A
WriPdK0018	TP15987	GAAGGTGACCAAGTTCATG**CT**AGTTAGTTGAAGCTGAGATGGGTTTT GAAGGTCGGAGTCAACGGATTGTTAGTTGAAGCTGAGATGGGTTTC	CAGCGGGTGTTTCGAAACGAAAACTA	A:G	G:G
WriPdK0019	TP17044	GAAGGTGACCAAGTTCATG**CT**GCGGTGGTTGCGAGGGT GAAGGTCGGAGTCAACGGATTGCGGTGGTTGCGAGGGC	CAGTCAGTTGCTGGTGCACCGA	A:G	G:G
WriPdK0020	TP16696	GAAGGTGACCAAGTTCATG**CT**CCGGTTCTTCCACAGAGGAT GAAGGTCGGAGTCAACGGATTCCGGTTCTTCCACAGAGGAG	GTTTTGGGGGAAACCTCTCCGAAAA	T:C	C:C
WriPdK0021	TP39442	GAAGGTGACCAAGTTCATGCTGTCAATGGGCCCTCCTATCACA GAAGGTCGGAGTCAACGGATTCAATGGGCCCTCCTATCACG	CCCTCGAAACATGTCAGGACATGAA	T:C	C:C

Table S3.2., continued.

Primer	SNP-bearing GBS Tag	Allele-specific primer sequences (5'-3')[1]	Common primer sequence (5'-3')	Nonpareil genotype	Lauranne genotype
WriPdK0022	TP3085	GAAGGTGACCAAGTTCATGCTGGCATTTGCTGGTTTCCTCGTATAT GAAGGTCGGAGTCAACGGATTGCATTTGCTGGTTTCCTCGTATAC	CACAGAAGTATAAAGTAACACATCAGCCAT	G:A	A:A
WriPdK0023	TP11297	GAAGGTGACCAAGTTCATGCTGACCTGCTCCGTCACGGCA GAAGGTCGGAGTCAACGGATTACCTGCTCCGTCACGGCG	ATTTGAAGGTTTCGGACTTCGGGCT	A:G	G:G
WriPdK0024	TP15894	GAAGGTGACCAAGTTCATGC**T**TGGCGTCGCCTCTGGACC GAAGGTCGGAGTCAACGGATTGTTGGCGTCGCCTCTGGACT	TCTGCCAAAGACGTAGCACACGTTT	T:C	C:C
WriPdK0025	TP16675	GAAGGTGACCAAGTTCATGCTGTTTGAATATCAAGACTTGATGAAATTGGA GAAGGTCGGAGTCAACGGA**TT**TGAATATCAAGACTTGATGAAATTGGC	ATCATCAACAAAAGAGGAGACCTGCTT	C:A	A:A
WriPdK0026	TP20130	GAAGGTGACCAAGTTCATGCTGGTTGCTCCATTGAGACTATGCG GAAGGTCGGAGTCAACGGA**TT**AGGTTGCTCCATTGAGACTATGCT	CTCCTAGCCAGCTCCTTTTGCAAAT	G:T	T:T
WriPdK0027	TP27369	GAAGGTGACCAAGTTCATGCTATAGCCGATGATGATCCGAAGTCT GAAGGTCGGAGTCAACGGAT**T**AGCCGATGATGATCCGAAGTCG	CATACAGCTTCCGCCGTTGATCAAT	A:C	C:C
WriPdK0028	TP28885	GAAGGTGACCAAGTTCATGC**T**GGTTATTAGTATTAGTTGGTGAAGTTGT GAAGGTCGGAGTCAACGGAT**T**GGTTATTAGTATTAGTTGGTGAAGTTGG	GGCTTGTGAGCCTCAAGCCACAA	A:C	C:C

Table S3.2., continued.

Primer	SNP-bearing GBS Tag	Allele-specific primer sequences (5'-3')[1]	Common primer sequence (5'-3')	Nonpareil genotype	Lauranne genotype
WriPdK0029	TP38441	GAAGGTGACCAAGTTCATGCTCCACTTGTTCTTCCTCACCGAG GAAGGTCGGAGTCAACGGATTCCACTTGTTCTTCCTCACCGAC	GGTGGAGAATGGCATCGTTGAACAT	G:C	C:C
WriPdK0030	TP4872	GAAGGTGACCAAGTTCATGC**T**CGAAACAGGAAATCCGGCACTC GAAGGTCGGAGTCAACGGAT**T**CGAAACAGGAAATCCGGCACTG	GATTGAGCCGCGCCAATATCCCTT	G:C	C:C
WriPdK0031	TP9584	GAAGGTGACCAAGTTCATG**CT**GCAAGGCTTCCTGCAACAAAC GAAGGTCGGAGTCAACGGA**TT**CTGCAAGGCTTCCTGCAACAAAA	CAGCCAGTGTCCTTGGAGCAAATTT	G:T	T:T
WriPdK0032	TP14993	GAAGGTGACCAAGTTCATGCTAAGGTTGCTGTACCACTTAGTATACTT GAAGGTCGGAGTCAACGGATTAAGGTTGCTGTACCACTTAGTATACTA	GCCCAACAAGAAGTACTGGGGAATT	T:A	A:A
WriPdK0033	TP861	GAAGGTGACCAAGTTCATGC**T**TGGGTTGCCCTGGACCACTAT GAAGGTCGGAGTCAACGGA**TT**GGGTTGCCCTGGACCACTAG	AAAGGCTGGCTACGTGAACGTGTAA	A:C	C:C
WriPdK0034	TP9301	GAAGGTGACCAAGTTCATG**CT**TTGCCTTTTGCTCAATAGCACC GAAGGTCGGAGTCAACGGATTCCTTTGCCTTTTGCTCAATAGCACT	GGTGTTGATAGGGAAGGAGCTCAAA	C:T	T:T
WriPdK0035	TP20502	GAAGGTGACCAAGTTCATGCTGAAACTGCCGAGATTCGCTTACT GAAGGTCGGAGTCAACGGATTGAAACTGCCGAGATTCGCTTACC	ATCGCCCCTGGAGCCCGATAT	G:A	A:A

Table S3.2., continued.

Primer	SNP-bearing GBS Tag	Allele-specific primer sequences (5'-3')[1]	Common primer sequence (5'-3')	Nonpareil genotype	Lauranne genotype
WriPdK0036	TP26236	GAAGGTGACCAAGTTCATGC**T**GTCACCAATACAAGAGGAGGAC GAAGGTCGGAGTCAACGGATTAGTGTCACCAATACAAGAGGAGGAT	TATGCAGAAGACGAAACCGGTGGAA	C:T	T:T
WriPdK0037	TP1517	GAAGGTGACCAAGTTCATGC**T**GGTCGTTGATGTGCTTTCCGTG GAAGGTCGGAGTCAACGGA**TT**GGTCGTTGATGTGCTTTCCGTC	GCAACAATGCCTTGGAGAACTCCTT	C:G	G:G
WriPdK0038	TP5970	GAAGGTGACCAAGTTCATGCTCCTGCTTTGAGACTTTTGTCCTTG GAAGGTCGGAGTCAACGGA**TT**CCCTGCTTTGAGACTTTTGTCCTTA	TGAGGAAGGCAAAGGCAAGAGCAAA	C:T	T:T
WriPdK0039	TP11568	GAAGGTGACCAAGTTCATGCTAGCCTTGATTTTCGATCCACTTGCA GAAGGTCGGAGTCAACGGATTCCTTGATTTTCGATCCACTTGCG	AGCCTGTCAAGGGTCTGGATAAGTT	G:A	A:A
WriPdK0040	TP15824	GAAGGTGACCAAGTTCATGCTCCGCCATCGGAGTAATCAGCT GAAGGTCGGAGTCAACGGATTCCGCCATCGGAGTAATCAGCG	TACGGTGGTGGTGGCGGCTAT	A:C	C:C
WriPdK0041	TP17599	GAAGGTGACCAAGTTCATG**CT**TTAGAGAACAGAAATGCCTCCAATGT GAAGGTCGGAGTCAACGGA**TT**TAGAGAACAGAAATGCCTCCAATGA	CAGCTATCAACAGAGCAGTTGCCAA	A:T	T:T
WriPdK0042	TP9629	GAAGGTGACCAAGTTCATGCTCCATAATTTGACTGGTGTGGTACTTA GAAGGTCGGAGTCAACGGATTCCATAATTTGACTGGTGTGGTACTTG	GGACAACTGACCATTAAAATTTGTGGAGAT	G:G	A:G

Table S3.2., continued.

Primer	SNP-bearing GBS Tag	Allele-specific primer sequences (5'-3')[1]	Common primer sequence (5'-3')	Nonpareil genotype	Lauranne genotype
WriPdK0043	TP15654	GAAGGTGACCAAGTTCATGCTCCACCGGAGGCATGCACAACT GAAGGTCGGAGTCAACGGATTCACCGGAGGCATGCACAACG	CAGCGGCCAACAACGCCAACAA	C:C	A:C
WriPdK0044	TP17917	GAAGGTGACCAAGTTCATGC**T**CGCCGCCCGCGGCC GAAGGTCGGAGTCAACGGATTGTCGCCGCCCGCGGCT	CCGAGGAGGACGGTGGCGA	C:C	T:C
WriPdK0045	TP19998	GAAGGTGACCAAGTTCAT**GCT**GAGTTACCTGCAAATGCATATCCA GAAGGTCGGAGTCAACGGATTCTGAGTTACCTGCAAATGCATATCCT	CTACAGCCATGCATTTTGGTGGTCTT	T:T	A:T
WriPdK0046	TP31655	GAAGGTGACCAAGTTCATGCTGGTGTTTGCTTGAAATGCCAAACC GAAGGTCGGAGTCAACGGATTCGGTGTTTGCTTGAAATGCCAAACT	CACACAAAAAGAGAAGCAATTCTTGGCTT	C:C	T:C
WriPdK0047	TP19885	GAAGGTGACCAAGTTCATGCTACTCCTGTTCTGTGCACAACCG GAAGGTCGGAGTCAACGGATTGACTCCTGTTCTGTGCACAACCT	GAGACACAAGCCCACTGAGACATTT	T:T	G:T
WriPdK0048	TP22397	GAAGGTGACCAAGTTCATGC**T**GGTGCTTAAACCCTTAAACCCCTA GAAGGTCGGAGTCAACGGA**TT**GGTGCTTAAACCCTTAAACCCCTT	GTCTACGAGACTTTGATAGTTTGGTTGAT	T:T	A:T
WriPdK0049	TP36168	GAAGGTGACCAAGTTCATGCTCAGCTGTGTTATACATTCGAGTGCA GAAGGTCGGAGTCAACGGATTAGCTGTGTTATACATTCGAGTGCG	CTTTGACATCCGACACTGTTGAGCTT	G:G	A:G

Table S3.2., continued.

Primer	SNP-bearing GBS Tag	Allele-specific primer sequences (5'-3')[1]	Common primer sequence (5'-3')	Nonpareil genotype	Lauranne genotype
WriPdK0050	TP38620	GAAGGTGACCAAGTTCAT**GCT**CATCATCTGACAAAACCCCACGA GAAGGTCGGAGTCAACGGATTCATCATCTGACAAAACCCCACGC	ATCATCTGTCACTTGAAAAATCAGAGCAA	C:C	A:C
WriPdK0051	TP6571	GAAGGTGACCAAGTTCATGCTAGGAACCCGGCGGACCGT GAAGGTCGGAGTCAACGGATTGGAACCCGGCGGACCGG	CAGCATCAATACCACGTCCCTCAT	C:C	A:C
WriPdK0052	TP36685	GAAGGTGACCAAGTTCATG**CT**GACATTGATGTGCCCAAGCCCTA GAAGGTCGGAGTCAACGGAT**T**GACATTGATGTGCCCAAGCCCTT	CTTTGCTTTTGTGTGGTGTTTCCCTTAAA	A:A	T:A
WriPdK0053	TP357	GAAGGTGACCAAGTTCATGCTCAAGTGCACCCAAAGAATTGGTAC GAAGGTCGGAGTCAACGGATTCCAAGTGCACCCAAAGAATTGGTAT	GCCGGCAGAATATCAGGGGGAT	C:C	T:C
WriPdK0054	TP8785	GAAGGTGACCAAGTTCATGCTCAAAACTTTTGGATGTGGAGTTGCAG GAAGGTCGGAGTCAACGGATTCAAAACTTTTGGATGTGGAGTTGCAC	CACAGATTGAGTCATAACTCCTCTCAT	C:C	G:C
WriPdK0055	TP14071	GAAGGTGACCAAGTTCATGCTGTGGTGTGACTGAAATGGTGGAC GAAGGTCGGAGTCAACGGATTGTGGTGTGACTGAAATGGTGGAG	CCACCGGAAAAAGTTCCAGGATCTT	G:G	C:G
WriPdK0056	TP18255	GAAGGTGACCAAGTTCATGC**T**CTCTTTGTCACTGCTCTTTCTTGC GAAGGTCGGAGTCAACGGATTATCTCTTTGTCACTGCTCTTTCTTGT	CCATGAGCTTGGCCTGAGAGGAA	T:T	C:T

Table S3.2., continued.

Primer	SNP-bearing GBS Tag	Allele-specific primer sequences (5'-3')[1]	Common primer sequence (5'-3')	Nonpareil genotype	Lauranne genotype
WriPdK0057	TP24705	GAAGGTGACCAAGTTCATGCTGGCTGGAAGCGGAACATTAGAAG GAAGGTCGGAGTCAACGGATTAAGGCTGGAAGCGGAACATTAGAAA	TGCAATATTTTCCTGTTAGGAGCAATTGAT	T:T	C:T
WriPdK0058	TP29336	GAAGGTGACCAAGTTC**ATGCT**CCAAAGGTTTCAACTGGAGCT GAAGGTCGGAGTCAACGGATTCTCCAAAGGTTTCAACTGGAGCC	TGCCACTAAACCTGACATTGTTGACAA	A:A	G:A
WriPdK0059	TP11093	GAAGGTGACCAAGTTCATGC**T**GTTTGCTTCCAATTTCCTGCTGAG GAAGGTCGGAGTCAACGGA**TT**GTTTGCTTCCAATTTCCTGCTGAT	ATCTTGTCATCACGATTCATAAATTGCCTT	G:G	T:G
WriPdK0060	TP9077	GAAGGTGACCAAGTTCATG**CT**AGCTAGCCAGGTTAACGTCAG GAAGGTCGGAGTCAACGGATTCTAGCTAGCCAGGTTAACGTCAC	CACGCATATAGACACACACATCATGTTTA	G:G	C:G
WriPdK0061	TP15284	GAAGGTGACCAAGTTCAT**GCT**GGAGGAGAAATTGGGGCCTTGT GAAGGTCGGAGTCAACGGATTGAGGAGAAATTGGGGCCTTGC	CGCTAGGCCGTTGCCATGACAA	A:A	G:A
WriPdK0062	TP29143	GAAGGTGACCAAGTTCAT**GCT**TGCAACGCAAGCACCCTCCA GAAGGTCGGAGTCAACGGA**TT**GCAACGCAAGCACCCTCCG	CGCCACCTGTAAAAGGAGGGTGAA	A:A	G:A
WriPdK0063	TP37734	GAAGGTGACCAAGTTCATG**CT**TGATTCGGTCAAGAATGAGAATTTG GAAGGTCGGAGTCAACGGAT**T**CCTTGATTCGGTCAAGAATGAGAATTTA	TGAGGTTGTGGTTGTGGCTGTTGTT	C:C	T:C

Table S3.2., continued.

Primer	SNP-bearing GBS Tag	Allele-specific primer sequences (5'-3')[1]	Common primer sequence (5'-3')	Nonpareil genotype	Lauranne genotype
WriPdK0064	TP9083	GAAGGTGACCAAGTTCATGC**T**CTCTTACGCATGCTTAGCAATGTTT GAAGGTCGGAGTCAACGGAT**T**CTCTTACGCATGCTTAGCAATGTTC	CTTCCATCTTCTGGAAACCGGACTT	G:G	A:G
WriPdK0065	TP16653	GAAGGTGACCAAGTTCATGCTGTCTCATTGGCCTTCTCGGCT GAAGGTCGGAGTCAACGGATTGGTGAGATTAAGTAGGACAGGGA	CAGCGTTTCACTTGTCGACTCTCAA	C:C	T:C
WriPdK0066	TP33954	GAAGGTGACCAAGTTCATGCTGTCTCATTGGCCTTCTCGGCT GAAGGTCGGAGTCAACGGATTGTCTCATTGGCCTTCTCGGCC	GCCAAGGAGGCCAGAGACACAA	A:A	G:A
WriPdK0067	TP36749	GAAGGTGACCAAGTTCATGC**T**GCCTTCAAGACTGTAAAAACCAACAG GAAGGTCGGAGTCAACGGA**TT**GCCTTCAAGACTGTAAAAACCAACAC	CCTTGCTTTTCAAGACGAGGGCAAA	G:G	C:G
WriPdK0068	TP5689	GAAGGTGACCAAGTTCATGC**T**TTCAGTTCCAGAGTCCAATCTCAG GAAGGTCGGAGTCAACGGAT**T**GTTTCAGTTCCAGAGTCCAATCTCAA	GTACTAGCTTCCTCAACTGAGTAAACTAT	C:C	T:C
WriPdK0069	TP10003	GAAGGTGACCAAGTTCATGCTCGAAGTTGCTGGAAAAAATGCCATG GAAGGTCGGAGTCAACGGATTCGAAGTTGCTGGAAAAAATGCCATC	CCATGTCCCTTGATTTTCCGGTGAT	C:C	G:C
WriPdK0070	TP17172	GAAGGTGACCAAGTTCATGCTAGGTTGCAGTATAGGAGTGGCC GAAGGTCGGAGTCAACGGATTAAGGTTGCAGTATAGGAGTGGCT	GACCCCACTCACCACCCTGTT	T:T	C:T

Table S3.2., continued.

