Characterization of Chalcone Synthase Genes in Red Raspberry Cultivar (*Rubus idaeus*)

IM-ERB PUNSOD

To my family and my sponsor for their supports and encouragements
and special to my two supervisors who can not stand with me at the end

Declaration

This thesis is the result of the author's original research. It has been composed by the author and has not been previously submitted for examination, which has led to the award of the degree.

Im-erb Punsod

Acknowledgements

I would firstly like to sincerely thank Dr. Elizabeth M. Ellis, my forth supervisor, who helped me to write my thesis and who provided excellent supervision and support during the most difficult time of my PhD study. Especially, I also offer very grateful thanks to Dr. Alistair Paterson and Dr. Eve M. Lutz, my first and third supervisors, for all kindness and everything given to me although they could not be here until the end. I am gratefully thankful to Dr. Julie Graham and Dr. Susan McCallum, at John Hutton institute, for kindly help and suggestions that were very useful for my work and which make me happily walk in the way of red raspberry quality improvements. Importantly, I would like to thank Dr. Rothwell Tate for his kindly help with supporting in DNA sequencing that was the main part of my thesis.

I also would like to thank to my dear friends in my Labs, specially to Walid who have faced the saddest time with me, and also to my other lovely friends including Asia, Dan and Alison who spent the last days in the office with me, as well as staffs on the fifth floor, technicians, officials, IT staff involved in my work and also cleaners who always made my days starting in early morning with friendly talking, even though the name of all people can not be written here, they are a part of this thesis as well.

Finally, I would like to thank all my dear Thai PhD students since the first day when I arrived until I left Glasgow, for giving help for my study and also sharing the days of happy times. A special thank to my workplace, NSRU, in Thailand for giving me the great chance and supporting the scholarship in studying abroad. Finally thanks to my family and my little sweetheart, Rugjung, without any worth word and I would like to say to myself with a half smile “ Finally my another long journey had been finished ”.

Contents

Table of Contents ..... *d*

List of Figures ..... *j*

List of Tables ..... *n*

List of Abbreviation ..... *o*

Summary ..... *s*

Chapter 1 General Introduction **1**

1.1 **Introduction of the Thesis** **2**

1.2 **Red raspberry *RUBUS IDAEUS*** **2**

1.2.1 Qualitative Traits **3**

1.2.2 Nutrient Content and Health Benefits **3**

1.2.3 Cultivars and Developments **4**

1.2.3.1 The European Red raspberry **6**

1.2.3.2 The North America Red raspberry **6**

1.2.4 Improvements in Quality Traits **9**

1.3 **Utilization of Genetic information in Plant Improvement** **10**

1.3.1 Marker-assisted Breeding **12**

1.3.2 Qualitative Trait Loci Approach **13**

1.3.3 Alleles and Mining Approach **14**

1.4 **Anthocyanins and their Biosynthesises** **15**

1.4.1 Fundamental of Biosynthesis **15**

1.4.2 Polyketide Syntheses **19**

1.4.3 Polyketide Synthase Gene Identification **21**

1.5 **The Scope and Aims of Thesis** **23**

Chapter 2 General Materials and Methodologies **24**

2.1 **Materials** **25**

2.1.1 BAC DNA of *R. iedea* cv. Glen Moy **25**

2.1.2 Plant Genomic DNA of *Rubus iedea* cv. Latham **25**

2.2 **Methodology** **26**

2.2.1 **Analysis of Bacterial Artificial Chromosome (BACs) (Chapter 3)** **26**

2.2.1.1 BAC DNA Isolation ……..... **26**
2.2.1.2 BAC DNA Restriction Digestion ……..... **27**
2.2.1.3 BAC PCR-based Cloning and Sequencing …… **28**
2.2.1.3.1 Polymerase Chain Reaction …….. **28**
2.2.1.3.2 Electrophoresis and visualization ……..… **29**
2.2.1.3.3 Gel extraction ……...……...……. **29**
2.2.1.3.4 Cloning reaction ……..…… **29**
2.2.1.3.5 Transformation …….…… **30**
2.2.1.3.6 DNA extraction ……..... **30**
2.2.1.3.7 Transformant analysis …….. **31**
2.2.1.4 Southern Blot Analysis …….…….... **32**
2.2.1.4.1 Restriction digestions ……..……. **32**
2.2.1.4.2 DNA transferring ……...……..……. **32**
2.2.1.4.3 Labelled probe preparing …….……..… **34**
2.2.1.4.4 Hybridization ……...……..… **35**
2.2.1.4.5 Immunological detection ……..…… **35**
2.2.2 **Fosmid library Construction (Chapter 4)** ……..... **36**
2.2.2.1 Fosmid Library Construction …….…….. **36**
2.2.2.1.1 Preparation of BAC DNA …..…….……. **37**
2.2.2.1.2 Recovery of the Size-Fractionated DNA …….…… **37**
2.2.2.1.3 DNA Precipitation …….……...… **37**
2.2.2.1.4 End-Repair Reaction ……... **37**
2.2.2.1.5 Ligation of Sized DNA Fragments ……..…… **38**
2.2.2.1.6 Fosmid Library Propagation …… **38**
2.2.2.1.7 Fosmid DNA Purification …… **39**
2.2.2.2 Fosmid Stability Assays ……..…… **40**
2.2.2.3 Fosmid Library Screening ……...…… **41**
2.2.2.3.1 Colony Hybridization Screening ……..... **41**
2.2.2.3.2 Polymerase Chain Reaction ……...…… **41**
2.2.2.3.3 Southern Blot Anaysis …….…..… **41**
2.2.2.4 Fosmid Ends Sequence Analysis and Annotation …….. **42**

2.2.3 **Assembly of Bacterial Artificial Chromosome (BAC) sequences and Construction of Physical Maps of PKS Gene Loci of Red Raspberry cv. Glen Moy (Chapter 5)** …….. **42**
2.2.3.1 Assembly of BAC sequences of Red Raspberry cv. Glen Moy **42**
2.2.3.2 Construction of Initial Physical Maps of PKS Gene Loci …… **43**
2.2.4 **Identification and Analysis of Open Reading Frames Polyketide Synthases Genes (Chapter 6)** …….…… **43**

Chapter 3 Characterization of Polyketide Synthase Genes of Red Raspberry cv. Glen Moy (BAC analysis) ……..…….. **45**
3.1 **Introduction** ……..…… **46**
3.1.1 Polyketide Synthas Genes in Raspberry and Related Fruits …… **46**
3.1.2 Bacterial Artificial Chromosomes ……..…….…… **47**
3.1.3 The Aims of the Chapter ……..… **47**
3.2 **Results** ……...…….. **47**
3.2.1 BAC DNA Restriction Digestion ……..…….…. **48**
3.2.2 Amplification of the PKS genes ……...… **49**
3.2.3 Characterization of the PKS genes ……... **52**
3.2.4 Detection of PKS genes by southern blotting analysis ……...…….. **57**
3.3 **Discussion** ……...……... **58**
3.4 **Conclusion** ……...…….. **59**

Chapter 4 Construction of a Fosmid sub-Library of BAC-derived Red raspberry *R. iedea* cv. Glen Moy DNA ……... **60**
4.1 **Introduction** ……..…….…….…….. **61**
4.1.1 Genomic DNA Libraries ……...……... **61**
4.1.2 Fundamentals of DNA Library Construction ……... **62**
4.1.2.1 DNA Fragment Preparation …….. **62**
4.1.2.2 Calculating the Number of Clones ……..……. **62**
4.1.2.3 Vectors for Library Construction ……. **63**
4.1.3 Use of Fosmid Libraries ……... **64**
4.1.4 The Aims of the Chapter …….. **65**

4.2 **Results** …… **65**
4.2.1 Fosmid Library Construction and Characterization ……. **65**
4.2.2 Fosmid Stability …… **67**
4.2.3 Fosmid Library Screening ……..…….…. **68**
4.2.3.1 Colony Hybridization Screening ……..……..….. **68**
4.2.3.2 Fosmid Library Screening by Polymerase Chain Reaction …. **71**
4.2.3.3 Fosmid Library Screening by Southern Blotting ..…… **72**
4.2.4 Fosmid End Sequencing and Sequence Analysis ……..……... **73**
4.3 **Discussion** ……..…… **74**
4.4 **Conclusion** ……...…….. **75**

Chapter 5 Assembly of Bacterial Artificial Chromosome (BAC) sequences and Construction of Initial Physical Maps of PKS Gene Loci of Red Raspberry cv. Glen Moy ……. **76**
5.1 **Introduction** ……..…….……..….. **77**
5.1.1 Concept of DNA Sequencing ……... **77**
5.1.1.1 Sanger Sequencing ……..……...…….. **78**
5.1.1.2 Second Generation Sequencing ……..….. **78**
5.1.1.3 Next Generation Sequencing ……..…….…. **79**
5.1.2 Sequencing and Sequence Assembly ……..…. **79**
5.1.3 Physical Map construction ……...…...…...…….. **81**
5.1.3.1 Construction of Physical Maps ……. **82**
5.1.3.2 Minimal Tilling Path and Validation …… **82**
5.1.4 Physical Maps of Raspberry and Related Fruits ……... **83**
5.1.5 The Aim of the Chapter ……...…..... **84**
5.2 **Results** ……...……..…... **84**
5.2.1 BAC 454 sequencing efficiency and Assembly Characterization …… **84**
5.2.2 Combination of BAC Paired-end Sequences ……...…… **88**
5.2.3 Genome Annotation of Combined 454 Sequences ……..... **91**
5.2.4 Combination of Fosmid Ended-pair Sequences ……...……. **91**
5.2.5 Prediction of Ordering Genes within the BAC sequences …… **94**
5.2.6 Comparison the Ordering Genes of Raspberry to Related fruits ……... **99**

5.2.7 Construction of Mini physical Maps ……. **100**
5.2.7.1 Physical Map derived from BAC24P12 **100**
5.2.7.2 Physical Maps derived from BAC29M05 …… **104**
5.2.7.3 Difference of the Mini Physical Maps ……. **106**
5.3 **Discussion** ……..... **108**
5.4 **Conclusion** ……..……..… **112**

Chapter 6 Identification and Analysis of Open Reading Frames from Red Raspberry cv. Glen Moy : Polyketide Synthase Genes ….……. **113**
6.1 **Introduction** ……..…….……... **114**
6.1.1 General Gene Structure ……...……...……... **114**
6.1.1.1 PKS Gene Structure ……..……...….. **115**
6.1.2 Genetic Conservation and Variation ……. **116**
6.1.3 The Aim of the Chapter ……...……. **117**
6.2 **Results** ……...……..….... **117**
6.2.1 Investigation of Open Reading Frames ……..….. **118**
6.2.1.1 PKS Gene Identification ……....…..……. **118**
6.2.1.2 Comparison of Protein Sequences ..…….……..….. **120**
6.2.1.3 Comparison of four PKS Proteins from cv. Glen Moy …… **121**
6.2.1.4 Detailed Comparison of Predicted PKS Proteins ……. **123**
6.2.1.5 Comparison of PKS Protein Sequence of cv. Glen Moy to Related Fruits ……....…… **128**
6.2.2 Analysis of PKS Nucleotide Sequences ……..…. **130**
6.2.2.1 Variation in Intronic Sequences …….…….. **130**
6.2.2.2 Regulatory Regions and Variations …… **130**
6.2.3 Variation in Other Genes…….…….. **131**
6.2.3.1 Identification of the Other Genes ……. **131**
6.2.3.1.1 Cytochrome P450 …….…… **131**
6.2.3.1.2 ABC transporter ……...…… **133**
6.2.3.1.3 Cycloartenal synthases …….…….... **137**
6.2.3.2 PCR Amplification of the Other Genes …….…….. **138**
6.3 **Discussion** ……...…….…….. **139**

6.4 **Conclusion** ……...…….. **142**

Chapter 7 Discussion and Conclusion ……..... **143**

7.1 **General diacussion** ……..……...…….. **144**

7.1.1 Details of BACs ……..…… **144**

7.1.2 Fosmid Sub-library Construction and Utilization ……. **145**

7.1.3 Completion of the Entire BAC Sequences …… **146**

7.1.4 Characterization of PKS/CHS genes in Raspberry …….. **147**

7.1.5 Initially Partial Physical Maps …….. **148**

7.2 **Overall Conclusions** ……...…….. **149**

7.3 **Future Works** ……..……..…..…..……..…… **149**

References **151**

Appendices **174**

List of Communications **202**

List of Figures

Chapter 1

Figure 1.1 Schema of an integrated view of genomic resources and genomics research in crop improvement using genetic and genomic approaches ……... **11**

Figure 1.2 An overview of the pathway and branch-points leading to the biosynthesis of the various flavoniods; anthocyanins and related compounds …….. **17**

Figure 1.3 Chemically basic structure of anthocyanin compounds ……..... **18**

Figure 1.4 Diagram of reactions of certain polyketide synthase (CHS and non CHS) and their main products ……..... **20**

Chapter 3

Figure 3.1 Fragment patterns of restriction digestion of the Glen Moy BAC clones (24P12; 29M05 and 31B12) ……..... **48**

Figure 3.2 Prediction of PCR products amplified using PKS2 gene of red raspberry cv. Royalty (PKS2Ri, AF292368.1) and PKS2 primers (prPKS2f and prPKS2r) ……...…….…... **51**

Figure 3.3 The amplified nucleotide sequences of BAC clones using PKS primers (PKS1; PKS2 and PKS3) ……... **52**

Figure 3.4 Comparison of raw nucleotide sequences of 1B185 subclone (5B185) amplified from BAC29M05 using PKS1 primer compared with PKS5 from GenBank (EF694718.1) …….. **56**

Figure 3.5 Southern blotting analysis of three BAC clones (24P12; 29M05 and 31B12) using DIG-labelled PKS1 probe ……. **57**

Chapter 4

Figure 4.1 Size of sheared DNA fragments from the three BACs to construct the randomly sheared fosmid libraries …… **66**

Figure 4.2 Stability assays on serial cultures for 1 and 100 propagations of fosmid clones constructed from the BACs …….. **67**

Figure 4.3 Colony hybridization screening of fosmid sublibraries constructed separately from the three BACs (FrB24; FrB29 and FrB31) …..… **70**

Figure 4.4 PCR screening of three fosmid sublibraries by PKS (PKS1; PKS2 and PKS3) primers ……..… **71**

Figure 4.5 A Southern blotting analysis of the three fosmid sublibraries using DIG-labelled PKS1 probe …….. **72**

Chapter 5
Figure 5.1 Schema of an assembly concept ……. **80**

Figure 5.2 An initial analysis of number of contigs and bases of the resulting 454 sequences of the three BACs (31B12, 29M05 and 24P12) from red raspberry cv. Glen Moy …… **86**

Figure 5.3 Characterizations of assemblies of resulting 44 sequences of the three BACs (31B12, 29M05 and 24P12) from red raspberry cv. Glen Moy …… **87**

Figure 5.4 A An association of ended-pair sequences (by Sanger sequencing) to resulting 454 sequences of BAC24P12 ……...…… **89**

Figure 5.4 B An association of ended-pair sequences (by Sanger sequencing) to resulting 454 sequences of BAC29M05 ……..……... **90**

Figure 5.5 Combination of paired-end sequences of Fosmid clones to resulting 454 sequences of BAC24P12 and BAC29M05 ……... **93**

Figure 5.6 A Schema of the order of genes on BAC29M05 and BAC24P12 derived from raspberry genome cv. Glen Moy ……... **96**

Figure 5.6 B Comparison of gene regions within an entire length of the maps derived from BAC24P12 (117 kb) and 29M05 (~170 kb) …….. **97**

Figure 5.7 Schema of an ordering fosmid clone set of the MTP of BAC24P12 …….. **100**

Figure 5.8 The schema of a mini constructed physical map of BAC24P12 with approximately 117 kb in length …….. **103**

Figure 5.9 The ordering of fosmid clones of the MTP of the assembled of BAC29M05 ……...…….. **104**

Figure 5.10 A mini constructed physical map for 98 kb of BAC29M05 ……... **107**

Chapter 6
Figure 6.1 Diagram of a general gene structure in eukaryotes ……. **115**

Figure 6.2 Schema of the four PKS gene regions on BAC24P12 and BAC29M05 …… **119**

Figure 6.3 Comparison of amino acid sequence of ORFs of PKS regions (RiPKSr1-4) on the two BACs (29M05 and 24P12) …….. **121**

Figure 6.4 Comparison of amino acid sequences in the ORFs of the two regions at RiPKSr1 (BAC24P12) and RiPKSr3 (BAC29M05) derived from raspberry cv. Glen Moy ……. **124**

Figure 6.5 Comparison of amino acid sequences in the ORFs of the two regions at RiPKSr1p (BAC24P12) and RiPKSr3p (BAC29M05) derived from raspberry cv. Glen Moy to **A**, PKS1, and **B**, PKS5 proteins of cv. Royalty and **C**, CHS5 proteins of cv. Meeker …… **125**

Figure 6.6 A Comparison at amino acid sequence level of RiPKSr3p (BAC29M05) to PKS1 of cv. Royalty and CHS6 of cv. Meeker using Sequencer CLC (viewer 6) …… **126**

Figure 6.6 B Comparison at amino acid sequence level of RiPKSr1p (BAC24P12) to PKS5 and CHS5 of cv. Royalty and CHS11 of cv. Meeker using Sequencer CLC (viewer 6) ……. **127**

Figure 6.6 C Comparison of amino acid sequences of RiPKSr1p and RiPKSr4p to PKS5 of cv. Glen Moy (RgPKS5) submitted in GenBank databases, PKS5 of cv. Royalty (RrPKS5) and CHS5 of cv. Meeker (RmCHS5) using Sequence viewer CLC software …… **128**

Figure 6.7 Comparison of Glen Moy PKS amino acid residues to two related fruits in the same family (Strawberry and prune) ……..... **129**

Figure 6.8 Comparison of amino acid sequences of the two regions, RiCYP450r1p (on BAC24P12) and RiCYP4502rp (on BAC29M05) respectively, and *Fragaria vesca* subsp. *Vesca* gene **132**

Figure 6.9 An alignment of amino acids of ORFs of ABC gene regions (RiABCr1p, on BAC12P12 and RiABCr2p on BAC29M05) …. **134**

Figure 6.10 Alignment of ORF of RiABCr1p (the 626 amino acids, on BAC24P12) to plants; hypothetical protein PRUPE sll0005-like of *Prunus persica*; uncharacterized protein sll0005-like of *Fragaria vesca* subsp. *Vesca* gene; protein ABC1 putative of *Ricinus communis* and ABC of Arabidopsis …… **135**

Figure 6.11 Alignment of ORF of RiABCr2p (the 846 amino acids, on BAC29M05) to plants; hypothetical protein PRUPE sll0005-like of *Prunus persica*; uncharacterized protein sll0005-like of *Fragaria vesca* subsp. *Vesca* gene; protein ABC1 putative of *Ricinus communis* and ABC of Arabidopsis ……..……...… **136**

Figure 6.12 An alignment of amino acid sequences of Cycloartenal synthases gene of RiCASr1 (BAC24P24) and RiCASr2 (BAC29M05) ……. **137**

Figure 6.13 Amplification of genomic DNA of red raspberry cv. Glen Moy and Latham using CYP450; ABC transporter and CAS primers … **138**

List of Tables

Table 1.1	Related raspberry species within the subgenus *Idaeobatus* ……..	**5**
Table 1.2	Cultivars and growing season of the European Red raspberries …….	**7**
Table 1.3	Polyketide Synthase Gene Identification …….	**22**
Table 2.1	Primers used in PCR amplifications of PKS genes …….	**28**
Table 2.2	Labelled DNA probes for Southern blotting and colony hybridization	**34**
Table 2.3 A	Functions of CYP450; ABC transporter and CAS genes ……	**44**
Table 2.3 B	Primers used for PCR amplifications of CYP450, ABC transporter and CAS genes …….	**44**
Table 3.1	Prediction of PCR amplifications using PKS genes ……	**50**
Table 3.2	Identity of the nucleotide sequences of the PKS-amplified fragments of subclones derived from BACs using BLASTn ……...	**54**
Table 3.3	Identity the deduced amino acid sequences of the PKS-amplified fragments of subclones derived from BACs using BLASTp ……..	**55**
Table 4.1	Vector systems and capacity of insert-size DNA for library construction ……..	**64**
Table 4.2	Number of positive fosmid clones screened by colony hybridization	**69**
Table 5.1	Gene sequences and coding proteins found in BAC29M05 and 24P12 ……...	**98**
Table 6.1	Software used for analysis PKS genes in BACs ……..	**117**
Table 6.2	Summary of PKS regions and their of encoding proteins on the two BACs ……	**120**
Table 6.3	Pairwise comparisons of all four PKS proteins on the two BACs …..	**122**

List of Abbreviations

Amino acid

Name	Code	Symbol
Alanine	Ala	A
Arginine	Arg	R
Aspartate	Asp	D
Asparagine	Asn	N
Cysteine	Cys	C
Glutamate	Glu	E
Glutamine	Gln	Q
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tryptophan	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V

DNA bases

A	Adenine
C	Cytosine
G	Guanine
T	Thymine

Standard units

^{o}C	Degree Celsius
g	gram
l	Litre
M	Molar
V	Volt
bp	base pairs
kb	kilo bases
Mb	Mega base pairs
ml	Millilitres
mM	Millmolar
µg	Micrograms
µl	Microlitres

Textual abbreviations and chemical formulae

Ab-QTL	Advanced backcross QTL
ACC	acetyl CoA carbokylase);
AcOH	Acetic acid
AFLP	Amplified fragment length polymorphism
amp	Ampicilin
ANS	Anthocyanins synthase
BAC	Bacterial Artificail chromosome
BAS	benzalacetone synthase
BBS	bibenzyl synthase
bHLH	basic helix loop helix
BLAST	Blast Local Alignment Search Tool
BLASTP	search protein database using a protein query
BLASTX	search protein database using a translated nucleotide query
BPS	benzophenone synthase
BUS	isobutyrophenone synthase
C3H	4-coumarate 3-hydroxylase
C4H	cinnamate 4-hydroxylase
CAS	cycloartenol synthase
CCR	cinnamyl-CoA reductase
cDNA	complementary DNA
CesA	Cellulose synthase
CHI	chalcone isomer
CHS	chalcone synthase
cM	centimorgan
contig	contiguous sequence
COS	conserved orthologous set
CTA	coumaroyl triacetic acid
CTAB	Hexaadecyltrimethylammonuim bromide
CTAS	Coumarytriacetric acid synthase
cv.	Cultivar
DFR	dihydroflavo 4-reductase
DIG	Digoxigenin
DNA	deoxyribonucleic acid
dNTP	deoxy-nucletide-tri phohate
DTT	Dithiothreiol
EDTA	ethylenediaminetetraacetic acid
eQTL	Expression QTL
EST	Expressed sequence tag
EST-SSR	Expressed sequence tag-derived SSR
EXP	expasin
F3H	flavanone 3-hydroxylase
FLS	flavonal synthase
HAC	Human Artificial Chromosomes
HCl	hydrochloric acid
IAA	Isoamyl alcohol
LD	linkage microsetellite

ILs	introgreesion lines
IPTG	*iso*propyl thiogalactoside
ISSR	Intersimple Sequence Repeat
kb	kilobases pairs
LAR	leucoaanthocyanin reductase
LB	Luria Bertani
LG	linkage group
LUP	Lupeol synthase
MAB	marker-assisted breeding
MAS	marker-assisted selection
$MgCl_2$	magnesium chloride
MYB	myeloblastosis
NaCl	sodium chloride
NaOH	sodium hydroxide
NCBI	National Center for Biotechnology Information
NH_4	Ammonium
ORFs	Open Reading Frame
P1AC	P1 Artificial Chromosome
PAL	pheammonia-lase
PCR	polymerase chain reaction
pH	Potential of Hydrogen
PKS	polyketide synthase
PVD	polyvinylpyrrolidone
QTL	quantitative trait loci
RAPD	randomly amplified polymorphic DNA
RFLP	restriction fragment length polymorphism
RNA	ribonucleic acid
RNAse	ribonuclease
SCRI	Scottish Crops Research Institute, Invergowrie
SNP	single nucleotide polymorphism
SRAP	Sequence-related Amplified Polymorphism
SSRs	simple sequence repeats
ST	stilbene
STRs	Short Tandem Repeats
STS	stilbene synthase
T_a	annealing temperature
TAC	Transformation-competent Artificial Chromosome
TAE	Tris-acetate-EDTA
TBE	Tris-borate-EDTA
TE	Tris-EDTA buffer
TILLING	targeted induced local lesion in genome
T_m	melting temperature
Tris HCl	tris-hydrochloric acid
TSS	Total soluble solid
TTG1	Transparent testa glabra 1
UFGT	UDPG-flavonoid-3-O-glucosyltrasferase
UV	ultra violet light
VPS	phlorisovalerophenone synthase

v/v	volume/volume
w/v	weight/volume
X-gal	bromo-chloro-indolyl-galactpyranoside
YAC	Yeast Artificial chromosomes

Summary

The biosynthesis of anthocyanin compounds involves type III polyketide synthases (PKSs): chalcone synthase (CHS), which is the first enzyme of the flavonoid synthetic pathways leading to synthesis of various anthocyanins. The European red raspberries are a commercially important fruit that needed both for fresh fruit consuming and food manufacturing. The fruits are a valuable source of nutraceuticals, notably anti-oxidants by anthocyanins that are pigments conferring impressive red colour to the fruits. To improve the quality in red raspberry through increasing anthocyanin content, the main objective was focused on investigation of the PKS genes that are an essential for the biosynthesis of anthocyanins. The aim of this thesis was to characterize the PKS/CHS gene regions within the three Bacterial Artificial Chromosome clones (BACs), which originated from the genome of the European red raspberries (*Rubus idaeus*) cv. Glen Moy (SCRI), using 454 sequencing technology assisted by a systematic construction of fosmid libraries. The two BACs (29M05 and 24B12) always presented similarities in almost analyses resulting from the presence of a 53 kb overlap that was revealed by assemblies of resulting 454 sequencing as it may be possible that there is no PKS/CHS gene in another one, BAC31B12. The assembly of BAC24P12 was completed, with an entire DNA length of 117 kb but BAC29M05 could not be completed along all its length of ~170 kb as the longest contig was 96 Kb provided the most important information of this BAC whilst containing a complete PKS1 gene sequence, currently identified as Naringenin chalcone-PKS. The alignments of assembled sequences also strengthened the similarities between these two BACs both at nucleotide and amino acid sequences. All four PKS/CHS gene regions found in the both BACs presented being multiple PKS/CHS gene family with conservation of an intron and two exons of gene structure as variations of those sequence confirmed being PKS/CHS pseudogenes in plant genome. The two informative sources generated in this thesis, including the fosmid sub-libraries, which are availably manageable sources for validating the BAC sequences, and physical map drafts provided better information of this chromosomal segment in this thesis, as they will be useful available tools for further studying in red raspberry genome.

CHAPTER 1

General Introduction

1.1 INTRODUCTION OF THE THESIS

This chapter provides a general background to the research that has been done in this thesis. As the thesis is based on a plant genomic study to search for and characterise the genes that are involved in red raspberry quality improvement, the introduction is divided into five main sections: Firstly, an introduction to the red raspberry and a brief explanation of cultivars and improvements is reviewed in section1.2; a background of approaches used in plant genomic studies and plant breeding development, particular to search for and analyse the genes influencing traits, anthocyanin content, is reviewed in section 1.3; an introduction to human health benefit and biosynthetic pathway of various flavonoid compounds as well as anthocyanins and polyketide synthases (PKS), key enzymes in anthocyanin biosynthesis, is given in section 1.4; finally, a brief concept and the aims of this thesis are reviewed in section 1.5.

