Copyright is owned by the Author of the thesis. Permission is given for a copy to be downloaded by an individual for the purpose of research and private study only. The thesis may not be reproduced elsewhere without the permission of the Author.

Comparative Genome Mapping of the Rosaceae

A thesis presented in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in

Plant Molecular Genetics and Genomics

Institute of Molecular BioSciences
Massey University
Palmerston North
New Zealand

Jill M. Bushakra

2012

ABSTRACT

Comparative genome mapping uses genetic map and DNA sequence alignment to assess genome conservation between two or more organisms. This study makes use of the recent genome sequence availability of four Rosaceae genera, and the development of new, and the expansion of existing, linkage maps to: 1) explore overall genome synteny between apple and strawberry; 2) assess homology between, and the degree of ancestral genome rearrangement among, four genera; and 3) compare genome synteny with respect to the production of anthocyanins between raspberry and strawberry.

The inter-tribal comparison of the genomes of apple and diploid strawberry, conducted by adding 56 newly developed orthologous markers to existing linkage maps, identified 21 regions of genomic synteny between the linkage groups of apple and strawberry. In addition, this work identified two each of potential translocations, inversions and insertions, and provided a set of orthologous markers that will be useful for orienting and anchoring other Rosaceae genome sequences.

Orthologous- and other DNA sequence-based markers were used in the construction of new linkage maps for *Rubus occidentalis* 96395S1 and *R. idaeus* ‘Latham’. The sequences from which the *Rubus* markers were designed were compared with the draft genome sequences of *Malus* × *domestica* ‘Golden Delicious’, *Fragaria vesca* ‘Hawaii 4’, and *Prunus persica* ‘Lovell’ to identify regions of orthology. This first comparison of *Rubus* linkage maps with other members of the Rosaceae identified a nearly 1:1 homology between the linkage groups of *Rubus* and *F. vesca*, as well as family-wide conservation among some linkage groups.

The F_1 progeny of *Rubus occidentalis* 96395S1 × *R. idaeus* ‘Latham’ was used to conduct a quantitative trait locus (QTL) analysis to explore the presence of associations between genotype and the variation in concentrations of anthocyanins in the fruit. Seven associations of traits with markers designed from the sequences of transcription factors and anthocyanin biosynthetic pathway genes were identified, providing opportunities for further fine-scale mapping, as well as cloning and expression analyses. The comparison of QTL maps of *Rubus* and *Fragaria* × *ananassa* suggests that homologous genomic regions may be important in the expression of various fruit quality traits.

ACKNOWLEDGEMENTS

It’s hard to believe that four years ago a PhD was furthest from my mind and yet, thanks to a brief comment by David Chagné at PAG 2008, here I am. I could not have completed this journey without the help and support of so many, but primarily I wish to thank my husband, John, for his faith, patience, support and commitment to me and my thirst for learning. We’ve been down another long road together.

Next, I would like to thank my advisors, especially David Chagné and Susan Gardiner for giving me this opportunity and always following through with their promises. Through this venture I have relied heavily on you both and you have always held me up. Thanks also to V. Vaughan Symonds and Emily Buck. Thank you also to Plant and Food Research (PFR) for funding from the Excellence Programme and to Tony Conner for the extension. The people of PFR have been very welcoming, encouraging and fun.

I would like to extend my appreciation to Michael McManus, Paul Dijkwell, and Brent Barrett for seeing my potential and providing helpful and constructive comments during my first two years. A special thanks to Nadia Ilyin for making it possible for John to work from NZ, and to Mom and Janis for taking care of the cats and for being okay with us being so far away.

My sanity was maintained with the help of my friends, especially Amy Watson, Toshi Foster, Claudia Wiedow, Cindy Skema, Fronny Plume, Mariam Sharland, Erica Prier and my family and friends back in California, especially Ed Show, Jennifer Wong-Izzo, and Albert Herrera for keeping me in the loop.

Finally, I would like to dedicate this thesis to my parents, who instilled in me a sense of adventure and willingness to try new things. I miss you both.

TABLE OF CONTENTS

ABSTRACT i
ACKNOWLEDGEMENTS ii
TABLE OF CONTENTS iii
LIST OF FIGURES viii
LIST OF TABLES xv
LIST OF ABBREVIATIONS xix
LIST OF COMMONLY REFERRED TO ROSACEAE SPECIES xxi
CHAPTER ONE 1
1 INTRODUCTION 2
- 1.1 Comparative genome mapping 2
- 1.2 Principles of comparative genome mapping 3
  - 1.2.1 Genetic linkage map construction 5
    - 1.2.1.1 Parental selection and population size for genetic mapping 7
    - 1.2.1.2 Molecular marker development 9
    - 1.2.1.3 Point mutations 9
      - 1.2.1.3.1 Restriction fragment length polymorphism 9
      - 1.2.1.3.2 Amplified fragment length polymorphism 11
      - 1.2.1.3.3 Random amplification of polymorphic DNA 12
      - 1.2.1.3.4 Sequence characterized amplified region 13
    - 1.2.1.4 High-resolution melting 14
    - 1.2.1.5 SNP chips and arrays 16
    - 1.2.1.6 Genotyping-by-sequencing 16
    - 1.2.1.7 Replication errors: simple sequence repeat 17
    - 1.2.1.8 Orthologous marker development 18
      - 1.2.1.8.1 Expressed sequence tags 19
      - 1.2.1.8.2 Conserved orthologous set 19
      - 1.2.1.8.3 Gene-based markers 20
    - 1.2.1.9 Linkage analysis of polymorphic markers 21
      - 1.2.1.9.1 Selective mapping for linkage analysis 21
  - 1.2.2 Summary 22
  - 1.2.3 Quantitative trait locus mapping in plants 22
    - 1.2.3.1 Segregating populations 25
    - 1.2.3.2 Phenotyping trait variation 25
    - 1.2.3.3 Linkage map construction 25
    - 1.2.3.4 QTL identification 25
    - 1.2.3.5 QTL significance 27
  - 1.2.4 Association mapping in plants 28
  - 1.2.5 Summary 29
  - 1.2.6 Comparative genome mapping in plants 29
    - 1.2.6.1 Previous comparative genome mapping studies in plants 29
    - 1.2.6.2 Previous comparative genome QTL mapping studies in plants 35
  - 1.2.7 Summary 37
- 1.3 A brief history of the Rosaceae 38