Primer	SNP-bearing GBS Tag	Allele-specific primer sequences (5'-3')[1]	Common primer sequence (5'-3')	Nonpareil genotype	Lauranne genotype
WriPdK0071	TP25503	GAAGGTGACCAAGTTCATG**CT**CTCTCTCAAGCCATGCAAGCAT GAAGGTCGGAGTCAACGGATT**T**CTCTCTCAAGCCATGCAAGCAC	GCCATGAGCTTAGGACTTGTTTGGAT	A:A	G:A
WriPdK0072	TP1262	GAAGGTGACCAAGTTCATGCTAAAGCTGAACTCTTCTCTGCCCTTT GAAGGTCGGAGTCAACGGATTAAAGCTGAACTCTTCTCTGCCCTTA	CCGGGGAAGCTTGGTCTTCCTT	T:T	A:T
WriPdK0073	TP7601	GAAGGTGACCAAGTTCATGCTGCAGCATACTATGGTACCCAGAG GAAGGTCGGAGTCAACGGATTGCAGCATACTATGGTACCCAGAC	GTGCATATAGTAAAAATACTCACCAGCCAT	C:G	C:C
WriPdK0074	TP2219	GAAGGTGACCAAGTTCATGC**T**CCTGCTGTGTTTATTGGTGGTGAT GAAGGTCGGAGTCAACGGATTCTGCTGTGTTTATTGGTGGTGAG	CAATAACACTTTCCAATCCCCCAAACAA	A:A	C:A
WriPdK0075	TP2822	GAAGGTGACCAAGTTCATGCTAAATTTCCATCCCTATCCTCCATGC GAAGGTCGGAGTCAACGGATTAAATTTCCATCCCTATCCTCCATGG	TGAAAGTTAGTGTGGCTTAGGTTGTCAAT	C:C	G:C
WriPdK0076	TP4478	GAAGGTGACCAAGTTCATGCTCCGAAGCCCACCACCACAGAT GAAGGTCGGAGTCAACGGATTCGAAGCCCACCACCACAGAC	AGCAGAAACAGAAACAGAAAATCATTCCAT	A:A	G:A
WriPdK0077	TP7440	GAAGGTGACCAAGTTCATGCTACCTGTCACGCACGTTATCAGC GAAGGTCGGAGTCAACGGATTAACCTGTCACGCACGTTATCAGA	GTGGTGCAAGTGGAGCGCTGTA	G:T	T:T

Table S3.2., continued.

Primer	SNP-bearing GBS Tag	Allele-specific primer sequences (5'-3')[1]	Common primer sequence (5'-3')	Nonpareil genotype	Lauranne genotype
WriPdK0078	TP8457	GAAGGTGACCAAGTTCATG**CT**ACTCGAGTTATGATTAAAGTAAGACAAG	CTCGCAGGCTGACCGCGCTT	C:C	T:C
		GAAGGTCGGAGTCAACGGATTCTACTCGAGTTATGATTAAAGTAAGACAAA			
WriPdK0079	TP11674	GAAGGTGACCAAGTTCATGC**T**ATCAACATTTCGGCCCGACCG	CTGGGGCCCACGAAAGAGGATA	G:G	C:G
		GAAGGTCGGAGTCAACGGA**TT**ATCAACATTTCGGCCCGACCC			
WriPdK0080	TP12629	GAAGGTGACCAAGTTCATGC**T**GAGGTCTCAAGTGTTTCTGCAAG	GCTCTTGCCGAAGACACCCCAA	G:G	C:G
		GAAGGTCGGAGTCAACGGAT**T**GAGGTCTCAAGTGTTTCTGCAAC			
WriPdK0081	TP11526	GAAGGTGACCAAGTTCATG**CT**GTAGAGTGGGGTTTCTCCC	CTTTGGTGGAGATTAGGAACCGCTT	T:T	C:T
		GAAGGTCGGAGTCAACGGATTCCTGTAGAGTGGGGTTTCTCCT			
WriPdK0082	TP15496	GAAGGTGACCAAGTTCATGCTCAGCCTATGAAGAGGCGTTCG	TCGAGCGCCGCCGCTCACA	G:T	T:T
		GAAGGTCGGAGTCAACGGAT**T**CCAGCCTATGAAGAGGCGTTCT			
WriPdK0083	TP16647	GAAGGTGACCAAGTTCATGC**T**TTAGGAATGGTGGCGGCAACTA	AAGAAGCCAAGAGCCTTGGGATCAA	A:G	G:G
		GAAGGTCGGAGTCAACGGA**TT**AGGAATGGTGGCGGCAACTG			
WriPdK0084	TP16703	GAAGGTGACCAAGTTCATGCTAGTTGGGTAGGCATTGGATTCG	GCTGTGTTGGCAGTTAATCGCCAAA	C:G	G:G
		GAAGGTCGGAGTCAACGGATTAGTTGGGTAGGCATTGGATTCC			

Table S3.2., continued.

Primer	SNP-bearing GBS Tag	Allele-specific primer sequences (5'-3')[1]	Common primer sequence (5'-3')	Nonpareil genotype	Lauranne genotype
WriPdK0085	TP18700	GAAGGTGACCAAGTTCATGCTCAGCTCCGGTGGTCTTCCA GAAGGTCGGAGTCAACGGATTCAGCTCCGGTGGTCTTCCG	TCTCCGGCTGGCTTCCCCAAA	A:G	G:G
WriPdK0086	TP1933	GAAGGTGACCAAGTTCATGCTCAAGGAGAGGCAGTCAATTGTGC GAAGGTCGGAGTCAACGGATTCAAGGAGAGGCAGTCAATTGTGG	CACAAGGGGCTAAATTCGAGTTGACTT	G:G	C:G
WriPdK0087	TP20147	GAAGGTGACCAAGTTCATGC**T**CGGAAGTCCAAACTCCAAAGCG GAAGGTCGGAGTCAACGGAT**T**CGGAAGTCCAAACTCCAAAGCC	GAAACTCTTTGACTTTTGACCATTCGGAA	C:G	G:G
WriPdK0088	TP31870	GAAGGTGACCAAGTTCATGCTGTAGATTTTACCTAGCCATGTCACG GAAGGTCGGAGTCAACGGATTCGTAGATTTTACCTAGCCATGTCACA	CCTAACATGAGCACATTGAGGTACAATT	G:A	A:A
WRIPdK0089	TP34218	GAAGGTGACCAAGTTCATGC**T**GGCGGATCAAACCACTTTAGCT GAAGGTCGGAGTCAACGGAT**T**GGCGGATCAAACCACTTTAGCG	CTGCGGCTGGCCACATACCATT	C:C	A:C
WriPdK0090	TP35046	GAAGGTGACCAAGTTCATGCTGAGCCTGAGGAGCCAAAACAC GAAGGTCGGAGTCAACGGATTGGAGCCTGAGGAGCCAAAACAA	GCTAACTGGAGGGATACCTTGTCTT	G:G	T:G
WriPdK0091	TP36434	GAAGGTGACCAAGTTCA**TGCT**CCCAACTGCCTCAACTGCC GAAGGTCGGAGTCAACGGATTCTCCCAACTGCCTCAACTGCT	GAGCAGTTGGCATGTATGGGCAATT	C:C	T:C

Table S3.2., continued.

Primer	SNP-bearing GBS Tag	Allele-specific primer sequences (5'-3')[1]	Common primer sequence (5'-3')	Nonpareil genotype	Lauranne genotype
WriPdK0092	TP37439	GAAGGTGACCAAGTTCAT**GCT**TGTTTCTGCAAGTTTTTCCAGCT GAAGGTCGGAGTCAACGGATTCTTGTTTCTGCAAGTTTTTCCAGCC	AAGAGAGAGCAGGAGCGCCAGA	A:A	G:A
WriPdK0093	TP37731	GAAGGTGACCAAGTTCATG**CT**AGCAACTCCACAGGCATGCT GAAGGTCGGAGTCAACGGAT**T**AGCAACTCCACAGGCATGCC	CTGAGGTTCTTGGGTTCTTCCAAGTT	C:C	T:C
WriPdK0094	TP37835	GAAGGTGACCAAGTTCA**TGCT**GATGAGAATGATGCCCACCA GAAGGTCGGAGTCAACGGATTCTGATGAGAATGATGCCCACCG	CTCAGTTGTTGAGTTCCAACTTCAATGAA	G:G	A:G
WriPdK0095	TP7187	GAAGGTGACCAAGTTCATGCTCAGCATGCAGATCAAATTTCACTACC GAAGGTCGGAGTCAACGGATTCAGCATGCAGATCAAATTTCACTACG	GGAACCTTACTGGACAGCATTGCAA	C:C	G:C
WriPdK0096	TP8083	GAAGGTGACCAAGTTCATGCTAGATCTTCGACACCATTGAGCATG GAAGGTCGGAGTCAACGGATTAGATCTTCGACACCATTGAGCATA	ATAAAATTTGGGGTGGAGGCCCGA	T:C	C:C
WriPdK0097	TP11059	GAAGGTGACCAAGTTCATGCTAGAGATTTTGACCCTGCCTGTGTT GAAGGTCGGAGTCAACGGATTGAGATTTTGACCCTGCCTGTGTC	CCGCAGCAGCCCCTCTCTCTT	G:G	A:G
WriPdK0098	TP11807	GAAGGTGACCAAGTTCATGC**T**TCGGGAAATACGAGCTGGGC GAAGGTCGGAGTCAACGGAT**T**GTTCGGGAAATACGAGCTGGGA	CGCCGCAGCCGAGGAGCTT	T:G	G:G

Table S3.2., continued.

Primer	SNP-bearing GBS Tag	Allele-specific primer sequences (5'-3')[1]	Common primer sequence (5'-3')	Nonpareil genotype	Lauranne genotype
WriPdK0099	TP13975	GAAGGTGACCAAGTTCATG**CT**TGCAGCCTTTATCCAGGCCG GAAGGTCGGAGTCAACGGAT**T**CTTGCAGCCTTTATCCAGGCCT	GATCTAGGATCGACGACGAAGACTT	T:T	T:G
WriPdK0100	TP13979	GAAGGTGACCAAGTTCATG**CT**GTGCAGCCTTTCATTCTATGTGA GAAGGTCGGAGTCAACGGAT**T**GTGCAGCCTTTCATTCTATGTGG	GTCTCTCCCTAACCGGCGATCAA	A:A	G:A
WriPdK0101	TP20195	GAAGGTGACCAAGTTCATG**CT**CAATGTTCTCAAGCTGAACCTCG GAAGGTCGGAGTCAACGGATTCTCAATGTTCTCAAGCTGAACCTCA	GTTCCATGCAGCTGGAATTTCTGCTT	T:T	T:C
WriPdK0102	TP23086	GAAGGTGACCAAGTTCATG**CT**GGTGTGAGAGAGTTTGGTCCT GAAGGTCGGAGTCAACGGAT**T**GGTGTGAGAGAGTTTGGTCCC	ATCTTCTCCACAGGTCGCAGCTTTT	G:A	A:A
WriPdK0103	TP23331	GAAGGTGACCAAGTTCATGC**T**TAATTGAAGCTGCAAACATTGTCCCA GAAGGTCGGAGTCAACGGA**TT**AATTGAAGCTGCAAACATTGTCCCT	CATAACCAAACATGCCAAGGACCCAA	T:T	T:A
WriPdK0104	TP27616	GAAGGTGACCAAGTTCATGC**T**GACTGGTTACTTAACTGGAAGCACT GAAGGTCGGAGTCAACGGATTACTGGTTACTTAACTGGAAGCACC	CTGAATGGGCTGCATCAGCAATGAT	G:G	G:A
WriPdK0105	TP33011	GAAGGTGACCAAGTTCATG**CT**TGCTGCGAAACCATCCTCGAAAA GAAGGTCGGAGTCAACGGA**TT**GCTGCGAAACCATCCTCGAAAC	GTTGCTTGGCGCAAGGGCTCAA	C:C	C:A

Table S3.2., continued.

Primer	SNP-bearing GBS Tag	Allele-specific primer sequences (5'-3')[1]	Common primer sequence (5'-3')	Nonpareil genotype	Lauranne genotype
WriPdK0106	TP33099	GAAGGTGACCAAGTTCATGCTCAGTAGTCTCATAGACCTGTGGTT GAAGGTCGGAGTCAACGGATTAGTAGTCTCATAGACCTGTGGTG	CGTGAGGCTATGTGGCTGCGAA	C:A	A:A
WriPdK0107	TP36190	GAAGGTGACCAAGTTCATGCTAAGCTGCTCAGTCAACGCCAC GAAGGTCGGAGTCAACGGATTGAAGCTGCTCAGTCAACGCCAT	CATATGGTCAACACTGGGAGTGTGAT	T:T	T:C
WriPdK0108	TP38827	GAAGGTGACCAAGTTCATG**CT**CTGTGCGTTGTTGCCTCCG GAAGGTCGGAGTCAACGGA**TT**CTCTGTGCGTTGTTGCCTCCA	AGAGATGGGGCTGCTGTGTTGTTT	T:T	T:C
WriPdK0109	TP4114	GAAGGTGACCAAGTTCAT**GCT**CTTTCGGGAGTATCAAACTCT GAAGGTCGGAGTCAACGGATTCTCTTTCGGGAGTATCAAACTCG	GCTAGAATACACTCCTTGCAGCACTA	C:A	C:C
WriPdK0110	TP4159	GAAGGTGACCAAGTTCATGCTGCACACTATTAAAGGGCTTGTGC GAAGGTCGGAGTCAACGGAT**T**GGCACACTATTAAAGGGCTTGTGA	CAACTGAATAGCAGCACTATTGTTACCTT	T:G	G:G
WriPdK0111	TP5223	GAAGGTGACCAAGTTCATGCTAGTTGCAGCAGGAGAAGTTATCCAA GAAGGTCGGAGTCAACGGA**TT**GCAGCAGGAGAAGTTATCCAG	GCGCAGGCTTAAACATTGTTGGGTT	A:A	G:A
WriPdK0112	TP9254	GAAGGTGACCAAGTTCATGC**T**GCAGCCAGAAAACTTTAGTGTCTC GAAGGTCGGAGTCAACGGA**TT**GCAGCCAGAAAACTTTAGTGTCTT	CGCATGCTTCTGGAGTTTTGTTGTTAAAA	C:C	T:C

Table S3.2., continued.

Primer	SNP-bearing GBS Tag	Allele-specific primer sequences (5'-3')[1]	Common primer sequence (5'-3')	Nonpareil genotype	Lauranne genotype
WriPdK0113	TP38839	GAAGGTGACCAAGTTCATGCTGCAAGGTGAGACGACGCGCT GAAGGTCGGAGTCAACGGATTCAAGGTGAGACGACGCGCC	CCGTCTTCCACGAGCTGCTGTT	G:A	G:G
WriPdK0114	TP15051	GAAGGTGACCAAGTTCATGCTGCCGGAATTCGACGACATCATC GAAGGTCGGAGTCAACGGATTCGCCGGAATTCGACGACATCATT	GAAATCGAGGTCGGTGGCCGTA	T:C	C:C
WriPdK0115	TP20847	GAAGGTGACCAAGTTCATGC**T**AGGTAGGGGACGGCAGCG GAAGGTCGGAGTCAACGGATTGTTAGGTAGGGGACGGCAGCA	CGTGGCAACGACCCGCGTAATA	T:C	C:C
WriPdK0116	TP17304	GAAGGTGACCAAGTTCATGC**T**GTGCTGCCACAAAGCCATCCG GAAGGTCGGAGTCAACGGAT**T**GTGCTGCCACAAAGCCATCCA	CTCAGGAGCAAGTGAGAGCAATGAA	T:C	T:T
WriPdK0117	TP155	GAAGGTGACCAAGTTCATGCTGATCTTGTTGTAGCACACTTGGC GAAGGTCGGAGTCAACGGAT**T**GGATCTTGTTGTAGCACACTTGGA	GGAACAGAGCATCAGCTTGGTCAAA	T:G	T:T
WriPdK0118	TP2738	GAAGGTGACCAAGTTCATGCTCAGCAATTGGCATTATTCTAGGAGGA GAAGGTCGGAGTCAACGGATTAGCAATTGGCATTATTCTAGGAGGG	CTAATAGAGGACCAGTTGCCAAAAATGAA	G:G	G:A
WriPdK0119	TP16180	GAAGGTGACCAAGTTCATGCTAGCAGGGATTGACACGTTGTGG GAAGGTCGGAGTCAACGGATTAGCAGGGATTGACACGTTGTGC	GAGCACATGAGTCACATTCCACCTT	G:G	G:C

Table S3.2., continued.

