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