1.4 Economic importance of Rosaceae genera 42
1.5 Genetic mapping in the Rosaceae 44
1.5.1....Obstacles to genetic mapping in the Rosaceae 44
1.5.2....Genetic map construction in the Rosaceae.. 45
1.5.3....Early linkage map construction in Rosaceae 46
1.5.4....The rise of simple sequence repeat molecular markers for Rosaceae studies48
1.5.5....Gene expression analyses and mapping in Rosaceae................................ 50
1.5.6....Summary 53
1.6 Quantitative trait locus mapping in Rosaceae.. 53
1.6.1....Apple 54
1.6.2....Strawberry 55
1.6.3....*Prunus* 55
1.7 Comparative genome mapping in the Rosaceae 56
1.7.1....Primer pair transferability between subtribes 59
1.7.2....Primer pair transferability within subtribes.. 64
1.7.3....Recent advances in genomics and technology for the Rosaceae 65
1.7.4....Summary 66
1.8 THESIS AIM AND OBJECTIVES 67
REFERENCES.. 69
CHAPTER TWO 86
2 Rosaceae conserved orthologous set (RosCOS) markers as a tool to assess genome synteny between *Malus* and *Fragaria* 87
2.1 Abstract.. 87
2.2 Introduction... 87
2.3 Materials and methods 92
2.3.1....Segregating populations used for RosCOS mapping in apple 92
2.3.2....RosCOS marker development for use with HRM in *Malus* 93
2.3.2.1 Marker Design Method 1: Intron-spanning............................... 93
2.3.2.2 Marker Design Method 2: SNP 94
2.3.2.3 PCR and HRM conditions 94
2.3.3....Data collection and linkage analysis 94
2.3.4....PCR fragment sequencing... 95
2.3.5....RosCOS genome coverage in apple 95
2.3.6....*In silico* detection of *Malus* SSR sequences in the *Fragaria vesca* WGS ... 95
2.3.7....*In silico* detection of strawberry gene-based markers in the GD apple draft whole genome sequence.. 96
2.3.8....Genetic map alignment and assignment of ancestral syntenic blocks 96
2.4 Results.. 97
2.4.1....RosCOS marker development for use with HRM in *Malus* 97
2.4.2....Confirmation of the genetic map position of RosCOS compared to the draft apple genome 99

2.4.3 ... Genome coverage of RosCOS in apple 99
2.4.4 ... Genome coverage of RosCOS in FVxFB 100
2.4.5 ... *In silico* detection of *Malus* SSR sequences in the *Fragaria vesca* genome 101
2.4.6 ... *In silico* detection of strawberry gene-based markers in the draft apple whole genome sequence 101
2.4.7 ... Genetic map synteny between *Malus* and *Fragaria vesca* 101
2.4.7.1 FVxFB LG II and apple homeologs LG 5 and 10, LG 8 and 15, and LG 7 107
2.4.7.2 FVxFB LG III and apple homeologs LG 3 and 11, and LG 17. 107
2.4.7.3 FVxFB LG V and apple homeologs LG 6 and 14, and LG 13 and 16 107
2.4.7.4 FVxFB LG VI and apple homeologs LG 12 and 14, and LG 17 108
2.5 Discussion 108
2.5.1 ... Orthologous marker development in Rosaceae 108
2.5.2 ... Genetic map synteny using RosCOS and the HRM technique 109
2.5.3 ... Ancestral chromosome contributions to *Malus* and *Fragaria* linkage groups 110
2.6 Conclusions 113
2.7 Acknowledgements 113
2.8 Authors' contributions 113
REFERENCES 115
2.9 Supplemental files 121
2.10 Addendum 136
CHAPTER THREE 137
3 Construction of black (*Rubus occidentalis*) and red (*R. idaeus*) raspberry linkage maps and their comparison to the genomes of strawberry, apple, and peach 138
3.1 Abstract 138
3.2 Introduction 138
3.3 Methods 142
3.3.1 ... Plant material and DNA extraction 142
3.3.2 ... Genetic markers 143
3.3.3 ... PCR and HRM conditions 146
3.3.4 ... Data collection and linkage analysis 146
3.3.5 ... Genetic map comparison to published Rosaceae genomes 146
3.3.6 ... Ancestral chromosome contribution determination 147
3.4 Results 147
3.4.1 ... *Rubus* genetic linkage map construction 147
3.4.2 ... Comparative genome mapping between *Rubus* and *Fragaria* 152
3.4.3 ... Comparative genome mapping between *Rubus* and *Malus* 156
3.4.4 ... Comparative genome mapping between *Rubus* and *Prunus* 157
3.4.5 ... Comparative genome mapping between *Rubus* and the Rosaceae ancestral genome 157

3.5 Discussion 159
3.5.1 Inter-specific *Rubus* genetic map construction 159
3.5.2 Development of orthologous markers for use in *Rubus* 160
3.5.3 Genetic map comparison between *Rubus* and other Rosaceae genera 162
3.6 Conclusions 163
3.7 Acknowledgements 164
3.8 Authors' contributions 164
REFERENCES 165
3.9 Supplemental files 169
CHAPTER FOUR 187
4 QTL involved in the modification of cyanidin compounds in black and red raspberry fruit 189
4.1 Abstract 189
4.2 Introduction 189
4.3 Methods 192
4.3.1 Plant material 192
4.3.2 Marker nomenclature and linkage groups 192
4.3.3 Experimental design, sample collection, extraction, preparation and analysis 193
4.3.3.1 Sample extraction, 2009: Solvent extraction method 194
4.3.3.2 UHPLC analysis of fruit polyphenolics 194
4.3.3.3 Compound quantification, 2009 195
4.3.3.4 Sample extraction, 2010 and 2011: Juice extraction method 195
4.3.3.5 HPLC analysis of fruit anthocyanins: 2010 and 2011 195
4.3.3.6 Confirmation of compound identity, 2010 and 2011 196
4.3.4 Statistical analyses 196
4.3.5 QTL map construction 197
4.3.6 QTL mapping 197
4.3.7 Epistasis calculations 198
4.3.8 Comparative gene analysis 198
4.4 Results 198
4.4.1 Description of traits 198
4.4.2 Descriptive statistics 201
4.4.3 Identification of QTL 207
4.4.3.1 Individual compounds 207
4.4.3.2 Ranking individuals 211
4.4.3.3 Total anthocyanin 212
4.4.4 Candidate genes co-locating with QTL 212
4.4.4.1 Association of QTL with markers derived from polyphenolic biosynthetic pathway genes and transcription factors 213
4.4.4.2 Association of QTL with markers derived from additional gene sequences 213

