Construction of a microsatellite based genetic linkage map of almond

By

Iraj Tavassolian

Thesis submitted in fulfilment of the requirements for the degree of

Doctor of Philosophy (PhD)

Discipline of Wine and Horticulture

School of Agriculture, Food and Wine

May 2008

Table of Contents

Table of Contents ..... *i*

List of Figures ..... *iv*

List of Tables ..... *vi*

Abstract ..... *vii*

Declaration ..... *xi*

Abbreviations ..... *xii*

Acknowledgments ..... *xvi*

Chapter 1 ..... *1*

1.1 General introduction **1**

1.2 Botanical classification of almond **2**

1.3 Economic importance and uses **3**

1.4 Genetic diversity and origin of almond **5**

1.5 Almond breeding **6**

1.6 Genetic markers **8**

1.6.1 Morphological markers 9

1.6.2 Biochemical markers 9

1.6.3 DNA markers 10

1.6.3.1 Hybridisation-based polymorphism 10

1.6.3.2 PCR-based polymorphisms 11

1.7 Application of the saturated linkage map in plant breeding **14**

1.8 Morphological traits in Almond **16**

1.9 Development of an Australian almond linkage map, prior to the **18**

1.9.1 Introduction 18

1.9.2 Mapping population 20

1.9.2.1 Introduction 20

1.9.2.2 Australian almond mapping population 20

1.9.3 Molecular and morphological markers used to develop the first generation of 23

1.9.4 Aims of study and expected outcome 24

Chapter 2 ..... *32*

2.1 Introduction **32**

2.2 Material and Methods **38**

2.2.1 Survey of published and unpublished microsatellites 38

2.2.2 Genomic DNA extractions 38

2.2.2.1 Introduction 38

2.2.2.2 Extraction method 40

2.2.2.3 Modified Lamboy and Alpha (1998) protocol macrotechnique 41

2.2.3 DNA quantification and PCR amplification 42

2.2.4 Segregation analysis and polymorphism detection 43

2.2.4.1 Fluorescent labelled microsatellite detection 43

2.2.4.2 Agarose gel electrophoresis 44

2.2.4.3 PAGE gel electrophoresis 45

2.3 Results **46**

2.3.1 Microsatellites isolated from almond 46
2.3.2 Microsatellites isolated from peach 48
2.3.3 Microsatellites isolated from other *Prunus* species 48

2.4 Discussion 50

Chapter 3 55

3.1 Introduction 55
3.1.1 Shell hardness 57
3.1.2 Kernel size 58
3.1.3 Testa colour and smoothness 59
3.1.4 Kernel taste 60
3.1.5 Percentage of doubles 61

3.2 Materials and Methods 63
3.2.1 Plant material and pomological traits measurement 63
3.2.1.1 Shell hardness and kernel weight assessment 64
3.2.1.2 Kernel size and shape 64
3.2.1.3 Testa characteristics 65
3.2.1.4 Kernel taste and amygdalin quantification 65
3.2.1.5 Percentage of doubles 66
3.2.2 Statistical analysis 66

3.3 Results 69
3.3.1 Shell hardness 69
3.3.2 In-shell weight 69
3.3.3 Kernel weight 70
3.3.4 Kernel length 70
3.3.5 Kernel thickness 70
3.3.6 Kernel width 71
3.3.7 Kernel shape 71
3.3.8 Testa colour and pubescence 71
3.3.9 Kernel taste and amygdalin quantification 72
3.3.10 Double kernel 76

3.4 Discussion 76
3.4.1 Shell hardness 85
3.4.2 In-shell weight 86
3.4.3 Kernel weight 87
3.4.4 Kernel size and shape 89
3.4.5 Testa colour and pubescence 90
3.4.6 Kernel taste and amygdalin quantification 92
3.4.7 Percentage of double kernels 94

3.5 Conclusion 95

Chapter 4 98

4.1 Introduction 98
4.1.1 Marker assisted selection 99
4.1.2 Map-based cloning and physical mapping 100
4.1.3 Comparative mapping 101
4.1.4 Construction of genetic linkage maps 101
4.1.4.1 Mapping population 101
4.1.4.2 Population genotyping 102
4.1.4.3 Linkage analysis 102
4.1.5 The *Prunus* linkage maps 104
4.1.5.1 The almond maps 105

4.2 Materials and Methods 108
4.2.1 Mapping population and DNA extraction 108
4.2.2 Microsatellites marker analysis 109
4.2.3 Verification of microsatellites as MAS for self-compatibility 109