Primer	SNP-bearing GBS Tag	Allele-specific primer sequences (5'-3')[1]	Common primer sequence (5'-3')	Nonpareil genotype	Lauranne genotype
WriPdK0120	TP19487	GAAGGTGACCAAGTTCATGCTAAATGCCGCTTTGAACCGCTCG GAAGGTCGGAGTCAACGGATTCAAATGCCGCTTTGAACCGCTCA	TGACAGAGGTGGCGCAGCTCTT	C:C	T:C
WriPdK0121	TP35343	GAAGGTGACCAAGTTCATGC**T**GGCGGCAACTTTCTCCATGTCT GAAGGTCGGAGTCAACGGATTGCGGCAACTTTCTCCATGTCG	CCGACGGCCTCCGCTGCTA	C:C	C:A
WriPdK0122	TP37464	GAAGGTGACCAAGTTCATGCTAGCTACGGAAAAACCGGCAACG GAAGGTCGGAGTCAACGGAT**T**CAGCTACGGAAAAACCGGCAACA	AGCTGCTGAACTCGGTCGGGTT	T:C	C:C
WriPdK0123	TP27203	GAAGGTGACCAAGTTCATGC**T**GGCAAGGCTGCAGGCTTCTCT GAAGGTCGGAGTCAACGGATTGCAAGGCTGCAGGCTTCTCC	CAAAGAGCATTTCGGCGCCGCAA	G:A	G:G
WriPdK0124	TP31757	GAAGGTGACCAAGTTCATGC**T**ATCTGACCCCATAAGCCTGGAAT GAAGGTCGGAGTCAACGGAT**T**CTGACCCCATAAGCCTGGAAC	AGCTCGGAGGGCTTGAGGACTT	G:A	G:G
WriPdK0125	TP32029	GAAGGTGACCAAGTTCATGC**T**CAGCATTACTTGCAGCCATTGC GAAGGTCGGAGTCAACGGAT**T**GTCAGCATTACTTGCAGCCATTGA	GACAAATTATGCTAGAGACGTTGCTCATT	T:G	G:G
WriPdK0126	TP24137	GAAGGTGACCAAGTTCATGC**T**CGGAAGTCCAAACTCCAAAGCC GAAGGTCGGAGTCAACGGAT**T**CGGAAGTCCAAACTCCAAAGCG	GAAACTCTTTGACTTTTGACCATTCGGAA	G:C	C:C

Table S3.2., continued.

Primer	SNP-bearing GBS Tag	Allele-specific primer sequences (5'-3')[1]	Common primer sequence (5'-3')	Nonpareil genotype	Lauranne genotype
WriPdK0127	TP9269	GAAGGTGACCAAGTTCATGCTATGGCCTGCGCCATAGCCG GAAGGTCGGAGTCAACGGATTCATGGCCTGCGCCATAGCCA	CCTACATTATGAAGAGCCCGTCCAT	T:C	C:C
WriPdK0128	TP6880	GAAGGTGACCAAGTTCATGC**T**CGCGAAGTAGTCTCTGTGGCC GAAGGTCGGAGTCAACGGAT**T**CGCGAAGTAGTCTCTGTGGCT	GTAGTCTTCCCCAGTCCTCCCAT	G:A	A:A
WriPdK0129	TP23628	GAAGGTGACCAAGTTCATGCTACATAGCAGCTAATAGATGGCGGA GAAGGTCGGAGTCAACGGATTCATAGCAGCTAATAGATGGCGGG	GCCCCAGTCACGCATACTTTTTCTT	T:C	T:T
WriPdK0130	TP10517	GAAGGTGACCAAGTTCATGC**T**TCTTTAAGGGCATCCATATCGACG GAAGGTCGGAGTCAACGG**ATT**TCTTTAAGGGCATCCATATCGACA	GTTGCCCGGCTCCTAACGCAT	T:C	T:T
WriPdK0131	TP11899	GAAGGTGACCAAGTTCATGCTGGAGTGGGAAATTCAAGAGAATAATG GAAGGTCGGAGTCAACGGAT**T**GGGAGTGGGAAATTCAAGAGAATAATA	CTTGGTGCAGCCGCAATGGCAT	C:C	T:C
WriPdK0132	TP12076	GAAGGTGACCAAGTTCA**TGCT**TTGACCACAACCTCCAGCTA GAAGGTCGGAGTCAACGGATTCTTTGACCACAACCTCCAGCTC	CATGAAGGAAGAGCTCGAGATCCTT	C:A	C:C
WriPdK0133	TP12086	GAAGGTGACCAAGTTCATGCTACTGTTCAATGGAAGTGGATGACAC GAAGGTCGGAGTCAACGGATTAAACTGTTCAATGGAAGTGGATGACAA	CTTGATGACCAGCAGTCTCGGATT	G:G	G:T

Table S3.2., continued.

Primer	SNP-bearing GBS Tag	Allele-specific primer sequences (5'-3')[1]	Common primer sequence (5'-3')	Nonpareil genotype	Lauranne genotype
WriPdK0134	TP12907	GAAGGTGACCAAGTTCATGC**T**CATGGAACGGACCAGCATCG GAAGGTCGGAGTCAACGGATT**T**GTCATGGAACGGACCAGCATCA	CTCTTCTCAGCCTACGCCTCCTT	T:T	T:C
WriPdK0135	TP13014	GAAGGTGACCAAGTTCATGCTATGTTCACTGCTTTGCTTCGATCG GAAGGTCGGAGTCAACGGATT**T**CATGTTCACTGCTTTGCTTCGATCA	CACTCCCCCTCTCTATTTCTCTGAT	T:T	T:C
WriPdK0136	TP13612	GAAGGTGACCAAGTTCATGCTCAGCATCAATGAGTGAAGAAGTGACT GAAGGTCGGAGTCAACGGATTAGCATCAATGAGTGAAGAAGTGACC	GATTTCTGCGGGGGCAGCCTTA	A:A	G:A
WriPdK0137	TP14084	GAAGGTGACCAAGTTCATGCTCCCCAAGCCAGCTCCTCCAT GAAGGTCGGAGTCAACGGATTCCCAAGCCAGCTCCTCCAC	GCAGCGAAACTTGACGATGAGCAA	A:A	G:A
WriPdK0138	TP14542	GAAGGTGACCAAGTTCATGC**T**GGGGAGAGGGCAGCGAGG GAAGGTCGGAGTCAACGGATT**T**GGGGAGAGGGCAGCGAGT	CCGCGAAGGCCTTCCATATTTTCAA	G:G	T:G
WriPdK0139	TP14709	GAAGGTGACCAAGTTCATGCTGCCTCAATAGTAACTCTCTCCAAAC GAAGGTCGGAGTCAACGGATTGGCCTCAATAGTAACTCTCTCCAAAT	CATTAAGCAGCAACTGGCATGGACAA	T:C	C:C
WriPdK0140	TP15125	GAAGGTGACCAAGTTCATG**CT**GCGAATCAAGAAGCGCGCC GAAGGTCGGAGTCAACGGATT**T**CTGCGAATCAAGAAGCGCGCT	GGATCGGAGAGAAAAGGAGAACCAA	C:C	T:C

Table S3.2., continued.

Primer	SNP-bearing GBS Tag	Allele-specific primer sequences (5'-3')[1]	Common primer sequence (5'-3')	Nonpareil genotype	Lauranne genotype
WriPdK0141	TP15324	GAAGGTGACCAAGTTCATG**CT**AAACTTATAGTCCATGCCTTTCTCACT GAAGGTCGGAGTCAACGGATTACTTATAGTCCATGCCTTTCTCACC	CAGCGCTGAGTTTGCATGGCGAT	G:A	G:G
WriPdK0142	TP32984	GAAGGTGACCAAGTTCATGCTACTCGTTTGTGTTCTTGACCCCT GAAGGTCGGAGTCAACGGATTCTCGTTTGTGTTCTTGACCCCC	CTGCCTTTTTGGAGCTCGATGACAA	A:A	G:A
WriPdK0143	TP15410	GAAGGTGACCAAGTTCATGCTCGAATAGCAAGATATTGCAGCGG GAAGGTCGGAGTCAACGGATTACCGAATAGCAAGATATTGCAGCGT	AAGAAATTGTTGAGGCATATGGGTGAGTT	G:G	T:G
WriPdK0144	TP1574	GAAGGTGACCAAGTTCATGCTATATAGGAACAGTCCCCATCCAG GAAGGTCGGAGTCAACGGATTCATATAGGAACAGTCCCCATCCAA	GAGCACAGAAGCAGCAAGAAATGCAA	C:C	T:C
WriPdK0145	TP15836	GAAGGTGACCAAGTTCATGCTATCCATGTCTGATCGAGCCCGA GAAGGTCGGAGTCAACGGAT**T**CCATGTCTGATCGAGCCCGG	TCTGGCTGCGGCATCTTGGCTA	G:A	G:G
WriPdK0146	TP15845	GAAGGTGACCAAGTTCATGCTCAGATTGTCCGGCTTCCTGTCT GAAGGTCGGAGTCAACGGATTAGATTGTCCGGCTTCCTGTCC	CGTATGAGCAGCGGCTCGGAAT	A:A	G:A
WriPdK0147	TP16042	GAAGGTGACCAAGTTCATGC**T**GAAGCAGCGGTCGTCAATCTCA GAAGGTCGGAGTCAACGGATTAAGCAGCGGTCGTCAATCTCC	CGTCTTCCTGCTCGTCTCAGCTT	C:A	C:C

Table S3.2., continued.

Primer	SNP-bearing GBS Tag	Allele-specific primer sequences (5'-3')[1]	Common primer sequence (5'-3')	Nonpareil genotype	Lauranne genotype
WriPdK0148	TP33604	GAAGGTGACCAAGTTCATGCTCAGCATTAAACTCTCCATCTCCATC GAAGGTCGGAGTCAACGGATTATCAGCATTAAACTCTCCATCTCCATA	GATTGCTGCGCCACGCAACTCAT	T:T	T:G
WriPdK0149	TP16224	GAAGGTGACCAAGTTCATGCTGGTTTGCTTTGAGAGACCTAATGCT GAAGGTCGGAGTCAACGGATTGTTTGCTTTGAGAGACCTAATGCC	GGGGCAGCGTAGTAAACCAACAAAA	G:A	G:G
WriPdK0150	TP16354	GAAGGTGACCAAGTTCATGCTGGACCAGTATGTTGCAGAGACCT GAAGGTCGGAGTCAACGGATTGGACCAGTATGTTGCAGAGACCA	CGTGCCCTCGCAGCGTCGT	T:A	T:T
WriPdK0151	TP16827	GAAGGTGACCAAGTTCATGCTAAACCCATCACATCTAATACTGGACAT GAAGGTCGGAGTCAACGGATTACCCATCACATCTAATACTGGACAC	TACTTGTATACAAGTTGGCTGTTGTCCAA	G:A	G:G
WriPdK0152	TP1721	GAAGGTGACCAAGTTCATGCTGGCTAAGCAGGAAGACATAATCCTT GAAGGTCGGAGTCAACGGATTGCTAAGCAGGAAGACATAATCCTC	ACCGCTTCCAACAGACTGACCATA	G:G	G:A
WriPdK0153	TP17874	GAAGGTGACCAAGTTCATGCTGAAGGCAGAAACCCACACAACG GAAGGTCGGAGTCAACGGATTGAAGGCAGAAACCCACACAACC	GGCGAGGCAGCTCAACCCAAAA	C:C	G:C
WriPdK0154	TP17942	GAAGGTGACCAAGTTCATGCTACCTTTCTTGGCCCACTCCC GAAGGTCGGAGTCAACGGATTGTACCTTTCTTGGCCCACTCCT	GGGGCAGCACAGAAGAATAGCAATT	T:C	C:C

Table S3.2., continued.

Primer	SNP-bearing GBS Tag	Allele-specific primer sequences (5'-3')[1]	Common primer sequence (5'-3')	Nonpareil genotype	Lauranne genotype
WriPdK0155	TP18235	GAAGGTGACCAAGTTCATGCTCCAAAGCCATCGAAAGCAGCTCA GAAGGTCGGAGTCAACGGATTCAAAGCCATCGAAAGCAGCTCG	GAGCAAACCACCTTCCCCTTTGTAT	G:G	G:A
WriPdK0156	TP18350	GAAGGTGACCAAGTTCATGCTAACAAGGTCTTGGAACTTTAGGCG GAAGGTCGGAGTCAACGGATTCAACAAGGTCTTGGAACTTTAGGCA	CTAAAAGGACCCAAAAGCAGCTCCAA	T:T	T:C
WriPdK0157	TP18466	GAAGGTGACCAAGTTCATGCTGGGCAAGGGGTGGGTCTCAA GAAGGTCGGAGTCAACGGATTGGCAAGGGGTGGGTCTCAC	TCGAGCAACAACCTAGGCGATTCTT	C:C	C:A
WriPdK0158	TP18642	GAAGGTGACCAAGTTCATGCTAACAATGCAGCCCAATCTACG GAAGGTCGGAGTCAACGGATTCTAACAATGCAGCCCAATCTACC	GACCAGCATCTTGCATTGTCTCCTT	G:G	G:C
WriPdK0159	TP18734	GAAGGTGACCAAGTTCATGCTATCAGCAGCTCCTATAGCCGC GAAGGTCGGAGTCAACGGATTCATCAGCAGCTCCTATAGCCGT	CAATGCAAGCGGCAGGCCAGAA	T:C	T:T
WriPdK0160	TP18874	GAAGGTGACCAAGTTCATGCTGAAAGAGCAGCTCGAGCCAAACA GAAGGTCGGAGTCAACGGATTAAAGAGCAGCTCGAGCCAAACG	CAGAAGCCCCATTTCTTTGTTGGGAT	G:A	A:A
WriPdK0161	TP19778	GAAGGTGACCAAGTTCATGCTAAAGCGAAGCATCGTCGTCCG GAAGGTCGGAGTCAACGGATTCAAAGCGAAGCATCGTCGTCCA	CCAAACCCGCAATTGCATGTGATGTA	T:C	T:T

Table S3.2., continued.

Primer	SNP-bearing GBS Tag	Allele-specific primer sequences (5'-3')[1]	Common primer sequence (5'-3')	Nonpareil genotype	Lauranne genotype
WriPdK0162	TP19878	GAAGGTGACCAAGTTCATGCTAAAGAAGGTTGGGTTTAATGGGGTTTT GAAGGTCGGAGTCAACGGATTAGAAGGTTGGGTTTAATGGGGTTTG	GAGCTCTGAGAACTCTCTCTTCGAA	C:A	A:A
WriPdK0163	TP19987	GAAGGTGACCAAGTTCATGCTCCTGATATCCAGCACATGCCTG GAAGGTCGGAGTCAACGGATTATACCTGATATCCAGCACATGCCTT	CTTTAAGAGACACAAGAGTAGCAACAGTT	T:G	G:G
WriPdK0164	TP20194	GAAGGTGACCAAGTTCATGC**T**GTGACGTCGTTTCCTTTCCGG GAAGGTCGGAGTCAACGGA**TT**GTGACGTCGTTTCCTTTCCGA	GCTCCGAATGGATCATATGATGCCTT	T:T	T:C
WriPdK0165	TP20430	GAAGGTGACCAAGTTCATGCTGGAAGTAGCAAAAGAGTGCTTCACA GAAGGTCGGAGTCAACGGATTGAAGTAGCAAAAGAGTGCTTCACG	GTCGAGAAGCAAGTGCCTTGTCATT	G:G	G:A
WriPdK0166	TP20488	GAAGGTGACCAAGTTCATGCTCGTATAACCGTTTGTGCAGCTGG GAAGGTCGGAGTCAACGGATTACGTATAACCGTTTGTGCAGCTGT	CAGGCTTGCAGGCGCTCCATA	T:G	T:T
WriPdK0167	TP20504	GAAGGTGACCAAGTTCATGCTCACCACCACCGCGTTGCACT GAAGGTCGGAGTCAACGGATTACCACCACCGCGTTGCACG	GGCGATGATGAGCCCTTATGAGAT	C:C	C:A
WriPdK0168	TP21444	GAAGGTGACCAAGTTCATGCTAATGGAGATACATCCTTCGCCAGT GAAGGTCGGAGTCAACGGATTGGAGATACATCCTTCGCCAGC	AAGTCTTCAAGAAATAGAGCTGCAGCTT	G:G	G:A

Table S3.2., continued.