4.4.5 ... Comparative gene analysis of the top of RLG2 214
4.5 Discussion 214
4.5.1 ... Individual anthocyanin concentrations in the F_1 progeny of 96395S1 × 'Latham' 214
4.5.2 ... Pearson's correlations coefficient analyses 214
4.5.3 ... 'Latham'*Rubus* linkage groups 2 and 7 and cyanidin modification........... 215
4.5.4 ... Comparison of QTL identified in the progeny of 96395S1 × 'Latham' with those identified previously in the progeny of 'Latham' × 'Glen Moy'...... 216
4.6 Conclusions 218
4.7 Authors' contributions 219
4.8 Acknowledgements 219
REFERENCES 220
4.9 Supplemental files 224
CHAPTER FIVE 243
5 Conclusions 244
5.1 Using the known to delve into the unknown 244
5.2 Ancestral genome contribution 245
5.3 Lineage-specific differences 245
5.4 Impressions and interpretations 246
5.5 Future work 250
REFERENCES 252
BIBLIOGRAPHY 254
APPENDIX: Reprints 289

LIST OF FIGURES

Fig. 1.1: Simple illustration of comparative genome mapping, synteny and colinearity. Markers on the chromosome from Genome A constitute a syntenic group, as do markers on the chromosome from Genome B (thick black lines). Colinearity is demonstrated by loci in the same order on both chromosomes. 4

Fig. 1.2: Relationship between orthologs and paralogs. An ancestral gene (grey) undergoes duplication in Species 0 to form paralogous genes A & B. A speciation event leads to Species 1 and 2. Genes A1 & A2 are orthologs, as are genes B1 & B2. A1 & B1 are paralogs, as are A2 & B2. However, A1 & B2 (and A2 & B1) are considered homologous, as they are not inherited by vertical descent (orthologs), nor are they duplicated within the same genome (paralogs). Figure copied from Koonin (2001) Figure 1. 5

Fig. 1.3: Illustration and explanation of meiotic recombination in a backcross 1 (BC1) progeny. F_1: first filial generation, the result of a cross between two parents; P: parent; A_1, a_1, B_2, b_2: loci; line color indicates parental descent; shading represents recombinant genotypes. 6

Fig. 1.4: Illustration of restriction fragment length polymorphism (RFLP) technique. (a) Two restriction endonucleases are used to digest DNA into fragments of various lengths. Horizontal lines represent homologous chromosomes. The lower chromosome has a difference in nucleotide sequence relative to the upper chromosome at a_{1+1}, thereby creating a fragment of a different length. (b) The hybridization pattern of the fragments is visualized on the gel. The two different sized fragments generated by enzyme A are shown as two bands of different molecular weight (the band of greater molecular weight is higher on the gel). This indicates heterozygosity at this locus. Enzyme B fragments are the same size and are visualized as a single band. This indicates homozygosity at this locus. The probed single copy region can also be used to detect homologous regions in other taxa. Figure modified from Botstein et al. (1980) Figure 1. 11

Fig. 1.5: Illustration of amplified fragment length polymorphism (AFLP) technique. Top: restriction fragment formed by the combination of enzymes *Eco*RI-*Mse*I with 5' protruding ends. Center: the same fragment with ligated adapters for the respective enzymes. Bottom: both strands of the fragment with their corresponding AFLP primers. The 3’ end of the primers and their recognition sequence in the *Eco*RI-*Mse*I fragment are highlighted. Figure copied from Vos et al. (1995) Figure 1. 12

Fig. 1.6: Gel demonstrating random amplification of polymorphic DNA (RAPD) technique. Amplification was performed on soybean *Glycine max* (M) and *G. soja* (S). The original random primer is shown in lanes 1-2. Each subsequent pair of lanes shows how the banding pattern changes with one change of nucleotide in the primer. Nucleotide change is boxed. The arrow on the right points to a polymorphic 0.65 kb band in the final pair of lanes (21-22). This experiment demonstrates that 10 nucleotide-long oligos can be used as PCR primers and detect polymorphism. It also illustrates that

nucleotide changes will determine fragment amplification. Figure copied from Williams et al. (1990) Figure 2. 13

Fig. 1.7: Illustration of high-resolution melting (HRM) technique. An example of genotyping two adjacent single nucleotide polymorphisms (SNP) within a 110 base pair (bp) fragment. (A) Original HRM curves showing two individuals each (replicated) for six genotypes (each color). (B) The same curves as shown in (A), after normalization. (C) Normalized curves after temperature shifting of each curve to overlay them on one of the AA (wildtype) curves between 5% and 10% fluorescence. (D) Fluorescence difference curves obtained by subtracting each curve in (C) from one wildtype curve. The temperature axis in (C) & (D) now reflects the difference in temperature relative to the wildtype, rather than an absolute temperature. The curves shown in (D) demonstrate the difference in shapes seen between homoduplexes (single, high peak shown in green and pink) and heteroduplexes (various shapes shown in yellow, blue and red). The black horizontal line is the wildtype. Genotype abbreviations (panel D): A: adenine; C: cytosine; S: either guanine or cytosine. Figure copied from Wittwer et al. (2003) Figure 3. 15

Fig. 1.8: Illustration of simple sequence repeat (SSR) technique. (a) DNA sequence from soybean with forward and reverse PCR primers shown with arrows; repeat motif is in bold. (b) Autoradiogram of a polyacrylamide gel separation of ^{33}P-labelled PCR products for a range of soybean DNA templates (one in each lane); outer-most lanes are size standards. The banding patterns shown in (b) illustrate the number of alleles per individual (bands per lane) and the number of alleles throughout the genus (variation across the gel). Figure after Powell et al. (1996) Figure 1. 18