1.2 RED RASPBERRY *RUBUS IDAEUS*

The Eurasian or red raspberry is commercially and widely grown in temperate regions, in at least 37 countries worldwide. While having a delicate skin, red raspberry has a powerful flavour and is colourful (red; black; purple or golden). The fruit is needed for both the fresh fruit market and food industries and can be frozen for some later uses. Red raspberry is produced for both fresh and food processing in the main regions; Russia; Europe (mostly in Poland; Hungary; Serbia; Germany; Italy and the United Kingdom) and the Pacific Coast of North America (British Columbia; Washington and Oregon). In eastern North America and many other countries such as New Zealand and Australia, it is produced only for fresh market (www.fruit.conell.edu). Raspberry is a very good source of nutrients and beneficial non-nutrients (Szajdek and Borowska, 2008 and Wolfgang and Jean, 2008), in particular polyphenolic compounds serve as phytochemical compounds that are considered good for human health (Weber and Lui, 2002).

1.2.1 Qualitative Traits

There are several physical qualities that are important in breeding raspberries. The first visual quality trait is the size of fruit, influencing acceptability and marketability of fresh fruit as large size is highly desirable. Physically fruit size depends on a combination of the number and size of drupelets (Jennings, 1988). Firmness is another key property of raspberries, indicating skin strength that relates to respiration rates and weight loss during storage and influences how long the fruits can maintain their shape and shelf life (Robbins and Sjulin, 1989). The colour is another physical property that is important in red raspberries resulting from carotenoids and flavonoids, in particular anthocyanins (Cseke *et al*., 2006). The color of carotenoids is masked by higher content of anthocyanins resulting in an impressive red color (Marinova *et al*., 2005). A darker color is preferred for fresh fruit markets and for the food industries, but heat or chemicals used in food manufacturing can easily alter the color (Wrolstad, 1976).

Other physical traits include taste and aromas leading to good flavour. As the precise chemical constituents directly influence these traits, an optimal balance between sugar and a principal acid in fruits makes their specific taste. Red raspberry contains a total sugar content of 5.34% w/w; ~10% total soluble solid content (TSS, °Brix); 213 volatile aroma,. compounds, and citric is the major acid with 2% w/w (Jennings, 1988; Purgar *et al*., 2012). However, only 9 from 213 volatiles contribute to flavour in red raspberries (Hadi, 2013). Wrolstad (1970) reported that flavour in cultivated raspberry is softer than in wild cultivars, leading to less pleasant fresh fruits, while heating or freezing could further decrease the flavour of fruits. Finally the presence of diseases and pests is the most important quality for consumer safety. Diseases caused by fungi; bacteria; viruses; pests; insects and nematodes all found in raspberry are reviewed in Jennings (1988) but they must not be found in fruits for consumer acceptability and safety reasons.

1.2.2 Nutrient Content and Health Benefits

Red raspberries are an excellent sources of dietary fibre, the highest contents up to 20% fibre per total weight (Cho *et al*., 2004), providing the fewest calories compared

to other fruits, with only 7.69 calories/gram of dietary fibre. They contain small content of vitamins B1; B2; B3; E; but significant amount of vitamin C with ~20 miligrams/100 grams. In addition, red raspberries have especially high level of ellagic acid that links to human health in cancer prevention (Dai *et al*., 2010). The fruits also have some minerals such as phosphorus; sodium; manganese; sodium; calcium; sulphur; chlorine; boron and high content of potassium at 130-221 miligrams/100 grams (Szajdek and Borowska, 2008; www.fruit.conell.edu).

Unlike pale-colored fruits and yellow raspberries, red raspberry contains significant amounts of anthocyanins which act as naturally occurring anti-oxidants, providing an ORAC value (oxygen radical absorbance capacity) of approximately 4900 per 100 gram that places red raspberry near the top of all fruits for antioxidant strength (Wang and Lin, 2000). Red raspberry has also carotenoids that are sources for vitamin A, with 40.6 IU per 100 gram, enhancing the immune system functions (Farges *et al*., 2012 and Chew and Park *et al*., 2004)) and sunburn protection (Korać and Khambholja, 2011). Handelman (2001) reported that although red raspberries are not high source of carotenoids their presence could contribute in the same way as anthocyanins.

1.2.3 Cultivars and Developments

Red raspberry belongs to the family *Rosacea* and subsequently subordinate to the genus *Rubus* being the most diverse in the plant kingdom. The genus *Rubus* contains over 700 species within 12 subgenera consisting of commercial fruits such as apple; peach; cherry; apricot; strawberry; blackberry, including arctic fruits and ornamentals. Whilst being classified into several subgenera, the *Eubatus* (blackberries) and *Idaeobatus* (raspberries) are the two significant subgenera of commercial fruits. As the subgenera *Eubatus* contains with very large number of species, *Idaeobatus* is composed of more than 200 species distributed in North and Eastern America, Asia, South Africa and Europe (Table 1.1). Numerous species of closely related fruits in the subgenus, *Idaeobatus* are called raspberries. The European red raspberries (*R. idaeus*); the North American red raspberry (*R. strigosus*) and the eastern American black raspberry (*R. occidentalis*) are the most significant raspberries for cultivation in these

main areas. In addition *R. idaeus* is related with several other species in different subgenera for instance *R. arcticus* (arctic raspberry); *R. nivalis* (snow raspberry); *R. odoratus* (flowering raspberry) and *R. sieboldii* (molucca raspberry) (Jennings, 1988). Many commercial cultivars are hybrids derived from the European and North American red raspberry as more than 20 cultivars of red raspberry were grown in both England and the USA such as Royalty; Malahat and Meeker (www.fruit.conell.edu).

Table 1.1 Related raspberry species within the subgenus *Idaeobatus.*

Species	Fruits
Rubus crataegifolius	Korean raspberry
Rubus gunnianus	Tasmanian alpine raspberry
Rubus idaeus	European red raspberry
Rubus leucodermis	Whitebark or Western raspberry, Blue raspberry, Black raspberry
Rubus occidentalis	Black raspberry
Rubus parvifolius	Australian native raspberry
Rubus phoenicolasius	Wine raspberry or Wineberry
Rubus rosifolius	West Indian raspberry
Rubus strigosus	American red raspberry
Rubus ellipticus	Yellow Himalayan Raspberry
Rubus arcticus	Arctic raspberry, subgenus Cyclactis
Rubus nivalis	Snow raspberry, subgenus Chamaebatus
Rubus odoratus	Flowering raspberry, subgenus Anoplobatus
Rubus sieboldii	Molucca raspberry, subgenus Malachobatu

Source: modified from Jennings (1988).

1.2.3.1 The European Red Raspberry

The European Red raspberry *R. idaeus* is a raspberry species native to Europe and Northern Asia, which is commonly cultivated in other temperate regions. The wild plants typically grow in forests at high altitudes in mountains in southern Europe and central Asia. In Scotland, to develop and increase raspberry cultivars, a number of raspberry varieties have been developed from the cultivars derived from Glen Clova, accounting for at least 70% of those in the UK. The varieties Glen Magna; Glen Ample; Glen Rosa and Glen Shee were all named after Scottish valleys ("Glens"), and were released in 1996. Glen Ample is the most successful cultivar accounting for one third of the total Scottish and UK crop. All these cultivars grow in early summer (e.g. Glen Moy; Glen Clova; Boyne and Titan) and some extend to the mid (e.g. Glen Ample and Glen Shee) and late summer (e.g. Glen Magna; Octavia and Coho) (Table1.2).

Glen Moy is one of the cultivars derived from Glen Clova and was released by the Scottish Crop Research Institute (SCRI) in 1981 (Jennings, 1988). This cultivar gives large fruits, and good yield but is susceptible to low temperature damage. Given its early growth and improved fruit size and flavor, Glen Moy is the original genetic material used in this thesis.

1.2.3.2 The North America Red Raspberry

R. strigosus is a raspberry species native to most of North America, being a popular fruit that produces dark red colour, with delicious flavour and good firmness while having an extended harvest period and tolerating several known viruses. The cultivar Latham was derived from crossing King and Loudon strains and was introduced commercially in 1914, thus becoming an important fruit cultivar at an early date in America, and later in parts of Eastern Europe (Jennings, 1988). It was one of the original cultivars that were improved by controlled breeding in the 1930s. Importantly, it is genetically approximate 60% similar to *R. idaeus* Glen Moy (Graham and McNicol, 1995). Whilst showing differences in certain traits, a general characteristic comparison between the cv. Glen Moy and Latham is shown in McCallum (2009).

Table 1.2 : Cultivars and growing season of the European Red raspberries

Colour and growing season of fruits	Selected important cultivars
Red, early summer	• Boyne • Fertődi Venus • Rubin Bulgarski • Cascade Dawn • Glen Clova • Glen Moy • Killarney • Malahat • Malling Exploit • Titan • Willamette
Red, mid summer	• Cuthbert • Lloyd George • Meeker • Newburgh • Ripley • Skeena • Cowichan • Chemainus • Saanich • Glen Ample • Glen Shee
Red, late summer	• Cascade Delight • Coho • Fertődi Rubina • Glen Prosen • Malling Leo • Octavia • Schoenemann • Glen Magna

Source: modified from Huxley (1992).

Table 1.2 : Cultivars and growing season of the European Red raspberries (continued).

Colour and growing season of fruit	Selected important cultivars
Red, fall, autumn	• Amity • Augusta • Autumn Bliss • Caroline • Fertődi Kétszertermő • Heritage • Josephine • Ripley • Summit
Gold/Yellow, fall, autumn	• Anne • Fallgold • Fertődi Aranyfürt • Goldenwest • Golden Queen • Honey Queen
Purple	• Brandywine • Royalty
Black	• Black Hawk • Bristol • Cumberland • Glencoe • Jewel • Munger • Ohio Everbearer • Scepter

Source: modified from Huxley (1992).

1.2.4 Improvements in Quality Traits

For almost two centuries, the raspberry has been developed starting in 1865. An essential objective in raspberry improvements has been to produce cultivars with high yield and other good traits in fruit including texture; size; weight; taste; flavour as well as resistance to diseases and pests. Breeders have crossed red raspberries within both their subgenus and a different subgenus to produce new hybrids providing the basis for further improvements. However, because many factors impact on fruit quality, particularly genetic diversity and the growth environment, qualitative traits may not be stable (Hartl and Jones, 2006). Making controlled hybrids to produce a new high quality cultivar takes a long time, between 10 and 15 years (Wenzel, 2006). Molecular techniques have been used extensively to help in improving the cultivars in a shorter time scale and achieving higher quality (Dale *et al.*, 1993). It is now essential to use molecular approaches for plant breeding.

Red raspberry is a diploid species with 7 chromosomes (2n=2x=14) and an estimated genome size of 275 Mb (Jennings, 1988 and Dale *et al.*, 1993), which is approximately twice times of the Arabidopsis (*A. thaliana*) genome (115-125 Mb) (Hu *et al*., 2011); equal to the diploid strawberry and peach genome (270 Mb) (Pontaroli *et al*., 2009) but smaller than melon (450 Mb) (Garcia-Mas *et al*., 2012) and also tomato genomes (950 Mb) (Michaelson *et al*., 1991 and Barone *et al*., 2008). Being commercial fruit that contains not too large a genome, raspberry has been a good model for genomic studies, which are aimed at plant breeding improvements. Conventional breeding and application of molecular breeding strategies have been reported for nearly 60 major gene traits in *R. idaeus* since the 1990's (Jennings, 1998; NCBI). The use of this information in improving quality traits and cultivars could be expanded, and used for other advantages.

To improve red raspberry quality for the United Kingdom fruit markets, breeders at SCRI have crossed *R. idaeus* (Glen Moy) and *R. strigosus* (Latham) using methods of plant breeding to produce hybrids for resistance to disease, together with higher yield and quality traits such as size; color; flavor and health benefits (Brennan and Graham, 2009). The genetic information for several traits has been studied and the results have

been published (Woodhead *et al*., 2010 and Graham *et al*., 2002; 2004; 2008; 2009). These are in the form of markers, quantitative trait loci (QTLs), and DNA sequences of genes known to be involved in quality traits. Seven genetic linkage groups (LGs) of red raspberry are now available (Graham *et al*., 2004; Kassim *et al*., 2008; 2009 and McCallum, 2009 2010; Paterson *et al*., 2013 and NCBI). This remarkable amount of information serves as tools that can now be used to achieve the desired qualities in red raspberry cultivars and also could be applied to other plants.

1.3 UTILIZATIONS OF GENETIC INFORMAION IN PLANT IMPROVMENT

The current concept for plant development is based on the integration of knowledge in conventional plant breeding with the use of molecular techniques known as Molecular Plant Breeding or MB (Wenzel, 2006). Using molecular techniques supports a better understanding of whole genomes and the function of genes in plants. Plant genomics has grown significantly in a short period, with genetic linkage maps, qualitative trait loci and genetic or molecular markers for the purpose of crop improvement by MB.

Genomics-Assisted Crop Improvement (Varshney *et al*., 2005; Varshney and tuberosa, 2007) or Genomics-Assisted Breeding (GAB) uses a range of applications within the MB concept. The main objective of GAB is to assess the availability of an entire genomic resource to effectively utilize the genetic information for developments in plant species. The approaches under GAB include: association genetics of complex traits; genetic maps; physical map; genome sequencing; Quantitative Trait loci (QTL) analysis; functional molecular marker; functional genomic; expression genetics; gene discovery; isolation of genes and allele mining etc. (Varshney *et al*., 2005), including transcriptomics and proteomics (Chen and Harmon, 2006) and metabolomics (Amiour *et al*., 2012; Fernie and Schauer, 2008; Cevallos-Cevallos *et al*., 2009). Varshney summarized the overview of the availability of genomic resources and genomic researches in crop plant species for effectively exploiting genomic researches for crop improvement in Figure 1.1, as the potential of GAB in improving in several cereal species is reported in Varshney (2006).

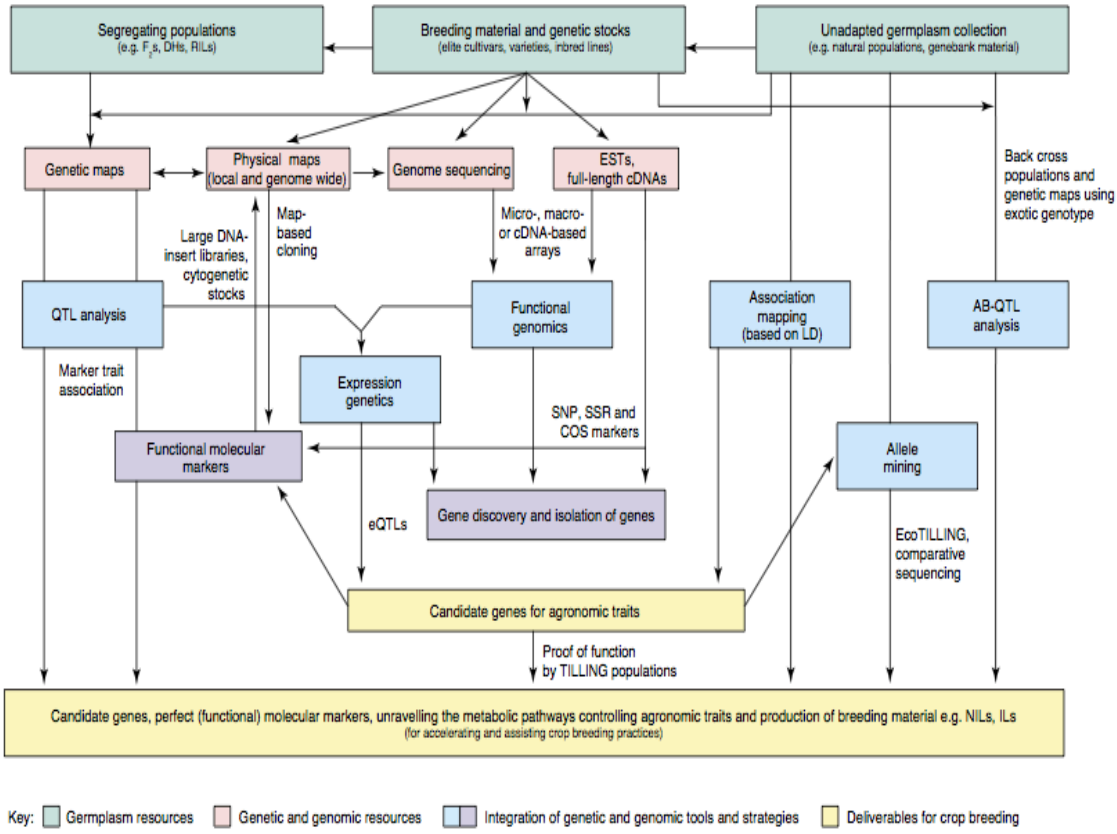


Figure 1.1 Schema of an integrated view of genomic resources and genomics research in crop improvement using genetic and genomic approaches. Several approaches are used, such as: genetic maps; physical map; genome sequencing; QTL and analysis; functional molecular marker; functional genomic; expression genetics; gene discovery; isolation of genes and allele mining etc. Abbreviation: Ab-QTL, advanced backcross QTL; eQTL, expression QTL; COS, conserved orthologous set; ESTs, expreesed sequence tags; ILs, introgreesion lines; LD, linkage microsetellite; TILLING, targeted induced local lesion in genome (Varshney *et al*., 2005).

1.3.1 Marker-assisted Breeding

A molecular marker, also called genetic marker is an identifiable trait that can be associated with a particular position on the genome. The major usefulness of a genetic marker is when linkage of the marker to a particular gene has been established. The presence of known markers linking to specific genes indicates the presence of those genes in plants, and this is very useful in plant breeding, specifically Marker-assisted Breeding (MAB) and Marker-assisted Selection (MAS) (Collard and Mackill, 2008; Morgante and Salamini, 2003 and Lande and Thompson, 1990). Despite there being different classes of markers, for most purpose, markers are divided into three general classes (Paterson, 1996a).

(1) **Morphological markers** are visually phenotypic traits such as flower colour, seed shape, growth habits or pigmentation.

(2) **Biochemical markers** are measurable compound molecules, serving as a sign of a biochemical or other abnormality such as hormones or enzymes.

(3) **DNA or molecular markers** are based on the presence of a DNA sequence on a chromosome, usually located in non-coding regions of DNA.

Both morphological and biochemical markers can be limited by influences of the growing environment, but they still have been used in the development of some plants (Balen, 2009). In contrast, molecular markers are unlimited, consequently leading to increased use in plant breeding and giving more unique markers. Since molecular markers are not the genes of interest, but are just acting as signals or tags that don't affect the genes of interest or the phenotype of traits (Winter and Kahl, 1995), they have become common tools for many applications to assess relationships of traits and inheritance to improve those traits in plants of interested (Dudley, 1993); diversity within germplasm and cultivar identity (Collard, 2005).

Molecular markers assist in indication the genomic location of specific genes, alleles of traits that could suggest the possibility of plant selection since at an early stage of generation and undesirable traits could be protected or decreased (Konzun, 2002). Analysing differences in genotypes with homologous or heterologous DNA sequences

by means of restriction fragment length polymorphism (RFLP); amplified fragment length polymorphism (AFLP) and randomly amplified polymorphic DNA (RAPD) lead to the generation of polymorphic markers. All these techniques are based on restriction digestion; polymerase chain reaction (PCR) and electrophoresis to identify RFLP have been extensively used since having more advantage than in AFLP and RAPD (Amsellem *et al.,* 2000 and Shulaev *et al*., 2008) but consuming time and might show the low level of polymorphism analysis in some plants (Silberstein *et al*., 1999).

Along with sequencing technology, several markers have been generated with short DNA sequences such as Simple Sequence Repeats (SSRs); expressed sequence tags (ESTs); Short Tandem Repeats (STRs); Sequence-related Amplified Polymorphism (SRAP); Intersimple Sequence Repeat (ISSR) and Singer Nucleotide Polymorphic (SNPs). Since not only being hypervariable; multiallelic; co-dominant; locus-specific but also distributing throughout the genome (Hartl and Jones, 2006), these markers have been used in many applications for linkage map construction (Diaz *et al*., 2011); physical map construction (Oeveren *et al*., 2011); hybrid seed purity testing (Dongre *et al*., 2011); cultivar identification (Wünsch and Hormaza, 2002) as well as allele mining (Chen *et al*., 2006). For instance, EST maskers have been used to select candidate genes implicated in traits of interest while the EST collections can be sources of SSRs and SNPs for further application of saturated genetic maps (Morgante *et al*., 2002 and Kantety *et al*., 2002) for crop plant species.

1.3.2 Quantitative Trait Loci Approach

Quantitative traits are phenotypes that are variable and which can be influenced by several genes and the environment (Russell, 2006). Quantitative Trait Loci (QTL) are locations of markers or genes linked to genes that control quantitative traits (Hartl and Jones, 2006). Both the number of QTLs and number of alleles on each QTL primarily impact on quantitative traits and environments may subsequently provide a wide range of phenotypic possibilities (Paterson, 1996a). QTL approach presents the relationship between positions of genes and traits of interest at both physiological and biochemical level under growing environments (Collins *et al.,* 2008; Kassim *et al*., 2009) and also provides useful information from genotype to expressed phenotype of the traits

(Asfaw *et al.,* 2012). Together with, identification of allelic variation of each QTL affecting plant phenotype can increase understanding and lead to trait improvements. Moreover, the closer a marker is to a gene within a QTL, the lower the chance of recombination occurs between the marker and the genes on the QTL; therefore if very close, they will usually be inherited together (Russell, 2006) and can be useful in applying for MAS. QTL mapping has been studied in organisms, together with commercial plants; apple (Newcomb *et al*., 2006; Chagné *et al*., 2012); tomato (Fei *et al*., 2006), grapevine (Cabezas *et al*., 2006) and other fruits in *Rosaceae* were reported in Kassim (2009).

1.3.3 Alleles and Mining Approach

An allele is an alternative form of gene coding or controlling a particular characteristic or trait (Russell, 2001). A strategy to finding naturally occurring allelic variation of candidate genes controlling traits is known as allele mining (Sangh and Dhananjay, 2013; Kumar *et al*., 2010). The common approach of allele mining is identification of the DNA sequence of gene of interest and comparison to known sequences, which are conserved in GeneBank or other information sources. While allele mining assists in discovery of beneficial alleles in plant genetic resources (Kabelka *et al.,* 2004) and development of allele-specific markers (Guo and Qiu *et al*., 2013) for use in MAS leading to development of new varieties in crop plants, it can trace the evolution of plants. Various means for allele identification include PCR (Wang *et al*., 2010; Ikegawa *et al*., 2002 and Nakitandwe *et al*., 2007); SNP resolution analysis to detect points of mutation at nucleotide base pair (Wang, 2008); Southern blotting for RFLP (Dirlewanger, 2004) or AFLP (Gramham *et al.*, 2004) and DNA chip array (DNA micro array) (Galbraith, 2006). However, PCR is the most frequently-used basic technique used for allele mining, being rapid and inexpensive as it subsequently can be used to isolate the alleles of related species either in same genus or family and other plants for trait development and biodiversity characterization. Cultivated crops such as rice (Latha, 2004); apple (Harada *et al*., 2000); pear (Costa *et al.,* 2008); and cherry (Lacis *et al*., 2008) have been developed using an allele mining approach.

1.4 ANTHOCYANINS AND THEIR BIOSYNTHESIS

Anthocyanins are a group of natural compounds that are responsible for the colours to various plants and microbes. In higher plants, they are responsible for most of the red to purple found in leaves; flowers; seed; root; stems and fruits (Lee and Gould, 2002; Wu *et al*., 2004). Anthocyanins also are secondary metabolites having a broad range of pharmacological properties for human health (Lea and Leegood, 1999 and Cseke *et al*., 2006). Being one of members of the flavonoid family, they are considered to have a potential beneficial effect on human health by having several activities such as anti-viral; anti-oxidant; anti-platelet; anti-tumor; anti-carcinogenic; anti-inflammatory and anti-allergic (Pal and Verma, 2013). Anthocyanins play a central role in protection against several human diseases, particularly due to anti-oxidant properties (Wolfgang and Jean, 2008). Together with several health benefits from anthocyanins have been reported in controlling diabetes by assisting to regulate blood glucose and decrease the risk of diabetes (Dembinska-Kiec *et al.,* 2008); preventing cancer by inhibiting the growth of cancer cells (Zafra-Stone *et al*., 2007); strengthening the veins; decreasing the risk of cardiovascular disease; improving vision (Miyake *et al*., 2012); retarding the effects of brain aging (Shukitt-Hale *et al*., 2008); preventing the loss of memory and motor skills (Beekwilder and Hall, 2005); protecting against infectious bacterial diseases and viral diseases.

1.4.1 Fundamental of Biosynthesis

Anthocyanins are produced by polyketide biosyntheses those are diverse pathways generating numerous aromatic polyketide compounds (Escarpa and Gonzales, 2000). Following the photosynthesis and carbohydrate catalytic pathway, the formation of aromatic polyketide compounds starts at the shikimate pathway in the vacuole of plant cells. Aromatic amino acids, phenylpropanoid and phenolic compounds are produced, then lead to the beginning of the synthetic pathways of quinones, tannins, lignin and flavonoids (Cseke *et al*., 2006). Approximately eight thousand phenolic compounds are synthesized and half of these are flavonoids (Harborne and Baxter, 1993). The mechanism for forming compounds is catalyzed by a family of polyketide synthases

(PKSs) and related transferases as compounds formed are the precursors in subsequent reactions (Heldt, 1997). The synthesis performs diverse activities alternating in starter molecules, number of chain extensions, mechanisms and cyclisations of condensation resulting in various products (Austin and Noel 2002). A few condensation cycles produces structurally simple polyketides as multiple cycles provide more complex compounds such as coumaroyl triacetic acid (CTA); chalcone (CH) and stilbene (ST).

Following the biosynthesis of phenylalanine in the phenylpropanoid pathway, the condensation reaction of one *p*-Coumaroyl CoA molecule and three malonyl-CoA that derived from the malonic acid/acetate pathway is catalyzed by the first enzyme of flavonoid syntheses, chalcone synthase (CHS), resulting in formations of chalcone derivatives as chalcone is the first necessary precursor for subsequent syntheses of anthocyanins; flavonols; flavanones; isoflavones; flanones and anthocyanins in closely related pathways (Bowsher *et al*., 2008). Anthocyanins are synthesized at the end of a branch-point of processed steps of former compounds and enzymes (Figure 1.2). More than five hundred different anthocyanins are synthesized in the pathways (Cseke *et al.,* 2006). The diversity of anthocyanin compounds makes them a very complex and provides a variety of colours.

The chemically basic structure of anthocyanins comprise of C6-C3-C6 skeletons of three rings: A; B and C. An aromatic ring A condensed from the three malonyl CoA molecules from malonic acid/acetate pathway as a heterocyclic ring B and partial of ring C derived from phenylpropanoid pathway, thus forming completed B ring with the three carbon bridge. Differences of adding substituent at the position 3'(R_1) and 5'(R_2) of the molecules result in a range of visible colours in numerous anthocyanins (Bowsher *et al*., 2008 and Romeo and Dixon, 1996). Some examples are shown in Figure 1.3 A and B. However, six common compounds are found in anthocyanins-containing fruit/vegetables including cyanidin (~50%), together with pelargonidin, peonidin and delphenidin each founding ~12%, followed by malvidin and petunidin each presenting ~7% (Castaneda-Ovando *et al*., 2009). Finding different anthocyanins in soft berries explains the reason why the fruits are sources for anti-oxidant properties (Figure 1.3 C).