Primer	SNP-bearing GBS Tag	Allele-specific primer sequences (5'-3')[1]	Common primer sequence (5'-3')	Nonpareil genotype	Lauranne genotype
WriPdK0169	TP21582	GAAGGTGACCAAGTTCATGCTATTTGCTCCTAACGCAGCTTCCAA GAAGGTCGGAGTCAACGGA**TT**GCTCCTAACGCAGCTTCCAC	GTTCCACTGGTAAGAATGCCACTTCAT	C:A	C:C
WriPdK0170	TP21619	GAAGGTGACCAAGTTCA**TGCT**CATACTTTACTCAACGCTTTTTCA GAAGGTCGGAGTCAACGGATTCTCATACTTTACTCAACGCTTTTTCG	GGCCGCGATCATGTGTTTTAGAGAA	G:A	A:A
WriPdK0171	TP21630	GAAGGTGACCAAGTTCATGCTCCAAGTGCACCCAAAGAATTGGTAT GAAGGTCGGAGTCAACGGATTCAAGTGCACCCAAAGAATTGGTAC	GGCCGGCAGAATATCAGGGGAT	G:G	G:A
WriPdK0172	TP22379	GAAGGTGACCAAGTTCATGCTAGCAAAACCTTGCGTGCGTGG GAAGGTCGGAGTCAACGGA**TT**GAGCAAAACCTTGCGTGCGTGA	CATAGTCATCATGATGTAGTTGCAGCTT	T:C	C:C
WriPdK0173	TP23364	GAAGGTGACCAAGTTCAT**GCT**GCGCCGGAGAAAGAAGACT GAAGGTCGGAGTCAACGGATTCTGCGCCGGAGAAAGAAGACG	CTTGACAGCACTCGCCTGGCAA	C:C	C:A
WriPdK0174	TP23737	GAAGGTGACCAAGTTCATGCTCGTCAACCGTGTTGAACTCAGAA GAAGGTCGGAGTCAACGGATTCGTCAACCGTGTTGAACTCAGAG	TCTCCGGCTGGGTTCAGAGCAT	G:A	G:G
WriPdK0175	TP24104	GAAGGTGACCAAGTTCATGCTCCAAATATGCTTAATTGTAGTGGCGATA GAAGGTCGGAGTCAACGGATTCAAATATGCTTAATTGTAGTGGCGATG	CTCCCTATGGCAGTACTACATCCAA	G:A	A:A

Table S3.2., continued.

Primer	SNP-bearing GBS Tag	Allele-specific primer sequences (5'-3')[1]	Common primer sequence (5'-3')	Nonpareil genotype	Lauranne genotype
WriPdK0176	TP24350	GAAGGTGACCAAGTTCATGCTCCCAAAACTGAGGACTCTAGGG GAAGGTCGGAGTCAACGGATTCCCCAAAACTGAGGACTCTAGGA	GTGGCCCTTCCCTTTGGTGTGAA	T:C	C:C
WriPdK0177	TP24448	GAAGGTGACCAAGTTCATGCTCGAAGTACAGTAGACTACAATGACAA GAAGGTCGGAGTCAACGGATTCGAAGTACAGTAGACTACAATGACAG	GCATGGCTAGCTGCACTCCTCAT	G:G	G:A
WriPdK0178	TP24844	GAAGGTGACCAAGTTCATGC**T**ACACGAACCCCGTTCCTTTCTC GAAGGTCGGAGTCAACGGAT**T**ACACGAACCCCGTTCCTTTCTA	AGAACACTCGCAGCTGCAATGCAT	G:G	T:G
WriPdK0179	TP24992	GAAGGTGACCAAGTTCATGCTGAGACGCAGAGACGGCTGCAA GAAGGTCGGAGTCAACGGATTGACGCAGAGACGGCTGCAG	CCGACCCACCGGCCACGAA	G:G	G:A
WriPdK0180	TP26493	GAAGGTGACCAAGTTCATGC**T**CCCATCGTCAGGCCGGTTTG GAAGGTCGGAGTCAACGGA**TT**CCCATCGTCAGGCCGGTTTA	GGCCAGCTGCAGAGTCGCCTA	G:G	T:G
WriPdK0181	TP26741	GAAGGTGACCAAGTTCATGCTAAGCATAGAAAAATGGCTGGCTGCA GAAGGTCGGAGTCAACGGATTAAGCATAGAAAAATGGCTGGCTGCT	CTGCATTGTGACTGGACATGTGAGAT	T:A	A:A
WriPdK0182	TP27036	GAAGGTGACCAAGTTCATGCTATGGAGATGGAAATCACATCGACG GAAGGTCGGAGTCAACGGAT**T**CATGGAGATGGAAATCACATCGACA	AGGATTTGATGCTGCAGTCGAGGAT	T:T	T:C

Table S3.2., continued.

Primer	SNP-bearing GBS Tag	Allele-specific primer sequences (5'-3')[1]	Common primer sequence (5'-3')	Nonpareil genotype	Lauranne genotype
WriPdK0183	TP2736	GAAGGTGACCAAGTTCATGCTATTGAAGGGTGATAGACTTGGCAC GAAGGTCGGAGTCAACGGATTCATTGAAGGGTGATAGACTTGGCAT	ATCGGAACTGCTACTTGTGAGACTTATTT	T:C	C:C
WriPdK0184	TP27576	GAAGGTGACCAAGTTCATGCTCCCAATTTGCTCGAGCTGGAG GAAGGTCGGAGTCAACGGATTGCCCAATTTGCTCGAGCTGGAA	GGGTATGCAAATATGGTACCGCAGAT	T:C	C:C
WriPdK0185	TP29641	GAAGGTGACCAAGTTCATGCTATGGAGCACCTCCTCTAGCCA GAAGGTCGGAGTCAACGGAT**T**GGAGCACCTCCTCTAGCCG	AGCTTTATGGGATTAGCTGCGGGAT	G:A	A:A
WriPdK0186	TP29895	GAAGGTGACCAAGTTCATGCTGCAGGAGATTCCCAAGTTGGG GAAGGTCGGAGTCAACGGATTGCAGGAGATTCCCAAGTTGGC	CCATGGCACTTCCACTAGCCGAT	G:C	C:C
WriPdK0187	TP29906	GAAGGTGACCAAGTTCATGCTGACAACGAACTTTGAAAGCAAGCAC GAAGGTCGGAGTCAACGGATTCGACAACGAACTTTGAAAGCAAGCAT	CGGCTTAGGCAGCCTCAAACCTT	C:C	T:C
WriPdK0188	TP22942	GAAGGTGACCAAGTTCATG**CT**CTCTCCGCCTCATACAAAAACAT GAAGGTCGGAGTCAACGGAT**T**CTCTCCGCCTCATACAAAAACAC	GAGTACACGCGGCGGTAGTACAT	G:A	A:A
WriPdK0189	TP2556	GAAGGTGACCAAGTTCATG**CT**TCTCTTAGGTGGTGGAGGAC GAAGGTCGGAGTCAACGGATTCTTCTCTTAGGTGGTGGAGGAG	CTCTTCTCAGACTTTCTTCGGCCAA	C:G	C:C

Table S3.2., continued.

Primer	SNP-bearing GBS Tag	Allele-specific primer sequences (5'-3')[1]	Common primer sequence (5'-3')	Nonpareil genotype	Lauranne genotype
WriPdK0190	TP32048	GAAGGTGACCAAGTTCATGCTATAATAAATTTCCCCATCAGCTGCCTA GAAGGTCGGAGTCAACGG**ATT**ATAATAAATTTCCCCATCAGCTGCCTT	TTTGCATGCCAAATTTGGTAGAGATCGAA	T:T	T:A
WriPdK0191	TP32410	GAAGGTGACCAAGTTCA**TGCT**GCCTCTTGCAGTACTTGAGC GAAGGTCGGAGTCAACGGAT**T**GCTGCCTCTTGCAGTACTTGAGT	GATACGATCTGACGGTGGACTTGAA	T:T	T:C
WriPdK0192	TP32743	GAAGGTGACCAAGTTCATGCTGGAAGCTGGCGAAGAGCGAT GAAGGTCGGAGTCAACGGATTGGAAGCTGGCGAAGAGCGAG	GAAGCGCTGCCTTCGCCTTGAA	C:C	C:A
WriPdK0193	TP33408	GAAGGTGACCAAGTTCATGCTGGTGGTGGAGGGCGGCG GAAGGTCGGAGTCAACGGATTGGTGGTGGAGGGCGGCA	GCACTCTCTCGCATGCTGCGAT	C:C	T:C
WriPdK0194	TP34047	GAAGGTGACCAAGTTCATGCTGGAAGCAAGGTCCTCTCGGG GAAGGTCGGAGTCAACGGATTCGGAAGCAAGGTCCTCTCGGA	AAACACTTGCGCCACAGCAGTGAA	C:C	T:C
WriPdK0195	TP34122	GAAGGTGACCAAGTTCATGCTCCTCCGGCTCCACCGCG GAAGGTCGGAGTCAACGGA**TT**CCTCCGGCTCCACCGCC	GAGGCTCCGGAGGTGGGCT	G:C	G:G
WriPdK0196	TP34523	GAAGGTGACCAAGTTCATGCTAGAAGGTTGGGTTTAATGGGGTTTG GAAGGTCGGAGTCAACGGATTAAAGAAGGTTGGGTTTAATGGGGTTTT	GAGCTCTGAGAACTCTCTCTTCGAA	T:G	T:T

Table S3.2., continued.

Primer	SNP-bearing GBS Tag	Allele-specific primer sequences (5'-3')[1]	Common primer sequence (5'-3')	Nonpareil genotype	Lauranne genotype
WriPdK0197	TP34564	GAAGGTGACCAAGTTCAT**GCT**CCAGTGATGGAGATTATCCT GAAGGTCGGAGTCAACGGATTGCTCCAGTGATGGAGATTATCCG	CATATGACCTTGGCGTCGGCGAA	A:A	C:A
WriPdK0198	TP34707	GAAGGTGACCAAGTTCATGCTAAATGCGCTGCGTGCCAAGCA GAAGGTCGGAGTCAACGGATTAAATGCGCTGCGTGCCAAGCT	CCACCCTTCGAATCTTGTATCCCAA	T:A	A:A
WriPdK0199	TP34743	GAAGGTGACCAAGTTCATGC**T**CCCGTCTACGGCTGCGTG GAAGGTCGGAGTCAACGGATTGATCCCGTCTACGGCTGCGTT	CTGAAGCTGGCGGAGTTGGTGTT	T:G	G:G
WriPdK0200	TP34810	GAAGGTGACCAAGTTCATGCTGATCGCACTCTCCAGCCAAGC GAAGGTCGGAGTCAACGGATTGATCGCACTCTCCAGCCAAGA	AAAGAATGGGCCGGGCTGCGTT	G:G	T:G
WriPdK0201	TP35291	GAAGGTGACCAAGTTCATGCTCAAAGGAGGAGTCTTGGTCCCA GAAGGTCGGAGTCAACGGATTAAAGGAGGAGTCTTGGTCCCC	TTAGACTGAGAGAAGCTTCTGGAGAAAT	C:A	A:A
WriPdK0202	TP36668	GAAGGTGACCAAGTTCATGCTGCATTTGCTGGTTTCCTCGTATAC GAAGGTCGGAGTCAACGGATTGGCATTTGCTGGTTTCCTCGTATAT	CACAGAAGTATAAAGTAACACATCAGCCAT	T:C	T:T
WriPdK0203	TP36935	GAAGGTGACCAAGTTCATGCTGAGGGCTCTCCTCCATTGCTC GAAGGTCGGAGTCAACGGATTAGAGGGCTCTCCTCCATTGCTT	GGTCTTCTTTAGACAAATCTGGAGCAATT	T:T	T:C

Table S3.2., continued.

Primer	SNP-bearing GBS Tag	Allele-specific primer sequences (5'-3')[1]	Common primer sequence (5'-3')	Nonpareil genotype	Lauranne genotype
WriPdK0204	TP37237	GAAGGTGACCAAGTTCATGCTCACGGTAGCTGCTCTGTCTCC GAAGGTCGGAGTCAACGGATTCACGGTAGCTGCTCTGTCTCG	GCCCTCGAGAGGGGTGCGAT	G:G	G:C
WriPdK0205	TP37395	GAAGGTGACCAAGTTCATGCTCAAGAAGCCTCCGGGTCTCC GAAGGTCGGAGTCAACGGATTAACAAGAAGCCTCCGGGTCTCA	GGCTCAGACAATCAAGGACCTCTT	T:G	G:G
WriPdK0206	TP37441	GAAGGTGACCAAGTTCATGCTCCTTGATTTTCGATCCACTTGCG GAAGGTCGGAGTCAACGGATTAGCCTTGATTTTCGATCCACTTGCA	AGCCTGTCAAGGGTCTGGATAAGTT	T:C	C:C
WriPdK0207	TP38360	GAAGGTGACCAAGTTCATGCTGGCTGCTGGTACCTGTTCTAGA GAAGGTCGGAGTCAACGGATTGGCTGCTGGTACCTGTTCTAGT	CAAGAGGAGATCTAGGTGCTCAGAT	T:A	T:T
WriPdK0208	TP38437	GAAGGTGACCAAGTTCATGC**T**TCTCTGGACTATGCAAGAACACGA GAAGGTCGGAGTCAACGG**ATT**CTCTGGACTATGCAAGAACACGT	GCAGCCTTGGCGTCATTTGCCAA	T:A	T:T
WriPdK0209	TP38516	GAAGGTGACCAAGTTCA**TGCT**GTAAAACCAAGCCTCTGTGTCG GAAGGTCGGAGTCAACGGATTCTGTAAAACCAAGCCTCTGTGTCT	CCACAGCTAAAACACCAATTAGGAAGTTA	T:T	T:G
WriPdK0210	TP38666	GAAGGTGACCAAGTTCATG**CT**CAGTAGGCATTCCATGGGCC GAAGGTCGGAGTCAACGGATTCTCAGTAGGCATTCCATGGGCA	GGCGTTACTCATGCAGAACAGAAGTA	T:T	T:G

Table S3.2., continued.

Primer	SNP-bearing GBS Tag	Allele-specific primer sequences (5'-3')[1]	Common primer sequence (5'-3')	Nonpareil genotype	Lauranne genotype
WriPdK0211	TP39254	GAAGGTGACCAAGTTCATGCTCATTATGCACAGTGTCATATGTTTCCT GAAGGTCGGAGTCAACGG**ATT**ATGCACAGTGTCATATGTTTCCC	GGGTTCTTGTAGCTCGGGATTGAT	A:A	G:A
WriPdK0212	TP39705	GAAGGTGACCAAGTTCATG**CT**CCTTGATAATCTCATCTTTGGTTAGA GAAGGTCGGAGTCAACGGATTCTCCTTGATAATCTCATCTTTGGTTAGT	AGCTCCTCAAGAACATGAATCACAAACAA	T:T	T:A
WriPdK0213	TP39811	GAAGGTGACCAAGTTCATGC**T**GAGCTTTGGTGCTGCTTGACC GAAGGTCGGAGTCAACGGA**TT**GAGCTTTGGTGCTGCTTGACT	GGGCTTATCTACCATGCATCCTTGAT	C:C	T:C
WriPdK0214	TP40103	GAAGGTGACCAAGTTCATGCTATAGCTGCTTGTGAGAAGAAGAATATC GAAGGTCGGAGTCAACGGATTCATAGCTGCTTGTGAGAAGAAGAATATT	GGTTTCTTCATACCGAGACCCGTAA	T:C	C:C
WriPdK0215	TP40450	GAAGGTGACCAAGTTCATGC**T**GTGGTGTTCAAGCTGCTTTGTCC GAAGGTCGGAGTCAACGGA**TT**GTGGTGTTCAAGCTGCTTTGTCT	GCGAGCACGGAGTTCAAAGAGAAAT	T:C	T:T
WriPdK0216	TP4873	GAAGGTGACCAAGTTCATGC**T**CGACCCAGTACCAAACAGCTCT GAAGGTCGGAGTCAACGGATTGACCCAGTACCAAACAGCTCC	CGAAGGCTTGGGAAGTTCACCTTTT	G:G	G:A
WriPdK0217	TP5114	GAAGGTGACCAAGTTCATGCTCAATGGCCTATTAATATACAGCAGCA GAAGGTCGGAGTCAACGGATTCAATGGCCTATTAATATACAGCAGCT	CGCTTGCATGCTCCACACACTAAAT	T:T	T:A

Table S3.2., continued.