Fig. 1.9: An example of quantitative trait locus (QTL) mapping in rice. Chromosome 8 of rice is represented as a vertical open bar with markers (names on left) distributed along its length. (a) The size and location of a homeologous segment of the wheat chromosome 7A based on markers in common; (b-j) various QTL for agronomic, disease resistance, flowering, and grain ripening characteristics. Asterisk (b-e, g, h) indicates traits mapped in a single population. No asterisk (f, i, j) indicates that the QTL were mapped in multiple populations using a common set of RFLP markers. This figure is an example of using comparative genome mapping to identify regions of similar function in two species (rice and wheat), and locates areas of interest for focused fine-scale mapping to detect underlying genes controlling traits of interest. Figure copied from McCouch & Doerge (1995) Figure 2. 24

Fig. 1.10: Distribution of a quantitative trait in parents and offspring. (A) The symmetrical shape of the data when the distribution of the trait is independent of the marker. (B) The asymmetrical shape of the data when the distribution of the trait varies according to a particular marker. The correlation coefficient (r) represents the tendency for pairs of numbers (in this case the trait measurement and the marker position) to vary together. Figure copied from Hartl & Jones (2009) Figure 18.13. 27

Fig. 1.11: Example of a QTL likelihood map and the construction of a one-LOD support interval. The peak of the QTL is located at 62.5 centimorgans (cM) with a confidence interval that spans from approximately 60 to 66 cM (dark bar and dashed lines), which equals the drop of 1 LOD. Figure copied from van Ooijen (1992) Figure 1. 28

Fig. 1.12: Conserved linkage between rice chromosome 4 (black) and maize chromosomes 2 and 10 (white). Maize chromosome 10 is inverted to clarify the relationship of it with the other chromosomes. Connecting lines show colinearity; solid bars indicate the approximate positions of centromeres; numbers in brackets following marker name or in white bars on chromosomes indicate homologous linkage group association. This figure illustrates the relationship of rice chromosome 4 to specific arms of the maize chromosomes (arms delimited by centromere). Colinearity is not completely conserved as three markers in the middle of rice 4 are homologous to maize chromosomes 4 and 5. Figure copied from Ahn & Tanksley (1993) Figure 3. 31

Fig. 1.13: Consensus maps for group 2 chromosomes of three Triticeae species. Conserved segments from homeologous rice, maize, and oat chromosomes are superimposed on the consensus map. Segments 1, 2, and 3 are conserved in all species compared. Symbols in bold type at the right of the linkage maps represent putative gene loci in Triticeae from two sources. Black arrows indicate centromeres. This figure demonstrates high degrees of colinearity across the grass subfamilies Ehrhartoideae (rice), Panicoideae (maize), and Pooideae (oats). Figure copied from Van Deynze et al. (1995) Figure 5...32

Fig. 1.14: Comparative mapping of a 1.5 megabase (Mb) *Arabidopsis thaliana* contig and its homeologs in *Brassica nigra*. (a) Physical map of a 1.5 Mb *A. thaliana* contig and the RFLP probes used on *B. nigra*. (b) Schematic genetic linkage map of the *B. nigra* linkage groups (LG) 2, 5, and 8 that are homologous to the *A. thaliana* contig (black boxes); grey boxes indicate homeology among the *B. nigra* LG. (c) The three *B. nigra* linkage groups in context. Figure copied from Lagercrantz et al. (1996) Figure 1........33

Fig. 1.15: Comparative genetic maps of potato and tomato based on RFLP. (A) Potato linkage map. (B) Tomato: hatched boxes represent intervals in which marker order is inverted between the two genomes. Centromeres are indicated by brackets. Figure copied from Bonierbale et al. (1988) Figure 1. 34

Fig. 1.16: Resistance gene organization in the Solanaceae. Black ring = tomato genome (T1-T12); grey ring = potato genome (I-XII); white ring = pepper genome (1-12). Solid lines perpendicular to the rings indicate chromosome ends; dashed perpendicular lines indicate borders of rearranged regions; black arrow heads indicate regions of inversion; ovals are drawn around R gene clusters. Black stars are added to show the two loci at which QTL for resistance to *Phytophthora* sp. in pepper (*phyt3*, *phyt1*) and potato (*phyt7*, *Pi1*) are located. Figure copied from Grube et al. (2000) Figure 1. 36

Fig. 1.17: The complete phylogeny of the Rosaceae divided into three subfamilies. Colored boxes highlight the genera addressed in this thesis. Phylogeny established by strict consensus of 226 most parsimonious trees constructed by analyzing ten genes or genomics regions (six nuclear and four chloroplast loci) in 91 taxa and and two cloned sequences in 54 taxa; bootstrap values above branches; Bayesian clade credibility below branches; arrows indicate groups that were supported by the Bayesian analysis but were not recovered in the strict consensus tree. Figure copied from Potter et al. (2007) Figure 1. 40

Fig. 1.18: Relationship of apple chromosomes based on a chromosome-by-chromosome comparison as determined by gene homology. Chromosome loss is hypothesized and shown here as the loss of portions of chromosomes 1 and 18, with the remainder of

chromosome 18 and part of chromosome 2 translocated to chromosome 15. Grey areas indicate missing duplicate counterparts; letters p & q indicate chromosome orientation. Figure modeled after Velasco et al. (2010). 41

Fig. 1.19: Alignment of ‘Emperor Francis’ (EF) and ‘New York 54 (NY) sweet cherry parental maps with the *Prunus* reference bin map [‘Texas’ almond × ‘Earlygold’ peach (T×E)]. Only markers in common between T×E and both EF and NY are shown on the T×E map. Solid connecting lines represent homology between the mapped markers; dashed connecting lines represent markers common between T×E and one sweet cherry parent. Shaded regions on the T×E linkage groups delimit the mapping bins. Boxed markers are anchors between EF and NY. Asterisks indicate deviation ($P<0.05$) from the expected chi-square segregation value. This figure illustrates the overall colinearity found in the genus *Prunus* as well as some rearrangements that may be unique to sweet cherry. Figure copied from Olmstead et al. (2008) Figure 1. 57