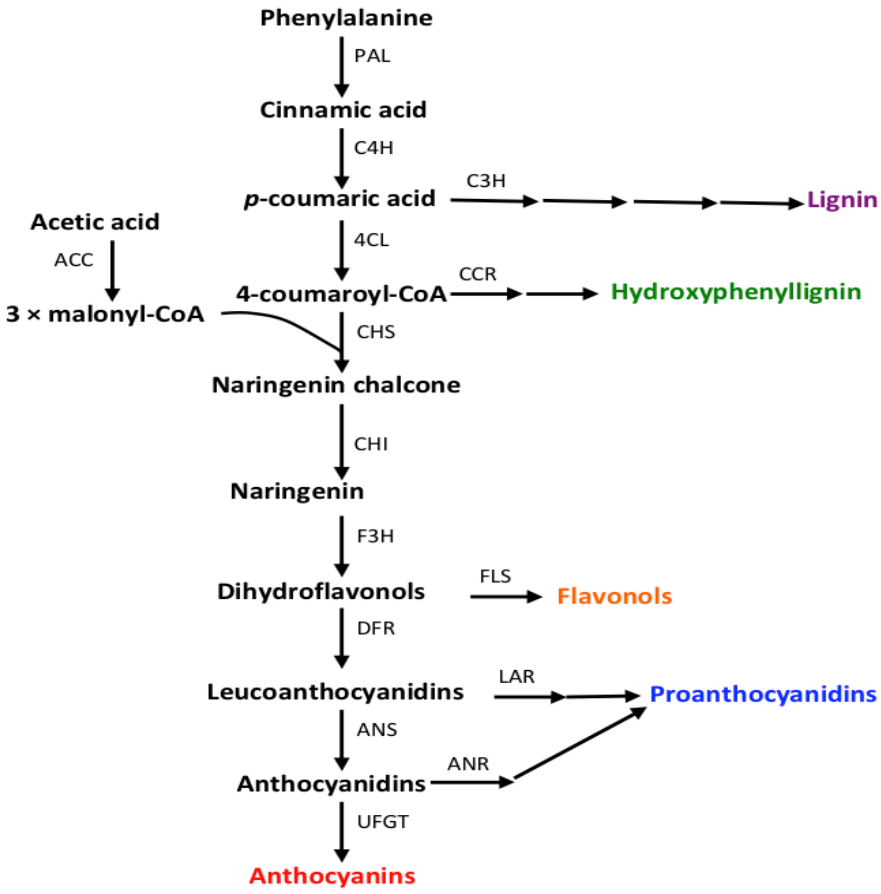
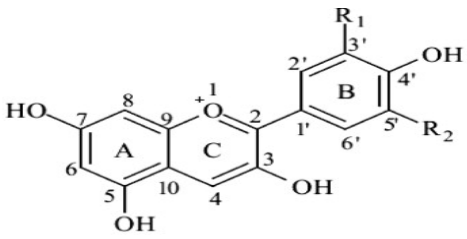


Figure 1.2 An overview of the pathway and branch-points leading to the biosynthesis of the various flavoniods; anthocyanins and related compounds. One *p*-Coumaroyl CoA molecule from phenylpropanoid pathway and three malonyl-CoA molecules that derived from the malonic acid/acetate pathway serve as precursors for condensation reaction catalyzed by chalcone synthase (CHS), leading to the formation of chalcone derivatives. PAL (pheammonia-lase); C4H (cinnamate 4-hydroxylase); 4CL (4-coumarate-CoA ligase); CHS (chalcone synthase or Naringenin-chalcone synthase); CHI (chalcone isomerase); F3H (flavanone 3-hydroxylase); DFR (dihydroflavo 4-reductase); LAR (leucoaanthocyanin reductase); ANS (anthocayanin synthase); ANR (anthocyanin reductase); UFGT (UDPG-flavonoid-3-O-glucosyltrasferase); C3H (4-coumarate 3-hydroxylase); CCR (cinnamyl-CoA reductase); FLS (flavonal synthase); and ACC (acetyl CoA carbokylase).

A



B

Anthocyanins	Substituent at position R1 and R2		Colour
Pelargonidin	3'-H,	5'-H	Orange-red
Cyanidin	3'-OH,	5'-H	Red
Peonidin	3'-OCH_3,	5'-OCH_3	Pink
Delphinidin	3'-OH,	5'-OH	Bluish-purple
Petunidin	3'-OCH_3,	5'-OH	Purple
Malvidin	3'-OCH_3,	5'-OCH_3	Reddish-purple

C

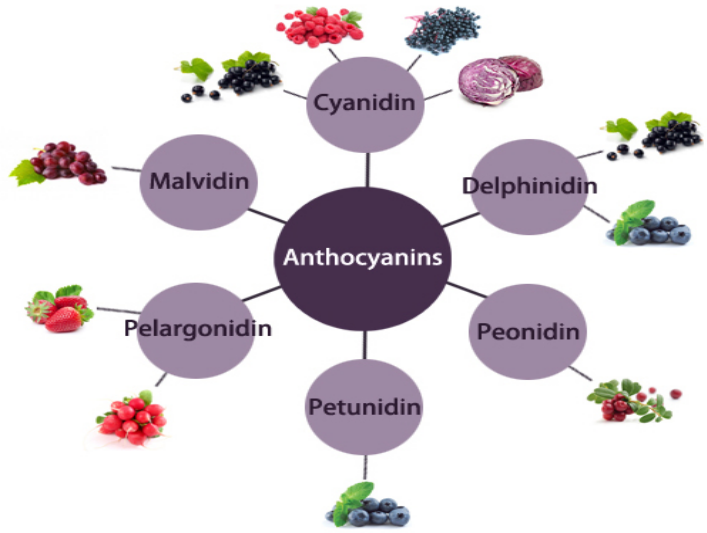


Figure 1.3 Chemically basic structure of anthocyanin compounds. **A**, Basic structure of anthocyanins comprises of rings: A; B and C rings (C6-C3-C6). **B**, differences of adding substituent at the position 3'(R_1) and 5'(R_2) of C ring of anthocyanin molecules resulting in range of visible colours. **C**, certain of Anthocyanins, such as cyanidin; delphenidin; malvidin; pelargonidin; peonidin and petunidin, are found in different soft berries (Bowsher *et al*., 2008 and Cseke *et al*., 2006).

1.4.2 Polyketide Synthases

Although polyketide synthases (PKSs) catalyze similar reactions, they are classified into three major types according to their architectural configurations as type I, II and III. Unlike Type I and II, Type III polyketide synthase have been found in plants, leading to them being known as plant polyketide synthases. PKSs play very diverse functions in the biosynthesis of plant polyketide derivatives, therefore numbers of PKS genes code polyketide synthases, differing in their catalytic functions. PKS1 and PKS5 have been known into be involved in chalcone derivatives pathways. Whilst being the first enzyme of the synthetic pathways of flavonoid derivatives, chalcone synthase (CHS) are well known representatives of this synthesis, and also PKS1 and PKS5. There is an entire CHS-gene family within most plants (Cseke *et al*., 2006).

The PKS family includes not only chalcone synthase (CHS) but also non-CHS type PKSs such as benzophenone synthase (BPS); benzalacetone synthase (BAS); bibenzyl synthase (BBS); phlorisovalerophenone synthase (VPS); isobutyrophenone synthase (BUS); coumaroyl triacetic acid synthase (CTAS) and stilbene synthase (STS) (Figure 1.4). The evidence in plants indicated that non-CHS-like PKSs evolved from CHS sequences several times independently during the long period of plant evolution. This process resulted in them being different from an ancestral gene, and influenced the diversity of these genes and enzymes. Because of being a very large family, numerous studies have revealed common features among the type III PKS enzymes and genes including sequence homology and gene structure, conserved amino acid residues involved in catalysis, enzymatic characteristics and reaction mechanism etc. Today there are over 435 partial or full chalcone synthase as well as chalcone synthase-like sequences reported in GeneBank.

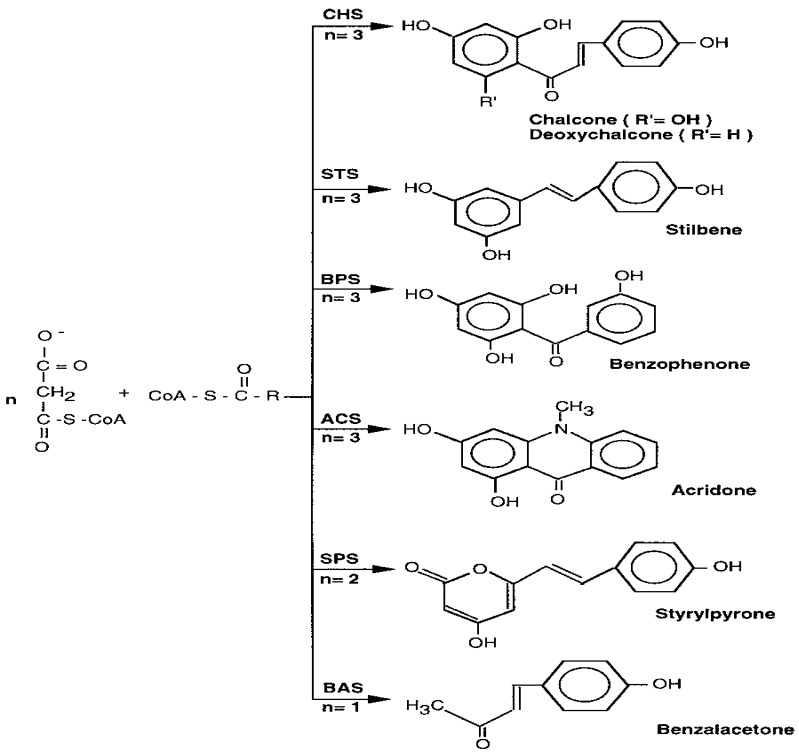


Figure 1.4 Diagram of reactions of certain polyketide synthase (CHS and non CHS) and their main products, includes chalcone synthase (CHS); stilbene synthase (STS); bibenzyl synthase (BPS); styrylpyrone synthase (SPS); acridone synthase (ACS) and benzalacetone synthase (BAS). N in the first substrate presents the number of malonyl-coA molecules as R in the second substrate are a *m*-hydroxybenzoic acid residue in the BPS reaction and a benzoic acid residue in the ACS reaction; a *p*-coumaryl residues in the CHS; STS; SPS and BAS reaction.

1.4.3 Polyketide Synthase Gene Identification

Following evolution from ancestor genes, PKS/CHS genes have become multiple genes (Hopwood, 1997). Although several PKS genes have been reported in plants, their functions have not been clarified. Comparative analysis of plant PKS reveals substantial similarity at both nucleotide and protein sequence level and gene identification must be conducted at high levels of similarity. All plant PKS consist approximately of 400 residue polypeptide chains that share >50% sequence identity. However, despite the considerable sequence similarity, PKS can exhibit very distinct catalytic capabilities. Because predicting the function of the individual PKS by sequence comparison alone is not reliable, expression analysis has been carried out, in parallel to sequence analysis. To identify, if PKS share >90% sequence identity, it is likely that their functions could be initially predicted by sequence comparison.

For instance, in *Gerbera hybrida*, the three PKS genes have been identified; gCHS1 and gCHS3 sharing 89% amino acid sequence identity showing a typical CHS-type activity while the gPS2 gene acting as a PS-type PKS (Pyrone synthase) shares 75% amino acid sequence identity to the first two (Eckermann *et al*., 1998). Schröder (1997) reported the two PKS genes in *Pinus srobus* share 87% amino acid sequence identity while the first gene (PStrCHS1) performs a typical CHS reaction but the second (PStrCHS2) catalyzes diketide derivative analogous to a CHS-intermediate rather than common CHS starter substrates. This proves that the catalytic capacity of PKS can not be predicted solely from sequence information. Today, there are at least eleven CHS gene sequences with high percentage identity, which are submitted in GenBank (NCBI), composed of CHS5; CHS6 and CHS11 that are identified as typical CHS gene while CHS1 and CHS2 genes are coding Naringenin chalcone and other CHS are CHS-like PKS (Table 1.3).

Several other PKS/CHS genes have been identified and studied for long term in plants such as pea (Harker *et al*., 1990); alfalfa (Junghans *et al*., 1993); soybean (Akada *et al*., 1995); morning glories (Durbin *et al.,* 2000 and Yang *et al*., 2004); rhubarb (Abe, *et al*., 2005); (Radhakrishan and Soniya, 2009); cannabis (Flores-Sanchez *et al*., 2010) and while smaller gene families have been cloned from species such as mustard

(Batschauer *et al*., 1991); gerbera (Helariutta *et al*., 1996); Mathhiola (Hemleben *et al*., 2004) and red raspberry (Zheng and Hrazdina, 2007 and Kassim, 2009), ginger (Radhakrishnan and Soniya, 2009), peony (Zhou *et al.*, 2011) In contrast, single PKS genes have been reported in Arabidopsis (Feinbaum and Ausubel, 1988 and Koch *et al*, 1999) and Antirrhinum (Ma *et al*., 2009).

Table 1.3 Polyketide Synthase Gene Identification

PKS gene	Identified gene/enzyme	Identified compound
PKS1	- CHS5, CHS6, CHS11 - CHS1, CHS2 - other CHS (CHS-like PKS)	- typical chalcone - naringenin chalcone - unknown function
PKS2	- unknown Function	- unknown function
PKS3	- CTAS (*p*-coumaryltriacetic acid synthase)	- coumaryltriacetic acid
PKS4	- BAS (benzalacetone synthase)	- benzalacetone
PKS5	- CHS (chalcone synthase)	- naringenin chalcone (exclusivly)

(from NCBI databases, Zheng, 2001; 2009 and Kumar and Ellis, 2003)

1.5 SCOPE AND AIMS OF THESIS

Being a commercially important fruit, red raspberry cv. Glen Moy has been the subject of breeding programs for many years to improve the quality traits of the fruit. Graham, plant breeder at SCRI, constructed the first genetic linkage map in 2004, thereafter the map has been under constant evolution. To improve the quality in red raspberry fruits through increasing anthocyanin content, the main objective of this thesis was focused on investigation of the type III Polyketide synthase genes (PKS) that are an essential for the biosynthesis of anthocyanins in red raspberry.

To access a better understanding and detailed information of the PKS genes identified previously, this thesis focused on characterizing a region of the red raspberry genome using 454 sequencing technology assisted by fosmid library and investigation of the PKS genes.

To achieve the proposal, there were five key objectives:

1. To characterize the polyketide synthases genes (PKSs) in selected BAC clones constructed from genomic DNA of the European red raspberry cv. Glen Moy (Chapter 3).
2. To construct a fosmid sub-library of BAC-derived red raspberry to be a more manageable source for further manipulation (Chapter 4).
3. To complete the sequence of BACs by assembly of resulting 454 sequence with assistance of the constructed fosmid sub-library. In addition, to construct physical maps derived from BAC assemblies, presenting the locations of the PKS genes within the BACs (Chapter 5).
4. Identification and analysis of Open Reading Frames to determine variation of PKS genes in red raspberry and comparison to other related fruits (Chapter 6).

CHAPTER 2

General Materials and Methodologies

2.1 MATERIALS

2.1.1 BAC DNA of *Rubus iedea* cv. Glan Moy

A large insert genomic bacterial artificial chromosome (BAC) library has been constructed from genomic DNA isolated from the European red raspberry cultivar Glen Moy at SCRI (Hein *et al.,* 2005). The library comprises an average insert size of approximately 130 kb and has been screened (Hein per. Com.) with chalcone synthase 11 (CHS11) probe generated from cv. Royalty (Kumar *et al*., 2001). The identified clones have been subsequently cloned into pT7Blue-3 vector DNA and screened for PKS genes sequences at the University of Strathclyde (Kassim, 2009) using primers derived from red raspberry cv. Royalty (Zheng *et al*., 2003). There were three BAC clones (31B12, 29M05 and 24P12) identified positive for PKS1, PKS2 and PKS3 which are the original material for this research.

2.1.2 Plant Genomic DNA of *Rubus iedea* cv. Latham

To prepare plant genomic DNA material, the DNA of red raspberry cv. Latham, which was grown at SCRI, has been extracted.

1) The 1 g of very fresh young leaves was ground in liquid nitrogen with pinch of sand, and added a spatula tip of polyvinylpyrrolidone (PVP). At the end of grinding, the 5 ml of the extraction buffer (2%CTAB, 100 mM Tris-HCl (pH 8), 1.4 M NaCl, 20 mM EDTA, 0.1% DTT) was added to the fine powder, then poured to a 15 ml tube.

2) The sample was placed in the water bath at 65°C for 30 minutes, and added with 7.5 ml of chorofrom/iso-amyl alcohol (24:1), then put on shaker for 15 minutes and centrifuged at 4500 rpm for 10 minutes.

3) The aqueous supernatant was filtered through muslin cloth into a clean tube. An equal volume of ice-cold iso-propanol was added, then mixed and incubated at room temperature for 15 minutes, and centrifuged for 20 minutes at 4500 rpm.

4) Finally, the supernatant was removed, and then dried the pellet for 5 minutes, resuspended in 0.5-1 ml of TE or water, then 1 µl of Rnase was added. The sample was run electrophoresis and DNA quantification.

2.2 METHODOLOGIES

Methodologies in this thesis are divided into five main parts responding to the five experiments that have been done in the thesis. Each section is separately explained for individual Chapter.

2.2.1 Analysis of Bacterial Artificial Chromosome (BACs) (Chapter 3)

2.2.1.1 BAC DNA Isolation

To prepare BAC DNA, BAC colonies were culture in 5 ml LB with chloramphenicol (25 µg/ml) at 37°C for 8 hr, then inoculated into 50 ml LB and cultured overnight, afterwards BAC DNA was isolated out using PhasePrep BAC DNA kits (NA0100, Sigma Aldrich).

1) An overnight culture of BAC clones was transfered (30-40 ml) and centrifuged at 4,000-5,000x*g* for 10 minutes. The medium supernatant was removed and 2 ml of Resuspension Solution was added to resuspend the pellet by pipetting up and down.

2) The 2 ml of Lysis Solution was added to the resuspended cells, then mixed immediately by gently inverting the tube 4-6 times, and incubated at room temperature for 5 minutes.

3) The 2 ml of chilled Neutralization Solution was added, mixed immediately and thoroughly by gently inverting the tube 6-8 times, together with incubated on ice for 5-10 minutes.

4) The lysate was centrifuged at 15,000x*g* for 20 minutes at 4°C to pellet the cell debris and the supernatant carefully was transfered with a pipette to a clean

centrifuge tube. To this was added 3.6 ml of room temperature 2-propanol, mixed thoroughly, centrifuged at 15,000x*g* for 20 minutes at 4°C.

5) The pellet was washed with 2 ml of room temperature 70% ethanol and centrifuged at the same speed and temperature for 5 minutes. The pellet was air-dried briefly for 5 minutes and resuspended in 650 µl of Elution Solution for 5 minutes, then 1 µl of the RNase Cocktail was added to the tube and incubated in a 60°C water bath for 10 minutes.

6) The samples were centrifuged briefly (1 minute) and transferred to a 1.5 ml microcentrifuge tube. To this was added 50 µl of sodium acetate Buffer Solution (3 M, pH 7.0) and mixed briefly, together with 120 µl of Endotoxin Removal Solution and mixed thoroughly by inversion for 30 seconds, then chilled on ice for 5 minutes.

7) The tube was warmed in a 37°C water bath for 5 minutes, centrifuged at maximum speed (~16,000x*g*) for 3 minutes at room temperature and carefully transfered the aqueous phase into a microcentrifuge tube. This was added 700 µl of room temperature DNA Precipitation Solution, mixed and centrifuged at maximum speed for 20 minutes at 4°C.

8) The supernatant was moved and the pellet was washed twice at room temperature with 70% ethanol, then was dried briefly (for 5 minutes) and resuspended in Elution Solution or sterile deionized water. Samples were run electrophoresis to check intensity of DNA and quantified by spectrophotometric analysis.

2.2.1.2 BAC DNA Restriction Digestion

BAC restriction digestions were setting up for each BAC clone with *Hind*III, *Xba*I and *Not*I in 1X buffer of 20 µl final volume reaction. The sample were incubated at 37°C for 2 hrs, then run electrophoresis in a 0.8 % agarose gel, prepared in TBE and added 2 µl of GelRed (nucleic acid gel stains, BIOTIUM), at 80 volt. At the final, the DNA fragments were estimated by comparing with 1 Kb ladder (Bioscience) under UV-transilluminator.

2.2.1.3 BAC PCR-based Cloning and Sequencing

2.2.1.3.1 Polymerase Chain Reaction

To amplify PKS genes (PKS1, PKS2 and PKS3) in BAC DNA using published PKS gene primers that were derived from Red raspberry cv. Royalty (Table 2.1) (Zheng *et al*., 2001 and Kasim, 2009), Polymerase Chain Reaction (PCR) began with setting up the following 50 µl PCR reaction composed DNA Template 10-100 ng, 5 µl of 10X PCR Buffer, 0.5 µl of 50 mM dNTPs, 1 µM each of Primers (~200 ng each), added water to a final volume of 49 µl and 1 µl of *Taq* Polymerase (1 unit/µl). Reaction were incubated at 94°C for 5 minutes to allow denaturation of template DNA, together with 30 amplification cycles carried out as follows; denature at 94°C for 1 minute, anneal at 56°C for 30 seconds, extend at 72°C, along with extension at 72°C for 10 minutes before holding at 4°C. The DNA segments can be visualized and estimated the relative size by comparing to DNA fragments of known length, a ladder.

Table 2.1 Primers used in PCR amplifications of PKS genes.

Primer	Target gene/ fragment	Accession no.	Primer sequence (5′-3′)	Predicted size (bps)
prPKS1f prPKS1r (PKS1.2*)	PKS1	AF292367.1	TCAGCCCAAGTCCAAAATC GCCACACTGTGAAGCACAAC	~800
prPKS2f prPKS2r (PKS2*)	PKS2	AF292368.1	AATTTCCCACGCAATCCTT GGCCTCGTTAAGGCTCTTCT	~700
prPKS3f prPKS3r (PKS3.2*)	PKS3	AF292369.1	AAAGAGGCTGCCACTAAGG TCAAGTTGAAGCTGCCACAC	~800

(*Original name, modified from : Kasim, 2009 and Zheng *et al*., 2001)

2.2.1.3.2 Electrophoresis and Visualization

DNA fragments obtained following PCR amplification were separated on the basic of their size or molecular weight by electrophoresis through an agarose gel. DNA fragments were prepared in 0.6% agarose in 1XTAE buffer and electrophoresis was performed at 80 volt then the PCR amplification was visualized and estimated the relative size by comparing to DNA fragments of known length, 1000 bp ladder (Promega). Each successfully amplified PCR product was separately extracted from the agarose gel using a scalpel blade in preparation for cloning.

2.2.1.3.3 Gel Extraction

Amplified DNA fragments were extracted from agarose gel using the manufacturer's guidelines (Qiagen II).

1) Within gel slice, DNA was weighed in a centrifuge tube, along with added 3 volumes of Buffer QX1 to 1 volume of gel slice. The sample was added 10 µl of QIAEX II and incubated at 50°C for 10 min whist mixing by vortex every 2 min to keep QIAEX II in suspension.

2) The sample was centrifuged at high speed for 30 seconds and carefully removed supernatant with a pipet, along with washed the pellet with 500 µl of Buffer QX1. The pellet was resuspened by vortexing and centrifuged for 30 seconds. All of the supernatant was discarded and the pellet was washed twice with 500 µl of Buffer PE, along with air-dried the pellet for 10–15 minutes or until the pellet became white.

3) The DNA sample was eluted by adding 20 µl of water within pH 7.5-8 and centrifuged for 30 seconds, then carefully pipeted the supernatant containing the purified DNA into a clean tube and the eluate was stored at –20°C.

2.2.1.3.4 Cloning Reaction

To subclone the Fragments from BAC, the TOPO® Cloning kit (Invitrogen) was utilized for Cloning and transformation. The DNA extracted from PCR product has

been cloned using the TOPO® Cloning reaction. The reactions were performed with a total volume of 6 µl, included 4 µl of fresh DNA from PCR product, 1 µl of salt solution, 1 µl of vector and sterile water respectively, together with mixed gently. The reactions were incubated at room temperature for 30 minutes and placed on ice for eventual transformation.

2.2.1.3.5 Transformation

The cloning reaction was added into 50 ul of chemically competent *E. coli*. (TOPO® Cloning kit) and mixed gently, along with incubated on ice for 5 to 30 minutes. The tube was heat-shocked at 42°C for 30 seconds without shaking and immediately transfered to ice, then added 250 µl of room temperature SOC medium, along with horizontal shaking the tube (200-300 rpm) at 37°C for 1 hr. Two different volumes (10-50 µl) of each transformation were separately spread on a prewarmed LB plates adding 50-100 µg/ml kanamycin and incubated overnight at 37°C. To analyze positive clones, 10 colonies randomly were collected and cultured overnight in 5 ml of LB containing 50 µg/ml kanamycin.

2.2.1.3.6 DNA Extraction

To extract DNA from transformants, 5 ml of an overnight culture (1–2 x 10^9 *E. coli*) was harvested using PureLink™ Quick Plasmid Miniprep Kit (Invitrogen).

1) The culture was centrifuged at 1,500xg at room temperature for 15 minutes and the pellet was completely resuspended in 250 µl of Resuspension Buffer containing with RNase. To this was added 250 µl of Lysis Buffer and mixed gently by inverting the tube 5 times, then incubated for 5 minutes at room temperature.

2) The 350 µl of Precipitation Buffer was added and immediately mixed by gently inverting the tube 5 times, and centrifuged at 12,000xg for 10 minutes. The supernatant was loaded onto the spin column and then centrifuged at 12,000xg for 1 minute, then discarded the flow-through.

3) The column was washed twice with 500 µl and 700 µl of Wash Buffer respectively, by incubating for 1 minute at room temperature, then centrifuged at 12,000xg for 1 minute, together with the flow-through was discarded.

4) The column was placed in a clean 1.5 ml recovery tube, then added 75 µl of water to the center of the column and incubated at room temperature for 5 minute. Finally, centrifuging at 12,000xg for 2 minutes eluted DNA and the purified DNA was analyzed using 0.6 % agarose gel electrophoresis in TAE buffer.

2.2.1.3.7 Transformant Analysis

To analyze the PKS gene inserted in transformants, some subclones were randomly collected. PCR has been setting up using the universal primers (M13 Forward (-20), M13 Reverse) (TOPO® Cloning reaction, Invitrogen) in combination with a designed PKS primer: PKS1, PKS2 and PKS3. At the same time, the restriction enzyme digestion was formulated in parallel to PCR screening

a) Transformant Analysis : Restriction Digestion

Restriction digestions were setting up cross BAC subclones. DNA of transformant clones was digested with Eco*RI* in 1X buffer of 20 µl final volume reaction. Then the reactions were incubated at 37°C for 2 hrs, while a 0.6 % agarose gel was prepared in TBE by adding with 2-3 µl of Syber green. The reactions were run at 80 volts, along with fragment DNA sizes were estimated by comparing with 1000 bps ladder using UV transilluminator.

b) Transformant Analysis : Polymerase Chain Reaction

To amplify PKS genes in subclones amplified from BAC DNA, PCR reactions were performed using the same PKS primer sets (Table 2.1) used for previous amplification. The PCR reactions were set at the same condition in 2.2.1.3.1.

A polymerase chain reaction cocktail was prepared for final 20 µl reaction volume consisting 4 µl of DNA (100 ng), 2 µl of 10X PCR Buffer, 0.5 µl of 50 mM dNTPs, 1

µl of designed PKS primer, 1 µl of Universal Primers (M13 Forward or M13 Reverse), 2 µl of MgCl (25Mm/µl), added 9.5 µl of PCR grade water and 1 µl of *Taq* Polymerase (1 unit/µl) (Promega), consequently. Then the reaction was begun with denaturation at 94°C for 2 minutes and amplified for 30 cycles (94°C for 1 minute, 55°C and 72°C for 1 minute, respectively). At the final extension, the reaction was incubated at 72°C for 10 minutes and hold at 4°C, then visualized by 0.6% agarose gel electrophoresis. The resulting performance of restriction analysis was compared in parallel to confirm the corrected PCR result.

2.2.1.4 Southern Blot Analysis

2.2.1.4.1 Restriction Digestions

Three BACs were analysed by restriction digestions composing with Eco*RI* in 1X buffer of 20 µl of final reaction volume, then incubated at 37°C, for 2 hrs meanwhile 0.6 % agarose gel was prepared in 1XTAE buffer. The DNA was run overnight at 35 volt, then with visualised using UV transilluminator.