Primer	SNP-bearing GBS Tag	Allele-specific primer sequences (5'-3')[1]	Common primer sequence (5'-3')	Nonpareil genotype	Lauranne genotype
WriPdK0218	TP5832	GAAGGTGACCAAGTTCAT**GCT**GGTGGTGACGCCCTTGTAG GAAGGTCGGAGTCAACGGATTCTGGTGGTGACGCCCTTGTAT	GTGCTGGTGGCCCAGAAGGTTT	T:T	T:G
WriPdK0219	TP6690	GAAGGTGACCAAGTTCATGCTCATGAGGCGCAGCATCATCAAAC GAAGGTCGGAGTCAACGGATTCATGAGGCGCAGCATCATCAAAT	CTCTCACATTGCTGGAGAAGTGGTT	T:C	T:T
WriPdK0220	TP6745	GAAGGTGACCAAGTTCATGCTCATTCATCAGCTAGATTGAGGTTGCT GAAGGTCGGAGTCAACGGATTCATCAGCTAGATTGAGGTTGCC	ATCAAGTGTGCTAACCACGCAGCAT	G:A	A:A
WriPdK0221	TP6994	GAAGGTGACCAAGTTCATGC**T**AAAATCAAAGTTTTCAGCAGCAGGACT GAAGGTCGGAGTCAACGG**ATT**AAAATCAAAGTTTTCAGCAGCAGGACA	GATTGACCACGTTCAGAGTCTACCAA	T:A	T:T
WriPdK0222	TP7711	GAAGGTGACCAAGTTCA**TGCT**ATCCCACCCTCAGCAGCA GAAGGTCGGAGTCAACGGATTCTATCCCACCCTCAGCAGCG	GTCCAGCTGGTCTTGTGGGCAA	G:G	G:A
WriPDK0223	TP797	GAAGGTGACCAAGTTCATGCTGCAAATGTCGGGCAGATGAATATGAA GAAGGTCGGAGTCAACGGATTCAAATGTCGGGCAGATGAATATGAG	AGGCTTAGGGCATTGTCCCAAGAA	G:G	G:A
WriPdK0224	TP8136	GAAGGTGACCAAGTTCATGCTGCGGTTTCCTTTCAACCAGATAGA GAAGGTCGGAGTCAACGGATTCGGTTTCCTTTCAACCAGATAGG	GATGCACATGCCGAGCTGGGTT	G:A	G:G

Table S3.2., continued.

Primer	SNP-bearing GBS Tag	Allele-specific primer sequences (5'-3')[1]	Common primer sequence (5'-3')	Nonpareil genotype	Lauranne genotype
WriPdK0225	TP8770	GAAGGTGACCAAGTTCATGCTAGCCACAGAAAGACACCATGGC GAAGGTCGGAGTCAACGGATTCAGCCACAGAAAGACACCATGGT	AGCCCTCAAGACCATCACAGTCTTT	C:C	T:C
WriPdK0226	TP8828	GAAGGTGACCAAGTTCATGCTATATATATATGTTAATAAGGAGAGGCAGCT GAAGGTCGGAGTCAACGGATTATATATGTTAATAAGGAGAGGCAGCC	GCAGAGGTGCTCAGTACAGCGTA	C:T	T:T
WriPdK0227	TP8999	GAAGGTGACCAAGTTCATGCTGAAGTAGCAAAAGAGTGCTTCACG GAAGGTCGGAGTCAACGGATTGGAAGTAGCAAAAGAGTGCTTCACA	GTCGAGAAGCAAGTGCCTTGTCATT	C:C	T:C
WriPdK0228	TP9096	GAAGGTGACCAAGTTCATGCTCACACAATGCTTCCTTTCACGCG GAAGGTCGGAGTCAACGGATTCACACAATGCTTCCTTTCACGCT	CAAACATGCAGCGCGTGTTATGGTA	T:G	G:G
WriPdK0229	TP9199	GAAGGTGACCAAGTTCATGCTGTCAATGCCTCAGTCTCTTCATCA GAAGGTCGGAGTCAACGGATTGTCAATGCCTCAGTCTCTTCATCT	GATGATGCAGTAACAACGGCTGCTT	T:T	T:A
WriPdK0230	TP9410	GAAGGTGACCAAGTTCATGCTAAAAGATGTGTCGGCAGCAGCG GAAGGTCGGAGTCAACGGATTAAAAAAGATGTGTCGGCAGCAGCA	CGTGTGAAATGTTTCGTGTCTTTGTTCTT	T:T	T:C
WriPdK0231	TP3887	GAAGGTGACCAAGTTCATGCTGGCAGCACGACGACGAAGAC GAAGGTCGGAGTCAACGGATTAGGCAGCACGACGACGAAGAT	CCTTGAGTTTATTATCCTCAAACCCGAAA	T:C	C:C

Table S3.2., continued.

Primer	SNP-bearing GBS Tag	Allele-specific primer sequences (5'-3')[1]	Common primer sequence (5'-3')	Nonpareil genotype	Lauranne genotype
WriPdK0232	TP16490	GAAGGTGACCAAGTTCATGC**T**TTTTGTTAGGGAACACAAGAAGTTCG GAAGGTCGGAGTCAACGGATTCATTTTTGTTAGGGAACACAAGAAGTTCA	GACTAAATGAAAGGGCGAAGCATTAAGTT	T:C	C:C
WriPdK0233	TP15129	GAAGGTGACCAAGTTCATG**CT**GAGAGACCACCAAAGCAGCG GAAGGTCGGAGTCAACGGAT**T**CTGAGAGACCACCAAAGCAGCT	GACCAGAGAAGGTATGGGCTTGAA	T:G	G:G
WriPdK0234	TP13076	GAAGGTGACCAAGTTCATGCTCCATAACAAGGTAGCAAGAGGGTA GAAGGTCGGAGTCAACGGATTCATAACAAGGTAGCAAGAGGGTG	CAGTGAATCTTCTCAATAAAGCTGTGGAT	G:A	A:A
WriPdK0235	TP605	GAAGGTGACCAAGTTCATGCTGGCCCACCTCAACCCATCTG GAAGGTCGGAGTCAACGGATTGGCCCACCTCAACCCATCTA	AGTGCTTCAGTGCAGCAAAGGCTAT	T:C	C:C
WriPdK0236	TP1293	GAAGGTGACCAAGTTCATGCTGCGCTGGCAGCAACCTAGTC GAAGGTCGGAGTCAACGGATTGCGCTGGCAGCAACCTAGTG	GAGTTCGAGCGCCGAGTTCCTA	G:C	C:C
WriPdK0237	TP1179	GAAGGTGACCAAGTTCATGC**T**ACTTGGGAGGCAGCAACCAAC GAAGGTCGGAGTCAACGGATTGTACTTGGGAGGCAGCAACCAAT	GGGTGATATCACTGTTGTTTTGTCATACAT	T:C	C:C
WriPdK0238	TP12290	GAAGGTGACCAAGTTCATG**CT**TGTTCGTTCCCAGTTCAGGC GAAGGTCGGAGTCAACGGAT**T**CTTGTTCGTTCCCAGTTCAGGT	GCCCTATCATCGTGGCCATCGAA	T:C	C:C

Table S3.2., continued.

Primer	SNP-bearing GBS Tag	Allele-specific primer sequences (5'-3')[1]	Common primer sequence (5'-3')	Nonpareil genotype	Lauranne genotype
WriPdK0239	TP15756	GAAGGTGACCAAGTTCATG**CT**GCGACGAGAACAGCTCCACC GAAGGTCGGAGTCAACGGAT**T**GCGACGAGAACAGCTCCACA	CGGCGGAGCGGGCCAGTT	T:G	G:G
WriPdK0240	TP1532	GAAGGTGACCAAGTTCATGCTCAAATTTGAGGCTTTGAGTAGGCTC GAAGGTCGGAGTCAACGGATTGCAAATTTGAGGCTTTGAGTAGGCTT	TTCAATTTGCTCACATTGTCCTCTGCTT	A:G	G:G
WriPdK0241	TP24362	GAAGGTGACCAAGTTCATGCTGAATGTGAGGCTCCAATCGCTC GAAGGTCGGAGTCAACGGAT**T**CGAATGTGAGGCTCCAATCGCTT	ATGGAATTTGAGGACGACGGCCTTT	A:G	A:A
WriPdK0242	TP20535	GAAGGTGACCAAGTTCATGCTCAACTCCAACCCCAAAAAGGCC GAAGGTCGGAGTCAACGGAT**T**CCAACTCCAACCCCAAAAAGGCT	CTGAGCTACAGAGAAAAGAAGAAAAGGTT	T:C	T:T
WriPdK0243	TP27979	GAAGGTGACCAAGTTCATGCTACGTGTTCTTGCTGCATGCTACAAA GAAGGTCGGAGTCAACGGATTACGTGTTCTTGCTGCATGCTACAAT	CTTCAAGCAGAACATGGTGGTCGTT	T:A	A:A
WriPdK0244	TP2602	GAAGGTGACCAAGTTCATGC**T**GAATCTCAATGGCTCTTCCCACAT GAAGGTCGGAGTCAACGGATTAATCTCAATGGCTCTTCCCACAC	CTTAGAGCTTGGGCGCAACCCTA	G:A	G:G
WriPdK0245	TP12538	GAAGGTGACCAAGTTCATGCTAAGTTCTTGTCTAACTTTTCTCGCCTT GAAGGTCGGAGTCAACGGATTGTTCTTGTCTAACTTTTCTCGCCTG	CAGGCCAATTAGTTGAGTAACTTCATCTT	A:A	C:A

Table S3.2., continued.

Primer	SNP-bearing GBS Tag	Allele-specific primer sequences (5'-3')[1]	Common primer sequence (5'-3')	Nonpareil genotype	Lauranne genotype
WriPdK0246	TP4228	GAAGGTGACCAAGTTCATGC**T**GATTTTCTGTTAGCAGCACTCTCCTA GAAGGTCGGAGTCAACG**GATT**TTCTGTTAGCAGCACTCTCCTT	GTTGGTGCTCGGGGATCAAATTCTA	A:A	T:A
WriPdK0247	TP6389	GAAGGTGACCAAGTTCATGCTACTGCCGCCATATTCTGCTGGT GAAGGTCGGAGTCAACGGATTCTGCCGCCATATTCTGCTGGC	TGTCTCTCAGCAGCATAGTGCCAAA	A:A	G:A
WriPdK0248	TP33363	GAAGGTGACCAAGTTCATGCTAACGACAGTACTGCTCTCTGAAACA GAAGGTCGGAGTCAACGGATTCGACAGTACTGCTCTCTGAAACG	CGCTCCGTTCTCTCGCCCCAA	A:A	G:A
WriPdK0249	TP45419	GAAGGTGACCAAGTTCA**TGCT**AAGCACCTGCACTTCTCTGG GAAGGTCGGAGTCAACGGATTCTAAGCACCTGCACTTCTCTGC	CTAACCAGCAGAGTCTCAGTGCAAA	C:C	G:C
WriPdK0250	TP36836	GAAGGTGACCAAGTTCAT**GCT**GCCTAACTTTTGACTAACAGTAG GAAGGTCGGAGTCAACGGATTAGCTGCCTAACTTTTGACTAACAGTAT	CGGGCCAAGGTTTTTTCATCTTGTTATAT	G:G	T:G
WriPdK0251	TP7579	GAAGGTGACCAAGTTCATGCTGAAGGTATTGTCCCGAAACACTTG GAAGGTCGGAGTCAACGGAT**T**GGAAGGTATTGTCCCGAAACACTTA	GGTCTCGTCTCAACAAAAGAGCTTGTA	C:C	T:C
WriPdK0252	TP26193	GAAGGTGACCAAGTTCAT**GCT**GCAGACACAGGCCGCT GAAGGTCGGAGTCAACGGATTCTGCAGACACAGGCCGCC	CAGAACTCATGTCTGGCATCCTGAA	A:A	G:A

Table S3.2., continued.

Primer	SNP-bearing GBS Tag	Allele-specific primer sequences (5'-3')[1]	Common primer sequence (5'-3')	Nonpareil genotype	Lauranne genotype
WriPdK0253	TP12428	GAAGGTGACCAAGTTCAT**GCT**TAGATGCCAGAGCCATCACG GAAGGTCGGAGTCAACGGATTGCTTAGATGCCAGAGCCATCACA	GAGCAGGGATAGCACAGTACATGAA	C:C	T:C
WriPdK0254	TP16651	GAAGGTGACCAAGTTCATGCTCGTTTCACGCTACACTTACAATGC GAAGGTCGGAGTCAACGGATTCGTTTCACGCTACACTTACAATGG	CGGGCATAGCACATGTCCGCAT	C:C	G:C
WriPdK0255	TP23267	GAAGGTGACCAAGTTCATGC**T**AACTCAAAATCAGTAACTGGCGGC GAAGGTCGGAGTCAACG**GATT**AACTCAAAATCAGTAACTGGCGGT	TGGCAGCCTCGTAACGCTGGTA	C:C	T:C
WriPdK0256	TP25017	GAAGGTGACCAAGTTCATGCTCACGTGGAGCTCTCTGAGGAA GAAGGTCGGAGTCAACGGATTCACGTGGAGCTCTCTGAGGAG	TAACCCAGTCGCTGCTGGCCTT	G:G	G:A
WriPdK0257	TP23919	GAAGGTGACCAAGTTCATGCTGTGCCTCATCACCGGACGG GAAGGTCGGAGTCAACGGATTGGTGCCTCATCACCGGACGT	CTGCCAGTTTGCTCACGGCGTT	G:G	T:G
WriPdK0258	TP13105	GAAGGTGACCAAGTTCATGCTGATTGGTTCTGAATACGACAGCGT GAAGGTCGGAGTCAAC**GGATT**GGTTCTGAATACGACAGCGG	CCTCCTCCGACCGCCTCCAA	A:A	C:A
WriPdK0259	TP14819	GAAGGTGACCAAGTTCATGCTATTCGAAGGAGAGCAATGCCCATA GAAGGTCGGAGTCAACGG**ATT**CGAAGGAGAGCAATGCCCATG	CTCTGCACAAGGTATTTGCAGCCAT	A:A	G:A

Table S3.2., continued.

Primer	SNP-bearing GBS Tag	Allele-specific primer sequences (5'-3')[1]	Common primer sequence (5'-3')	Nonpareil genotype	Lauranne genotype
WriPdK0260	TP23124	GAAGGTGACCAAGTTCATGCTCGCATCGTCTGCCACCGC GAAGGTCGGAGTCAACG**GATT**GCGCATCGTCTGCCACCGT	GCACGTTAGCGAGGAGCACCAA	C:C	T:C
WriPdK0261	TP37130	GAAGGTGACCAAGTTCA**TGCT**GAGTGTGGACTCTCCTATGCT GAAGGTCGGAGTCAACGGAT**T**GAGTGTGGACTCTCCTATGCC	CTGCTCTGACCCATAATGGCAACAA	A:A	G:A
WriPdK0262	TP5236	GAAGGTGACCAAGTTCA**TGCT**CCTACTTGAGCAGCAGGAG GAAGGTCGGAGTCAACGG**ATT**GCTCCTACTTGAGCAGCAGGAT	GCTCTGTTCTCCTACCCAGATAAGAT	G:G	T:G
WriPdK0263	TP36487	GAAGGTGACCAAGTTCATGCTGAGAAGCTCCATATGCAAAGGCG GAAGGTCGGAGTCAACGGATTAAGAGAAGCTCCATATGCAAAGGCA	CCGGCTCATCTGAATGTTTTGTGATAAAT	C:C	T:C
WriPdK0264	TP29814	GAAGGTGACCAAGTTCA**TGCT**CTTGAAGTGCTGCCATCG GAAGGTCGGAGTCAACGGAT**T**CTGCTCTTGAAGTGCTGCCATCT	CCAGCTTCACTCTCCAAACCCCAT	G:G	T:G
WriPdK0265	TP28045	GAAGGTGACCAAGTTCATG**CT**ACGATATGGAGCCAGGTCACAG GAAGGTCGGAGTCAACGGAT**T**ACGATATGGAGCCAGGTCACAA	GAACTTGCTGCATGGGCCCCTA	T:T	T:C
WriPdK0266	TP38730	GAAGGTGACCAAGTTCATGCTAATGAACTAGTTGAAGCGGATAGAGAA GAAGGTCGGAGTCAACGGATTAATGAACTAGTTGAAGCGGATAGAGAT	TCAAGAAAGTCCCAGCACATAACATCTTA	T:T	T:A

Table S3.2., continued.