Fig. 1.20: Alignment of *Eriobotyra*, *Malus* and *Pyrus* linkage groups (LG) sharing at least five linked SSR from different SSR-based maps. (A) LG5; (B) LG9; (C) LG12; (D) LG16; Apple maps represented by: F: ‘Fiesta’; D: ‘Discovery’; F×T: ‘Fiesta’ × ‘Totem’; Pear maps represented by B: ‘Bartlett’; H: ‘Housui’; Loquat represented by A: ‘Algerie’ and Z: ‘Zaozhong-6’. This figure illustrates the overall colinearity among the genera, with some rearrangements that may be specific to loquat. Figure copied from Gisbert et al. (2009) Figure 2. 58

Fig. 1.21: Detection of polymorphism by six apple-derived expressed sequence tag-simple sequence repeat (EST-SSR) markers across eight apple cultivars (lanes 1-8 for each marker). Only markers CN918509 and CN857658 exhibit polymorphism in apple; all other markers appear monomorphic. Gel photograph copied from Gasic et al. (2009b) Figure 1........61

Fig. 1.22: Gel demonstrating the amplification of apple-derived expressed sequence tag-simple sequence repeat (EST-SSR) marker CO414802 across different Rosaceae species. Repeat type $(GGA)_7$; size predicted from apple EST database is 142 base pairs (bp); M, 1kb DNA standard; *lanes 1-6* pear; *7-9* rose; *10-17* strawberry; *18-21* apricot; *22-26* European plum; *27-31* almond; *32-36* Japanese plum; *37-40* peach; and *41-42* apple. This gel demonstrates that this marker, while able to amplify a band in many species, is uninformative for mapping. Gel photograph copied from Gasic et al. (2009b) Figure 2. 61

Fig. 2.1: Simplified and abbreviated Rosaceae phylogeny, highlighting economically important genera (modeled after Potter et al. 2007). Branches approximate shared ancestry and are not to scale nor all-inclusive. Base chromosome numbers (x) are indicated. Phylogeny illustrates the relative distance between *Malus* (apple) and *Fragaria* (strawberry). 89

Fig. 2.2: Relationship of apple chromosomes based on a chromosome-by-chromosome comparison as determined by gene homology (modeled after Velasco et al. 2010). Grey areas indicated missing duplicate counterparts. Letters p & q indicate chromosome orientation. 90

Fig. 2.3: Syntenic genomic regions between *Malus* and *Fragaria vesca* as determined by genetic mapping of 56 Rosaceae conserved orthologous set (RosCOS) markers in

common between the two genera (see also Supplemental Fig. S2.1 and Supplemental Tables S2-1, S2-2 and S2-5). Linkage maps of *Fragaria vesca* × *F. bucharica* (FVxFB) linkage groups (LG) I-VII and apple integrated LG 2-8 and 10-17 are shown. Connecting lines illustrate numbers and locations of RosCOS markers in common. Color blocks and notations above connecting lines indicate the ancestral relationship based on Vilanova et al. (2008) and Illa et al. (2011). For example, A1.1 indicates derivation from Ancestral chromosome 1 (A1) and one instance of occurrence (.1). Solid black vertical lines indicate regions in which RosCOS markers were not mapped in either apple or strawberry. Presence of an * followed by ancestral derivation indicates support from *in silico* location of apple orthologs and their approximate position on the FVxFB LG. 105

Fig. 3.1: Simplified and abbreviated Rosaceae phylogeny. The diagram represents the genera included in this study. Branches approximate shared ancestry and are not to scale nor all-inclusive. Base chromosome numbers (*x*) are indicated. The phylogeny illustrates the relative distances between the sub-families Rosoideae and Spiraeoideae, and between the genera within each sub-family. 140

Fig. 3.2: Parental pedigrees. (A) Partial pedigree of black raspberry parent 96395S1 (*Rubus occidentalis*). Great grandparents V32 and V23 are full siblings; grandparents 88408GO-5 and 88408RQN10 are full siblings; R indicates the individual was produced by reciprocal cross. Individuals used in the crosses 88408 and 88407 are half-siblings as they are both derived from 'Jewel'. Parents 92361AC10 and 92365BE10 share the same pollen donor (88407RTN11). (B) Known pedigree of red raspberry parent 'Latham' (*Rubus idaeus*)........ 143

Fig. 3.3: Genetic linkage maps of *Rubus occidentalis* 96395S1 and *R. idaeus* 'Latham'. Original linkage group (OLG) nomenclature after Graham et al. (2004) is given in parentheses, followed by proposed *Rubus* linkage group (RLG) nomenclature corresponding with *Fragaria* LG nomenclature. Map distances are in centimorgans (cM). RLG2 through RLG7a are presented for 96395S1; RLG1 through RLG7a & 7b are presented for 'Latham'. Connecting lines indicate markers in common. 151

Fig. 3.4: Comparison of *Rubus idaeus* 'Latham' genetic linkage map with *Fragaria*, *Malus*, and *Prunus*. Markers for which sequence was available for BLAST analysis are arranged based on the 'Latham' linkage map order; map distance is in centimorgans (cM). *Rubus* original linkage group (OLG) is based on numbering scheme by Graham et al. (2004). Colors indicate ancestral chromosome contribution (see Fig. 3.5 for color key based on Vilanova et al. 2008). 154

Fig. 3.5: Ancestral chromosome (A1-A9) contribution to the genomes of *Rubus* linkage groups (RLG1-RLG7), *Fragaria* linkage groups (FLGI-FLGVII), *Prunus* groups (G1-G8), and *Malus* linkage groups (MLG1-MLG17). 155

Fig. 4.1: Scatter plot for compounds cyanidin 3-O-sophoroside (C3S) and cyanidin 3O-glucoside and cyanidin 3O-sambubioside (C3G_C3Sb) in 2009. The subset of F_1 progeny of *Rubus occidentalis* 96395S1 × *R. idaeus* 'Latham' segregates into two groups according to genotype at locus RubFruitE4 on 'Latham' *Rubus* linkage group (RLG) 2. Set A (open diamond symbols) consists of 61 heterozygous ("lm") individuals; Set B (solid cross symbols) consists of 94 homozygous ("ll") individuals, where "ll" and "lm" represent allele combinations. The correlation coefficient for the

entire data set is r = 0.08; the correlation coefficient for Set A is r = 0.57, and the correlation coefficient for set B is r = 0.85. Arrows indicate progeny that group opposite of their genotype. 205