2.2.1.4.2 DNA Transferring

After separating the digested DNA by electrophoresis, the gel was transferred to a filter membrane by capillary transfer (applied from Sambrook *et al*., 1989). DNA transferring included denaturation and neutralization steps.

1) To denature DNA, the gel was removed unwanted areas with a clean scalpel blade and soaked the gel in several volumes of denaturing solution (1.5 M NaCl, 0.5 NaOH), for 45 minutes with constant, gentle agitation.

2) The gel was briefly rinsed in deionized water, and neutralized by soaking in several volumes of neutralization buffer (1 M Tris (pH 7.4), 1.5 M NaCl) at room temperature for 30 minutes with the same condition. Following this, changed the neutralization buffer and continue soaked for 15 minutes.

3) At the same time, cut two sheets of Whatman 3MM paper to form supports that were longer and wider than platform, then soaked in transfer buffer and placed these over the platform.

4) Again, two pieces of 3MM paper and a piece of nitrocellulose filter were cut as the same size of the gel in both dimensions, and then soaked 3MM in transfer buffer until thoroughly wet. At the same time, nitrocellulose filter were floated until wet completely, and then immersed in transfer buffer for at least 5 minutes.

5) For the gel, after neutralization, the buffer was removed. The gel was inverted so that underside turn to uppermost, and then placed on the platform of a capillary transfer system filled with transfer buffer (20XSSC). This gel was centred on the wet 3 MM papers. The wet membrane filter was set on the top of the gel, along with wet two pieces of 3MM papers.

6) Preparing a stack of paper towels (5-8 cm high) about smaller than 3 MM sheets, put on wet 3MM papers and also placed the glass plate on top of the stack and weight with a 500 g weight.

7) The DNA was transferred for 8-24 hrs at room temperature. Following this, the paper towels and 3MM papers were removed, the membrane and the gel were turn over, and then put on a dry 3MM paper. Before separating the gel and the membrane, mark the position of the gel slots with a pencil.

8) For the gel, it was peeled and discarded while filter was soaked in 5XSSC at room temperature for 5 minutes. After removing the solution, to fix the DNA on membrane, this was covalently cross-linked by exposuring to a 312 nm ultraviolet light transilluminator.

9) The membrane was placed, DNA side down, on a piece of clingfilm, and exposed for 3 minutes, and then briefly rinsed in double distilled water and air-dried on paper towels for 30 minutes. The filter could be used immediately or stored dry at 4°C until required.

2.2.1.4.3 Labelled Probe Preparing

For Southern blotting, the experiment, including probe preparing, hybridization and immunological detection, was followed using DIG High Prime DNA Labelling and Detection Starter Kit I (ROCHE). To prepare the labelled DNA probes for Southern blotting, the nine amplified DNA using PKS primers (Table 2.1) has been labelled with Digoxigenin-11-dUTP using DIG-High Prime (Table 2.2).

1) The DNA was prepared for 1 µg, then added with autoclaved, double distilled water up to 16 µl in final reaction volume. This was denatured by heating in a boiling-water bath for 10 minutes and quickly chilled in an ice or water bath.

2) The 4 µl of the Mix DIG-High Prime (vial 1) was added to the denatured DNA, mixed and centrifuged briefly. This was incubated at 37°C for 4-6 hrs or overnight whereas longer incubations (up to 20 hrs) will increase the yield of DIG-labelled DNA.

3) The reaction was stopped by adding 2 µl of 0.2 M EDTA (pH 8.0) and/or by heating to 65°C for 10 min.

Table 2.2 Labelled DNA probes for Southern blotting and colony hybridization.

Probe generated	**Gene/ amplified fragment**	**Original BAC**
PKS1p31 PKS1p29 PKS1p24	PKS1	31B12 29M05 24P12
PKS2p31 PKS2p29 PKS2p24	PKS2	31B12 29M05 24P12
PKS3p31 PKS3p29 PKS3p24	PKS3	31B12 29M05 24P12

2.2.1.4.4 Hybridization

DIG-labeled probes (from 2.2.1.4.3) was used for hybridization to membrane blotted DNA according to standard methods. The DIG Easy Hyb working solution was prepared by adding carefully 64 ml sterile double distilled water in two portions to the DIG Easy Hyb Granules (bottle 7), dissolved by stirring immediately for 5 minutes at 37°C.

1) The appropriate volume of the DIG Easy Hyb working solution (10 ml/100 cm^2 membrane) and the membrane were pre-hybridized in an appropriate container at 37-42°C for 30 minutes with gentle agitation. As DIG-labeled DNA probe (about 25 ng/ml) was denatured by boiling in heat box for 5 minutes and rapidly cooling in ice.

2) Then the probe was added to pre-heated DIG Easy Hyb (3.5 ml/100 cm^2 membrane), and mixed thoroughly. The pre-hybridization solution was pour off, therefore adding the probe/hybridization mixture to membrane, along with incubated for 4 hrs or overnight with gentle agitation.

3) After hybridization, the membrane was washed twice in ample 2XSSC, 0.1% SDS at 15-25°C for 5 minutes under constant agitation. Then washed twice in 0.5XSSC, 0.1% SDS at 65-68°C for 15 minutes under constant agitation. DIG Easy Hyb containing DIG-labeled probe could be stored at -15 to -25°C and be reused several times when freshly denatured at 68°C for 10 min before use.

2.2.1.4.5 Immunological Detection

To immunologically detect the DNA, the hybridized probes were detected with anti-digoxigenin-AP and visualized with the colorimetric substrates.

1) For the beginning of detection, the membrane was rinsed briefly for 1-5 minutes in Washing buffer (0.1 M Maleic acid, 0.15 M NaCl;pH 7.5 (20°C), 0.3% (v/v) Tween 20), and then incubated for 30 minutes in 100 ml of 1x Blocking solution (preparing a 1x working solution by diluting 10x Blocking solution (vial 6) 1:10 with Maleic acid buffer).

2) The membrane was continuously incubated in 20 ml of Antibody solution for 30 minutes (preparing by centrifuging Anti-Digoxigenin-AP (vial 4) at 10000 rpm for 5 minutes, pipeted the amount carefully from the surface, and then diluted 1:5000 (150 mU/ml) in 10xBlocking solution).

3) Following this, the membrane was washed twice with 100 ml of Washing buffer for 15 minutes. This was move into 20 ml Detection buffer (0.1 M Tris-HCl, 0.1 M NaCl; pH 9.5 (20° C)) for 2-5 minutes.

4) After this, the membrane was soaked in 10 ml of freshly prepared Color substrate solution (preparing by adding 200 µl of NBT/BCIP stock solution (vial 5) to 10 ml of Detection buffer) in an appropriate container in the dark. The membrane has been exposed to light for short time periods to monitor color development.

5) After the desired spot or band intensities was achieved, the reaction was stopped by washing the membrane for 5 minutes with 50 ml of sterile double distil water or with TE-buffer (10 mM Tris-HCl, 1 mM EDTA, pH8.0).

2.2.2 Fosmid library Construction (Chapter 4)

Methodologies for fosmid library construction in Chapter 4 are divided into two mains areas with methods for construction including the checking of stability of the libraries generated, and then fosmid clones screening

2.2.2.1 Fosmid library Construction

To generate a completely unbiased primary fosmid library which maintained a single copy in the clones, the library was constructed in the pEpiFOS™-5 vector using the EpiFOS Fosmid Library Production Kit (Epicentre, Madison, USA). Since there were three BAC clones, three subsequent libraries were constructed using the same method.

2.2.2.1.1 Preparation of BAC DNA

BAC DNA was prepared as described in Chapter 2, and 2.5 µg was randomly sheared by manual mechanical shearing in a 200 µl small pipette tip. DNA was aspirated and expelled through the pipette for 50,100, 150 and 200 times.

2.2.2.1.2 Recovery of the Size-Fractionated DNA

Samples were loaded onto 1% low melting agarose gel in 1X TAE buffer without ethidium bromide and run overnight at 30-35 Volts. To recover the size-fragmented DNA while protecting the sample from UV light, the marker DNA lanes were cut off the gel and separately visualized under UV light and the position marked on the gel. The entire gel was reassembled, then the position of 36 kb was compared with lanes containing DNA of interested. The region containing the appropriately-sized fragment DNA was excised from the gel and the DNA extracted using Qiagen II (in chapter 2).

2.2.2.1.3 DNA Precipitation

To the sample was added to 1/10 volume of 3 M sodium acetate (pH 7.0) and mixed gently. To this were added 2.5 volumes of absolute ethanol and then mixed thoroughly by gentle inversion. The sample was incubated at room temperature for 10 minutes, then centrifuged for 20 minutes. The supernatant was removed and the pellet was washed twice with cold, 70% ethanol, by careful inversion. The sample was air-dried for 5-10 minutes, and the DNA pellet was gently resuspended in water or TE Buffer.

2.2.2.1.4 End-Repair Reaction

This step was to generate blunt-ended, 5'-phosphorylated end of fractionated DNA. The end-repair reaction was composed of 10 µl of sheared DNA (5 µg DNA), 2 µl of End-Repair 10X Buffer, 2 µl of 2.5 mM dNTP Mix, 2 µl of 10 mM ATP, 1 µl of End-Repair Enzyme Mix (to achieve blunt-ended DNA prior to DNA ligation) and water to 20 µl of final volume reaction. The reaction was incubated at room temperature for 45 minutes, then incubated in a water bath at 70°C for 10 minutes.

2.2.2.1.5 Ligation of Sized DNA Fragments

The concentration of DNA (0.25 µg, 0.01 pmol of 40 Kb DNA) was formulated to ligate to the pEpiFOS-5 fosmid vector (0.5 µg, 0.1 pmol). The reactions were performed with a 10 µl total reaction volume, included 1 µl of concentrated DNA, 1 µl of vector (0.5 µg/µl), 1 µl of 10X Fast-link ligation Buffer, 1 µl of 10 mM ATP, 1 µl Fast-Link DNA Ligase and sterile water to 10 µl of final volume. The ligation was mixed thoroughly by tapping and incubated at room temperature for 2 hours and then transferred to 70°C for 10 minutes to stop the reaction.

2.2.2.1.6 Fosmid Library Propagation

a) Bacterial Culture Preparation

The EPI100-T1R bacteria (from the Fosmid Library Production Kit) were streaked onto a Luria Agar plate (LA) without antibiotic and incubated overnight at 37°C. Following this, a colony was inoculated into 5 ml of Luria Broth (LB) and incubated at 250 rpm at 37°C. Subsequently, to prepare the bacteria for accepting the packaging reactions, 5 ml of the EPI100-T1R overnight culture was inoculated into 50 ml of LB with 10 mM $MgSO_4$ and grown with shaking at 37°C.

b) Fosmid Packaging Reaction

A tube of the MaxPlax Lambda Packaging Extracts (Fosmid Library Production Kit) was thawed on ice to a ligation reaction (2.2.2.1.5). The 25 µl (one-half) of each MaxPlax Lambda packaging Extract was immediately added to each ligation reaction. A remaining 25 µl of the MaxPlax Packaging Extract was returned back to the -70°C freezer and rapidly frozen. The reaction was mixed by pipetting several times, and then briefly centrifuged and incubated at 30°C for 90 minutes. Following incubation, the remaining 25 µl of MaxPlax Lambda Packaging Extract was added to the reaction and incubated for an additional 90 minutes. At the end of incubation, Phage Dilution Buffer (PDB) was added to 1 ml final volume of the reaction and mixed gently. Finally, 25 µl of chloroform was added to the tube and mixed gently and stored at 4°C.

c) Titering the Packaged Fosmid Clones

The titering was carried out by serially diluting the 1 ml of packaged phage particles (2.2.2.1.6/b) into PDF in sterile microcentrifuge tubes, following: 10^2; 10^4; 10^5 and 10^6. Then, 10 µl of each dilution was individually added to 100 µl of the prepared EPI100-T1R culture (2.2.2.1.6/a) This incubated at 37°C for 20 minutes and the infected EPI100-T1R cells were spread on an LA plate with chloramphenicol (12.5 mg/ml) and incubated at 37°C overnight. Following this, colonies were counted and the titer of the packaged phage particle was calculated.

d) Plating and Selecting the Fosmid Clones

The packaged phage particles from each set of ligation reactions were diluted with PDB to required dilution. Subsequently, the 10 µl of dilution was mixed with 90 µl EPI100-T1R cell and the cultures were incubated at 37°C for 20 minutes, then the infected cultures were spread on LA plates with chloramphenicol (+12.5 mg/ml) and incubated overnight at 37°C.

e) Induction of Fosmid Clones to High Copy Number

Following plating, all fosmid clones were individually inoculated into 5 ml of LB medium with chloramphenicol (12.5 mg/ml) and grown overnight at 37°C. To induce to high copy number, 200 µl of overnight cultures were inoculated in 800 µl of LB with chloramphenicol and 1 µl of induction solution (EPICENTRE® Biotechnologies). The cultures were grown at 37°C with vigorous shaking for 5 hours and afterwards inoculated into 5 ml of LB with chloramphenicol and grown for 12-16 hrs at 37°C.

2.2.2.1.7 Fosmid DNA Purification

1) The overnight cultures were transferred to a 1.5 ml microcentrifuge tube and centrifuged at 15,000xg for 3 minutes. Then the supernatant was discarded, and 200 µl of chilled of FosmidMAX Solution1 was added to the pellet, and completely resuspended by vigorous vortexing. To this the 400 µl of FosmidMAX Solution2 was added and mixed by inverting 2-3 times very gently, together with incubated at room temperature for 5 minutes.

2) 300 µl of chilled of FosmidMAX Solution3 was added and mixed by inverting 2-3 times very gently, then incubated on ice for 15 minutes. Again, the lysate was centrifuged at high speed at 4°C for 15 minutes.

3) The supernatant was slowly removed to a microcentrifuge tube without disturbing the pellet. The 540 µl or 0.6 volumes of room temperature isopropanol were added and mixed thoroughly by inverting 4-6 times.

4) The sample was precipitated by centrifugation at 4°C for 15 minutes and the isopropanol was removed and the pellet was air-dried at room temperature for 3-5 minutes, then re-suspended in 250 µl of TE buffer by tapping and swirling the tube. To this 250 µl of chilled FosmidMAX Solution 4 was added and mixed thoroughly by tapping the tube, together with incubated on ice for 15 minutes. Again the sample was centrifuged for 15 minutes, then the supernatant transferred to a microcentrifuge tube.

5) 1 ml of absolute ethanol was added to the recovered supernatant and mixed gently by inverting 4-6 times. Following this, the precipitated fosmid DNA was collected by centrifugation at maximum speed for 15 minutes at 4°C, then the ethanol was carefully pipetted off without disrupting the pellet.

6) The DNA pellet was air-dried briefly at room temperature for 3-5 minutes, resuspended by adding 25 µl of sterile deionized water, and then incubated at room temperature for 10 minutes.

7) At the final step, 1 µl of diluted RiboShredder RNase Blend was added to the tube and incubate at 37°C for 30 minutes. Fosmid DNA was analyzed by agarose gel electrophoresis and to check the intensity and quantity of DNA by spectrophotometer analysis, and stored at –20°C.

2.2.2.2 Fosmid Stability Assays

Stability assays was to analyze stability of the fosmid libraries during propagations. Eighteen individual fosmid clones from a library were randomly selected, and then serially cultured for 6 days according to Kim *et al*. (1992). In brief, each single clone

was picked and inoculated in 5 ml of LB medium, then grown overnight at 37°C at 220 rpm while these cultures were designated as time 0. Each day the cultures were diluted to 10^{-6} and regrown overnight to represent 20 generations a day. These were done for 5 days to recovery 100 generations. The extracted fosmid DNA from the generation 0 and 100 were digested with EcoRI, and then electrophoresis on a 1% agarose gel. The restriction fragment patterns were compared side by side between the two generations.

2.2.2.3 Fosmid library Screening

2.2.2.3.1 Colony Hybridization Screening

Hybridization was conducted on a 22 x 22 cm filter membrane (2.2.1.3.2). The 2 µl of cells from each clone was spotted onto the membrane in duplicate. The screening was performed by hybridization using three PKS probes generated from amplified PCR products using PKS primers set used in 2.2.1.4.4 and 5 to carried out hybridization.

2.2.2.3.2 Polymerase Chain Reaction

PCR reactions were performed using the PKS primer sets (PKS1, PKS2, and PKS3) used for previous PCR amplification (Chapter 3). PCR reactions consisted of DNA Template 100 ng, 5 µl of 10X PCR Buffer, 0.5 µl of 50 mM dNTPs, 1 µM each of Primers (~200 ng each), water to a final volume of 49 µl and 1 ul of *Taq* Polymerase (1 unit/µl). The reactions were incubated at 94°C for 5 minutes to denature a template DNA, together with 30 amplification cycles carried out as follows; denaturing at 94°C for 1 minute, annealing at 56°C for 30 seconds, extending at 72°C, along with final extension at 72°C for 10 minutes. The products were visualized by agarose gel electrophoresis and the amplified size estimated by comparing to a ladder.

2.2.2.3.3 Southern Blot Analysis

Fosmid clones were digested with *Hin*d III, along with electrophoresis and transferred to the membranes, following the steps in 2.2.1.4.4 and 5. The fragments of DNA were hybridized using DIG High Prime labelled PKS-DNA probes amplified from PKS-PCR products using the primers used for PCR screening.

2.2.2.4 Fosmid Ends Sequence Analysis and Annotation

The fosmid clones were end-sequenced with two specific primers for the fosmid vector (pCC1FOS provided by EpiCentre), a 27-mer, pCC1/pEpiFOS forward primer (5'-GGATGTGCTGCAAGGCGATTAAGTTGG-3') and a 26-mer, pCC1/pEpiFOS reverse primer (5'-CTCGTATGTTGTGTGGAATTGTGAGC-3') (Appendix A.1) by Sanger sequencing, Source Bioscience, Cambridge, UK. End-pair Sequences of clones were subjected to sequence similarity searches using BLASTn and BLASTx.

2.2.3 Assembly of Bacterial Artificial Chromosome (BAC) Sequences and Construction of Physical Maps of PKS Gene Loci of Red Raspberry cv. Glen Moy (Chapter 5)

Methodologies for Chapter 5 include the two main parts, following assembly of BAC 454 sequences and construction of initial physical maps of PKS Gene Loci on BAC assembled sequences.

2.2.3.1 Assembly of BAC Sequences of Red Raspberry cv. Glen Moy

To complete DNA sequences of the three BAC clones from red raspberry cv. Glen Moy, a set of the BACs (24P12, 29M05 and 31B12) was sequenced with the Roche 454 GS FLX Titanium system (Source BioScience, Cambridge, UK). The resulting 454 sequence files were analyzed and subsequent combined, with paired-ends of those BACs (from Kassim, 2009) and then paired-ends of fosmid clones sequenced by Sanger method (Source BioScience, Cambridge, UK, in Chapter 4), using Sequencer software (version 5.0.1, U.S.A). Finally, the BAC assembled sequences were analyzed with Blastn, Blastx, tBlastx and Blastp to identify the insert DNA sequences of red raspberry in these three BACs.

2.2.3.2 Construction of Initial Physical Maps of PKS Gene Loci

To construct a draft of physical map derived from the two BACs (24P12 and 29M05), every 2 kb of assembled nucleotide sequences were annotated to search for Open Reading Frames (ORF) using softwares including the gene prediction program for Eukaryotes (GeneMark.hmm-E* and GeneMark-E), FGENESH and TSSP (Softberry test online). BLAST, Sequencer sequence analysis and Sequences viewer CLC.

The ORFs of PKS and all genes identified in were placed in the order based on the assembled sequences along the length of the BACs. Fosmid clones assisted the assemblies of the two BACs (from 2.2.3.1) were chosen to determine their overlaps for constructing the map whilst being clones of the Minimal Tiling Path (MTP) for validation the map drafts in future using restriction digestion (with some rare-cutter and frequency-cutter enzymes) and fragment prediction by Cut Map (Sequencer software, version 5.0.1, U.S.A) including SynGene Gen tools (GeneSnap version 6.08.04) to identify the restriction fragment lengths.

2.2.4 Identification and Analysis of Open Reading Frames Polyketide Synthases Genes (Chapter 6)

To initial investigate the resulting assembled sequences from 2.2.3 (Chapter 5), the sequences of the BAC were searched for the detailed information by identifying and analysing the ORF along their sequences.

To investigate the raspberry PKS gene clusters placed on the physical map drafts in the BACs (from 2.2.3), the regions containing the ORFs of PKS genes were used to identify their sequences using BLAST and compared to other related fruits in the same family using Sequences viewer CLC.

Additional, to confirm the information on the BAC DNA sequences, the other three genes were chosen for PCR including Cytochrome P450 gene; ATP-binding cassette transporter gene and Cycloartenol synthase gene (Table 2.3 A). The amplification of the three BACs and extracted genomic DNA of red raspberry cv. Glen Moy (GM) and

Latham (L) using CYP450; ABC transporter and CAS primers were set at the same condition in 2.2.1.3.1. The reactions were incubated at 94°C for 5 minutes, and followed with 30 amplification cycles carried out as follows: denature at 94°C for 1 minute; anneal at 64°C for 1 minute; extend at 72°C and extension at 72°C for 10 minutes before holding at 4°C. The primer sequences were shown in Table 2.3 B.

Table 2.3 A Functions of CYP450; ABC transporter and CAS genes.

Name of gene/enzyme/protein	Function of Genes/Protein
Cytochrome P450 (CYP450)	Catalyzes the oxidation of organic substrates including metabolic intermediates, for instance lipids and steroidal hormones
ATP-binding cassette transporter (ABC transporter)	Implicated encompasses polar auxin transport, lipid catabolism, xenobiotic detoxification, disease resistance, and stomatal function
Cycloartenol synthase (CAS)	Catalyzes the reaction producing Cycloartenol and participates in biosynthesis of steroids

Table 2.3 B Primers used for PCR amplifications of CYP450; ABC transporter and CAS genes.

Target gene/ fragment	Primer name	Primer sequence (5'-3')
CYP450	prCYPf	GCTTGCATCTTCGCCGGTGC
	prCYPr	CTTGGGTGGGTGAAGCTTTGCC
ABC transporter	prABC1f	CCTCCGAAGAACCAAACCCAAG
	prABC1r	TGGGAGAGCGAGAGGGAGGG
CAS	prCASf	GCCCAGCATCGACGAGGGAC
	prCASr	ACGTAAATTGATCTGGGCCCGGT

CHAPTER 3

Characterization of Polyketide synthase Genes (PKSs) in Red Raspberry cv. Glen Moy (BAC analysis)

3.1 INTRODUCTION

Because of significant similarity both in nucleotide and amino acid sequences, the relationship between PKS/CHS gene sequences and enzyme activities have not yet been clarified completely. In this chapter, the PKS genes (in the three BACs derived from red raspberry genome) involved in the biosynthetic pathway of anthocyanins will be initially investigated for similarity of both nucleotide and amino acid sequences.

3.1.1 Polyketide Synthase Genes in Raspberry and Related Fruits

Despite distinction in substrates and reaction mechanisms of plant-specific PKSs, the diverse PKS sequences revealed highly similar sequence regions. In a *Rubus* genus, PKS gene consists of at least eleven members and ten PKS genes (*Ripks1-10*) sharing 82–98% nucleotide sequence identity were identified in the *Rubus idaeus* genome and submitted in NCBI (Kumar and Ellis, 2003). Although some genes are composed of only short amino acid sequences, five PKS genes with complete sequences are identical to different genes/enzymes (Table 1.3). Functional investigations suggested that RiPKS1 is a chalcone synthase, RiPKS2 is of unknown function and RiPKS3 is identified as *p*-coumaryltriacetic acid synthase (CTAS). RiPKS4 is benzalacetone synthase (BAS) and RiPKS5 is another chalcone synthase synthesizing naringenin chalcone exclusively.

Zheng (2001) identified the three PKS genes (RiPKS1, RiPKS2 and RiPKS3) in red raspberry R. *idaeus* cv. Royalty. Subsequently RiPKS4 and RiPKS5 were identified (2009). The three RiPKS (PKS1, 2 and 3) coding full length of 391 aa share 91–99% amino acid sequence identity and they also share considerable sequence homology with other related fruits in the same family with 94–95% amino acid sequence identity to strawberry CHS and at 86–90% with apple CHS. In genomic DNA of raspberry cv. Glen Moy and Latham, RiPKS1, RiPKS2 and RiPKS3 genes were detected with 391 amino acid residues as characterization showed a high degree of relatedness with cv. Royalty (Kassim, 2009). Moreover, the RiPKSs also share major similarity with plant families such as *Arabidopsis* and other plants with lesser portion of CHS type PKSs (~78–87% amino acid sequence identity) including with other non-CHS type PKSs.

3.1.2 Bacterial Artificial Chromosomes

To study the genome of organisms, constructing of a library of genomic material is a necessary first step. BAC (bacterial artificial chromosomes) has become an essential tool for genomic study with the stability of insert (approximately 140-300 kb), their ease of manipulation and propagation compared to other vectors. They can provide maximal information and utility for further understanding the sequences of genes of interest, organization and arrangement of genes in a chromosome. This also leads to the further construction of a genome-wide physical map, positional cloning and then genome sequencing, particularly for higher eukaryotes (Hamiton *et al*., 1996; Korban *et al*., 2002). Construction of BAC libraries have become an essential strategy creating the powerful resources of genomics in various organisms including plants such as rice (Ammiraju *et al.,* 2000); *Brassica rapa* L. (Mun *et al*., 2008); barley (Schulte *et al*., 2011); apple (Han *et al*., 2007 and 2011); strawberry (Bonet *et al*., 2009); citrus (Terol *et al*., 2008); tomato (Datema *et al*., 2008); papaya (Ming *et al*., 2001).

3.1.3 The Aims of the Chapter

The aim of this chapter was to investigate the PKS gene family in the three positive BACs from red raspberry (*Rubus idaeus*) cv. Glen Moy genome library in order to identify PKS/CHS genes that may be involved in the production of anthocyanins, being important antioxidants and flavour compounds.

3.2 RESULTS

To construct significantly useful tools for genome studies in red raspberry (*Rubus idaeus*) cv. Glen Moy having a genome size (275 Mbp), the first genomic DNA library, a BAC library was constructed with average inserted size of approximately 130 kbs (Hein *et al*., 2005). The genomic DNA fragments were cloned and screened (Hein pers. com.) with a chalcone synthase 11 (CHS11) probe (Kumar and Ellis, 2001), resulting in the thirty BAC clones acted positive signals. These clones were screened by PCR using primer set for PKS1, PKS2 or PKS3 designed from the cv.

Royalty genomic sequence and subsequently the three positive BAC clones were analysed (31B12; 29M05 and 24P12) (Kasim, 2009).

3.2.1 BAC DNA Restriction Digestion

In this chapter, an initial analysis of the insert DNA in the three PKS-positive BACs from Glen Moy (31B12; 29M05 and 24P12) was carried out by restriction digestion with several restriction enzymes and electrophoresis. DNA fragment patterns revealed both similarities and differences (Figure 3.1). The two BACs (29M05 and 24P12) showed many shared fragments with *Hind*III; *Xba*I and *Not*I, indicating the possibility of overlapping clones as a different pattern was observed between BAC31B12 and the other two BACs

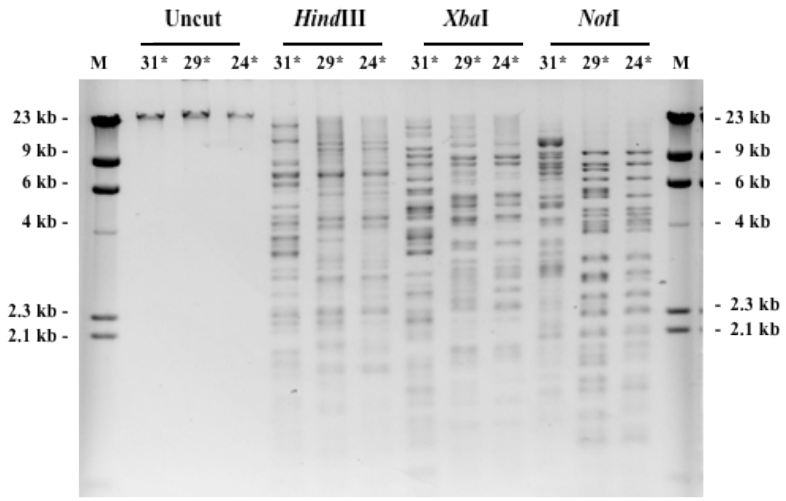


Figure 3.1 Restriction fragment patterns of the Glen Moy BAC clones (31*, 31B12, 25*, 25M05 and 24*, 24P12) by *Hind*III, *Xba*I and *Not*I separated by electrophoresis in an 0.8% agarose gel, compared with 23 Kb λ*Hind*III Marker.