Primer	SNP-bearing GBS Tag	Allele-specific primer sequences (5'-3')[1]	Common primer sequence (5'-3')	Nonpareil genotype	Lauranne genotype
WriPdK0267	TP10006	GAAGGTGACCAAGTTCATGCT**T**CCATTTTGGAGCCCTAGAAGAAGATA GAAGGTCGGAGTCAACGGATTCATTTTGGAGCCCTAGAAGAAGATG	GCGGCACTGCTGATGCAAATATCAT	G:G	G:A
WriPdK0268	TP1484	GAAGGTGACCAAGTTCATGCTACTGCAGGCTCAACTGGACGAA GAAGGTCGGAGTCAACGGATTCTGCAGGCTCAACTGGACGAG	GGCAAGACCAGCTGAGCTTCCTA	G:G	G:A
WriPdK0269	TP30199	GAAGGTGACCAAGTTCAT**GCT**GCCCACACCGGCGAGAA GAAGGTCGGAGTCAACGGAT**T**GCCCACACCGGCGAGAG	TCGTTCCCTTCACCGCGAAAAACAA	A:A	G:A
WriPdK0270	TP25717	GAAGGTGACCAAGTTCATGCTAAAAGGCTCGCTGCACCTTCC GAAGGTCGGAGTCAACGGATTAAAAGGCTCGCTGCACCTTCG	TCCCCCGAAGCCTCCAAAATTTGAT	C:C	G:C
WriPdK0271	TP1054	GAAGGTGACCAAGTTCATGCTGATGTTGGTGCCTCCAAGAGCA GAAGGTCGGAGTCAACGGATTGATGTTGGTGCCTCCAAGAGCT	GTTGGTGCTTCCCTTGGGCTGTT	A:A	T:A
WriPdK0272	TP20631	GAAGGTGACCAAGTTCATGCTCAGCACAGTCACATCCTTCACG GAAGGTCGGAGTCAACGGATTGCAGCACAGTCACATCCTTCACA	CCAGTAGGCTTTACTTGCTGAGGAT	T:T	T:C
WriPdK0273	TP11106	GAAGGTGACCAAGTTCATGCTGCGGCGTTGAACGCTGTGTAT GAAGGTCGGAGTCAACGGATTGCGGCGTTGAACGCTGTGTAC	TTCTCGAACACGGCAGCCCGAA	A:A	G:A

Table S3.2., continued.

Primer	SNP-bearing GBS Tag	Allele-specific primer sequences (5'-3')[1]	Common primer sequence (5'-3')	Nonpareil genotype	Lauranne genotype
WriPdK0274	TP18097	GAAGGTGACCAAGTTCATGC**T**GTGCTTTGTTCTTGGGGGTGG GAAGGTCGGAGTCAACGGAT**T**ATTGTGCTTTGTTCTTGGGGGTGA	GATAACGATTCGCAGCAGCTCAGAA	T:T	T:C
WriPdK0275	TP23743	GAAGGTGACCAAGTTCAT**GCT**GCAACAAAATAACCACGCCAGA GAAGGTCGGAGTCAACGGATTCTGCAACAAAATAACCACGCCAGT	TGGAAGCTGATCCTCTAGTCACCAT	T:T	T:A
WriPdK0276	TP18179	GAAGGTGACCAAGTTCATGC**T**CTTCAAAATCCAGAAAGTGGAGCC GAAGGTCGGAGTCAACGGA**TT**ATCTTCAAAATCCAGAAAGTGGAGCT	CCGACCCAAAATGATTTCTGCCCTT	T:T	T:C
WriPdK0277	TP15122	GAAGGTGACCAAGTTCAT**GCT**CGAGTCGTGGCTCGAG GAAGGTCGGAGTCAACGGATTGGCTCGAGTCGTGGCTCGAT	GCGTCGTGTTCTGGAGCTCGAA	G:G	T:G
WriPdK0278	TP15412	GAAGGTGACCAAGTTCATG**CT**CAACAGAGCAGCGGAAGAGA GAAGGTCGGAGTCAACGGAT**T**CAACAGAGCAGCGGAAGAGG	GGTTAGCCACGAGTTCAACTGCAAT	G:G	G:A
WriPdK0279	TP2696	GAAGGTGACCAAGTTCATGCTCAAAGGGAAAGGAAATGGACTCG GAAGGTCGGAGTCAACGGA**TT**GCAAAGGGAAAGGAAATGGACTCT	TTAGGTGGAACCAAGATTTTTCGTCCTTT	G:G	T:G
WriPdK0280	TP33276	GAAGGTGACCAAGTTCATGCTCACCACCTTAGGTTTCCACCTC GAAGGTCGGAGTCAACGGATTACACCACCTTAGGTTTCCACCTA	GGAAATTAGGCAACCCCATGAGGAA	G:G	T:G

Table S3.2., continued.

Primer	SNP-bearing GBS Tag	Allele-specific primer sequences (5'-3')[1]	Common primer sequence (5'-3')	Nonpareil genotype	Lauranne genotype
WriPdK0281	TP10404	GAAGGTGACCAAGTTCATGCTCCGATGGTTGTTTTCGAACAATCTGT GAAGGTCGGAGTCAACGGATTCGATGGTTGTTTTCGAACAATCTGC	GCAGCCACAATGATTAAAGATAGCAACAA	A:A	G:A
WriPdK0282	TP23837	GAAGGTGACCAAGTTCATGC**T**TGGCAGCCACGACTATAAGGG GAAGGTCGGAGTCAACGGA**TT**TGGCAGCCACGACTATAAGGA	CAAGGTGTTCTTCCTCCTCCTAGTA	C:C	T:C
WriPdK0283	TP13896	GAAGGTGACCAAGTTCATGCTACCTCTCCAGGACTACGAGCTT GAAGGTCGGAGTCAACGGATTCCTCTCCAGGACTACGAGCTC	GGCCGACCCGATCTGATCGAA	A:A	G:A
WriPdK0284	TP4297	GAAGGTGACCAAGTTCATG**CT**TTCTCCACATACACCATTCTAGC GAAGGTCGGAGTCAACGGA**TT**CTTTCTCCACATACACCATTCTAGA	GAGCAAATGCAGCACTGGATGGAAA	T:G	G:G
WriPdK0285	TP35652	GAAGGTGACCAAGTTCAT**GCT**GGTGCTGCGGCAAGTTCCTC GAAGGTCGGAGTCAACGGAT**T**GGTGCTGCGGCAAGTTCCTA	GCTGCCCTGCCCTGCCGAA	T:G	G:G
WriPdK0286	TP28761	GAAGGTGACCAAGTTCATGCTGAGATTCCATCAAAAGCAAGTGTTTTG GAAGGTCGGAGTCAACGGAT**T**GGAGATTCCATCAAAAGCAAGTGTTTTA	GATGGCAGAAGGACGCTGCCAA	T:T	T:C
WriPdK0287	TP16954	GAAGGTGACCAAGTTCATG**CT**ACATTTGCAGAATGCTGCATGCG GAAGGTCGGAGTCAACGGA**TT**CTACATTTGCAGAATGCTGCATGCA	CTCCAAATCCGCCGGGAACCAA	C:C	T:C

Table S3.2., continued.

Primer	SNP-bearing GBS Tag	Allele-specific primer sequences (5'-3')[1]	Common primer sequence (5'-3')	Nonpareil genotype	Lauranne genotype
WriPdK0288	TP34627	GAAGGTGACCAAGTTCAT**GCT**GCGGCTGCCATGGGTTGATAC GAAGGTCGGAGTCAACGGAT**T**GCGGCTGCCATGGGTTGATAT	CTCAGCTCTCTTCATGGAGCAACTT	T:C	C:C
WriPdK0289	TP2161	GAAGGTGACCAAGTTCATG**CT**AGAAAAGGAAGAATGGTGCTAAGAGAT GAAGGTCGGAGTCAACGGATTGAAAAGGAAGAATGGTGCTAAGAGAC	CAGCAAGCGATGAGATGTCTAATGAAAA	G:A	G:G
WriPdK0290	TP39659	GAAGGTGACCAAGTTCATGCTCCATGGCTGCTTCTGCTTCTC GAAGGTCGGAGTCAACGGATTCCATGGCTGCTTCTGCTTCTG	CAAGCCAAAGCCACGAGAGATGTAA	G:C	G:G
WriPdK0291	TP5787	GAAGGTGACCAAGTTCATGC**T**GGCGAAGAAGAGGAAGCTAGACA GAAGGTCGGAGTCAACGGATTGCGAAGAAGAGGAAGCTAGACG	CCAACCGGCAAAACGACACCCAT	C:T	C:C
WriPdK0292	TP9076	GAAGGTGACCAAGTTCATGC**T**TTCAAAAGCATGTGTCAAGAGTACCA GAAGGTCGGAGTCAACGGA**TT**CAAAAGCATGTGTCAAGAGTACCG	GAAGTGAGCTTCTCTGGGCTTCAAA	G:A	A:A
WriPdK0293	TP26398	GAAGGTGACCAAGTTCATGCTCAGATGAAGCATATATGTTACACAACAG GAAGGTCGGAGTCAACGGATTCCAGATGAAGCATATATGTTACACAACAA	AGCCAGAGAAAGCAGGCACTTACTT	T:C	T:T

[1]Allele-specific primers include tails (underlined) that are complementary to FRET cassettes in the KASP™ Master Mix added to the 5' end. The nucleotides that overlapped between the tail and allele-specific primers are in bold text.

Table S3.3. SNP-bearing sequences from Nonpareil that were used in comparative mapping with the peach sequence assembly. SNP-bearing sequences (query), length of the query sequence, expected values (E-value), score, percentage of identity (PID), start and stop points of each query, mapped peach scaffold (Pp), start and stop points of the mapping sequences are shown.

Query	Length of the query	E-value	Score	PID %	Align length	Query start	Query stop	Match	Match start	Match stop
TP21619	173	2e-61	236	97	134	40	173	Pp1	6,900.910	6,901,041
TP15291	64	0.009	40	95	24	31	54	Pp5	9,790,190	9,790,167
TP16224	131	2e-61	236	97	131	1	131	Pp1	795,926	795,796
TP17599	64	5e-29	127	100	64	1	64	Pp1	590,639	590,576
TP38360	151	1e-75	283	98	151	1	151	Pp1	1,690,391	1,690,241
TP8083	64	5e-29	127	100	64	1	64	Pp1	4,010,970	4,011,033
TP2736	177	6e-96	351	100	177	1	177	Pp1	7,498,976	7,499,152
TP3887	121	6e-58	224	98	121	1	121	Pp1	9,487,011	9,487,131
TP22379	130	7e-61	234	97	130	1	130	Pp1	15,924,617	15,924,746
TP29895	173	4e-91	335	99	173	1	173	Pp1	10,575,132	10,575,304
TP18674	64	5e-29	127	100	64	1	64	Pp1	12,672,576	12,672,639
TP19987	131	3e-66	252	99	131	1	131	Pp1	24,881,438	24,881,568
TP33068	166	7e-80	297	97	166	1	166	Pp1	26,181,339	26,181,504
TP16675	64	1e-26	119	98	64	1	64	Pp1	29,815,793	29,815,856
TP7440	64	5e-29	127	100	64	1	64	Pp1	31,383,252	31,383,315
TP34627	176	1e-90	333	98	176	1	176	Pp1	32,080,604	32,080,779
TP14993	64	5e-29	127	100	64	1	64	Pp1	31,997,048	31,996,985
TP19778	170	2e-89	329	99	170	1	170	Pp1	33,475,996	33,476,165
TP2161	148	2e-67	256	97	145	1	145	Pp1	38,073,780	38,073,924
TP7601	139	5e-71	268	99	139	1	139	Pp1	35,899,835	35,899,697
TP22942	159	3e-73	276	96	159	1	159	Pp1	35,207,847	35,207,689
TP13360	64	5e-29	127	100	64	1	64	Pp1	36,942,615	36,942,552

Table S3.3., continued.

Query	Length of the query	E-value	Score	PID %	Align length	Query start	Query stop	Match	Match start	Match stop
TP20130	64	1e-26	119	98	64	1	64	Pp1	33,885,257	33,885,320
TP24104	150	4e-75	281	99	150	1	150	Pp2	2,757,920	2,758,068
TP38441	65	3e-21	101	98	59	1	59	Pp2	3,546,460	3,546,403
TP21582	196	1e-102	373	98	196	1	196	Pp2	244,114	244,309
TP37441	251	1e-125	448	98	250	2	251	Pp2	17,248,264	17,248,512
TP11568	65	7e-25	113	98	65	1	65	Pp2	17,248,444	17,248,381
TP4872	65	7e-25	113	98	65	1	65	Pp2	19,550,244	19,550,244
TP32567	65	7e-25	113	98	65	1	65	Pp2	20,504,819	20,504,756
TP2556	175	2e-83	309	97	176	1	175	Pp2	22,407,261	22,407,436
TP9584	65	3e-18	92	96	58	1	58	Pp2	22,550,238	22,550,294
TP5970	65	7e-25	113	98	65	1	65	Pp2	23,224,193	23,224,256
TP16490	152	4e-81	301	100	152	1	152	Pp2	23,869,675	23,869,826
TP15836	172	3e-88	325	98	172	1	172	Pp2	25,658,777	25,658,948
TP40450	176	5e-81	301	97	176	1	176	Pp2	26,062,524	26,062,350
TP27576	201	1e-103	375	99	201	1	201	Pp2	26,071,654	26,071,853
TP15129	155	2e-77	289	98	154	1	154	Pp2	23,258,509	23,258,356
TP16647	65	7e-25	113	98	65	1	65	Pp2	26,155,427	26,155,490
TP26236	65	7e-25	113	98	65	1	65	Pp2	27,750,366	27,750,303
TP9968	65	2e-22	105	96	65	1	65	Pp2	28,927,147	28,927,210
TP39442	65	7e-25	113	98	65	1	65	Pp2	28,777,011	28,777,074
TP14389	65	2e-22	105	96	65	1	65	Pp3	250,694	250,631
TP861	65	2e-22	105	96	65	1	65	Pp3	1,506,285	1,506,222
TP23086	65	7e-25	113	98	65	1	65	Pp3	2,809,011	2,809,074
TP40103	113	5e-65	248	98	113	1	113	Pp3	4,235,021	4,235,153
TP13076	128	4e-50	198	94	128	1	128	Pp3	5,189,108	5,188,981
TP605	174	2e-89	329	98	174	1	174	Pp3	7,310,900	7,310,727
TP38437	147	2e-73	276	99	147	1	147	Pp3	14,721,884	14,722,029
TP23737	132	2e-64	246	99	132	1	132	Pp3	18,145,149	18,145,019

Table S3.3., continued.

Query	Length of the query	E-value	Score	PID %	Align length	Query start	Query stop	Match	Match start	Match stop
TP21192	65	7e-25	113	98	65	1	65	Pp3	14,597,211	14,597,148
TP15894	65	5e-23	107	98	62	1	62	Pp3	21,849,871	21,849,931
TP28885	65	7e-25	113	98	65	1	65	Pp3	23,392,571	23,392,508
TP13153	65	7e-25	113	98	65	1	65	Pp3	23,315,562	23,315,625
TP18874	185	4e-91	335	97	185	1	185	Pp3	21,994,990	21,994,806
TP18734	148	6e-74	278	99	148	1	148	Pp3	23,653,549	23,653,403
TP1293	153	5e-53	208	92	149	1	149	Pp3	27,082,725	27,082,577
TP6745	128	3e-57	222	97	128	1	128	Pp3	23,979,215	23,979,341
TP27369	65	2e-22	105	96	65	1	65	Pp3	25,943,241	25,943,304
TP39659	177	4e-91	335	98	177	1	177	Pp3	27,140,581	27,140,405
TP1179	168	7e-77	287	97	168	1	168	Pp4	3,127,554	3,127,718
TP31870	65	7e-25	113	98	65	1	65	Pp4	1,895,399	1,895,336
TP17942	147	3e-63	242	96	146	1	146	Pp4	4,319,792	4,319,648
TP3083	65	2e-22	105	96	65	1	65	Pp4	4,319,668	4,319,731
TP34523	169	2e-86	319	99	165	5	169	Pp4	6,362,189	6,362,025
TP19878	175	5e-78	291	99	155	21	175	Pp4	6,362,020	6,362,173
TP6994	176	3e-27	123	97	74	15	87	Pp1	29,506,717	29,506,644
TP12290	171	5e-90	331	99	171	1	171	Pp4	10,256,581	10,256,751
TP20078	65	7e-25	113	98	65	1	65	Pp4	11,045,837	11,045,774
TP16042	173	2e-86	319	98	173	1	173	Pp4	11,117,839	11,117,668
TP15756	157	4e-75	281	98	150	8	157	Pp4	11,513,496	11,513,645
TP33099	65	7e-25	113	98	65	1	65	Pp4	18,843,139	18,843,202
TP26741	156	1e-78	293	99	156	1	156	Pp4	22,295,158	22,295,004
TP23628	129	3e-60	232	98	129	1	129	Pp5	609,337	609,210
TP20147	65	2e-22	105	96	65	1	65	Pp5	4,029,085	4,029,022
TP9269	93	1e-36	153	96	93	1	93	Pp5	3,972,835	3,972,926
TP15051	120	2e-57	222	99	120	1	120	Pp5	3,937,483	3,937,365
TP24137	114	3e-44	178	95	114	1	114	Pp5	4,028,992	4,029,104

Table S3.3., continued.