Fig. 4.2: Ranking of individuals based on the percentage of cyanidin 3-O-sophoroside (C3S) of the total anthocyanin (ACY) accumulated in 2009. This graph suggests that, generally, percent C3S accumulation in 2009 (diamonds) agreed with percent C3S accumulation in 2010 (squares). Percent C3S accumulation in 2011 was the most variable (triangles), with the low and high extremes trending to the opposite of 2009. 212

Fig. 4.3: Six cyanidin-based anthocyanin compounds found in red and black raspberry fruit. The addition of sugar moieties (above the arrows) modifies cyanidin 3-O-glucoside (C3G). 215

Supplemental Fig. S2.1: Synteny between *Fragaria vesca* × *F. bucharica* (FVxFB) and apple as determined by genetic mapping of 56 Rosaceae conserved orthologous set (RosCOS) markers. Colors indicate ancestral relationships; *in silico* homolog of apple orthologs of strawberry gene-based markers is provided as additional support. 129

Supplemental Fig. S3.1: Comparison between *Rubus idaeus* 'Latham' genetic map and *Fragaria* physical map constructed using BLAST analysis. 'Latham' map distance is measured in centimorgans (cM); *Fragaria* linkage group (FLG) map distance is measured in megabase pairs (Mbp). 'Latham' linkage groups arranged in proposed order (RLG) with original numbering (OLG) in parentheses. Connecting lines indicate markers in common. 169

Supplemental Fig. S3.2: Comparison between *Rubus idaeus* 'Latham' genetic map and *Malus* physical map constructed using BLAST analysis. 'Latham' map distance is measured in centimorgans (cM); *Malus* linkage group (MLG) map distance is measured in megabase pairs (Mbp). 'Latham' linkage groups arranged in proposed order (RLG) with original numbering (OLG) in parentheses. Connecting lines indicate markers in common. 173

Supplemental Fig. S3.3: Comparison between *Rubus idaeus* 'Latham' genetic map and *Prunus* physical map constructed using BLAST analysis. 'Latham' map distance is measured in centimorgans (cM); *Prunus* linkage group (PG) map distance is measured in megabase pairs (Mbp). 'Latham' LG arranged in proposed order (RLG) with original numbering (OLG) in parentheses. Connecting lines indicate markers in common. 177

Supplemental Fig. S4.1: Residual plots from simple fixed effects analysis for traits for which epistasis analyses were conducted. 227

Supplemental Fig. S 4.2: Graphical summaries of the descriptive statistics of each trait detected in the F_1 progeny of *Rubus occidentalis* 96395S1 × *R. idaeus* 'Latham' in 2009, 2010 and 2011. 231

Supplemental Fig. S 4.3: Spearman's ranking scatter plots for anthocyanin compounds detected in the sub-set of 155 F_1 progeny of *Rubus occidentalis* 96395S1 × *R. idaeus* 'Latham' for 2009, 2010, and 2011. 234

Supplemental Fig. S 4.4: Scatter plots for anthocyanin compounds detected in the subset of 155 F_1 progeny of *Rubus occidentalis* 96395S1 × *R. idaeus* ‘Latham’. 235

Supplemental Fig. S 4.5: Linkage maps of *Rubus idaeus* ‘Latham’ and *R. occidentalis* 96395S1 displaying quantitative trait loci (QTL) for variation in anthocyanin trait expression. QTL were mapped to all ‘Latham’ *Rubus* linkage groups (RLG) except RLG4. QTL were mapped to RLG 3, 5, and 6 for 96395S1...238

Supplemental Fig. S 4.6: Ranking of 155 individual F_1 progeny of *Rubus occidentalis* 96395S1 × *R. idaeus* ‘Latham’ by their percent accumulation of each of the compounds analyzed in 2009 (diamonds), compared with their accumulation in 2010 (squares) and 2011 (triangles). 240

Supplemental Fig. S 4.7: Linkage map of *Rubus idaeus* ‘Latham’ displaying quantitative trait loci (QTL) for variation in total anthocyanin trait expression. QTL were mapped to just two ‘Latham’ *Rubus* linkage groups (RLG). 241

LIST OF TABLES

Table 1-1: Population types commonly used for genetic mapping in plants. 8

Table 1-2: Summary of the percentage hybridization success of restriction fragment length polymorphism (RFLP) probes derived from barley, oat and rice to identify homologous loci on seven grass genera (Van Deynze et al. 1995). Wheat, barley and oats belong to subfamily Pooideae; rice is from Ehrhartoideae; and maize, sorghum and sugarcane are from Panicoideae. 32

Table 1-3: World-wide production (2008, in tonnes) and value (2007, in US$) of eleven Rosaceae fruit crops. 42

Table 1-4: Economically important species of Rosaceae. Table after Hummer & Janick (2009) Table 2. 43

Table 1-5: Genetic maps constructed in Rosaceae genera using predominantly first generation molecular markers. Examples of first generation markers are restriction fragment length polymorphism, isozyme and morphological. 47

Table 1-6: Summary of the most prevalent marker types used for cultivar identification in the Rosaceae. 49

Table 1-7: Summary of the most prevalent marker types used for genetic map construction or augmentation of existing maps in the Rosaceae. 50

Table 1-8: An example of cross-generic simple sequence repeat (SSR) primer pair amplification and detection of a polymorphic product during construction of an *Eriobotrya japonica* (loquat) linkage map. Primer pairs developed from *Malus* (apple), *Pyrus* (pear) and *Prunus* spp. were assessed. 64

Table 2-1: Description of apple segregating populations used for Rosaceae conserved orthologous set (RosCOS) genetic mapping. 93

Table 2-2: Rosaceae conserved orthologous set (RosCOS) marker distribution and density on *Fragaria vesca* × *F. bucharica* (FVxFB I-VII) linkage groups (LG) in centimorgans (cM). Length of FVxFB LG and coverage of RosCOS mapped in apple is included to give approximate equivalent to the genome coverage of apple. Bioinformatics was used to located RosCOS markers on the 'Golden Delicious' (GD) apple genome, and the DNA sequence information was used to design primer pairs suitable for use with high-resolution melting (HRM) analysis. 98