3.2.2 Amplification of the PKS Genes

To subclone regions containing the Glen Moy PKS genes in the raspberry genome, each BAC was amplified by PCR using PKS primers (PKS1; PKS2 and PKS3) derived from Red raspberry (*Rubus idaeus*) cv. Royalty, GeneBank by Kassim (2009) (Table 2.1). The set of PKS primers was designed from conserved regions of known aromatic polyketide synthase, sharing high sequence homology of PKS gene family in various plants. The reactions were amplified at an annealing temperature of 56°C for 30 seconds. Prediction of sizes of expected amplified products using PKS gene primers and Accession no. of their originals was shown in Table 3.1 as all PKS genes contain an intron and two exons (the first one has a length of ~180-200 bp and the second one ~1,000 bp), the expected fragments were in the second exon. For instance, figure 3.2 showed prediction of nucleotide sequences of PKS2 with ~700 bp in the second exon (from at ~ 520 to1250 bp of ~1611 bp in total).

PCR products were successfully amplified from all three BACs, (Figure 3.3). The results indicated that the PCR products for PKS1 and PKS2 primers had similar sizes (approximately 800 bp), while the PKS3 products were slightly larger (approximately 1 kb). To directly analyze the PKS-amplified fragments, they were subcloned and screened by PCR using the universal primers (M13 Forward (-20) or M13 Reverse) in combination with each PKS primer as well as digested with *Eco*RI in parallel. Nine subclones derived from PCR products used PKS1; 2 and 3 primers from all three BACs were prepared for sequencing. The PKS-amplified fragment was ~700 to 800 bp. The nucleotide sequences amplified from all PKS primers showed matching in the second exon. Successful PCR amplification in all BACs suggested the possibility of the presence of a PKS gene family in the BACs. However, it seemed that the PKS-amplified fragments were a bit longer than the predicted one (Table 2.1 and 3.1), this might result from difference of base sequences between these two cultivars.

Table 3.1 Prediction of PCR amplifications using PKS genes.

Gene/ fragment	Accession no. of gene sequences used for primer design	Length of PKS gene sequences (bp)	Predicted size of PCR amplifications (bp)
PKS1	AF292367.1	2,432	~780 (at 1,169-1,971)
PKS2	AF292368.1	1,611	~700 (at 515-1,253)
PKS3	AF292369.1	2,461	~839 (at 1,200-2,040)

(Modified from: Kasim, 2009 and NCBI)

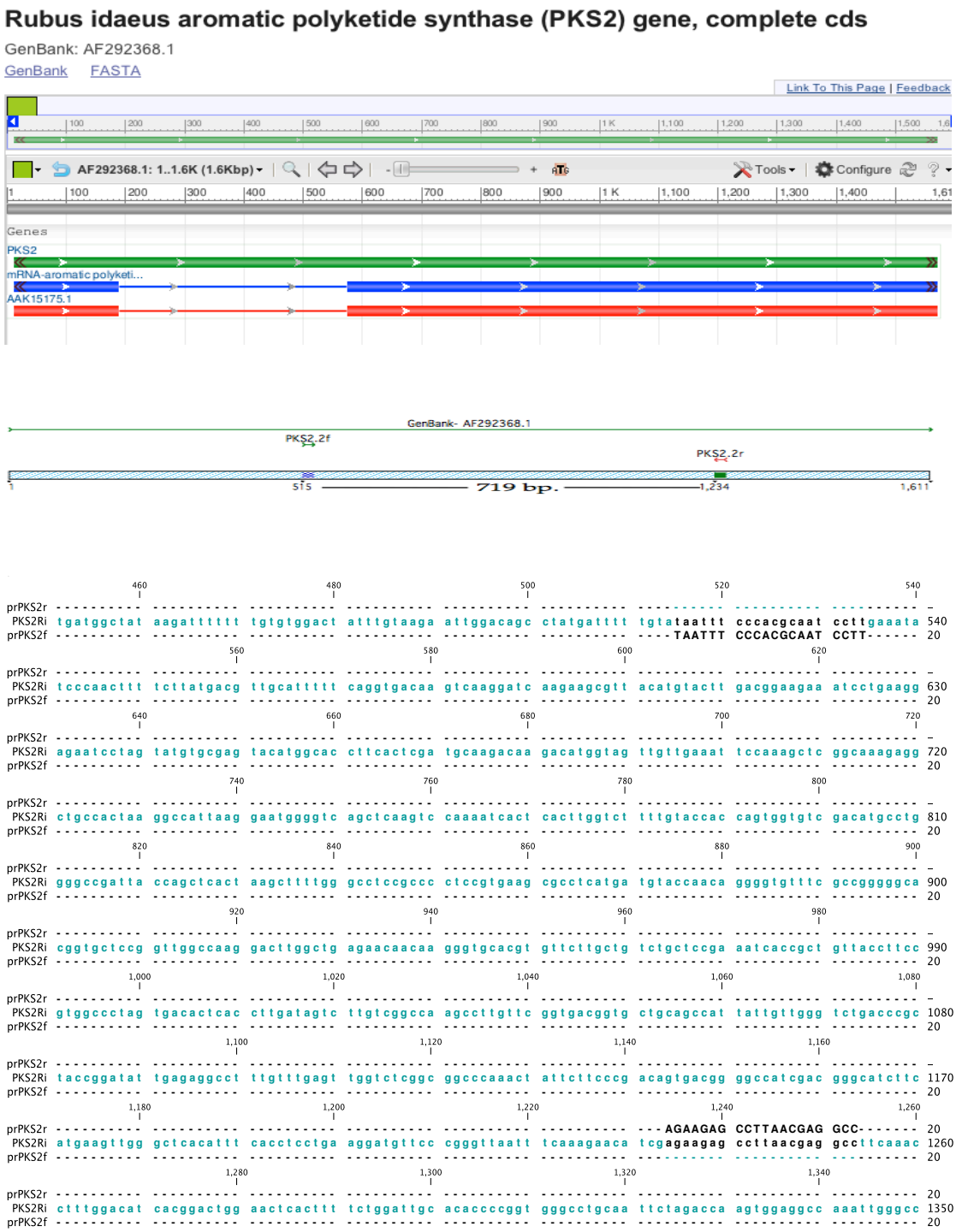


Figure 3.2 Prediction of PCR products amplified using PKS2 gene of red raspberry cv. Royalty (PKS2Ri, AF292368.1) and PKS2 primers (prPKS2f and prPKS2r). PKS2 gene contains ~1,611 bp with an intron and two exons. The amplified fragment was in the second exon of PKS2 with size ~700 bp (from at ~ 520 to1250 bp).

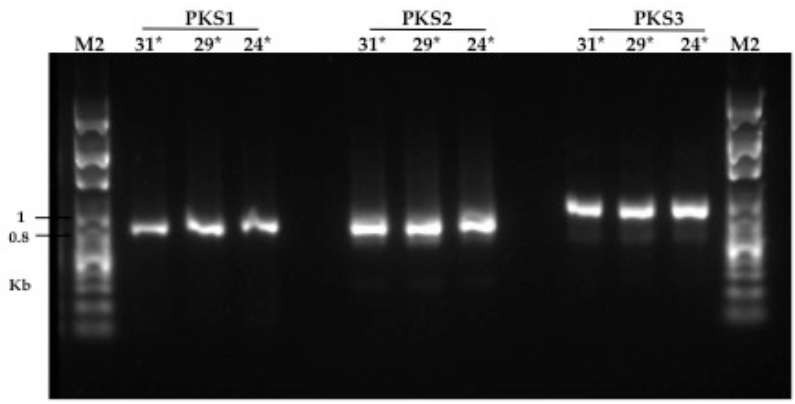


Figure 3.3 The amplified nucleotide sequences from all three pairs of PKS primers in BACs (31*, 31B12, 29*, 29M05 and 24*, 24P12) using PKS primers derived from Red raspberry (*R. idaeus*) cv. Royalty (Kasim, 2009). The amplified PKS1 and 2 sequences were ~800 bp and PKS3 was slightly larger (~1 kb). The 50 µl of the PCR reactions were denatured at 94°C for 5 minutes, then amplified following denaturing at 94°C for 1 minute, annealing at 56°C for 1 minute and extension at 72°C, using PKS1.2, PKS2.2 and PKS3.2 compared with 1 kb Molecular standard.

3.2.3 Characterization of the PKS Genes

To identify the partial PKS gene sequences of the three BAC, the amplified PCR products were sequenced using Sanger sequencing (Source Bioscience, UK). The nucleotide and deduced amino acid sequences analyses of subclones (sequence comparision; BLASTn; BLASTx; TBLASTx and BLASTp) identified them as PKS genes (Table 3.2 and 3.3). The majority of amplified fragments that were characterised using BLASTn showed 99% similarity to nucleotide sequence identity. They matched the PKS gene, complete cds from *R. idaeus* cv. Glen Moy (EU862821.1)(Kasim *et al*., 2008) and the PKS1 gene from *R. idaeus* cv. Royalty (Accession no. AF292367.1, NCBI). In addition, they also were identified as the gene sequences corresponding to PKS primers (Table 3.2).

To analyse using BLASTp, the majority of clones showed the same identification using BLASTn but less similarity, at 94-99% nucleotide sequence identity (Table 3.3). Using BLASTx, the clones also presented similarity 96-100% amino acid sequences identity of aromatic polyketide synthase from cv. Royalty (AA15174.1) and PKS3 (AA15176.1) (Zheng *et al*., 2001). They are also identified as CHS6 from cv. Meeker (AAM90652.1) that contains highly similar sequences with PKS1, 98-100% amino acid sequence identity (no result shown).

Following BLASTp, a few clones were identified as sequences of leucoanthocyanidin dioxygenase (LDOX) gene coding leucoanthocyanidin compound (colourless) at less than 40% identity of amino acid sequence (no result showed). Interestingly, all clone (amplified using different primers) shared 98-100% amino acid sequence (222 to 282 residues) to aromatic polyketide synthase (PKS1, AAK15174.1/Q9AU11.1); PKS2 (AF292368.1/Q9AU10.1) and PKS3 (AA15176.1/Q9AU09.1) from cv. Royalty as well as CHS6 (AAM90652.1) from cv. Meeker (Kumar and Ellis, 2003) (Table 3.3).

One clone (1B185 from BAC 29M05) amplified using PKS1 primer was identified as PKS5 gene sequences and other related genes. Following BLASTn, the clone showed 97-98% nucleotide sequence identity, corresponding to the PKS5 (EF694718.1) cv. Royalty (Zheng and Hrazdina, 2008); CHS5 sequence (AF400565.1) and CHS11 (AF400566.1) from cv. Meeker (Kumar and Ellis, 2003) (Table 3.2). The comparison of raw nucleotide sequences that matched with PKS5 using CLC sequencer viewer 6, is shown in Figure 3.4. Whist presenting 99% identity of both PKS1 (Q9AU11.1) and PKS3 (Q9AU09.1), together with 98% identical to PKS2 (Q9AU10.1) from cv. Royalty (Table 3.3), this indicated high similarity of both nucleotide and amino acid sequences between PKS and CHS genes in related fruits. However, this comparison is an initial analysis, it still needs further investigation.

Table 3.2 Identity of the nucleotide sequences of the PKS-amplified fragments of subclones derived from BACs using BLASTn

Primer/ Gene fragment	BAC	Clone	Raw nucleotide Identity (%)	Identical Gene (Number of nucleotides)	Accession no. of identical gene
PKS1	31*	1B171	811/812 (99%) 810/812 (99%) 808/812 (99%)	PKS (3101) PKS (2432) PKS3 (2461)	EU862821.1 AF292367.1* AF292369.1
	29*	1B185	778/797 (98%) 777/797 (97%) 774/797 (97%)	CHS5 (1482) PKS5 (2211) CHS11 (1507)	AF400565.1 EF694718.1 AF400566.1
	24*	1B191	811/813 (99%) 810/813 (99%) 808/813 (99%)	PKS (3101) PKS (2432) PKS3 (2461)	EU862821.1 AF292367.1* AF292369.1
PKS2	31*	2B172	735/739 (99%) 734/739 (99%) 731/739 (99%)	PKS (3101) PKS (2432) PKS2 (1611)	EU862821.1 AF292367.1* AF292368.1
	29*	2B182	735/739 (99%) 734/739 (99%) 731/739 (99%)	PKS (3101) PKS (2432) PKS2 (1611)	EU862821.1 AF292367.1* AF292368.1
	24*	2B192	857/860 (99%) 856/860 (99%) 854/860 (99%)	PKS (3101) PKS (2432) PKS3 (2461)	EU862821.1 AF292367.1* AF292369.1
PKS3	31*	3B173	854/861 (99%) 853/861 (99%) 851/861 (99%)	PKS (3101) PKS (2432) PKS3 (2461)	EU862821.1 AF292367.1* AF292369.1
	29*	3B183	859/860 (99%) 858/860 (99%) 856/860 (99%)	PKS (3101) PKS (2432) PKS3 (2461)	EU862821.1 AF292367.1* AF292369.1
	24*	3B193	736/739 (99%) 735/739 (99%) 732/739 (99%)	PKS (3101) PKS (2432) PKS3 (2461)	EU862821.1 AF292367.1* AF292369.1

(31*, BAC31B12; 29*, BAC29M05; 24*, BAC24P12 and AF292367.1*, PKS1 gene sequence from *R. idaeus* cv. Royalty)

Table 3.3 Identity the deduced amino acid sequences of the PKS-amplified fragments of subclones derived from BACs using BLASTp

Primer/ Gene fragment	BAC	Clone	Amino acid Identity (%)	Identical Gene (Number of amino acids)	Accession no. of identical gene
PKS1	31*	1B171	251/252 (99%) 250/252 (99%) 248/252 (98%)	PKS1 (391) PKS3 (391) PKS2 (391)	Q9AU11.1 Q9AU09.1 Q9AU10.1
	29*	1B185	251/252 (99%) 250/252 (99%) 248/252 (98%)	PKS1 (391) PKS3 (391) PKS2 (391)	Q9AU11.1 Q9AU09.1 Q9AU10.1
	24*	1B191	251/252 (99%) 250/252 (99%) 248/252 (98%)	PKS1 (391) PKS3 (391) PKS2 (391)	Q9AU11.1 Q9AU09.1 Q9AU10.1
PKS2	31*	2B172	226/226 (100%) 224/226 (99%) 223/226 (99%)	PKS1 (391) CHS6 (391) CHS (389)	Q9AU11.1 AAM90652.1 AEC13058.1
	29*	2B182	225/226 (99%) 223/226 (99%) 222/226 (98%)	PKS1 (391) CHS6 (391) CHS (389)	Q9AU11.1 AAM90652.1 AEC13058.1
	24*	2B192	249/251 (99%) 248/251 (99%) 245/251 (98%)	PKS1 (391) PKS3 (391) CHS6 (391)	Q9AU11.1 Q9AU09.1 AAM90652.1
PKS3	31*	3B173	260/271 (96%) 258/262 (98%) 256/271 (94%)	PKS1 (391) PKS3 (391) CHS6 (391)	Q9AU11.1 Q9AU09.1 AAM90652.1
	29*	3B183	284/285 (99%) 282/285 (99%) 280/285 (98%)	PKS1 (391) PKS3 (391) CHS6 (391)	Q9AU11.1 Q9AU09.1 AAM90652.1
	24*	3B193	225/226 (99%) 223/226 (99%) 222/226 (98%)	PKS1 (391) CHS6 (391) CHS (389)	Q9AU11.1 AAM90652.1 AEC13058.1

(31*, BAC31B12; 29*, BAC29M05; 24* and BAC24P12)

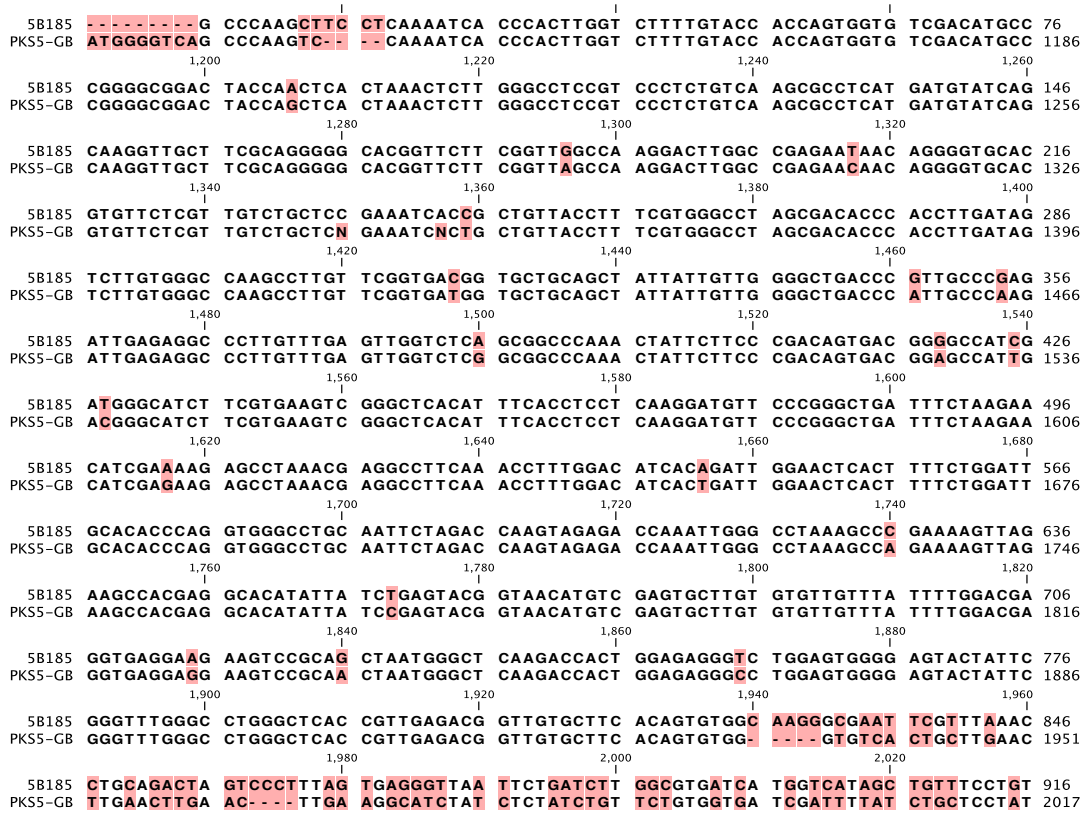


Figure 3.4 Comparison of raw nucleotide sequences of 1B185 subclone (5B185) amplified from BAC29M05 using PKS1 primer compared with PKS5 from GenBank (EF694718.1) using CLC sequencer viewer 6. Differences from each other are shown in pink shading.

3.2.4 Detection of PKS Genes by Southern Blotting Analysis

To detect the PKS genes in the three BAC clones, southern blotting analysis was carried out by digesting with *Eco*RI, *Hind*III and double enzymes and hybridized at high stringency. The DIG-labelled probes were generated from amplified products using primers to designed PKS1 from Glen Moy DNA, approximately 800 bp and leave analyzed at high stringency. The Southern blotting showed similarity of hybridizing bands of BAC29M05 and 24P12 while there was no signal obtained from BAC31B12 (Figure 3.5). This indicated the existence of genes with similar sequence to PKS1 in both BAC clones. The probe revealed the two fragments differing in size, approximately 7 and 11 kb with *Hind*III and 2 and 4 kb with *Eco*RI, respectively whist showing multiple fragments with double enzymes.

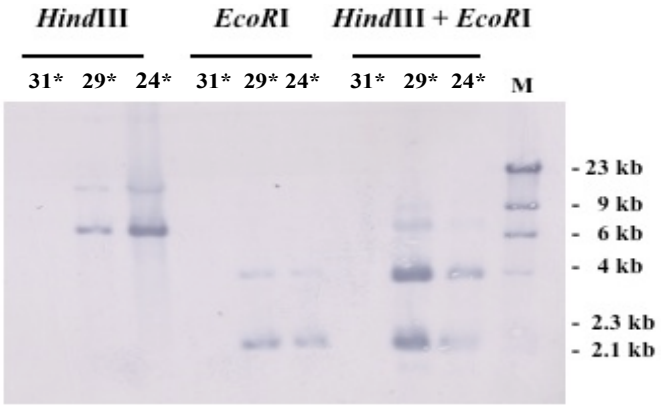


Figure 3.5 Southern blotting analysis of BAC clones (31*, 31B12, 29*, 29M05 and 24*, 24P12) using DIG-labelled PKS1 probe. Each BAC clone was digested with *Hind*III, *Eco*RI and double enzyme. The hybridization mixture to membrane blotted DNA were incubated at 42°C for overnight, then washed twice in 2XSSC, 0.1% SDS at 15°C for 5 minutes, along with stringency washed twice in 0.5XSSC, 0.1% SDS at 65°C for 15 minutes. Labelled molecular standards are 23 kb DNA ladder (ROCHE) indicated on the right side.

Zhenge (2001) reported that Southern blotting hybridized signals of two out of three PKS genes (PKS1, PKS2 or PKS3) in raspberry genomic DNA were 2-3 kb fragments with the *Eco*RI digestion. As the representative restriction map of two PKS genes (PKS5 and PKS11) of *Rubus* containing the both start and stop codon presented a 2 kb fragment of the digestion with these double enzymes (*Eco*RI and *Hind*III) (Kumar and Ellis, 2003). Therefore this indicated that there should be sequences of these PKS genes in the two BAC clones from raspberry cv. Glen Moy in this research.

3.3 DISCUSSION

Comparison of the PKS sequences in this chapter confirmed that the PKS genes family is highly similar both at nucleotide and amino acid sequences. However the slight differences between some nucleotide sequences can lead to protein-coding region exchanges and result in designating the diverse PKS genes. In this chapter, a set of PKS primers were designed from conserved regions of known aromatic PKSs from red raspberry cv. Royalty. Through sequence homology with known PKS genes, a plant-specific PKS gene presence was indicated.

Based on different results of restriction digestion and Southern blotting hybridization, analysis of BAC31B12 showed positive results in PCR amplification using PKS primer, together with identities of its sub-clone sequences as PKS gene sequences, but its restriction digestion fragments revealed a difference from those of the other two BACs, including absence of hybridized signals. However, in the absence of negative controls in PCR amplification (Figure 3.3), there is a possibility in that there are some false-positive amplifications of the BAC31B12, resulting from contamination with template in the reaction.

As PKS1 and PKS5 genes cod typical chalcone synthases and Naringenin-chalcone synthase (CHS), PKS3 is a *p*-coumarate triacetic acid synthase (CTAS) but PKS2 is inactive naringenin-chalcone synthase. The initial analysis of BAC DNA demonstrated the presence of a cluster of PKS genes in raspberry cv. Glen Moy genome, showing a set of PKS genes (PKS1; PKS2; PKS3 and PKS5). According to identification as CHS

genes (CHS5; CHS6 and CHS11), these CHSs are in the gene cluster catalysing the reactions of anthocyanin production at the first step of flavonoid synthetic pathway. The sequences of PKS genes share a higher percent sequence identity with CHS-type PKS (82-91% amino acid sequence identity), the PKS sequences in the BACs were then identified as both subfamilies of PKS and CHS genes. Base on similarity of DNA, amino acid sequences and also functional activities among PKS genes (PKS1; CHS5; CHS6; CHS11 and PKS5) (Zheng, 2001; 2009 and Kumar and Ellis, 2003), in this Chapter, it was a possibility of the existence of the set of several genes involved in the anthocyanin biosynthetic pathway in these BACs.

Identification of the sequence of the clone amplified using PKS1 primer as PKS5 gene confirmed high similarity between both genes whilst both of them catalysing the same pathway of the first reaction of flavonoid synthesises pathways, leading to anthocyanin biosynthesis. It indicated the possibility in presence both PKS1 and PKS5 genes in these BACs. The coding sequences of the PKS genes (PKS1; 2; 3 and 5) consists of ~1173 bp in size, containing an intron and two exons that codes 391 amino acids of different proteins. Although absolutely identified PCR sequences were approximately 800 bp, (being too short while PKS gene is ~1200 bp or more), there were difficulties in the identification of genes from this DNA region. Therefore an identification of whole gene region is necessary for further study.

3.4 CONCLUSION

General characteristics of the three BACs from red raspberry cv. Glen Moy genomic library were reviewed in this Chapter, as it demonstrated the similarity between the two BAC clones (29M05 and 24P12). Although BAC31B12 provided positive results with PKS primers, there were differences between this BAC and the other two BACs. All these BAC clones contained several PKS/CHS gene sequences that might involve in the anthocyanin biosynthetic path way as they are important sources that introduces to further study focusing on the PKS genes in the red raspberry genome. However, as PKS/CHS genes share considerably similarity, additional information is needed to investigate the actual sequences in these three BACs.

CHAPTER 4

Construction of a Fosmid Sub-library of BAC-derived Red Raspberry *R. idaeus* cv. Glen Moy DNA

4.1 INTRODUCTION

The Bacterial Artificial Chromosomes (BAC) clones from red raspberry *R. idaeus* cv. Glen Moy containing the PKS gene sequences (PKS1; CHS6 and PKS5; CHS5 and CHS11) described in Chapter 3 have provided some useful information. However the BAC clones contained large sized insert of DNA (approximately 130 kb), they were not appropriate for further studies aimed at manipulating the DNA and detail analysis. In addition, the complete genome sequences of the BAC clones have not been determined. Therefore the construction of a sub-library containing more manageable sizes of BAC DNA was needed. This chapter focuses on the construction of a sub-library in fosmids as well as the screening for PKS genes. The fosmid library prepared will represent being a new source for further studying for the PKS genes in raspberry.

4.1.1 Genomic DNA Libraries

DNA libraries are collections of clones containing all the DNA fragments from one source or a single organism of interest (Pierce, 2002). Libraries are different in terms of the source of DNA from which they are made, depending on the objectives for which the library will be used. Generally, there are two main types of library: those made from genomic DNA, a single chromosome (chromosome library) or the entire genome (genomic library); or those made from complementary DNA (cDNA library) that is derived from mRNA (Klung and Cumming, 2003). Ideally, the library contains at least one copy of every DNA sequence in the genome, chromosome or BAC clone of interest. The library needs to contain a sufficient number of clones to ensure that all DNA sequences in the source DNA are represented in the library. Genomic libraries with large DNA inserts are essential for complete genome sequencing, physical mapping, and positional gene cloning including analysis of gene structure as well as function. As sub-libraries with smaller DNA inserts are used for smaller genomes and single chromosome libraries. In this Chapter the source of DNA is the genomic DNA in BAC clones derived from plant genome, so can be considered as a genomic DNA library. As next generation sequencing technologies are revolutionizing the field of genomics, traditional genomic DNA libraries are still required for correct genome

assembly and can be extended to functional studies for understanding DNA regulatory elements and other applications.

4.1.2 Fundamentals of DNA Library Construction

4.1.2.1 DNA Fragment Preparation

To construct the DNA libraries, after isolating the genomic DNA of interest from an organism or chromosome, there are three methods that can be used to prepare the smaller DNA fragments.

(1) Digestion with restriction endonucleases. Although this is a common method, there is a disadvantage in that if the gene of interest contains one or more restriction sites for the enzymes used, the gene will be split in to two or more fragments and be separately cloned resulting in difficult in identifying the complete sequence.