Query	Length of the query	E-value	Score	PID %	Align length	Query start	Query stop	Match	Match start	Match stop
TP6880	115	1e-49	196	97	115	1	115	Pp5	2,846,847	2,846,960
TP155	129	1e-62	240	98	129	1	129	Pp5	5,961,912	5,962,040
TP17304	100	4e-43	174	98	100	1	100	Pp5	9,774,755	9,774,853
TP16696	65	3e-21	101	98	59	1	59	Pp5	9,349,036	9,348,979
TP38839	129	3e-60	232	98	129	1	129	Pp5	9,673,297	9,673,424
TP20847	98	3e-44	178	98	98	1	98	Pp5	9,496,567	9,496,663
TP33011	65	2e-22	105	96	65	1	65	Pp5	11,980,884	11,980,821
TP5787	163	2e-80	299	98	163	1	163	Pp5	13,482,964	13,483,126
TP31757	133	5e-65	248	98	133	1	133	Pp5	13,111,614	13,111,746
TP1532	177	6e-96	351	100	177	1	177	Pp5	13,120,997	13,120,821
TP15987	65	3e-24	111	98	64	2	65	Pp5	14,927,161	14,927,223
TP9301	65	7e-25	113	98	65	1	65	Pp5	17,176,598	17,176,661
TP6690	161	7e-77	287	98	161	1	161	Pp5	15,766,285	15,766,126
TP9076	135	4e-49	228	96	135	1	135	Pp5	15,953,801	15,953,667
TP9096	168	1e-78	293	97	168	1	168	Pp5	15,400,939	15,400,773
TP37464	90	1e-39	163	98	90	1	90	Pp5	14,891,176	14,891,088
TP32029	152	4e-69	262	97	152	1	152	Pp5	14,736,499	14,736,649
TP29093	65	7e-16	84	94	58	1	58	Pp5	14,105,485	14,105,541
TP27203	118	1e-46	186	95	118	1	118	Pp5	14,205,931	14,205,815
TP34707	188	3e-95	349	98	188	1	188	Pp6	5,427,298	5,427,112
TP24350	174	4e-46	252	94	174	1	174	Pp6	3,321,379	3,321,206
TP24362	181	2e-95	349	99	180	1	180	Pp6	5,320,034	5,319,855
TP11807	64	1e-26	119	98	64	1	64	Pp6	3,166,228	3,166,165
TP4114	65	7e-25	113	98	65	1	65	Pp6	7,263,429	7,263,492
TP4159	65	2e-22	105	96	65	1	65	Pp6	6,429,017	6,429,080
TP20488	166	2e-80	299	98	163	4	166	Pp6	6,414,188	6,414,349
TP15324	182	8e-83	307	97	182	1	182	Pp6	16,850,001	16,849,820
TP16417	65	7e-25	113	98	65	1	65	Pp6	19,487,636	19,487,699

Table S3.3., continued.

Query	Length of the query	E-value	Score	PID %	Align length	Query start	Query stop	Match	Match start	Match stop
TP34383	65	6e-07	54	88	59	3	61	Pp1	40,260,227	40,260,170
TP20502	65	1e-17	89	93	65	1	65	Pp6	24,693,594	24,693,531
TP11297	65	7e-25	113	98	65	1	65	Pp6	26,828,835	26,828,898
TP10517	166	4e-85	316	99	166	1	166	Pp6	26,907,812	26,907,648
TP34743	176	1e-90	333	99	176	1	176	Pp6	30,686,759	30,686,585
TP8828	178	2e-80	299	98	163	1	163	Pp6	30,418,165	30,418,003
TP18700	64	1e-26	119	98	64	1	64	Pp7	7,891,722	7,891,659
TP12076	164	2e-83	309	98	164	1	164	Pp7	146,665	146,828
TP26398	169	2e-86	319	98	167	1	167	Pp7	7,920,297	7,920,405
TP15496	64	1e-26	119	98	64	1	64	Pp7	13,087,657	13,087,594
TP16298	130	4e-53	105	93	122	1	122	Pp7	14,302,633	14,302,762
TP20535	170	5e-87	162	98	168	1	168	Pp7	14,356,831	14,357,000
TP27979	168	1e-90	168	100	168	1	168	Pp7	14,166,103	14,165,936
TP16827	166	5e-87	321	99	166	1	166	Pp7	14,305,337	14,305,172
TP2602	163	1e-75	142	96	158	1	158	Pp7	14,914,034	14,913,872
TP35291	159	7e-83	307	99	159	1	159	Pp7	16,799,981	16,800,139
TP16354	193	1e-103	375	99	193	1	193	Pp7	16,960,341	16,960,149
TP8136	160	2e-83	309	99	160	1	160	Pp7	20,227,564	20,227,405
TP14709	174	4e-94	345	100	174	1	174	Pp7	20,284,636	20,284,809
TP1517	64	1e-26	119	98	64	1	64	Pp7	21,196,885	21,196,948
TP3085	64	1e-26	119	98	64	1	64	Pp7	21,833,997	21,834,060
TP36668	149	3e-63	242	97	138	12	149	Pp7	21,834,103	21,833,966
TP4297	163	1e-87	163	100	163	1	163	Pp8	4,611,842	4,611,680
TP15824	65	4e-20	98	96	61	1	61	Pp8	3,170,330	3,170,389
TP34122	172	1e-90	333	99	172	1	172	Pp8	2,388,632	2,388,803
TP35652	151	4e-04	23	96	59	33	59	Pp4	9,984,770	9,984,744
TP17044	64	1e-20	99	100	50	1	50	Pp8	2,672,088	2,672,137
TP6073	64	2e-06	52	96	30	35	64	Pp4	3,011,364	3,011,335

Table S3.3., continued.

Query	Length of the query	E-value	Score	PID %	Align length	Query start	Query stop	Match	Match start	Match stop
TP29641	172	1e-90	333	99	172	1	172	Pp8	11,921,092	11,921,263
TP24404	65	7e-25	113	98	65	1	65	Pp8	14,623,377	14,623,440
TP37395	167	3e-79	295	97	165	1	165	Pp8	15,870,009	15,870,173
TP16703	64	7e-25	113	98	64	1	64	Pp8	16,520,536	16,520,472
TP10517	166	4e-85	316	99	166	1	166	Pp6	26,907,812	26,907,648
TP34743	176	1e-90	333	99	176	1	176	Pp6	30,686,759	30,686,585
TP8828	178	2e-80	299	98	163	1	163	Pp6	30,418,165	30,418,003
TP18700	64	1e-26	119	98	64	1	64	Pp7	7,891,722	7,891,659
TP12076	164	2e-83	309	98	164	1	164	Pp7	146,665	146,828
TP26398	169	2e-86	319	98	167	1	167	Pp7	7,920,297	7,920,405
TP15496	64	1e-26	119	98	64	1	64	Pp7	13,087,657	13,087,594
TP16298	130	4e-53	105	93	122	1	122	Pp7	14,302,633	14,302,762
TP20535	170	5e-87	162	98	168	1	168	Pp7	14,356,831	14,357,000
TP27979	168	1e-90	168	100	168	1	168	Pp7	14,166,103	14,165,936
TP16827	166	5e-87	321	99	166	1	166	Pp7	14,305,337	14,305,172
TP2602	163	1e-75	142	96	158	1	158	Pp7	14,914,034	14,913,872
TP35291	159	7e-83	307	99	159	1	159	Pp7	16,799,981	16,800,139
TP16354	193	1e-103	375	99	193	1	193	Pp7	16,960,341	16,960,149
TP8136	160	2e-83	309	99	160	1	160	Pp7	20,227,564	20,227,405
TP14709	174	4e-94	345	100	174	1	174	Pp7	20,284,636	20,284,809
TP1517	64	1e-26	119	98	64	1	64	Pp7	21,196,885	21,196,948
TP3085	64	1e-26	119	98	64	1	64	Pp7	21,833,997	21,834,060
TP36668	149	3e-63	242	97	138	12	149	Pp7	21,834,103	21,833,966
TP4297	163	1e-87	163	100	163	1	163	Pp8	4,611,842	4,611,680
TP15824	65	4e-20	98	96	61	1	61	Pp8	3,170,330	3,170,389
TP34122	172	1e-90	333	99	172	1	172	Pp8	2,388,632	2,388,803
TP35652	151	4e-04	23	96	59	33	59	Pp4	9,984,770	9,984,744
TP17044	64	1e-20	99	100	50	1	50	Pp8	2,672,088	2,672,137

Table S3.3., continued.

Query	Length of the query	E-value	Score	PID %	Align length	Query start	Query stop	Match	Match start	Match stop
TP6073	64	2e-06	52	96	30	35	64	Pp4	3,011,364	3,011,335
TP29641	172	1e-90	333	99	172	1	172	Pp8	11,921,092	11,921,263
TP24404	65	7e-25	113	98	65	1	65	Pp8	14,623,377	14,623,440
TP37395	167	3e-79	295	97	165	1	165	Pp8	15,870,009	15,870,173
TP16703	64	7e-25	113	98	64	1	64	Pp8	16,520,536	16,520,472

Table S3.4. SNP-bearing sequences from Lauranne that were used in comparative mapping with the peach sequence assembly. SNP-bearing sequences (query), length of the query sequence, expected values (E-value), score, percentage of identity (PID), start and stop points of each query, mapped peach scaffold (Pp), start and stop points of the mapping sequences are shown.

Query	Length of the query	E-value	Score	PID %	Align length	Query start	Query stop	Match	Match start	Match stop
TP37835	64	3e-24	111	96	64	1	64	Pp1	1,460,776	1,460,839
TP36749	64	3e-24	111	96	64	1	64	Pp1	3,692,168	3,692,105
TP12538	175	1e-87	323	98	175	1	175	Pp1	4,247,266	4,247,440
TP35046	64	5e-29	127	100	64	1	64	Pp1	2,560,495	2,560,558
TP4228	178	5e-87	321	97	178	1	178	Pp1	4,545,704	4,545,881
TP39705	155	4e-78	291	98	155	1	155	Pp1	6,809,615	6,809,769
TP23364	185	6e-96	351	98	185	1	185	Pp1	11,416,183	11,415,999
TP19885	64	1e-26	119	98	64	1	64	Pp1	14,644,471	14,644,534
TP37237	166	5e-87	321	99	166	1	166	Pp1	23,421,612	23,421,777
TP18255	64	5e-29	127	100	64	1	64	Pp1	26,159,752	26,159,689
TP32743	192	6e-93	341	97	192	1	192	Pp1	27,103,583	27,103,392
TP6389	161	3e-79	295	98	161	1	161	Pp1	32,867,098	32,867,258
TP33363	179	2e-80	299	96	179	1	179	Pp1	34,923,017	34,922,839
TP38827	64	2e-25	115	98	62	1	62	Pp1	36,479,517	36,479,578
TP36168	64	5e-29	127	100	64	1	64	Pp1	35,516,185	35,516,248
TP15845	190	5e-75	281	94	188	3	190	Pp1	40,981,162	40,981,346
TP45419	172	8e-86	317	98	172	1	172	Pp1	39,284,068	39,283,897
TP36836	175	5e-81	301	97	175	1	175	Pp1	29,461,374	29,461,202
TP10003	64	5e-29	127	100	64	1	64	Pp1	33,890,605	33,890,542
TP20430	172	3e-88	325	98	172	1	172	Pp1	34,687,316	34,687,145
TP899	172	8e-86	317	98	172	1	172	Pp1	34,687,133	34,687,304
TP4873	170	3e-82	305	98	166	5	170	Pp2	17,690,631	17,690,466

Table S3.4., continued.

Query	Length of the query	E-value	Score	PID %	Align length	Query start	Query stop	Match	Match start	Match stop
TP1574	158	7e-80	297	98	158	1	158	Pp3	10,063,437	10,063,280
TP2219	64	1e-26	119	98	64	1	64	Pp2	14,876,337	14,876,400
TP34047	173	7e-77	287	95	173	1	173	Pp2	3,800,804	3,800,976
TP7187	64	5e-29	127	100	64	1	64	Pp2	1,305,069	1,305,132
TP7579	176	3e-88	325	98	176	1	176	Pp2	13,538,648	13,538,473
TP15654	64	3e-24	111	96	64	1	64	Pp2	15,347,526	15,347,463
TP1933	64	1e-26	119	98	64	1	64	Pp2	18,645,356	18,645,293
TP12629	64	2e-22	105	98	57	1	57	Pp6	13,221,592	13,221,648
TP11674	64	1e-26	119	98	64	1	64	Pp2	20,715,858	20,715,921
TP34564	164	2e-83	309	98	164	1	164	Pp2	21,973,255	21,973,418
TP38620	64	1e-26	119	98	64	1	64	Pp2	23,259,710	23,259,773
TP39254	173	9e-89	327	98	173	1	173	Pp2	24,554,697	24,554,869
TP24448	191	1e-102	371	99	191	1	191	Pp2	24,751,826	24,752,016
TP14542	188	3e-95	349	98	188	1	188	Pp2	25,474,310	25,474,497
TP8770	170	2e-89	329	99	170	1	170	Pp2	26,131,462	26,131,293
TP26193	174	2e-89	329	98	174	1	174	Pp2	29,832,756	29,832,583
TP6571	64	1e-23	109	96	63	1	63	Pp2	27,747,940	27,747,878
TP12428	175	2e-92	339	99	175	1	175	Pp2	27,831,238	27,831,064
TP25503	64	1e-26	119	98	64	1	64	Pp2	27,937,242	27,937,305
TP16651	156	3e-76	285	98	156	1	156	Pp2	27,939,291	27,939,446
TP23267	156	3e-76	285	98	156	1	156	Pp2	27,939,291	27,939,446
TP25017	166	3e-82	305	98	166	1	166	Pp2	24,397,506	24,397,341
TP11899	169	2e-86	319	98	169	1	169	Pp2	24,337,812	24,337,644
TP17917	64	5e-29	127	100	64	1	64	Pp2	24,529,047	24,528,984
TP9410	127	3e-63	242	99	126	1	126	Pp3	532,635	532,760
TP37734	64	5e-29	127	100	64	1	64	Pp3	104,211	104,148
TP18235	181	4e-91	335	98	181	1	181	Pp3	7,607,982	7,608,162
TP18350	175	2e-92	339	99	175	1	175	Pp3	7,038,473	7,038,299

Table S3.4., continued.

Query	Length of the query	E-value	Score	PID %	Align length	Query start	Query stop	Match	Match start	Match stop
TP4478	64	7e-25	113	98	61	4	64	Pp3	9,870,146	9,870,206
TP34810	158	1e-72	274	96	158	1	158	Pp3	12,905,340	12,905,497
TP21444	181	1e-81	303	96	181	1	181	Pp3	9,445,618	9,445,798
TP5114	159	1e-56	220	93	160	4	159	Pp6	14,058,279	14,058,121
TP17874	163	1e-87	323	100	163	1	163	Pp3	10,063,072	10,063,234
TP23919	162	5e-87	321	100	162	1	162	Pp3	16,305,895	16,305,734
TP13105	173	4e-91	335	99	173	1	173	Pp3	18,131,390	18,131,218
TP13014	185	9e-86	317	96	187	1	185	Pp3	19,893,710	19,893,896
TP12086	188	3e-95	349	98	188	1	188	Pp3	19,677,623	19,677,436
TP18642	142	2e-70	266	98	142	1	142	Pp3	23,640,562	23,640,421
TP32984	185	3e-95	349	99	184	1	184	Pp3	25,854,309	25,854,491
TP36190	64	5e-29	127	100	64	1	64	Pp3	21,733,914	21,733,977
TP32410	161	1e-81	303	98	161	1	161	Pp4	24,947,235	24,947,075
TP28761	161	2e-86	319	100	161	1	161	Pp4	24,947,540	24,947,700
TP36434	64	2e-13	75.8	97	42	15	56	Pp4	24,289,504	24,289,463
TP14819	159	1e-75	283	97	159	1	159	Pp4	15,853,339	15,853,497
TP20195	64	1e-26	119	98	64	1	64	Pp4	14,977,134	14,977,071
TP39811	162	3e-82	305	98	162	1	162	Pp4	11,509,614	11,509,453
TP29336	64	5e-29	127	100	64	1	64	Pp4	10,739,574	10,739,637
TP37731	64	1e-26	119	98	64	1	64	Pp4	7,705,879	7,705,942
TP33604	125	1e-62	240	99	125	1	125	Pp4	5,589,483	5,589,359
TP23124	168	3e-88	325	100	164	1	164	Pp4	5,679,004	5,679,167
TP37130	152	4e-81	301	100	152	1	152	Pp4	5,589,307	5,589,458
TP5236	161	1e-81	303	98	161	1	161	Pp4	4,586,068	4,585,908
TP21630	176	9e-89	327	98	177	1	176	Pp4	4,404,788	4,404,612
TP357	64	1e-26	119	98	64	1	64	Pp4	4,404,653	4,404,716
TP20504	151	4e-78	291	99	151	1	151	Pp4	4,106,122	4,106,272

Table S3.4., continued.