4.2.4 Linkage analysis and map construction 109

4.3 Results 111

4.3.1 Construction of the second Australian linkage maps for almond 111

4.3.1.1 NP linkage map 111

4.3.1.2 LA linkage map 111

4.3.1.3 The integrated linkage map 112

4.3.2 Comparison among the new maps and previous Australian almond maps 112

4.3.3 Marker assisted selection test for *S*-alleles 113

4.3.4 Comparative mapping with other *Prunus* maps 113

4.4 Discussion 114

Chapter 5 134

5.1 Polymorphism detected by microsatellites 136

5.1.1 Comparison of microsatellite detection systems 136

5.1.2 Genotyping of mapping population by microsatellites 137

5.2 Pomological analysis of the mapping population 138

5.2.1 Phenotypic and inheritance studies of the shell hardness, kernel size and weight 139

5. 2.2 Analysis of kernel flavour inheritance and amygdalin contents 140

5.2.3 Phenotypic variation of double kernels 141

5.2.3.1 Benefit of double kernels in genetic studies 142

5.3 Almond saturated map 143

5.3.1 Comparative mapping of the integrated map with *Prunus* reference map 144

5.3.2 Marker assisted selection for *S*-alleles 144

5.4 Future Prospect of the almond map 146

References 148

APPENDIX 179

List of Figures

Fig. 1.1 Almond world production during 2000-2006.....4

Fig. 1.2 The world top producers of almond.....4

Fig. 1.3 main almond production area in Australia 5

Fig. 1.4 Map of the world showing the origin for almond [*Prunus* dulcis (1)] and different relative Prunus species. 6

Fig. 1.5 The controlled cross hybridisation between NP and LA in Waite Campus within the bird cage and branches sealed in insect proof nets.....22

Fig. 1.6 A linkage map of the diploid almond, developed from the markers present in the male parent 'Lauranne' 28

Fig. 1.7 A linkage map of the diploid almond, developed from the markers present in the female parent 'Nonpareil' 29

Fig. 1.8 Integrated linkage map of the diploid almond cross of LA × NP.....31

Fig. 2.1 Microsatellite polymorphism in some progeny amplified by marker EMPA004 49

Fig. 2.2 Electrophorogramme of fluorescent labelled alleles detected by loci BPPCT013 51

Fig. 3.1 Birds are very problematic, mostly in paper / soft-shell cultivars.....58

Fig. 3.2 Variation of testa colour in almond.....60

Fig. 3.3 Average of traits over three year's measurements (2005-2007). A. Percentage of double kernels, B. Shell hardness; C. In-shell weight and D. Kernel weight. 74

Fig. 3.4 The graphs depict the correlation between in-shell and kernel weight among the mapping population of cross 'Nonpareil' × 'Lauranne' in the average value of three years.....75

Fig. 3.5 The graph shows the high correlation between kernel width and length and low correlation between kernel thickness and the other kernel size traits in a segregating population. 75

Fig. 3.6 A. B. C Percentage of progeny traits over three years for A. Shell Hardness, B. Kernel Thickness and C. Kernel Shape 77

Fig. 3.7 A. B. C Distribution of percentage frequency of the population NP × LA grouped into two classes for A. Testa colour, B. Testa pubescence and C. Kernel taste 78

Fig. 3.8 Frequency of total progeny showing 0-100% doubles in 10% increment categories over three years (2005-2007) 80

Fig. 3. 9 Percentage of double kernels in two categories over three years (2005-2007) 80

Fig. 3.10 Double Kernel produced in almond. Compared with single kernel, (illustrated on right) double kernels are deformed and misshapen. 82

Fig. 3.11 The variation of morphological traits in nuts and kernels of the segregating population of NP × LA.....82

Fig. 3.12 Chromatogram of amygdalin in slightly bitter kernels 83

Fig. 3.13 Chromatogram of amygdalin in sweet kernels 83

Fig. 4.1 Approximate position of 28 major genes mapped in different populations of *Prunus*. 106

Fig. 4.2 Segregation of microsatellite UDP98-412 on Self-compatible and incompatible almond cultivars 114

Fig. 4.3 Genetic linkage map of Nonpareil. 117

Fig. 4.4 Genetic linkage map of Lauranne. 119

Fig. 4. 5 Comparison of the IN map with almond RxD and *Prunus* reference map. 126

List of Tables

Table 1.1 Objectives of Almond Breeding Programs 7

Table 1.2 Polymorphic SSR markers used by Gregory, (2004).....26

Table 1.3 Polymorphic SSR markers screened by Mnejja (2004). 27

Table 2.1 Microsatellites: designation, number of screened markers and origin.....39

Table 2.2 Microsatellite name and fluorescine label.....44

Table 2.3 Summary of microsatellite marker analysis results 47

Table 3.1 Descriptor list for ten traits scored over the three years on the F_1 progeny. 68

Table 3.2 Morphological characters measured on F_1 hybrid population of cross between 'Nonpareil' × 'Lauranne', mean value for data collected for three years (2005-2007).....73

Table 3.3 Chi-square test (χ^2) for expected frequency 3:1 or 1:1 of kernel taste, colour, smoothness and double kernels in the population NP × LA.....79