(2) Partial digestion. This method is based upon a limitation in the amount of the enzyme used or the time of digestion. Being slightly different from common digestion, this means that only a proportion of the available restriction sites will actually be cut with the enzyme, resulting in a population of overlapping DNA fragments.

(3) Mechanical shearing. Unlike directly using restriction enzymes, DNA fragments can be generated by passing DNA through a syringe needle to make a population. However, it is necessary to add an enzymatic manipulation step to the construct the appropriate ends of the DNA fragments to allow the molecule to be inserted into the cloning vector. Theoretically, partial digestion and mechanical shearing provide a set of overlapping genomic fragments.

4.1.2.2 Calculating the Number of Clones

Based on the theory of library construction, a few clones of an overlapping library will contain the entire gene of interest and some will contain a part of the gene, but most contain fragments that have no part of the gene of interest. Moreover, there is chance of losing some portions of DNA during DNA preparation. To generate a complete

DNA library, the number of clones that are required to represent an entire sequences in the genome depends on the size of the starting genome material being fragmented, and the average size of the DNA fragments inserted into the vector used (Russell, 2006).

Theoretically, the probability of recovering a particular gene in the library can be calculated by the formula:

$N=\ln(1-P) / \ln(1-f)$.

Where P is the desired probability, f is the proportion of the starting material contained in a single clone, and N is the required number of clones to be screened.

4.1.2.3 Vectors for Library Construction

A vector is the means of carrying the fragments for cloning and library construction. Since the genome size varies among various organisms, the cloning vector needs to be selected to allow a realistic number of clones to be screened. Basically, a vector type with large insert capacity should be chosen for large genomes so that the number of clones is minimized. The developments of vectors and capacity of insert-size DNA well support for library construction of different insert-size DNA from large (with ~2000 kb) to small fragments (with ~10-15 kb); e.g. Human Artificial Chromosomes (HAC); Yeast Artificial chromosomes (YAC); Bacterial Artificial Chromosomes (BAC); P1 Artificial Chromosome (P1AC); cosmid; fosmid systems; Transformation-competent Artificial Chromosome (TAC), together with plasmids. While differing in host cells used for cloning the DNA such as yeast or bacteria etc, using these systemic vectors also bases on the maintainability the inserted DNA in host cells. The capacities of insert-size DNA of vector systems are summarised in Table 4.1. Whilst a library of large genomic fragments, approximately 50 to >1500 kb, can be constructed using YAC or BAC at the beginning of any genome sequencing project, the smaller DNA fragments can be inserted into fosmids or cosmids at the subsequent step.

Table 4.1 Vector systems and capacity of insert-size DNA for library construction

Vector Type	Capacity (thousands of bases, kb)
Human Artificial chromosome (HAC)	2000
Yeast artificial chromosomes (YAC)	300 to >1500
Bacterial artificial chromosomes (BAC)	50 to 300
P1 Artificial Chromosome (P1AC)	125-150
Cosmids and Fosmids	35 to 45
Transformation-competent Artificial Chromosome (TAC)	< 30
Phages (lambda)	25
Plasmids	15

4.1.3. Use of Fosmid Libraries

Fosmids have an advantage over the YAC or BAC as well as cosmid systems in terms of capability of stably propagating DNA fragments, which makes it ideal for complex genomes, and is faster, easier, and much cheaper than other system such as the BAC library system (Kim, 1992 and Tomaso and Weissman, 2003). The vector used in fosmid system is based on the bacterial F-plasmid origin of replication and partitioning mechanism, thus allowing cloning of large insert DNA fragments whilst containing only one copy of insert DNA per host cell (EPICENTRE, biotechnologies). This system provides low copy number resulting in higher stability. Fosmids are efficient in maintaining a complex genome over 100 generations (Kim *et al*., 1992).

On the other hand, although the primary disadvantage of fosmid libraries is insert size limitation to about 35-45 kp, fosmid clones contain more information than the smaller library clones constructed using the other vectors e.g. plasmid. However, fosmid libraries demonstrate their utility for applications such as being sub-library of BACs to fill the gabs between BACs or complete whole sequence, and performing to isolate functional elements within the genome and structural variation studies. These have subsidized the success of sequencing many genomes for instance plants (Cheng *et al*.,

2005 and Meyer *et al*., 2008), animals (Humphray *et al.,* 2007 and Kim *et al.,* 2003) and humans since 2000 (Shizuya *et al*., 1992).

4.1.4 The Aims of the Chapter

The aim of the work described in this chapter are to construct three fosmid libraries using the genomic DNA present in each of the three BACs identified in Chapter 3. The main stages are:

- Preparation of fragmented DNA from each of the BACs
- Ligating the DNA fragments into the fosmid, packaging and transfecting into bacteria
- Screening the fosmid library for the presence of PKS genes by hybridization, PCR and Southern blotting

The clones can then be analyzed in further detail.

4.2 RESULTS

To support further understanding of the organization of Aromatic Polyketide Synthase (PKS) in red raspberry, in this Chapter the three randomly sheared fosmid libraries (FrB31; FrB29 and FrB24) were separately constructed from the three BACs (31P12; 29M05 and 24P12) of red raspberry *R. idaeus* cv. Glen Moy, as described in Materials and Methods, along with fosmid clones containing PKS gene were screened.

4.2.1 Fosmid Library Construction and Characterization

To prepare the DNA for the fosmid libraries, the first step was to fragment the BAC DNA into size approximately 36-40 kb in length. Samples were analyzed by agarose gel electrophoresis, showing that the sheared DNA ranged in size from ~23 to 50 kb (Figure 4.1). Size-fractionated DNA of 36 kb was excised from agarose gel and then used to prepare the library, described in detail in Chapter 2. Following construction of the fosmid library, 45 randomly selected fosmid clones were isolated and analyzed. This showed that the majority of clones contained inserts between 32 to 39 kb with an

average insert size 36 kb, meaning the shearing and size-fractionation had been successful.

Fosmid library 1 (FrB31, derived from BAC31B12) was composed of 648 clones and had a total insert size of 23.33 Mb. Fosmid library 2 (FrB29 from BAC29M05) consisted of estimated 333 clones with an entire insert size of 11.99 Mb and fosmid library 3 (FrB24) contained 429 clones with an insert size of 15 Mb. Following these results, all three libraries consisted of 1,410 clones as the entire-length sequence of these three libraries was equal to 50.76 Mb (1,410×36 kb), corresponding to 0.18 genome equivalents, assuming a *R. idaeus* genome size of 257 Mb.

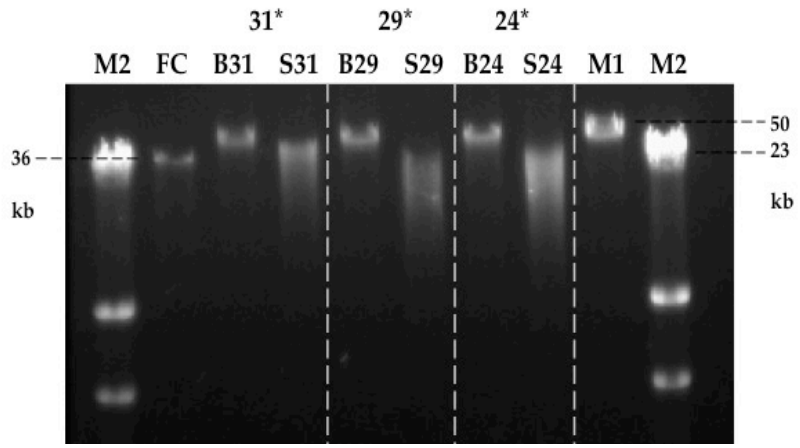


Figure 4.1 Size of sheared DNA fragments from the three BAC (31*, 31B12; 29*, 29M05; 24*, 24P12) used to construct the randomly sheared fosmid libraries of red raspberry *R. idaeus* cv. Glen Moy, compared with the 36 Kb fosmid control DNA, M1 and M2 (molecular weight marker 50 and 23 kb) by 0.8% agarose gel electrophoresis (B, BAC DNA; S, sheared DNA fragments).

4.2.2 Fosmid Stability

To analyze stability of the fosmid libraries, eighteen randomly selected clones were grown for serial cultures of approximately 100 generations covering a period of 6 days and then restriction digestion patterns between the original and final generation were compared. Following electrophoresis, the restricted patterns using *Eco*RI of the both generation were similar in all selected clones (Figure 4.2). After long-term culturing it was demonstrated that *R. idaeus* cv. Glen Moy DNA in the fosmid system is stable during propagations. Therefore this new library could be used in raspberry genomic studies continuously.

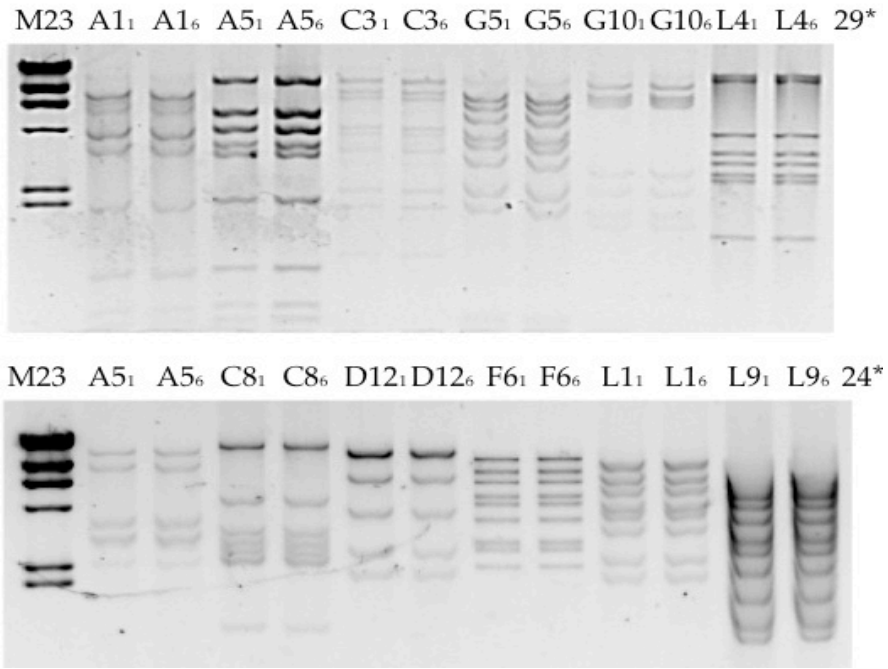


Figure 4.2 Stability assays on serial cultures for 100 propagations of the randomly selected clones from the library FrB29 (29*) and FrB24 (24*). Six clones collected on day 1 (1, the first generation) and day 6 (6, the one hundred generation) was digested with *Eco*RI, then electrophoresis on 0.8 % agarose gel and compared with 23 kb DNA marker on the left.

4.2.3 Fosmid Library Screening

Using the formula N=ln (1−P)/ln (1−f) where f is the original proportion contained in a single BAC clone (~130 kb). With average insert size of approximately 36 kb, P=36. For the desired probability at 99%, the number of clones to be screened (N) was calculated as 130 clones. To identify those fosmid clones containing the PKS genes, ~150 clones from each sub-library were randomly selected and grown for screening to confirm containing PKS genes.

4.2.3.1 Colony Hybridization Screening

Colony hybridization screening aimed to primarily select the clones containing PKS genes, which are derived from each BAC. In total, 450 clones from the all three sub-libraries (150 clones each BAC), were screened for PKS gene regions by hybridization (described in Chapter 2, Table 2.2) using nine PKS probes separately which were generated from PCR fragments amplified from each BAC using PKS primers set (from Chapter 3).

In summary, using the PKS gene probes, generated from all three BACs, sixty clones hybridized with the different PKS gene probes, giving a positive rate of 13.3% (Table 4.1). Positive clones were identified in all libraries, responding to the probes used. The number of clones per probe ranged from two to eighteen (6.66 on average). However, some of the probes did not hybridize any clone in the libraries (Table 4.2), such as the PKS1 probe generated from the original BAC31B12 (PKS1p31) and BAC29M05 (PKS1p29) as well as PKS3 probe from BAC24P12 (PKS3p24). This hybridization screening revealed the proportion of clones expected containing PKS genes regions in each library. The PKS1 probe (PKS1p24), generated from the BAC 24P12, identified in total thirty-two fosmid clones (Figure 4.3). As all three PKS2 probes hybridized to a total of twenty-four clones and only four clones hybridized the two PKS3 probes (PKS3p31 and PKS3p29). No figure shown for hybridization using PKS2 and PKS3 probes. However, it is possible that some of these hybridized clones could be artifacts, and then they need further screening in the next steps.

Table 4.2 Number of positive clones screened by colony hybridization within fosmid libraries

Target gene/ fragment	**Original BAC**	**Probe**	**FrB31** (Number of positive clones)	**FrB29** (Number of positive clones)	**FrB24** (Number of positive clones)
PKS1	31B12 29M05 24P12	PKS1p31 PKS1p29 PKS1p24	0 0 6	0 0 8	0 0 18
PKS2	31B12 29M05 24P12	PKS2p31 PKS2p29 PKS2p24	0 0 0	0 4 0	4 6 10
PKS3	31B12 29M05 24P12	PKS3p31 PKS3p29 PKS3p24	0 2 0	2 0 0	0 0 0

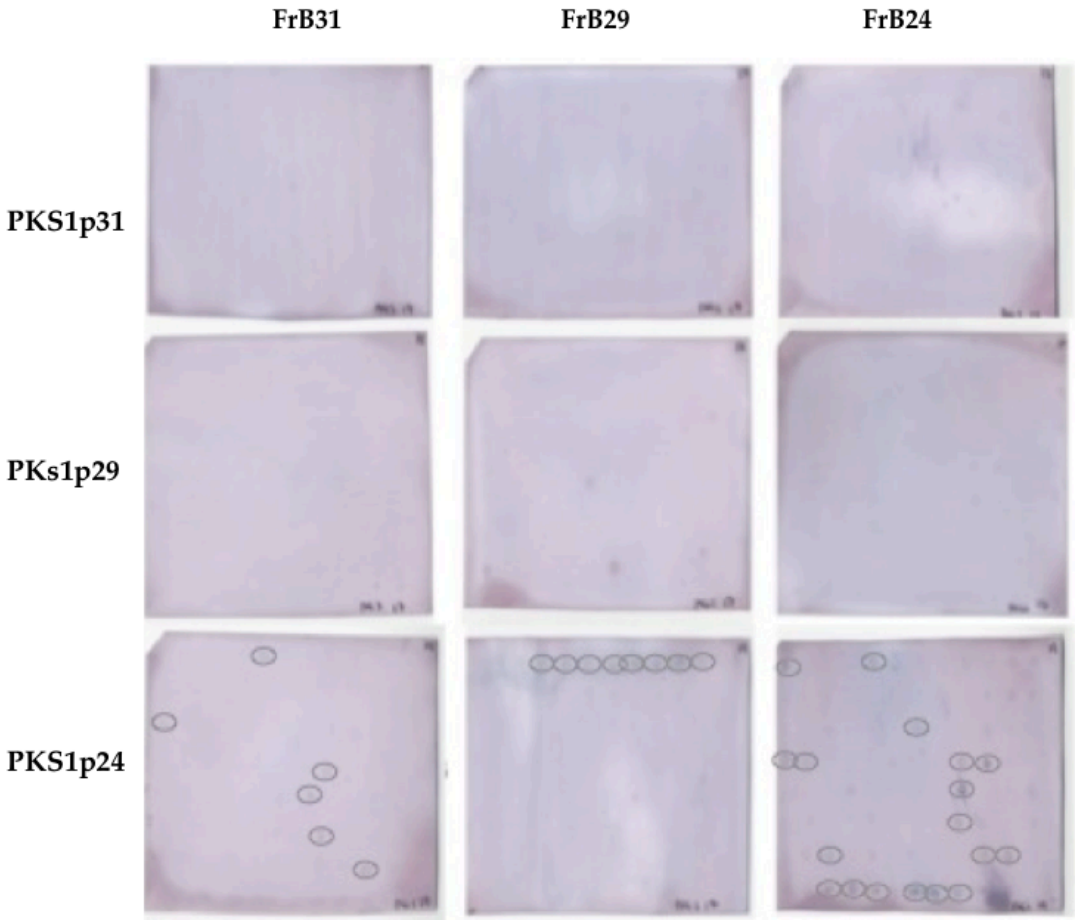


Figure 4.3 Colony hybridization screening of fosmid sublibraries (FrB31, FrB29 and FrB24) constructed separately from the three BACs (31B12, 25M05 and 24P12). Clones were screened using three DIG-labelled PKS1 probe (PKS1p31, PKS1p29 and PKS1p24), generated from amplified PCR products of the BACs (31B12; 25M05 and 24P12). The hybridization reaction were mixed with the membranes and incubated at 42°C for overnight, then washed twice in 2XSSC, 0.1% SDS at 15°C for 5 minutes, along with stringency washed twice in 0.5XSSC, 0.1% SDS at 65°C for 15 minutes. Only the probe PKS1p24 derived from the BAC24P12 can screen in total thirty-two fosmid clones of all three fosmid libraries.

4.2.3.2 Fosmid Library Screening by Polymerase Chain Reaction

In order to confirm the presence of a PKS gene in a positive clone identified by colony hybridization, the clones were subsequently screened by PCR using the sets of PKS primers previously used (Chapter 3). PCR screening yielded seven and ten positive clones from FrB29 and FrB24 respectively, presenting as similar sizes, approximately 700-800 bp (Figure 4.4). Notably, all amplicons were the same sizes as expected from the amplification in BACs (Table 3.1). Two libraries (FrB29 and FrB24) confirmed the presence of PKS genes although different primers amplified different clones while some clones were amplified with all three PKS primers. In total, seven out of fourteen positive clones of library FrB29 and ten clones of FrB24 presented the amplified fragments with this PKS primer set but without amplification in FrB31. Confirmation by PCR indicated the presence of PKS gene clusters in the fosmid clones.

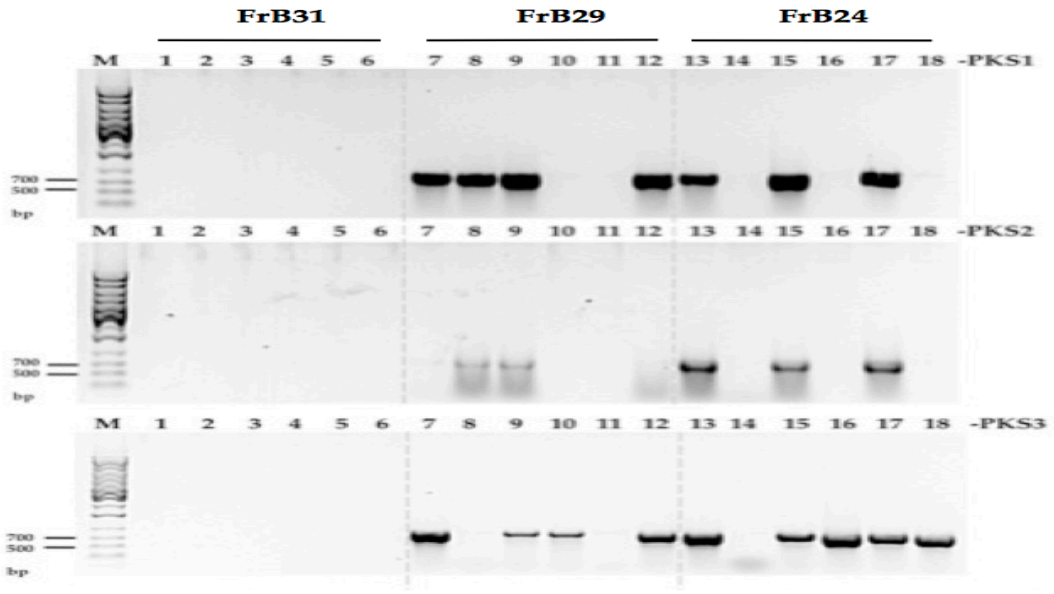


Figure 4.4 PCR screening of the three fosmid libraries (FrB31; FrB29 and FrB24) by a set of PKS primers (PKS1; 2 and 3), compared with 1 kb Molecular standard marker. All PCR products were as the same size (~700-800 bp) in the two libraries (FrB29 and FrB24) but without amplification in FrB31. The 20 µl PCR Reaction were denatured at 94°C for 5 minutes, then amplified following denaturing at 94°C for 1 minute, annealing at 56°C for 1 minute and extension at 72°C.

4.2.3.3 Fosmid Library Screening Southern Blot Analysis

In parallel with PCR screening, the positive clones from colony hybridization were analysed by Southern blotting. DNA was prepared from sixty clones, and digested with *Hind*III, then hybridized using a set of DIG-labelled PKS probes. The Southern blotting confirmed that several clones carried PKS genes (Figure 4.5). Three and six clones of library FrB29 and FrB24 revealed hybridization with the PKS1 probes but there was no positive reaction with FrB31 clones. The hybridized clones not only showed a similar pattern of multiple fragments but also showed the two fragments of approximately 7 and 11 kb in length previously revealed in the Southern blotting in Chapter 3. However, some clones could not hybridise with the probes that are used for previous colony hybridization screening, indicating they were false-positive clones. This might result from the insert DNA in the BAC was contaminated with chloroplast or mitochondrial DNA, thus containing only partial sequences the DNA. While fosmid library could have the contamination of mitochondrial or chloroplast DNA sequences with up to 1.5 % of the library (Ammiraju *et al*., 2005).

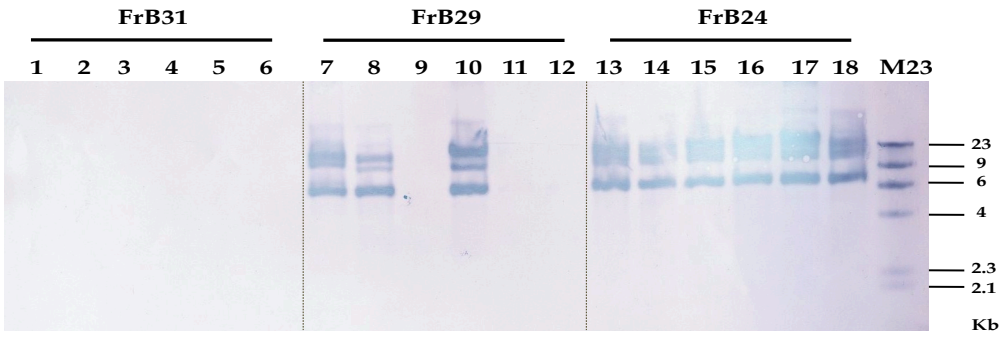


Figure 4.5 A Southern blotting analysis of the three fosmid libraries (FrB31; FrB29 and FrB24) using DIG-labelled PKS1 probe. Clones were digested with *Hind*III and hybridized with probe, then incubated at 42°C for overnight. Following, they were washed twice in 2XSSC, 0.1% SDS at 15°C for 5 minutes and stringency washed twice in 0.5XSSC, 0.1% SDS at 65°C for 15 minutes. Labelled molecular standards are 23 kb DNA ladder (Roach) indicated on the right side.

4.2.4 Fosmid End Sequencing and Sequence Analysis

To carry out a preliminary assessment of the raspberry DNA derived from the three BACs in the fosmid libraries, a total of 42 clones (at least 13-15 clones from each library) were randomly selected from all libraries for end-paired sequencing using the two specific primers for the fosmid vector described in Section 2.2.2.1 and appendix A.1.

Approximately 800-1,000 bp of sequence was obtained for each of the ends. Using BLASTn and BLASTx searches, it was possible to determine that the sequences in some clones were homologous with raspberry genome sequences. With a total of forty-two clones or eighty-four ends, four clones of the libraries (FrB29 and FrB24) were homologous with raspberry genome sequences but none of the end-sequence of clones from FrB31 was homologous. However, another forty-two clones were selected and end-sequenced for additional confirmation. In summary, four out of forty-two clones of the new series fosmid libraries were found to contain the expected regions of the PKS genes in these libraries. These four homologous ended-sequences presented a total length of 14.4 kb (4x36 kb) covering 0.005 % of red raspberry genome (275 Mb).

Like FrB29, the one clone of FrB24 was homologous with *R. idaeus* aromatic poly ketide synthase genes while another clone was analogous to TTG1 that was found in BAC25D10 of the same library derived from cv. Glen Moy and has been mapped on chromosome 1 in GeneBank by Graham in 2011 (NCBI). TTG1 gene (Transparent testa glabra 1) can control the physiological processes in plant both accumulating and not accumulating flavonoids (Walker *et al*., 1999). By associating with MYB and bHLH transcription factors, this complex protein impacts on anthocyanins production.

However, finding of the fosmid ended-sequences in this chapter still is a partial of the sequences (~800 bp of each end, out of ~36-40 kb in total length of each fosmid clone sequences), as it needs a longer sequence to be further identified. In addition, the other three clones of FrB24P12 library (from BAC24P12) were homologous with the two related plants in the same family: *R. glaucus* (blackberry) and *Rosa rugosa* (rose) as well as with at least thirteen clones of this library were analogue with different plants such as *Lupinus angustifolius* (lupin), *Zea mays* (maize), *Pinus pinaster* (pine) and

cellulose synthase genes sequence of *Vitis vinifera* (grape). In contrast, only one out of all end-sequences from FrB31 library was homologous with *R. glaucus* (blackberry). This may indicate difference of genetic information between BAC31B12 and the other two.

4.3 DISCUSSION

A fosmid library is a useful tool in genome analysis, giving the high cloning efficiency with relatively small manageable insert size as well as being high stability making them suitable for subsequent manipulations. Moreover, generation of the different DNA fragments by randomly mechanical shearing provided unbiased libraries. In this Chapter the fosmid sub-libraries were successfully constructed and characterized.

As fosmid libraries were screened by colony hybridization, PCR and southern blotting analysis respectively, screening clones of the libraries not only reduced the number of unwanted clones but also confirmed the true-positive clones containing PKS genes. Although colony hybridization firstly observed several positive colonies across the library, only some of those clones subsequently presented the PKS genes. While occurrence of false-positive signals in colony hybridization is difficult to screen the clones, it suggested a requirement of high stringency wash by increasing the duration (number/time) of washing step to remove the non-specifically bound DNA in false-positive clones.

Following Southern blotting, unwanted clones still were decreased. However, there was no difference in the hybridized fragment patterns using PKS1 probe, it is possible that these positive clones contained the same gene. In addition, the similar patterns may still refer to the same allele of the PKS1 gene, thus being necessary to carry out the additional analysis to confirm the actual sequence. However it is now known that nine positive-PKS clones carry PKS gene sequences and this amount is sufficient for further genomic study focusing on PKS genes in raspberry. In addition, the resulting data in Figure 4.4 (PCR screening) and 4.5 (Southern blotting analysis) increase the possibility that some of the data shown earlier in Figure 3.3 is a contamination artifact and the BAC31B12 does not contain any PKS like genes.

Although end sequences showed the presence of *R. idaeus* PKS gene sequences in the two libraries (FrB29 and FrB24), there was blackberry *Rubus glaucus* PKS gene sequence in FrB24 and FrB31, as well as sequences of Asian rose *Rosa rugosa* and different plants. This demonstrated similarity of some sequences not only in BACs derived from raspberry but also to other related plants both in the same and farther distant families such as lupin (*Lupinus angustifolius*), maize (*Zea mays*) and pine (*Pinus pinaster*). Since *R. idaeus* PKS gene sequences were found in the two sub-libraries, they can be selected to more specific study in the future. Additional, finding the sequence of *R. idaeus* transparent testa glabra1 (TTG1) gene coding WD40 protein that is involved in the regulation of the flavonoid pathway is useful information for further studying in gene regulation and expression.