Query	Length of the query	E-value	Score	PID %	Align length	Query start	Query stop	Match	Match start	Match stop
TP9083	64	1e-26	119	98	64	1	64	Pp4	4,205,734	4,205,671
TP31655	64	5e-29	127	100	64	1	64	Pp4	4,205,732	4,205,795
TP34218	64	5e-23	107	100	54	1	54	Pp4	3,658,072	3,658,019
TP36487	166	3e-82	305	98	166	1	166	Pp4	2,448,003	2,447,838
TP19998	64	7e-22	103	95	64	1	64	Pp4	3,295,682	3,295,619
TP29814	178	4e-94	345	99	178	1	178	Pp5	12,179,653	12,179,476
TP19487	105	3e-53	208	100	105	1	105	Pp5	12,254,295	12,254,399
TP29143	64	5e-29	127	100	64	1	64	Pp5	14,175,203	14,175,266
TP35343	109	3e-53	208	99	109	1	109	Pp5	11,445,292	11,445,184
TP1721	166	3e-82	305	98	166	1	166	Pp5	14,159,663	14,159,498
TP9077	64	7e-22	103	96	64	1	64	Pp5	14,000,811	14,000,749
TP36685	64	1e-26	119	98	64	1	64	Pp5	13,868,124	13,868,187
TP28045	168	8e-86	317	98	168	1	168	Pp5	14,175,252	14,175,419
TP24992	175	1e-87	323	98	175	1	175	Pp5	11,255,600	11,255,774
TP27036	156	1e-78	293	98	156	1	156	Pp5	9,703,229	9,703,384
TP26493	164	5e-87	321	100	162	1	162	Pp5	10,954,815	10,954,654
TP9629	64	3e-24	111	96	64	1	64	Pp5	10,550,893	10,550,830
TP797	171	3e-85	315	98	171	1	171	Pp5	3,645,223	3,645,393
TP16954	142	2e-70	266	98	142	1	142	Pp5	5,288,302	5,288,161
TP16180	122	8e-42	170	100	86	1	86	Pp5	3,913,098	3,913,183
TP8785	64	5e-29	127	100	64	1	64	Pp5	3,217,881	3,217,944
TP2738	114	4e-56	218	99	114	1	114	Pp5	2,265,718	2,265,605
TP23331	64	1e-26	119	98	64	1	64	Pp6	6,107,660	6,107,597
TP36935	173	6e-93	341	100	172	2	173	Pp6	404,032	403,861
TP38730	171	2e-80	299	97	171	1	171	Pp6	11,663,702	11,663,532
TP7711	180	5e-81	301	96	180	1	180	Pp6	12,923,474	12,923,653
TP11093	64	5e-29	127	100	64	1	64	Pp6	11,687,374	11,687,311

Table S3.4., continued.

Query	Length of the query	E-value	Score	PID %	Align length	Query start	Query stop	Match	Match start	Match stop
TP5832	159	7e-83	307	99	159	1	159	Pp6	14,681,717	14,681,875
TP33408	163	3e-85	315	99	163	1	163	Pp6	20,389,910	20,390,072
TP10006	169	7e-77	287	96	169	1	169	Pp6	23,404,154	23,403,986
TP29906	154	1e-38	161	90	134	21	154	Pp4	14,735,805	14,735,676
TP38666	159	7e-83	307	99	159	1	159	Pp6	24,304,375	24,304,217
TP16653	64	1e-14	79.8	92	56	9	64	Pp6	25,247,359	25,247,414
TP11059	64	3e-24	111	96	64	1	64	Pp6	25,023,993	25,023,930
TP9199	160	1e-78	293	98	160	1	160	Pp6	28,214,538	28,214,379
TP1484	163	2e-80	299	98	163	1	163	Pp6	28,195,564	28,195,726
TP2822	64	1e-26	119	98	64	1	64	Pp6	28,044,930	28,044,867
TP32048	155	2e-80	299	99	155	1	155	Pp6	25,740,980	25,740,826
TP30199	155	1e-75	283	98	155	1	155	Pp7	20,554,896	20,554,742
TP25717	171	2e-92	339	100	171	1	171	Pp7	19,931,987	19,931,817
TP13979	65	7e-25	113	98	65	1	65	Pp7	20,178,313	20,178,250
TP1054	172	8e-86	317	98	172	1	172	Pp7	17,598,511	17,598,340
TP12907	168	5e-81	301	97	168	1	168	Pp7	16,971,086	16,971,253
TP15284	64	2e-10	65.9	89	57	5	61	Pp4	9,955,719	9,955,663
TP2019	176	5e-78	291	96	175	2	176	Pp7	15,398,870	15,399,044
TP20631	163	1e-87	323	100	163	1	163	Pp7	14,756,553	14,756,391
TP33954	64	1e-26	119	98	64	1	64	Pp7	14,469,048	14,468,985
TP22397	64	1e-26	119	98	64	1	64	Pp7	14,709,322	14,709,259
TP11106	161	5e-84	311	99	161	1	161	Pp7	13,373,155	13,372,995
TP18097	174	2e-77	289	95	174	1	174	Pp7	13,947,923	13,947,750
TP37439	64	1e-26	119	98	64	1	64	Pp7	15,179,402	15,179,339
TP11526	64	2e-19	95.6	93	64	1	64	Pp7	12,276,204	12,276,141
TP38516	174	3e-79	295	96	173	2	174	Pp7	11,296,698	11,296,526
TP13612	164	8e-86	317	99	164	1	164	Pp7	11,131,254	11,131,091

Table S3.4., continued.

Query	Length of the query	E-value	Score	PID %	Align length	Query start	Query stop	Match	Match start	Match stop
TP13975	64	5e-29	127	100	64	1	64	Pp8	10,358,708	10,358,645
TP23743	181	3e-79	295	96	181	1	181	Pp8	2,372,919	2,373,097
TP18179	162	3e-82	305	98	162	1	162	Pp8	10,969,869	10,969,708
TP15410	154	2e-73	276	98	147	8	154	Pp8	12,182,454	12,182,308
TP24844	162	3e-82	305	98	162	1	162	Pp8	13,823,084	13,822,923
TP15122	173	5e-84	311	97	173	1	173	Pp8	16,519,916	16,520,088
TP15125	147	2e-73	276	98	147	1	147	Pp8	18,861,531	18,861,385
TP27616	64	5e-29	127	100	64	1	64	Pp8	21,832,434	21,832,371
TP15412	174	9e-92	337	99	174	1	174	Pp8	19,668,055	19,668,228
TP14071	64	1e-26	119	98	64	1	64	Pp8	20,838,353	20,838,290
TP5689	64	5e-29	127	100	64	1	64	Pp8	18,502,714	18,502,777
TP5223	64	1e-26	119	98	64	1	64	Pp8	19,723,712	19,723,649
TP2696	176	2e-95	349	100	176	1	176	Pp8	19,540,160	19,539,985
TP24705	64	3e-24	111	96	64	1	64	Pp8	18,057,366	18,057,429
TP33276	163	5e-87	321	100	162	1	162	Pp8	17,600,194	17,600,033
TP9254	64	5e-29	127	100	64	1	64	Pp8	15,780,123	15,780,186
TP14084	154	3e-82	305	100	154	1	154	s_141*	5,899	5,746
TP17172	64	5e-29	127	100	64	1	64	Pp8	13,300,793	13,300,730
TP18466	167	3e-42	172	100	87	1	87	Pp8	13,056,238	13,056,152
TP10404	157	8e-55	214	95	144	15	157	Pp8	12,870,656	12,870,519
TP23837	165	7e-77	287	96	165	1	165	Pp8	12,683,550	12,683,386
TP8457	64	2e-25	115	98	62	3	64	Pp8	2,005,708	2,005,647
TP1262	64	2e-10	65.9	84	73	1	64	Pp8	2,541,858	2,541,786
TP13896	179	2e-92	339	98	179	1	179	Pp8	497,293	497,471

*s_141-scaffold_141

APPENDIX 4

A year in the life of an almond tree

An annual cycle of an almond tree has many stages and is full of beauty. It is a long journey that a nut tree takes to make its way to become a human food. The stages in the almond annual cycle are dormancy, bloom, pollination, nut growth and maturation, harvest and processing.

S4.1 Dormancy

Almond trees are dormant over the winter, around May to July in Australia and around November to February in the USA. When the weather is cool vegetative buds are in a resting period and flower buds start to differentiate and store up nutrients for next year's crop (Fig. S4.1). At the end of dormancy, emergence of flowers followed by leaves can be observed.

Fig. S4.1 An almond tree in dormancy. Only the 'skeleton' of the tree can be seen from afar.

S4.2 Bloom

Flowering usually occurs from late July to early September in Australia and from late February to early March in the USA. During this period emergence of floral buds on shoots or spurs occurs and those buds burst into light pink and white blossoms. Nonpareil is among the first to bloom, while other cultivars, such as Carmel and Mission, bloom later. Generally blossoms can be observed in alternate rows in the orchard. This is because almonds are not self-fertile and always two or three cultivars are inter-planted in an orchard.

The important blooming stages are pink bud (Fig. S4.2), popcorn (Fig. S4.3), full bloom (Figs. S4.4 and S4.5), petal fall (Fig. S4.6) and post-petal fall (Fig. S4.7).

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

Fig. S4.3 Popcorn stage. Flower is not fully open and petals pop up as in popcorn.

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

S4.3 Full bloom

Full bloom refers to the point at which the majority of flowers in the orchard are fully open (Fig. S4.5). By this time some will be past full bloom while others will be at earlier stages. The proportion of flowers that are fully open when the orchard is at full bloom can vary substantially depending on winter chilling. During high chilling years, about 80% of the blossoms reach full bloom at the same time. In low chilling years the proportion may be below 50%.

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

S4.4 Pollination

Many almond trees are not self-pollinating, so populations of bees (Fig. S4.6) are brought to the orchard to carry pollen and initiate crop development. Under Australian conditions five to eight hives of bees per hectare should be distributed evenly throughout the orchard to obtain satisfactory fruit set.

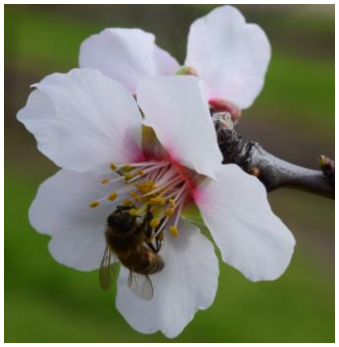


Fig. S4.6 Pollination in almond. Honey bees act as pollen carriers from one flower to the other.

S4.5 Petal fall

Petal fall refers to the stage at which most blossoms in the orchard have dropped their petals and only sepals, styles, stigmas, and stamens are left (Fig. S4.7).

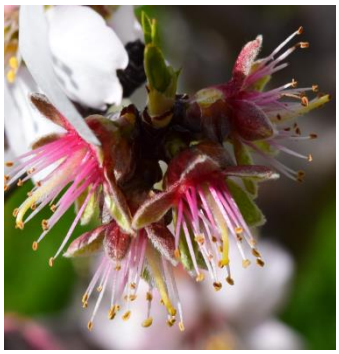


Fig. S4.7 Flowers at petal fall stage.

S4.6 Post-petal fall

Post-petal fall refers to the stage when a majority of the sepals (calyx, shuck) on the remains of flower blossoms are dry, senescing, and turning brown (Fig. S4.8).

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

S4.7 Nut growth and maturing

In Australia, this normally occurs from September to December. Following petal fall the leaves, new shoots and greyish-green fruit begin to grow rapidly. The hulls which cover the growing nuts continue to mature and harden and toward the end of this period the kernel begins to increase in weight while maturing (Fig. S4.9).

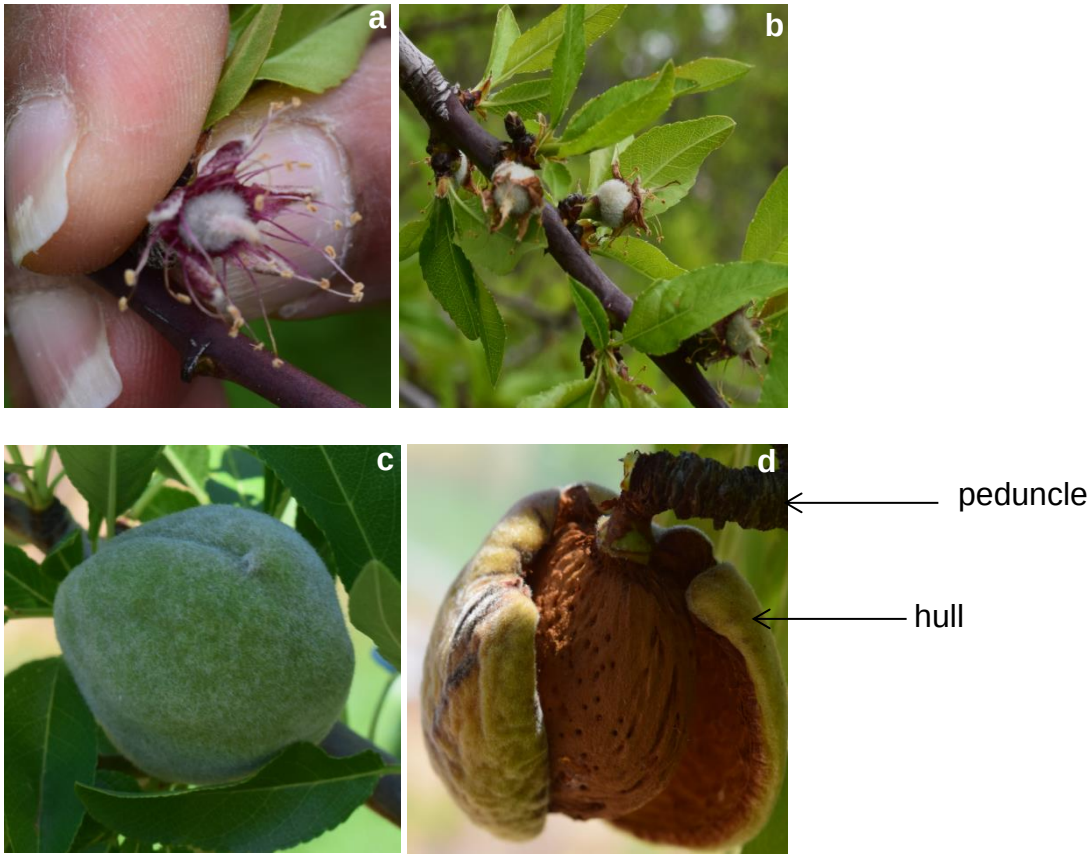


Fig. S4.9 Fruit set in almond: swollen ovary after fertilisation (a), emerging fruits, with sepals and pistils still attached to the fruits (b), an immature fruit (c) and a mature fruit with a split hull (d).

Once the fruit has finished growing, the hull begins to split during summer, from early January. Over the next month the split widens and opens completely. The almond shell is now visible through the split in the hull and the nut begins to dry (Fig. S4.10). The junction between the stem of the whole fruit and the tree weakens and the fruit is ready for harvest. In the USA, these changes can be observed from July to early August.

Fig. S4.10 Nuts are ready to harvest. The hull and the peduncle have turned brown (a), almond nut is loosely attached to the stem (b).

S4.8 Harvest

Harvest occurs between February and April in Australia, when the nut is at an acceptable moisture level (< 10%). Mechanical harvesting requires orchard floors to be clear of large weeds and foreign materials. Shakers are used to vibrate the tree trunk and the fruit (hull, shell and nut) falls to the orchard floor. After drying naturally (usually about 10 days on the orchard floor), fruits are swept into rows and picked up ready for storage. In the USA, this happens from mid October through November.

S4.9 Processing and storage

After harvest, almonds go to a huller/sheller where the nuts pass through a roller to remove the hull, shell and any remaining debris. Then, through the handler for sizing, where the almond kernels drop into separate bins according to size. Almonds are kept in controlled storage conditions to maintain

quality. Almonds can be further processed for manufacturing purposes and supplied as slivered, sliced, diced, split, left whole or ground for almond meal/flour depending on application.