Table 2-3: Rosaceae conserved orthologous set (RosCOS) marker performance based on method of primer design. Categories were based on marker performance in one or more apple populations. 98

Table 2-4: Apple parents from five segregating populations needed to establish a minimum of two Rosaceae conserved orthologous set (RosCOS) markers per linkage group (LG). RosCOS markers were mapped on 15 of the total 17 apple LG. 100

Table 2-5: Rosaceae conserved orthologous set (RosCOS) marker coverage, in centimorgans (cM), of apple consensus linkage groups (LG). 100

Table 2-6: Distribution of RosCOS markers mapped on apple homeologous linkage groups (LG) and their *Fragaria vesca* × *F. bucharica* (FV×FB) orthologous LG. 103

Table 2-7: Distribution of 56 Rosaceae conserved orthologous set (RosCOS) markers common between FVxFB linkage groups (LG) and apple homeologs. FVxFB LG are listed on the y-axis, apple paralogs on listed on the x-axis. Each cell is populated with the number of orthologous RosCOS markers. Bold font indicates contiguous markers. 103

Table 3-1: Results and sources of primer sequences used for marker development and molecular mapping in *Rubus occidentalis* 96395S1 × *R. idaeus* 'Latham'. 145

Table 3-2: Linkage group descriptions for the parental maps of *Rubus occidentalis* 96395S1 and *R. idaeus* 'Latham'. 152

Table 3-3: Summary of the number and occurrence of markers in common between *Rubus occidentalis* 96395S1 × *R. idaeus* 'Latham' and *Fragaria*, *Malus*, and *Prunus* genomes as determined by BLAST analyses. 156

Table 3-4: Summary of transferability of markers derived from *Fragaria* DNA sequence to raspberry and blackberry. 162

Table 4-1: Derivation of markers mapped on *Rubus occidentalis* 96395S1 × *R. idaeus* 'Latham' that were designed from transcription factors and genes for enzymes involved in the polyphenolic biosynthetic pathway (Bushakra et al. 2012a). 192

Table 4-2: Derivation of markers mapped on *Rubus occidentalis* 96395S1 × *R. idaeus* 'Latham' that were designed from genes from pathways other than the polyphenolic pathway, and other sequences (Bushakra et al. 2012a)..193

Table 4-3: Anthocyanin compounds extracted from purple raspberry fruit collected from 155 F_1 progeny of *Rubus occidentalis* 96395S1 × *R. idaeus* 'Latham', analyzed with ultra high performance liquid chromatography (UHPLC) in 2009 and HPLC in 2010 and 2011, and used for quantitative trait locus (QTL) analysis. 199

Table 4-4: Mean proportions of anthocyanins detected in the F_1 progeny of *Rubus occidentalis* 96395S1 × *R. idaeus* 'Latham' in 2009, 2010, and 2011. 200

Table 4-5: Summary of growing conditions at Plant & Food Research, Motueka Research Centre for the *Rubus occidentalis* 96395S1 × *R. idaeus* 'Latham' parents and F_1 progeny. 201

Table 4-6: Mean phenotypic values and basic statistical parameters for the five traits analyzed in 155 F_1 progeny of *Rubus occidentalis* 96395S1 × *R. idaeus* 'Latham' in 2009 and 2011, and the progeny and parents in 2010...201

Table 4-7: Variance components for five analyzed anthocyanin compounds for three years (2009, 2010, 2011) detected in 155 F_1 progeny of *Rubus occidentalis* 96395S1 × *R. idaeus* 'Latham'. Values are based on the mean percentage of the total anthocyanin accumulation for each year. 203

Table 4-8: Pearson's correlation coefficient calculations and statistical significance (I-value) for each pair of the six anthocyanin compounds analyzed for 2009 and 2010 in 155 F_1 progeny of *Rubus occidentalis* 96395S1 × *R. idaeus* 'Latham'. 204

Table 4-9: Pearson's correlation coefficients calculated for each pair of compounds that segregated into two distinct groups in the F1 progeny of Rubus occidentalis 96395S1 × R. idaeus 'Latham'. The 61 individuals comprising Set A are heterozygous ("lm") at locus RubFruitE4 on 'Latham' *Rubus* linkage group (RLG) 2. 206

Table 4-10: Pearson's correlation coefficients calculated for each pair of compounds that segregated into two distinct groups in the F1 progeny of Rubus occidentalis 96395S1 × R. idaeus 'Latham'. The 94 individuals comprising Set B are homozygous ("ll") at locus RubFruitE4 on 'Latham' *Rubus* linkage group (RLG) 2. 206

Table 4-11: Comparison of Set A and Set B for the four compounds that segregated into two distinct groups in the F_1 progeny of *Rubus occidentalis* 96395S1 × *R. idaeus* 'Latham' during correlation analysis. 207

Table 4-12: Mapping and significance information for five anthocyanin compounds mapped in Rubus occidentalis 96395S1 and R. idaeus 'Latham', % variation explained per locus, and the nearest mapped marker. 209

Table 4-13: Proportions of the total anthocyanin for each compound detected in the analyzed set of F1 progeny of Rubus occidentalis 96395S1 × R. idaeus 'Latham' in 2009 and 2010. Proportions of anthocyanins from populations of black raspberry cultivars, red raspberry 'Latham' × 'Glen Moy', and purple raspberry are provided for comparison. 211

Table 5-1: Genome regions conserved in terms of marker order among *Rubus* (RLG1-RLG7), *Fragaria* (FLGI-FLGVII), *Prunus* (PG1-PG8) and *Malus* (MLG1-MLG17). Markers on RLG linkage groups are genetically mapped and distances spanned by the markers are in centimorgans (cM); markers on the other three genera are physically mapped and distances spanned by the markers are in megabases (Mb). Shaded rows indicate three regions conserved in all four genera...249

Table 5-2: Distribution of ancestral chromosomes (A1-A9) among the linkage groups (LG) of *Rubus*, *Fragaria*, *Malus*, and *Prunus*. 250

Supplemental Table S2-1: Rosaceae conserved orthologous set (RosCOS) marker design notes for mapped markers. Each *Fragaria vesca* × *F. bucharica* (FVxFB) linkage group (LG) is discussed along with information on the bin (LG:bin) in which the RosCOS marker is located, its orthologous 'Golden Delicious' (GD) apple LG, and additional comments. 121