In theory, the 14 end-paired sequences of each library could cover all the genomic DNA of individual BAC (14 clones x ~36 kb in total length of fosmid clones, formula in 4.1.2.2). However some of the end-pair sequences can be identified as BAC vector sequences (vector used to constructed BAC library, pT7blue-3, Merck, 3.8 kb) rather than raspberry sequences, with approximately 2.9% (3.8 kb out of 130 kb). Notably, almost all end-sequences of analyzed clones of FrB31 were homologous with BAC vector sequence while they may contain DNA sequences of PKS and other genes of raspberry genome receiving from the beginning of BAC library construction. Unlike the other two BACs, the FrB31 library derived from BAC31B12 has not provided any useful information for PKS genes in raspberry. However, some further analysis might need to search for the remaining information of the BAC.

4.4 CONCLUSION

The new fosmid sub-libraries that are constructed from BAC clones provided further useful information presenting the similarities between BAC24P12 and BAC29M05. In conjunction with the end-paired sequences from the libraries can assist the BACs with sequence completion, they therefore could be a new beneficial tool for studying in the entire BAC sequences and also support a better understanding and utilization of the raspberry genomic information in the future.

CHAPTER 5

Assembly of Bacterial Artificial Chromosome (BAC) sequences and Construction of Initial Physical Maps of PKS Gene Loci of Red Raspberry cv. Glen Moy

5.1 INTRODUCTION

The fosmid library constructed in Chapter 4 from the three BAC clones derived from red raspberry *Rubus idaeus* cv. Glen Moy, is not only more manageable size but also provides a readily available source of information for further study. Although the BAC clones are known to contain the PKS/CHS genes (Chapter 3), their entire sequences have not been completely analyzed yet. The fosmid libraries and BAC clones have been prepared from the same original part of red raspberry genome; consequently they will contain similar sequences. While current-sequencing technologies have been developed so far and succeeded at genome assemblies in various organisms, they can be used to assist the completion of sequence analysis for this part of the red raspberry genome. This Chapter describes the results of utilizing the sequence technologies to combine the genetic sources and construct physical map drafts derived from genome of raspberry cv. Glen Moy to present knowledge that could improve a further understanding of the PKS genes in raspberry genome.

5.1.1 Concept of DNA Sequencing

DNA sequencing is an essential molecular technique in determining the order of the four nucleotides (A, adenine; G, guanine; T, thymine and C, cytosine) in a strand of DNA, providing the significant genetic information contained in a DNA molecule. Depending on the different methods used, sequencing can reveal the sequence of a particular gene including intergenic sequences, introns and promoter sequences as well as larger genetic regions such as clusters of genes or operons and full chromosomes or entire genomes and leads to the precise location of gene sequences (Brown, 2002); identification of genes and regulatory elements or mutations and other interindividual variations (Harlt and Jones, 2006) and comparison of homologous genes in different species (Frazer *et al.,* 2003).

As first generation sequencing is still required, new innovative sequencing techniques become the second and the next generation with advances in rapidity, simplicity and reliability of the process of sequencing (Pareek *et al.,* 2011). By DNA sequencing methods, the *Arabidopsis thaliana* genome sequence was the first plant genome

successfully sequenced and published. Subsequently the rice (*Oryza sativa)*, papaya (*Carica papaya*) (Ming *et al.*, 2008) and maize (*Zea mays*) (Schnable *et al*., 2009) genomes have been sequenced while red raspberry genome sequencing is underway (Graham pers comm).

5.1.1.1 Sanger Sequencing

Sanger sequencing is one of the original methods of first generation DNA sequencing, which was designed by Fred Sanger in the 1970s (Klug and Cummings, 2003) and has been used for decades. The principle of Sanger sequencing technique is based on the random incorporation of dideoxynucleotides (ddNTPs) into newly synthesized DNA strands. The ddNTPs contain a hydrogen group on the 3' carbon instead of a hydroxyl group (OH) to prevent the further incorporation of nucleotides into a chain, while a phosphodiester bond cannot be formed. It also is referred to as "chain termination sequencing" or "dideoxy sequencing" (Fletcher *et al*., 2006). This technique produces sequences up to 700-1,000 bp in length of sequenced fragments. However, although Sanger sequencing presents slow sequencing speed and high costs (Claros *et al*., 2012), it still is in common use as its accuracy is good and best approach for sample checks of sequences e.g. a plasmid.

5.1.1.2 Second-generation Sequencing

Second-generation sequencing is a group of new sequencing technologies with advances of higher speed, simplicity and reliability of the process of sequencing to determine nucleotide sequences of large genes as well as an entire genome. Based on the concept of advanced sequencing techniques, most of them generate numbers of short sequence reads (~20 to 200 bp), meaning that these approaches need assembly algorithms for combining reads to create longer sequences, making them different from Sanger method (Pareek *et al.*, 2011). Second-generation sequencing techniques have been widely used nowadays for instance, the Genome Analyzer from Illumina (www.illumina.com) that yields ~4 Gb of bases per an hour (26-150 bp read length, average as 100 bp); the Genome Sequencer FLX+/454 from Roche (www.454.com) that produces over 100,000 reads of 400-800 bases per an hour run (~4-8 Gb); and the

Applied Biosystems SOLiD from Life bioscience (by Oligo Ligation and Detection) (www.appliedbiosystems.com) that produces 0.4-12.5 Gbp of short reads (up to 75 bp) per a running hour. These techniques have been reviewed (Shendure and Hanlee, 2008; Morozova and Marra, 2008 and Ba*, et al.,* 2011) as they have enables to the rapid and efficient development of genomic studies. Using these techniques genome sequences projects have progressed both in plants and animals

5.1.1.3 Next Generation Sequencing

Although both Sanger and other techniques of second-generation sequencing still are powerful methods, new sequencing technologies (or next generation sequencing) such as Heliscope™ Single Molecule Sequencer; Single-Molecule Real-Time (SMRT™) Sequencer and Ion Personal Genome Machine™ are becoming new available methods with higher capacity for producing longer sequence reads up to 1000 bp whilst taking shorter time and lower costs per instrument run (Mardis *et al.,* 2008 and Schadt *et al.,* 2010). Moreover, some techniques for instance Single-molecule real-time sequencing (Pacific Bio) can produce long sequence of 5,000 bp while the maximum read length can be up to ~22,000 bp.

5.1.2 Sequencing and Sequence Assembly

An assembly is combination of a hierarchical sequence collection that represents the reconstruction of the original from which sequences of DNA came. This strategy is an essential step to sequencing to consolidate the sequence results. In Figure 5.1, the assembly process is based on breaking down the original sequence into small fragments, then sequencing and reassembling the sequences by aligning and merging of multiple short sequences called reads into contigs which later built into scaffolds. The contigs provide a multiple sequence alignment of reads to give a consensus sequence. As scaffolds provide the order, orientation of contigs and also the sizes of the gaps between contigs. Eventually, if the scaffold contains adequate information, it would be assembled into full length of chromosome sequence of that organism.

Size and accuracy of contigs and scaffolds are characters that demonstrate the quality of the assembly. Size can be evaluated by statistical analysis of (1) maximum length; (2) average length; (3) combined total length and (4) N50 (length the shortest contig in the set of the largest contigs representing at least 50% of the total length of assembly) as measuring accuracy is difficult. However, combination and detailed comparison of resulting sequences that are derived from the same original part can make exactness. A key is the association of resulting long sequences from second-generation sequencing assisted by paired-end sequences and other resulting sequences by Sanger method to complete the entire sequences. Assembly approaches have been used in plant genome researches as at least one single sequence of ~80,000 known plant species are reported in GenBank (Claros, 2012) and genome sequences of most economically important crops are summarised in Hamilton and Buell (2012) and Proost (2011).

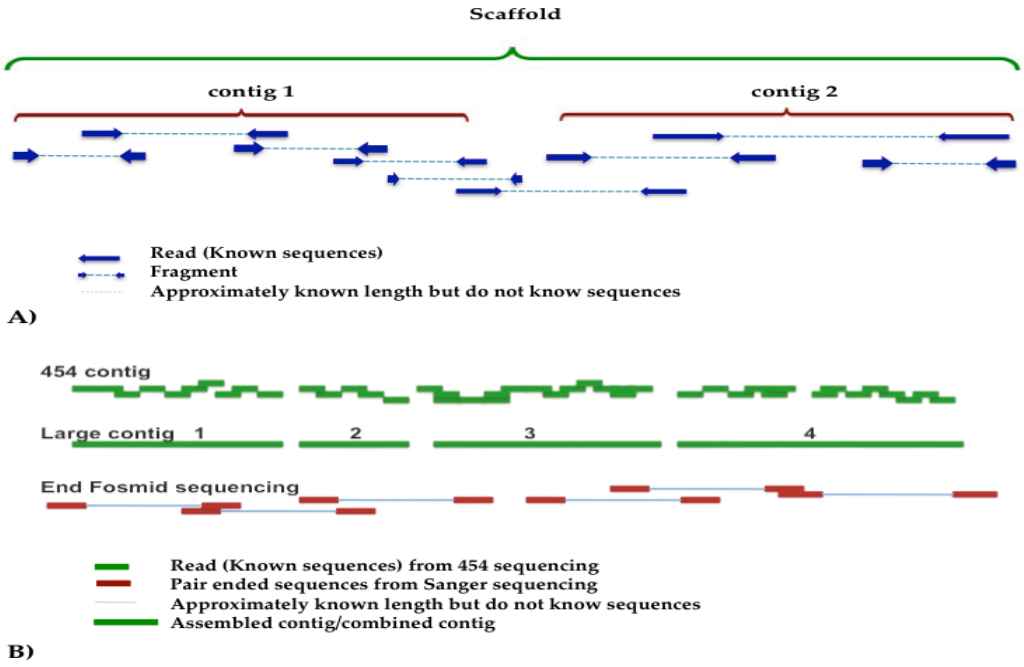


Figure 5.1 Scheme of an assembly concept. **A**, Reads of sequence combine into small fragments and reassembled into contigs, then assembled contigs can be linked together to create scaffolds. **B**, Association of resulting sequences from Sanger method and next generation sequencing (454 sequencing) to create long contigs and scaffolds.

5.1.3 Physical Map Construction

A physical map is a schematic diagram that represents the linear order of genes within a chromosome or genome (Harlt and Jones, 2006). In contrast, genetic linkage maps are based on the frequency of gene recombination events, and are presented in centiMorgans (cM) (Pierce, 2002) but physical maps are generated using the physical distances of genes, measured in nucleotide base pairs as Kilobases (Kb) to Gigabases (Gb) (Mayer *et al*., 2012). As genetic maps (linkage map) displaying the chromosomal locations of the genes along the individual chromosome (Fletcher *et al*., 2006), the physical map assists with the exactness at nucleotide sequence level (Russell, 2006). The maps can represent approximately hundreds Kb of chromosomal segments to an entire chromosome. Moreover, maps also represent the likelihood of coinheritance of genes as this depends on the distances between those genes (Harlt and Jones, 2006). Therefore an accurate physical map is essential in predicting coinheritance.

The main objective of physical mapping follows on from sequencing and leads to detailed molecular information supporting the improvements of traits of interest. In plants, physical maps have been constructed for *Arabidopsis thaliana* (Mozo *et al*., 1999), apple (Lauer and Seemulier, 2000 and Han *et. al*., 2007), peach (Ahmad *et al*., 2011), papaya (Yu *et al*., 2009), rice (Cheng *et al*., 2005), barley (Schulte *et al.,* 2011), soybean (Ha *et al*., 2012) and melon (González *et al*., 2010). Physical maps also have been generated for many species including human (Bentley *et al*., 2001) and other mammals (Lewin *et al.,* 2009), chicken (Ren *et al.,* 2003) and microorganisms (Wu *et al*., 2009).

The current genomic study focusing on the PKS genes within the two BACs derived from genome of cv. Glen Moy could support the previous information obtained for the PKS genes and certain other genes in the same region of chromosome. Construction of physical maps derived from these BACs is useful for obtaining accurate distances and chromosomal locations of genes, and provide available information for further studies and applications. In this Chapter, the construction of the physical maps is described.

5.1.3.1 Construction of Physical Maps

Before the advent of large-scale sequencing projects, physical maps were constructed using either marker-based mapping (Pourarenani *et al*., 2013) or a DNA restriction mapping (fingerprinting) approach (Wallis *et al*., 2004 and Fleury *et al*., 2010). Marker-based mapping methods such as STS-based mapping are laborious and use more than 5,000 markers (Philippe *et al*., 2013). Although DNA fingerprinting method, which performed by restriction mapping, has been replaced it, markers still have been mapped on physical framework to further marker utilizations. By DNA restriction mapping fingerprinting, fragments of DNA are digested by restriction endonucleases followed by agarose or capillary electrophoresis (Schulte *et al.*, 2006 and Meyers *et al.,* 2004). Based on this information, the restriction map is used to determine whether clones overlap. While DNA fingerprinting is developed so far using softwares to analyse clone overlaps, a current use is High Information Content Fingerprinting (HICF) based on restriction digestion and capillary electrophoresis, (Nelson *et al*., 2005 and Ariyadasa, 2011). Former physical maps can be created through YAC (Saji *et al*., 2001) or PAC (Amemiya *et al*., 2001) that were replaced by more suitable vectors. However, BAC-based physical mapping is the most common approach for restriction map-based fingerprinting in numerous projects (Cheng *et al*., 2005; Schulte *et al*., 2011 and Xu *et al.,* 2011) including using fosmid and smaller clone (~10 kb) to close gaps of physical maps generated (Yang *et al*., 2003).

5.1.3.2 Minimal Tiling Path and Validation

Minimal Tiling Path (MTP) refers to a minimum number of overlapping clones that carry on the entire length of the physical map (Green, 2001). Sequencing of all clones of the MTP validates the maps, thus confirming the clone overlaps and importantly detailed molecular information of the map constructed. Clone-by-clone MTP approach is commonly use to finish the process of genome sequencing in various genome projects for instance, BAC-by-BAC MTP in potato; chinese cabbage; maize; rice; barrel medic; trefoil (Feuillet *et al.,* 2011 and Fang *et al*, 2010). However, difficulties in detection of the overlapping depend on difference in size and number of clones such as BAC (130-200 kb); fosmid (36-40 kb) or smaller clones (~10 kb), and restriction

enzymes used (rare or frequent-cutter enzymes) that impact on size and number of fragments generated by those restriction digestion.

5.1.4 Physical Maps of Raspberry and Related Fruits

From the first genomic study of the European red raspberry, 'Glen Moy', a set of three genetic linkage maps of the red raspberry *Rubus idaeus* was constructed from crossing the European cv. Glen Moy and the North American cv. Latham (Glen Moy x Latham) (Graham *et al*., 2004). Several linkage groups (LGs) were mapped on chromosomes. Certain structural genes (F3'H; FLS; DFR; IFR; OMT and GST) and transcription factors (bZIP; bHLH; FRUITE4 and MYB), involved in determining anthocyanin content, colour and some fruit qualities such as fruit ripening, were mapped on LG1; 2; 3; 4 and 6 (Kassim *et al.,* 2009; McCallum, 2009 and Graham *et al*., 2006; unreported data). Importantly, Kassim (2009) mapped a PKS1 (CHS) gene on LG7, and genes involving in production of fifteen antioxidant compounds were mapped to LG4 (seven) and LG1 (eight) respectively (Kassim *et al*, 2008). The production of anthocyanins is a complex trait therefore construction of a physical map is the obvious step to access the targeted chromosome regions leading to critical role of those genes.

Today it is known that R. *idaeus* contains about 30,500 genes that are arranged ~1 gene per 5.7 kb in gene-rich regions. Physical maps have been constructed and developed for key linkage map regions of several fruit quality genes (Woodhead and McCallum, personal communication), including other traits. Some physical maps have been prepared particularly for linkage groups 2, 3 and 6 (Graham *et al*., 2006, 2011, Graham unreported data).

In other fruits in the family Rosaceae, physical maps are in progress in several fruits while finished for peach and apple (http://www.rosaceae.org/). The peach physical map covers approximately 303 Mb in physical length that is longer than its genome size of 270 Mb (Zhebentyayeva *et al.*, 2006; 2008) while the physical map of apple is 927 Mb that spans ~177 Mb from genome size (750 Mb) (Han *et al.*, 2007). Together, physical maps of strawberry are in progress and gene neighbourhoods were examined in strawberry (Bonet *et al.*, 2009 and Davis *et al.*, 2010).

5.1.5 Aim of the Chapter

The aim of this chapter composted of two main sections. Firstly, to assemble an entire BAC sequences. Resulting 454 sequencing of each BAC and fosmid clones (Chapter 4) were used to complete the sequences of the BACs. Secondly, to construct physical maps derived from the BAC sequences and includes determination of fosmid clones of the minimal tiling path (MTP) within the map to develop an available tool for further related applications of the Europe red raspberry cv. Glen Moy.

5.2 RESULTS

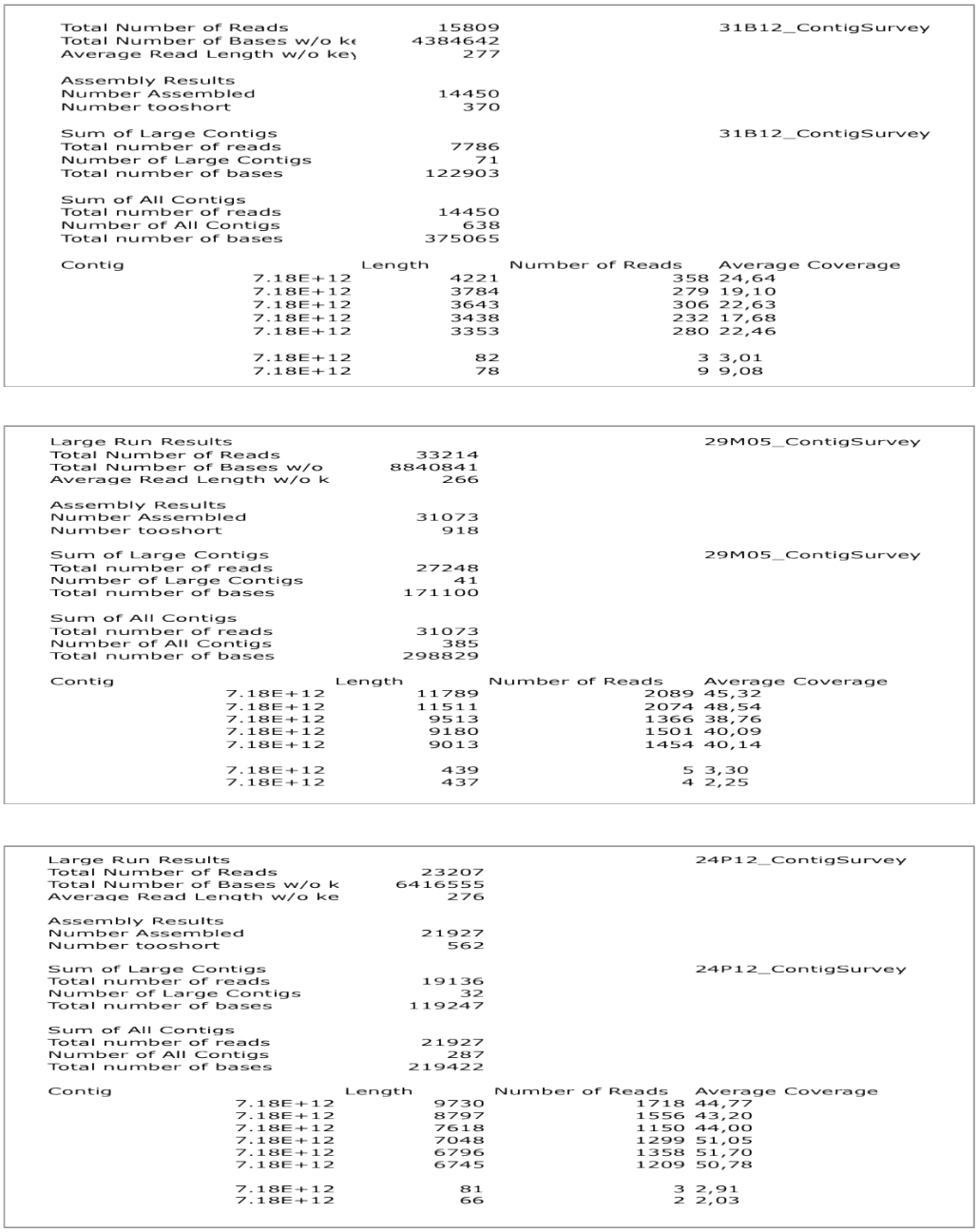
In order to complete DNA sequences of ABC clones from red raspberry cv. Glen Moy, a set of the three BAC clones (24P12, 29M05 and 31B12) was sequenced with the Roche 454 GS FLX Titanium system (Source BioScience, Cambridge, UK). The resulting sequence files were analyzed and subsequent combined with paired-ends of those BACs and then fosmid paired end sequences (from Chapter 4) using Sequencer software (version 5.0.1, U.S.A). In addition, the assembled sequences of each BAC have been used to construct physical maps, the open reading frames (ORFs) of genes identified in BACs (24P12 and 29M05) were placed in order along BAC length, and fosmid clones assisted the assemblies of the BACs were determined the overlaps for additional confirmation of these informative maps.

5.2.1 BAC 454 Sequencing Efficiency and Assembly Characterization

After 454 sequencing, the number of reads and contigs produced by the three runs of the three BACs are shown in Figure 5.2. The resulting 454 sequencing produced 15,809; 33, 214 and 23,207 reads respectively, giving an average number of 24,076 reads per BAC. Average of read lengths of each BACs differed between 277; 266 and 276 nucleotides and mean of all reads was 273 nucleotides. The numbers of large contigs were equal to 71; 41 and 32 of BAC31B12; 29M05 and 24P12, representing a total of 123; 171 and 119 kb, respectively. The total number of contigs from all three BACs was 1,310 contigs, meaning that on average 436.6 contigs were obtained from

each BAC. The length of all contigs ranged from 66 bp for BAC31B12 to 11,789 bp for 29M05, with an average 675 bp. The total number of nucleotides for all three BAC was 884,316 bp, representing 2.26 equivalent length of all three BAC sequences (with an average size of 130 kb per BAC) and ~3.2% of red raspberry genome (275 Mb).

Each BAC assembly was characterized (Figure 5.3). The best initial assembly was for BAC24P12, representing the longest sequence of 117 kb, consisting of 252 fragments. While BAC29M05 consisted of a set of four contigs, resulting from the combination of 19; 59; 65 and 171 fragments and representing 7.3; 24; 46 and 95 kb in length respectively. In contrast, more than twenty-three combined contigs were obtained in assembly BAC31P12, ranging from 417 bp to 33.4 kb in length and representing a total of ~165 kb. Notably, only eight contigs were ranging 9 to 33 kb. In conclusion, the results demonstrated that the best assembly was of BAC24P12 presenting the longest contig of 117 kb, covering 90% of the BAC sequence (~130 kb). While BAC29M05 represented several contigs, the largest one was 95 kb, covering 73% of BAC sequences. Because the sequence of BAC31B12 was divided into several contigs, it seemed that the resulting 454 sequences were not efficient for combining an entire sequences, then it was difficult to examine the actual size of DNA insert in this BAC.

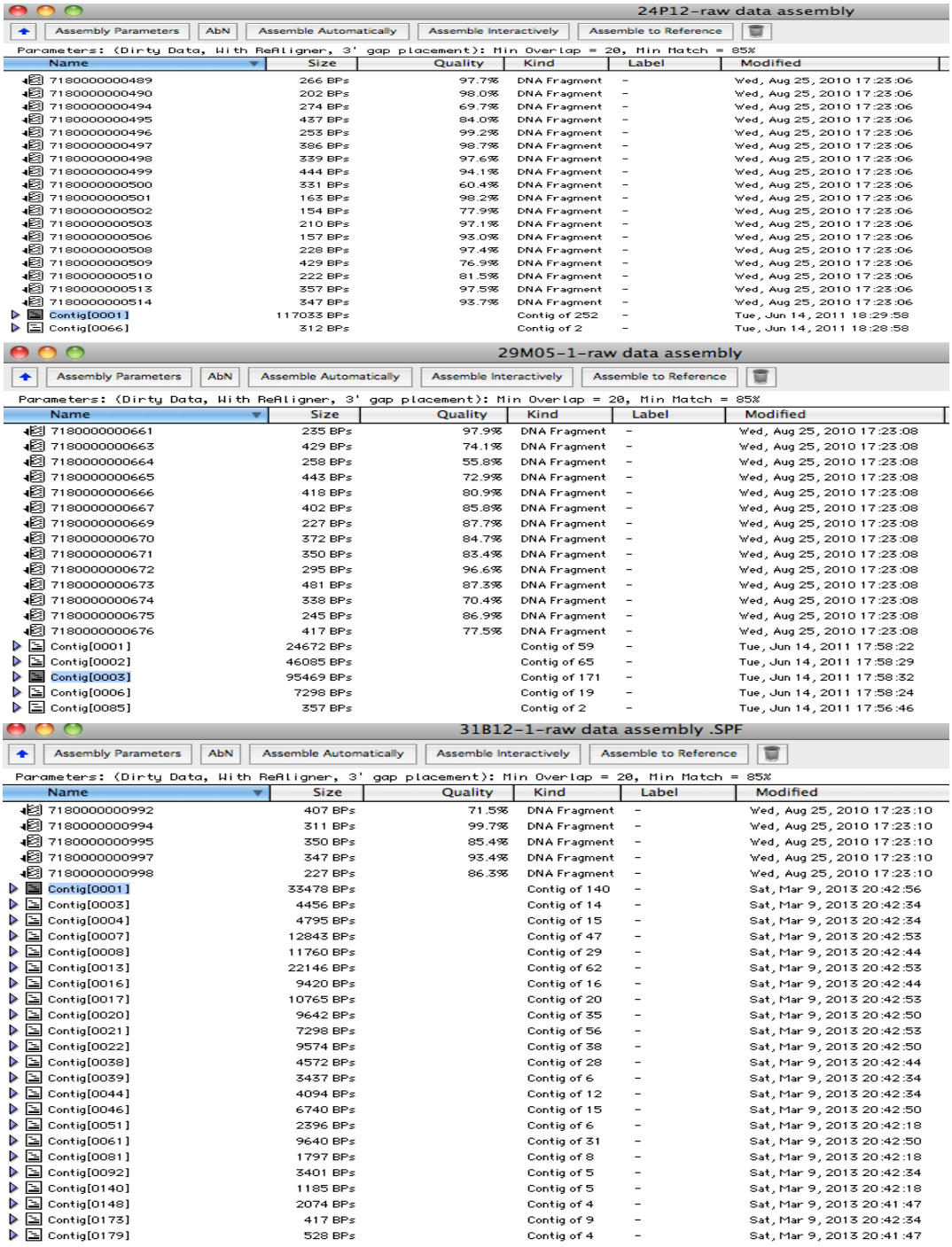


Total Number of Reads	15809			31B12_ContigSurvey
Total Number of Bases w/o k	4384642			
Average Read Length w/o ke	277			
Assembly Results				
Number Assembled	14450			
Number tooshort	370			
Sum of Large Contigs				31B12_ContigSurvey
Total number of reads	7786			
Number of Large Contigs	71			
Total number of bases	122903			
Sum of All Contigs				
Total number of reads	14450			
Number of All Contigs	638			
Total number of bases	375065			
Contig		Length	Number of Reads	Average Coverage
	7.18E+12	4221	358	24,64
	7.18E+12	3784	279	19,10
	7.18E+12	3643	306	22,63
	7.18E+12	3438	232	17,68
	7.18E+12	3353	280	22,46
	7.18E+12	82	3	3,01
	7.18E+12	78	9	9,08

Large Run Results				29M05_ContigSurvey
Total Number of Reads	33214			
Total Number of Bases w/o	8840841			
Average Read Length w/o k	266			
Assembly Results				
Number Assembled	31073			
Number tooshort	918			
Sum of Large Contigs				29M05_ContigSurvey
Total number of reads	27248			
Number of Large Contigs	41			
Total number of bases	171100			
Sum of All Contigs				
Total number of reads	31073			
Number of All Contigs	385			
Total number of bases	298829			
Contig		Length	Number of Reads	Average Coverage
	7.18E+12	11789	2089	45,32
	7.18E+12	11511	2074	48,54
	7.18E+12	9513	1366	38,76
	7.18E+12	9180	1501	40,09
	7.18E+12	9013	1454	40,14
	7.18E+12	439	5	3,30
	7.18E+12	437	4	2,25

Large Run Results				24P12_ContigSurvey
Total Number of Reads	23207			
Total Number of Bases w/o k	6416555			
Average Read Length w/o ke	276			
Assembly Results				
Number Assembled	21927			
Number tooshort	562			
Sum of Large Contigs				24P12_ContigSurvey
Total number of reads	19136			
Number of Large Contigs	32			
Total number of bases	119247			
Sum of All Contigs				
Total number of reads	21927			
Number of All Contigs	287			
Total number of bases	219422			
Contig		Length	Number of Reads	Average Coverage
	7.18E+12	9730	1718	44,77
	7.18E+12	8797	1556	43,20
	7.18E+12	7618	1150	44,00
	7.18E+12	7048	1299	51,05
	7.18E+12	6796	1358	51,70
	7.18E+12	6745	1209	50,78
	7.18E+12	81	3	2,91
	7.18E+12	66	2	2,03

Figure 5.2 An initial analysis of number of contigs and bases of the resulting 454 sequences of the three BACs (31B12, 29M05 and 24P12) from red raspberry cv. Glen Moy. There were a total of 1,310 contigs, ranging from 66 bp to 11,789 bp in length, with an average 675 bp. The numbers of large contigs were equal to 71; 41 and 32 representing 122.9; 171 and 119 kb in BAC31B12; 29M05 and 24P12, respectively.