Table 3.4 Broad sense heritability of traits in the population NP × LA for three years (2005-2007) 79

Table 4.1 Distribution of mapped markers and their coverage across linkage groups of 'Nonpareil' (NP), 'Lauranne' (LA) and Integrated (IN) maps.....116

Abstract

Almond (*Prunus dulcis*) is the most important nut crop in terms of world production. Due to its health benefit and high nutritional value the consumption and world supply of almond is increasing. To remain competitive in the world market, the Australian almond breeding program was established to produce cultivars with better adaptation to Australian conditions. As part of this program an almond mapping population consisting of 93 F_1 progeny derived from a cross between the American cultivar 'Nonpareil' (NP) and the European self-compatible cultivar 'Lauranne' (LA) was produced to construct the genetic linkage maps. The first almond linkage map developed prior to the commencement of this project failed to produce the eight linkage groups similar to the basic chromosome number of almond (x = 8) and many large gaps were also observed on the linkage groups. Therefore, more markers were needed to saturate the maps.

Microsatellite markers are considered one of the best choices for mapping studies. 195 microsatellite markers isolated from *Prunus* species were obtained from published papers or by personal communication. Polymorphism was revealed by three different methods, and in general, polyacrylamide gel electrophoresis (PAGE) compared to the fluorescent labelled marker detection using an automated DNA sequencer or agarose gel electrophoresis, showed the most efficient and cost effective method of genotyping. A subset of 54 markers which produced reliable and easily interpretable polymorphic bands was selected to screen the whole mapping population. Microsatellites originally isolated from almond species showed the highest rate of amplification and polymorphism followed by peach microsatellites and the least informative markers were isolated from cherry. It seems that the level of

transportability and usefulness of microsatellite markers is related to the genetic distance of the closely related species. Almond and peach belong to the same subgenus (Amygdalus) and other *Prunus* species are classified in Prunophora subgenus.

The nut, or kernel, is the commercial part of the almond tree, thus to improve the quality of fruit an understanding of environmental influence, heritability and correlation of traits is required. Pomological and quality characters such as: shell hardness, kernel size, shape, taste, pubescence, colour, and percentage of doubles were measured during three consecutive years (2005-2007) on the total mapping population, but data analysis (ANOVA) was performed only on trees that survived for all three years. Most of the traits showed high broad-sense heritability and kernel shape showed the highest heritability of $H^2 = 0.92$ suggesting high genetic control of this trait. Occasionally larger kernels than either parent were found in the progeny indicating potential for improvement of this trait even with smaller kernel size parent that encompass many desirable characters. High correlation was also found between the in-shell and kernel weight ($r = 0.74$), kernel length / kernel width ($r = 0.67$), kernel weight to kernel length ($r = 0.78$) and kernel width ($r = 0.80$). This correlation estimation pointed out in this study indicates that the improvement of one character may result the progress in another trait. Neither of the parents in the mapping population had bitter or obvious slightly bitter taste but slightly bitter kernels were observed among the progeny. Amygdalin was assumed to be responsible for bitter taste in almond; therefore we measured the amount of amygdalin in sweet and slightly bitter kernel progeny by HPLC. However, the results showed that amygdalin exists in sweet kernels as well. Although the average amount of amygdalin in slightly bitter kernels (20.34 mg kg^{-1} FW) was higher than sweet kernels (3.67 mg kg^{-1} FW), some

sweet kernels had higher amounts of amygdalin suggesting the impact of other components on slightly bitter kernel. The highest variability within the traits was observed in the percentage of double kernel, which showed the highest standard error. Strong environmental effects, particularly low temperature at pre-blossom time is speculated to produce much higher double kernels.

Three genetic linkage maps, one for each parent and an integrated map were constructed by the addition of 54 new microsatellite markers to the previous dataset. All the data was scored and coded according to the coding system necessary by JoinMap3 which was used for map construction. 131 markers including microsatellite, ISSR, RAPD, SCAR and *S*-allele markers were placed on the integrated map covering 590.7 cM with the average density of 4.5 cM/marker. The minimum number of six microsatellite markers was placed on linkage group 8 and the linkage group 1 which is the longest linkage group has 14 microsatellite markers. Comparative mapping study with other *Prunus* maps, especially with the highly saturated reference map showed complete synteny and minor changes in the order of four markers on linkage groups compared with *Prunus* reference map. The conservation of molecular marker order observed in this study supports the idea of looking at *Prunus* genome as a single genetic system and practical application of this similarity would be in cross-transportability of microsatellite markers from well developed linkage maps to the less studied species in *Prunus*. Ten microsatellite loci placed on our map have not been reported before and could be used to improve the density of other *Prunus* maps, especially the reference map.