Supplemental Table S2-2: Forward and reverse primer pair sequences for all mapped Rosaceae conserved orthologous set (RosCOS) markers. Primer names indicate the RosCOS sequence used for primer pair design, the linkage group for which the marker was designed, the exons in which the primers were placed or if the target was a SNP; a and b indicate that more than one primer pair was designed for that locus. 122

Supplemental Table S2-3: Twelve markers placed *in silico* on the ‘Golden Delicious’ (GD) apple whole genome sequence, but which mapped genetically to a different apple linkage group (LG), and the *Fragaria vesca* x *F. bucharica* (FVxFB) LG and bin (LG:bin) to which the marker is genetically mapped. The two best recombination frequency (Rf) and log of odds (LOD) scores and their associated frame-work markers show support for the genetic placement. Primer names indicate the RosCOS sequence used to design the primer pair, the LG for which the marker was designed, the exons in which the primers were placed or if the target was a SNP; a and b indicate that more than one primer pair was designed for that locus..125

Supplemental Table S2-4: Apple simple sequence repeat (SSR) markers compared with *Fragaria vesca* scaffold sequences. The best 13 of 103 markers [56, 57] are indicated by having the lowest E-values using BLASTN analysis. Forward (F) and reverse (R) apple primer sequence BLASTN results are given for each marker. In no case was the SSR motif present in the *Fragaria vesca* scaffold target sequence. 126

Supplemental Table S2-5: Strawberry gene-based markers genetically mapped *Fragaria vesca* × *F. bucharica* (FVxFB) compared with the ‘Golden Delicious’ apple whole genome sequence. 127

Supplemental Table S3-1: Redesigned and newly designed primer pair sequences of markers polymorphic in *Rubus occidentalis* 96395S1 x *R. idaeus* ‘Latham’........................180

Supplemental Table S3-2: Marker design information. Information on origin and type of sequence, locations in strawberry, apple and peach genomes, expected values (E-values), and percent identity for each marker mapped in *Rubus occidentalis* 96395S1 × *R. idaeus* ‘Latham’...182

Supplemental Table S 4-1: Spearman’s correlation coefficient calculations and statistical significance (p-value) for each pair of the six anthocyanin compounds analyzed for two consecutive years in 155 F1 progeny of Rubus occidentalis 96395S1 × R. idaeus ‘Latham’...242

LIST OF ABBREVIATIONS

A	Adenosine
a/a/p	*Amygdalus*/*Armeniaca*/*Prunocerasus*
ACY	Anthocyanins
AFLP	Amplified fragment length polymorphism
Amplicon	Amplified product of PCR
BLAST	Basic local alignment search tool
bp	Base pair
C	Cytosine
c/l/p	*Cerasus*/*Laurocerasus*/*Padua*
C3G	Cyanindin 3-*O*-glucoside
C3GR	Cyanindin 3-*O*-2^{G}-glucosylrutinoside
C3R	Cyanindin 3-*O*-rutinoside
C3S	Cyanindin 3-*O*-sophoroside
C3Sb	Cyanindin 3-*O*-sambubioside
C3XR	Cyanindin 3-*O*-2^{G}-xylosylrutinoside
cDNA	Complementary DNA
cM	Centimorgans
Contig	Contiguous sequence
COS	Conserved orthologous set
DNA	Deoxyribonucleic acid
dsDNA	Double stranded DNA
ESI	Electrospray interface
EST	Expressed sequence tag
FISH	Fluorescence *in situ* hybridization
FV×FB	*Fragaria vesca* × *F. bucharica*
FW	Fresh weight
g	Gram
G	Guanine
GDD	Growing degree days
gDNA	Genomic DNA
gSSR	Genomic SSR
HG	Homeologous group
HPLC	High performance liquid chromotography
HRM	High-resolution melting
IM	Interval mapping
indel	Insertion or deletion
kb	Kilobase
K-S DMax	Kolmogorov-Smirnov Dmax test
LG	Linkage group
LOD	Logarithm of odds
M.9×R5	‘Malling 9’ × Robusta 5

MAB	Marker assisted breeding
Mb	Megabase
μg	Microgram
mL	Milliliter
MQM	Multiple-QTL model
mRNA	Messenger RNA
MS	Mass spectrometry
my	Million years
nm	Nanometers
P3R	Pelargonidin 3-*O*-rutinoside
PCR	Polymerase chain reaction
pg	Picogram
QTL	Quantitative trait locus
R gene	Resistance gene
RAPD	Random amplification of polymorphic DNA
Rf	Recombination frequency
RFLP	Restriction fragment length polymorphism
RLG	*Rubus* linkage group
SCAR	Sequence characterized amplified region
SD	Standard deviation
s.e.	Standard error
SEM	Standard error of the mean
SNP	Single nucleotide polymorphism
sp.	Species
SSR	Simple sequence repeat
STS	Sequence tagged site
T	Thymine
T×E	‘Texas’ × ‘Earlygold’
TS	Transgressive segregant
UHPLC	Ultra high performance liquid chromotography
unigene	Unique gene
x	Base chromosome number

LIST OF COMMONLY REFERRED TO ROSACEAE SPECIES

Scientific name	Chromosome number	Common name
Cydonia oblonga	2*n*=34	quince
Eriobotyra japonica	2*n*=34	loquat
Fragaria bucharica	2*n*=14	
Fragaria vesca	2*n*=14	alpine strawberry
Fragaria × ananassa	8*n*=56	cultivated strawberry
Malus × domestica	2*n*=34	apple
Prunus armeniaca	2*n*=16	apricot
Prunus avium	2*n*=16	sweet cherry
Prunus ceracus	2*n*=16	sour cherry
Prunus domestica	2*n*=16	European plum
Prunus dulcis	2*n*=16	almond
Prunus persica	2*n*=16	peach
Pyrus communis	2*n*=34	European pear
Pyrus pyrifolia	2*n*=34	Asian pear
Rosa sp.	2*n*=14	rose
Rubus idaeus	2*n*=14	red raspberry
Rubus occidentalis	2*n*=14	black raspberry
Rubus sp.	4*n*=28	blackberry/hybrid berry