This study contributed to the better understanding of the mode of inheritance and environmental effect on morphological traits and the effect of amygdalin on kernel taste. The most saturated microsatellite based almond linkage map developed

in this study can serve as a framework for future almond breeding program in Australia and benefit *Prunus* improvement programs internationally.

Declaration

This work contains no material which has been accepted for the award of any other degree or diploma in any university or tertiary institution and, to the best of my knowledge contains no material previously published or written by another person, except where due reference has been made in the text.

I give consent for my thesis to be made available for photocopying and loan when deposited in the University library.

Iraj Tavassolian

May, 2008

Abbreviations

A	Amygdalin
AFLP	Amplified fragment length polymorphism
ANOVA	Analysis of variance
BAC	Bacterial artificial chromosome
BC	Before Christ
bp	Base pair
BSA	Bulked segregant analysis
χ^2	Chi-squared
ºC	Degrees celsius
cM	centi-Morgan
CsCl	Cesium choloride
CTAB	Cetyltrimethylammonium bromide
cv.	Cultivar
CAPs	Cleaved amplified polymorphic sequence
DArT	Diversity arrays technology
DETC	Diethylidithiocarbamic acid
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleotide triphosphate
DPSTC	Double pseudo-testcross
EDTA	Ethylene diamine tetraacetic acid
EST	Expressed sequence tag
FAM	6-carboxy-fluorescine
F_1	First familial generation

F_2	Second familial generation
FAM	6-carboxyfluorescein
FAO	Food and agriculture organisation
FISH	Fluorescence in-situ hybridisation
FLB	formamide / bromophenol blue
FW	Fresh weight
g	gram
G	Linkage group
GDR	Genome database for Rosaceae
H	Broad sense inheritability
HCl	Hydrochloric acid
HEX	Hexacholoro-6-flourescine
HEX	Hexachloro-6-carboxy-fluorescein
HPLC	High performance liquid chromatography
ISSR	Inter simple sequence repeat
Kb	Kilobase
LA	Lauranne
LD	Linkage disequilibrium
LOD	Logarithm of odds ratio
LSD	Least significant difference
M	Molar
MAS	Marker assisted selection
min	minute
mg	milligram
µL	Microlitres

mL	Millilitres
mM	Millimolar
mm	Millimetre
NH4Ac	Ammonium acetate
NP	Nonpareil
NSW	New South Wales
PAGE	Polyacrylamide gel electrophoresis
PCR	Polymerase chain reaction
pH	Power of hydrogen
%	Percent
QTL	Quantitative trait loci
®	Registered trademark
r	coefficient of correlation
RAPD	Randomly amplified polymorphic DNA
RFLP	Restriction fragment length polymorphism
RHS	Royal Horticultural Society
SCAR	Sequence characterised amplified region
SD	Standard deviation
SDS	Sodium dodecylsulphate
SE	Standard error
sec	second
SNP	Single nucleotide polymorphism
sp.	Species
SSR	Simple sequence repeats
TBE	Tris borat- EDTA buffer

TE	Tris-EDTA buffer
TEMED	N,N,N',N'- tetramethylethylenediamone
T × E	‘Texas’ × ‘Earlygold’
U	Enzyme unit
USDA	United state department of agriculture
V	Volts
var.	Variety
v/v	Volume per volume
w/v	Weight per volume
YAC	Yeast artificial chromosome

Acknowledgments

This thesis would not been written without the invaluable assistance provided by many people listed below. If I missed anyone I sincerely apologise. I am deeply grateful to Dr Chris Ford for his peer supervision, constructive comments and endless support and encouragement. I always remember him as a role model in my next career. I express my appreciation to Dr Michelle Wirthensohn for all her patience, expert advice and enthusiasm. She has really brilliant mind and was so helpful to me and my wife. I also extend my gratitude to Prof. Margaret Sedgley who was very inspiring and her constant interest gave me the determination about the project from the start.

I wish to specially thank Dr Shubiao Wu for his important assistance on scoring of the microsatellite markers and comments on a draft of the thesis manuscript. Many thanks also to Prof. Raffaele Testolin, Dr Pere Arús and Dr Werner Howad for providing me the sequence of some unpublished microsatellite markers. My thank goes to Dr Ian Riley for statistical assistance and helpful advice during study in Australia.

I am also grateful to many other friends: Esmail, Abbas and Nasser for joyful conversations and sharing their time with me. Many thanks to the colleagues Dr Tricia Franks, Dr Matthew Hayes, Vanessa Melino, Venetia Joscelyne and Crista Burbidge for lively and lovely atmosphere they provide during my lab work.

I gratefully acknowledge Ministry of Science, Research and Technology of Iran for financial support of this study. Finally I want to express my greatest thanks to my wife, Samaneh Poormohammad, and my parent for their patience, constant

encouragement, support and trust that they placed in me throughout this study. Without their help I would never been able to do this research.