24P12-raw data assembly

Assembly Parameters | AbN | Assemble Automatically | Assemble Interactively | Assemble to Reference

Parameters: (Dirty Data, With ReAligner, 3' gap placement): Min Overlap = 20, Min Match = 85%

Name	Size	Quality	Kind	Label	Modified
7180000000489	266 BPs	97.7%	DNA Fragment	-	Wed, Aug 25, 2010 17:23:06
7180000000490	202 BPs	98.0%	DNA Fragment	-	Wed, Aug 25, 2010 17:23:06
7180000000494	274 BPs	69.7%	DNA Fragment	-	Wed, Aug 25, 2010 17:23:06
7180000000495	437 BPs	84.0%	DNA Fragment	-	Wed, Aug 25, 2010 17:23:06
7180000000496	253 BPs	99.2%	DNA Fragment	-	Wed, Aug 25, 2010 17:23:06
7180000000497	386 BPs	98.7%	DNA Fragment	-	Wed, Aug 25, 2010 17:23:06
7180000000498	339 BPs	97.6%	DNA Fragment	-	Wed, Aug 25, 2010 17:23:06
7180000000499	444 BPs	94.1%	DNA Fragment	-	Wed, Aug 25, 2010 17:23:06
7180000000500	331 BPs	60.4%	DNA Fragment	-	Wed, Aug 25, 2010 17:23:06
7180000000501	163 BPs	98.2%	DNA Fragment	-	Wed, Aug 25, 2010 17:23:06
7180000000502	154 BPs	77.9%	DNA Fragment	-	Wed, Aug 25, 2010 17:23:06
7180000000503	210 BPs	97.1%	DNA Fragment	-	Wed, Aug 25, 2010 17:23:06
7180000000506	157 BPs	93.0%	DNA Fragment	-	Wed, Aug 25, 2010 17:23:06
7180000000508	228 BPs	97.4%	DNA Fragment	-	Wed, Aug 25, 2010 17:23:06
7180000000509	429 BPs	76.9%	DNA Fragment	-	Wed, Aug 25, 2010 17:23:06
7180000000510	222 BPs	81.5%	DNA Fragment	-	Wed, Aug 25, 2010 17:23:06
7180000000513	357 BPs	97.5%	DNA Fragment	-	Wed, Aug 25, 2010 17:23:06
7180000000514	347 BPs	93.7%	DNA Fragment	-	Wed, Aug 25, 2010 17:23:06
Contig[0001]	117033 BPs		Contig of 252	-	Tue, Jun 14, 2011 18:29:58
Contig[0066]	312 BPs		Contig of 2	-	Tue, Jun 14, 2011 18:28:58

29M05-1-raw data assembly

Assembly Parameters | AbN | Assemble Automatically | Assemble Interactively | Assemble to Reference

Parameters: (Dirty Data, With ReAligner, 3' gap placement): Min Overlap = 20, Min Match = 85%

Name	Size	Quality	Kind	Label	Modified
7180000000661	235 BPs	97.9%	DNA Fragment	-	Wed, Aug 25, 2010 17:23:08
7180000000663	429 BPs	74.1%	DNA Fragment	-	Wed, Aug 25, 2010 17:23:08
7180000000664	258 BPs	55.8%	DNA Fragment	-	Wed, Aug 25, 2010 17:23:08
7180000000665	443 BPs	72.9%	DNA Fragment	-	Wed, Aug 25, 2010 17:23:08
7180000000666	418 BPs	80.9%	DNA Fragment	-	Wed, Aug 25, 2010 17:23:08
7180000000667	402 BPs	85.8%	DNA Fragment	-	Wed, Aug 25, 2010 17:23:08
7180000000669	227 BPs	87.7%	DNA Fragment	-	Wed, Aug 25, 2010 17:23:08
7180000000670	372 BPs	84.7%	DNA Fragment	-	Wed, Aug 25, 2010 17:23:08
7180000000671	350 BPs	83.4%	DNA Fragment	-	Wed, Aug 25, 2010 17:23:08
7180000000672	295 BPs	96.6%	DNA Fragment	-	Wed, Aug 25, 2010 17:23:08
7180000000673	481 BPs	87.3%	DNA Fragment	-	Wed, Aug 25, 2010 17:23:08
7180000000674	338 BPs	70.4%	DNA Fragment	-	Wed, Aug 25, 2010 17:23:08
7180000000675	245 BPs	86.9%	DNA Fragment	-	Wed, Aug 25, 2010 17:23:08
7180000000676	417 BPs	77.5%	DNA Fragment	-	Wed, Aug 25, 2010 17:23:08
Contig[0001]	24672 BPs		Contig of 59	-	Tue, Jun 14, 2011 17:58:22
Contig[0002]	46085 BPs		Contig of 65	-	Tue, Jun 14, 2011 17:58:29
Contig[0003]	95469 BPs		Contig of 171	-	Tue, Jun 14, 2011 17:58:32
Contig[0006]	7298 BPs		Contig of 19	-	Tue, Jun 14, 2011 17:58:24
Contig[0085]	357 BPs		Contig of 2	-	Tue, Jun 14, 2011 17:56:46

31B12-1-raw data assembly .SPF

Assembly Parameters | AbN | Assemble Automatically | Assemble Interactively | Assemble to Reference

Parameters: (Dirty Data, With ReAligner, 3' gap placement): Min Overlap = 20, Min Match = 85%

Name	Size	Quality	Kind	Label	Modified
7180000000992	407 BPs	71.5%	DNA Fragment	-	Wed, Aug 25, 2010 17:23:10
7180000000994	311 BPs	99.7%	DNA Fragment	-	Wed, Aug 25, 2010 17:23:10
7180000000995	350 BPs	85.4%	DNA Fragment	-	Wed, Aug 25, 2010 17:23:10
7180000000997	347 BPs	93.4%	DNA Fragment	-	Wed, Aug 25, 2010 17:23:10
7180000000998	227 BPs	86.3%	DNA Fragment	-	Wed, Aug 25, 2010 17:23:10
Contig[0001]	33478 BPs		Contig of 140	-	Sat, Mar 9, 2013 20:42:56
Contig[0003]	4456 BPs		Contig of 14	-	Sat, Mar 9, 2013 20:42:34
Contig[0004]	4795 BPs		Contig of 15	-	Sat, Mar 9, 2013 20:42:34
Contig[0007]	12843 BPs		Contig of 47	-	Sat, Mar 9, 2013 20:42:53
Contig[0008]	11760 BPs		Contig of 29	-	Sat, Mar 9, 2013 20:42:44
Contig[0013]	22146 BPs		Contig of 62	-	Sat, Mar 9, 2013 20:42:53
Contig[0016]	9420 BPs		Contig of 16	-	Sat, Mar 9, 2013 20:42:44
Contig[0017]	10765 BPs		Contig of 20	-	Sat, Mar 9, 2013 20:42:53
Contig[0020]	9642 BPs		Contig of 35	-	Sat, Mar 9, 2013 20:42:50
Contig[0021]	7298 BPs		Contig of 56	-	Sat, Mar 9, 2013 20:42:53
Contig[0022]	9574 BPs		Contig of 38	-	Sat, Mar 9, 2013 20:42:50
Contig[0038]	4572 BPs		Contig of 28	-	Sat, Mar 9, 2013 20:42:44
Contig[0039]	3437 BPs		Contig of 6	-	Sat, Mar 9, 2013 20:42:34
Contig[0044]	4094 BPs		Contig of 12	-	Sat, Mar 9, 2013 20:42:34
Contig[0046]	6740 BPs		Contig of 15	-	Sat, Mar 9, 2013 20:42:50
Contig[0051]	2396 BPs		Contig of 6	-	Sat, Mar 9, 2013 20:42:18
Contig[0061]	9640 BPs		Contig of 31	-	Sat, Mar 9, 2013 20:42:50
Contig[0081]	1797 BPs		Contig of 8	-	Sat, Mar 9, 2013 20:42:18
Contig[0092]	3401 BPs		Contig of 5	-	Sat, Mar 9, 2013 20:42:34
Contig[0140]	1185 BPs		Contig of 5	-	Sat, Mar 9, 2013 20:42:18
Contig[0148]	2074 BPs		Contig of 4	-	Sat, Mar 9, 2013 20:41:47
Contig[0173]	417 BPs		Contig of 9	-	Sat, Mar 9, 2013 20:42:34
Contig[0179]	528 BPs		Contig of 4	-	Sat, Mar 9, 2013 20:41:47

Figure 5.3 Characterizations of assemblies of resulting 44 sequences of the three BACs (31B12, 29M05 and 24P12) from red raspberry cv. Glen Moy. An assembly BAC24P12 composed a long contig of 117 kb as BAC29M05 composed of a set of four scaffolds, representing 7.3; 24; 46 and 95 kb in length. BAC31B12 was divided into twenty-three contigs.

5.2.2 Combination of BAC Paired End Sequences

If the BAC presented several contigs, it is possible that real size of insert DNA could equal the total of all lengths. An association with paired end sequences of BAC can assist in determining this proposed assumption. To determine an overall insert length of each BAC, end pairs of the three BACs were sequenced by Sanger sequencing (Source Bioscience, UK), then combined with the resulting 454 sequences (Figure 5.4 A). The association of the end sequences of the BACs with 454 sequencing data resulted in a slight difference from initial assemblies.

The sequences of both ends of, BAC24P12, separately merged at the beginning and the end of the 117 kb contig, strengthening the evidence that this BAC contained a complete segment of ~117 kb of insert DNA from red raspberry. Apart from this combination, it was not easy to estimate the real size of insert DNA of the other two BACs (29M05 and 31B12). Both ends of BAC29M05 were found separately in the two contigs, 24 and 7.3 kb (Figure 5.4 B). It was possible that this BAC carried only three contigs, totaling approximately 120 kb. However it is also possible that this BAC consists of all four contigs, meaning that it contained a total insert DNA of ~130 to 175 kb. Similar to the BCA29M05, both ends of the BAC31P12 were separated in two contigs, of 22.6 kb and 20 kb (data not shown).

Therefore BAC24P12 can be assumed to carry on the insert DNA of ~117 kb and BAC29M05 contains ~120 or 175 kb of red raspberry cv. Glen Moy. But it failed to reveal an insert size of BAC31B12. However, there were at least 3 gaps between the four contigs of BAC29M05. While there must be several gaps in BAC31B12 assembly, the constructed fosmid libraries (Chapter 4) are expected to extend or fill the gaps of these sequence contigs.

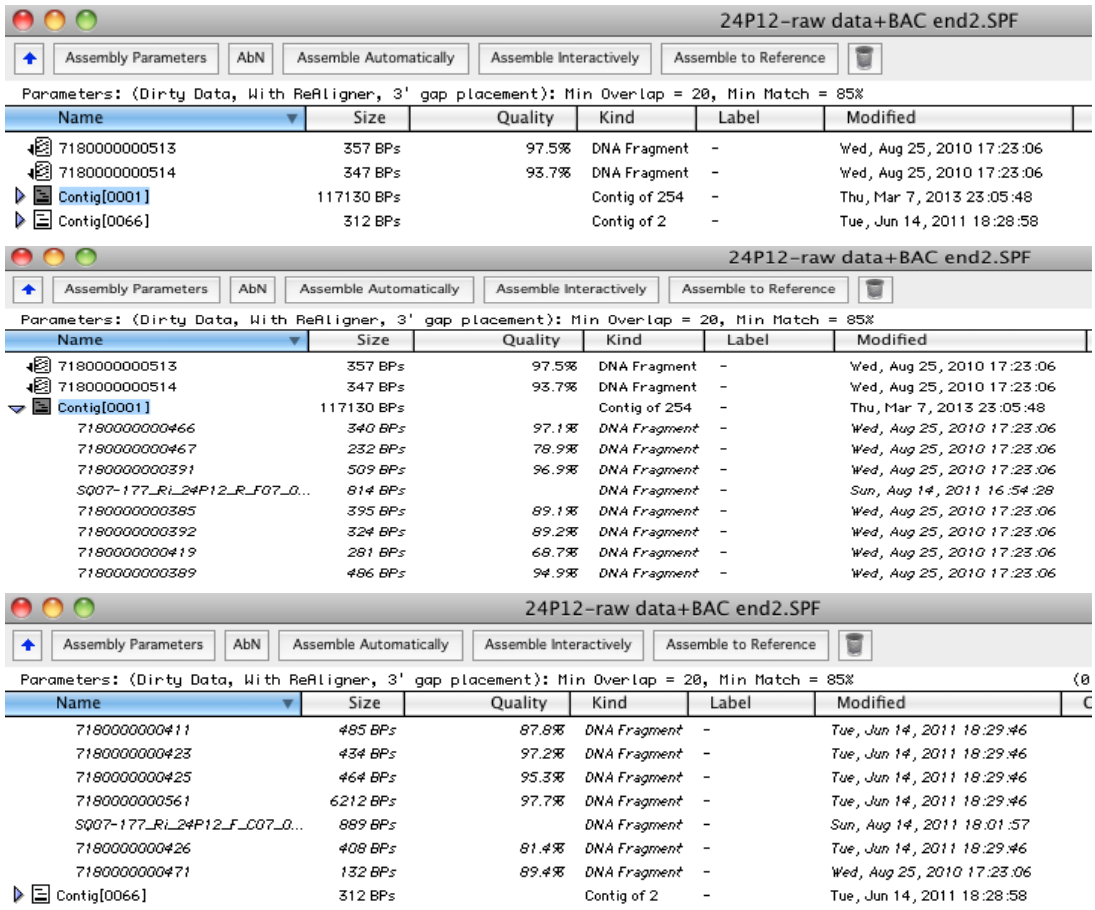


Figure 5.4 A An association of paired end sequences (by Sanger sequencing) to resulting 454 sequencing data of BAC24P12. The both ends of BAC24P12 merged at the beginning and end of the 117 kb contig, thus confirming presence ~117 kb of insert DNA from red raspberry.

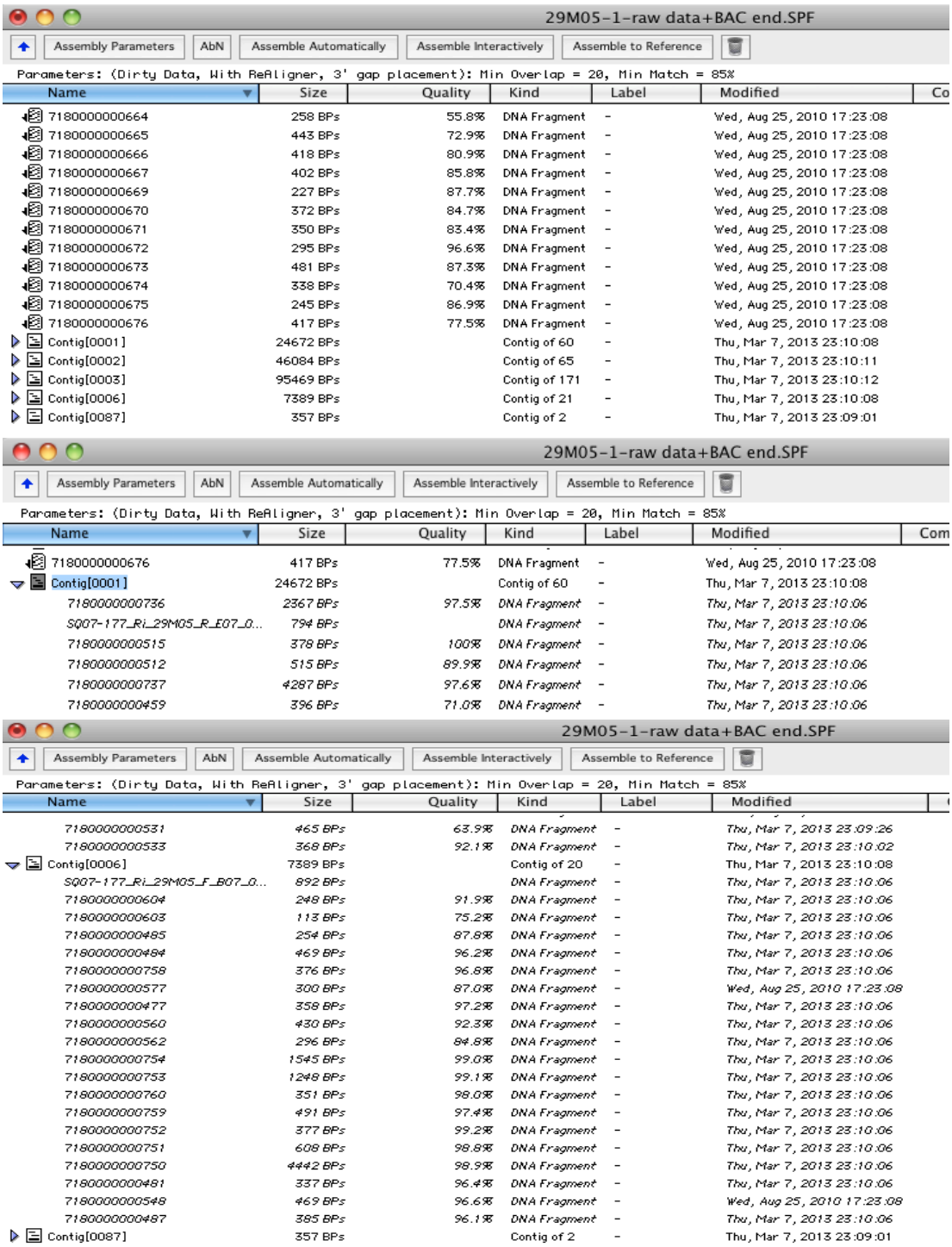


Figure 5.4 B An association of paired end sequences to resulting 454 sequencing data of BAC29M05. The both ends of BAC29M05 were separately in the two contigs, 24 and 7.3 kb. The BAC probably consisted of either only three contigs or all four of them, meaning that it may contain insert DNA ~130 to 175 kb.

5.2.3 Genome Annotation of Combined Sequences

All assembled contig sequences of the BACs were evaluated by BLASTn alignments. Of these twenty-eight divided contigs, four presented significant identity with known genome sequences of red raspberry.

From the 117 kb in length of the combined sequences of BAC24P12, it was separated into sixteen smaller contigs (~7.3 kb) to manageably analyze. Blast alignment of two contigs showed that the sequences were identical to PKS/CHS genes of raspberry *R. idaeus* (Figure 5.5 A), being very similar as the two contigs of BAC29M05.

Based on aligning of both BAC24P12 and 29M05, the first contig was homologous with *R. idaeus* PKS gene and was 99% (3083/3090 bp) identical to *R. idaeus* Aromatic Polyketide Synthase gene, cDNA (Accession no. EU862821.1)(Appendix A.2) whilst also matching with CHS6 gene (Accession no. AF400567.1) with 96 % (1253/1300 bp) identity. The second contig showed 92% identical (510/552 of ~19,800 bp) to the TTG1 gene of *R. idaeus* genomic DNA (Accession no. HM579852.1). While other ten contigs were homologous with blackberry; grape; prune and other plants (no results shown). Differently, only three out of the twenty-three contigs of BAC31B12 were identified as having similarity to blackberry genome with (~242/278 bp) 89% of nucleotides identical but none were similar to raspberry gene sequences. In summary, detailed analysis of the 454 assembled sequences demonstrated the presence of PKS/CHS-related genes in the both BACs (24P12 and 29M05) of red raspberry.

5.2.4 Combination of Fosmid Ended-pair Sequences

In order to assist with the assemblies of the BAC contigs, the fosmid end sequences (from Chapter 4) were combined with the 454-derived sequences of the BACs. A total of eighty-four paired-end sequences of forty-two selected fosmid clones were aligned to the resulting 454 sequences. Separately, twenty-eight end sequences, averaging approximately 800 bp, were added to each BAC, then analyzed using Sequencer.

Eighteen fosmid end sequences were associated with the resulting 454 sequences of BAC24P12 while several ends were not. Six ended-pairs were associated with ~80 kb of the contig of 117 kb. Importing of more paired ends were necessary for further covering the remaining ~38 kb (Figure 5.5 A). Although another fourteen fosmid ended-pairs were subsequently added, only four pairs were associated in but they did not added in region needed.

The combination of fosmid ends in BAC29M05 made twelve ends distributed in all four contigs of the BAC. Six ends combined in the contig of 96 kb, representing two ended pairs corresponding to ~58 kb (Figure 5.5 B), then confirmation was needed for ~37 kb. However, there were at least two paired ends, separately merging between two contigs as these clones may support a connection of these two contigs and helping to close gaps.

Nevertheless, based on the results of combining fosmid ends to contigs, the three gaps of BAC29M05 could be closed with two fosmid clones (L1 and D12) with predicted gap size ~6-7 kb between contig no.2 and 3. While it should be shorter than 2 kb between contig no. 2 and 4. This confirmed that the inclusion of fosmid paired end sequences enhanced the quality of the initial assembly obtained using 454 sequencing.

For BAC31B12, only twelve ends were combined in the 454 sequences while several ends were out. Five pairs showed linking of some contigs. Since all contigs of this BAC were smaller than 30 kb, there was not any pair of fosmid clones found in the same contig. With at least fourteen contigs had no association of any ends. Although several fosmid ends were added to help assembly of this BAC sequences, due to insufficient 454 sequencing and no correlation between each contig, it was difficult to complete the assembly of these sequences, and it is not possible to predict the order of genes within this BAC in next step.

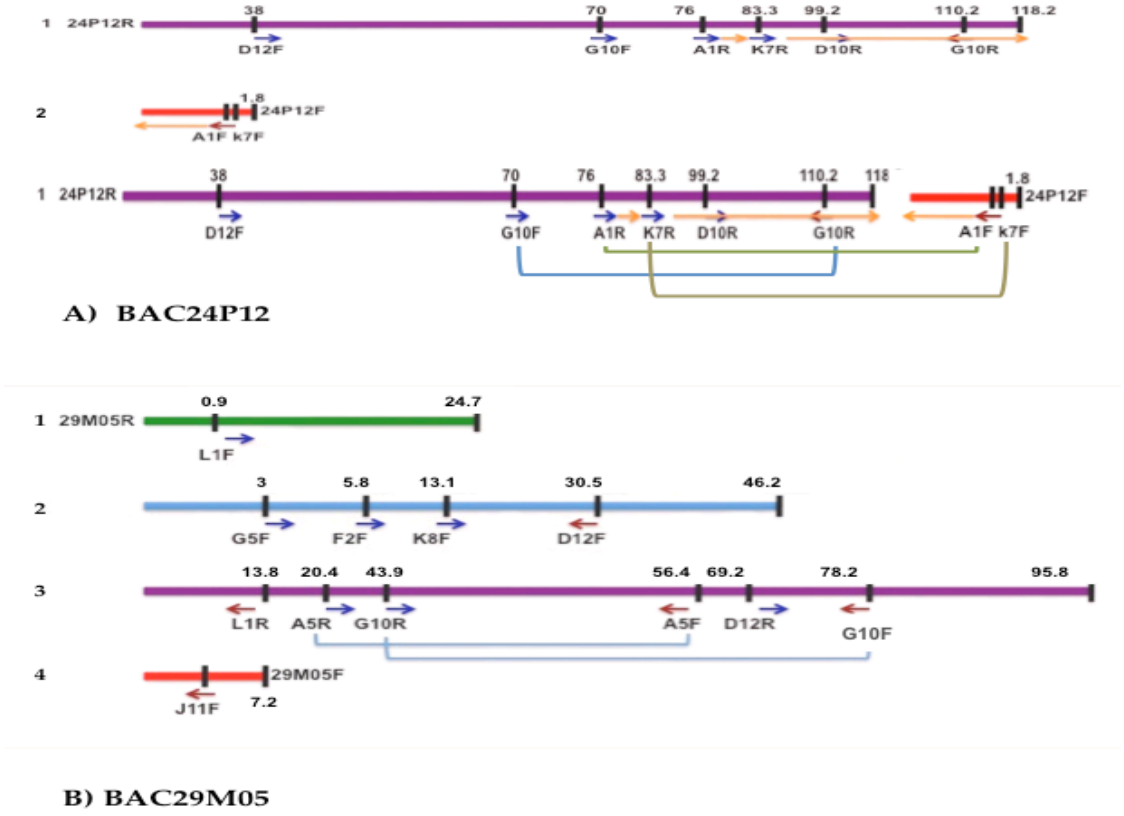


Figure 5.5 Combination of paired-end sequences of Fosmid clones to resulting 454 sequences of BAC24P12 and BAC29M05. **A**, seven paired-end sequences associated to resulting 454 sequences of BAC24P12, corresponding to ~80 kb out of ~118 kb of contig no.1 and ~1.8 kb of contig no.2, showing one ended-pairs of clone G10 was on contig no3 while two ended-pairs of two clones (A1 and k7) were on different contigs. **B**, twelve paired-end sequences combined to all four contigs of BAC29M05 and only two ended-pairs of two clones (A5 and G10) merged with contig no. 3, covering ~60 kb in length of sequence of BAC29M05. As several sequenced-ends of different pairs were out from resulting 454 sequences of BAC each, thus without showing their position in theses Figures. F and R, forward and reverse primers used for amplifying the sequences while blue and red arrows indicted the direction of known sequences amplified along BAC sequences and orange arrows referred to possible direction for further sequencing to cover the gap between the two contigs of BAC24P12.

5.2.5 Prediction of Ordering Genes within the BAC Sequences

According to BAC sequence assemblies, to predict the ORFs within assemblies of the BACs (BAC24P12 and BAC29M05), nucleotide sequences was analyzed using the gene prediction program for Eukaryotes in finding a gene start site (GeneMark.hmm-E* and GeneMark-E*, (http://exon.gatech.edu/eukhmm.cgi)). Following the detailed annotation of assembled BAC sequences using BLASTx, the open reading frames (ORFs) found in each BAC (BAC24P12 and 29M05) were ordered and drafts of physical maps were constructed (Figure 5.6 A).

The BLASTx query of the BAC assemblies identified PKS-like genes in four regions. These are represented as RiPKSr1-4. While the first two (RiPKSr1 and RiPKSr2) were on BAC24P12, another two (RiPKSr3 and RiPKSr4), were on BAC29M05. An assembly including 117 kb of BAC24P12 presented the PKS-related gene regions at 73-75 kb (RiPKSr1) and 96-99 kb (RiPKSr2). The BAC29M05 sequence assembly appeared to contain another two PKS-related genes regions at 51-53 kb (RiPKSr3) and 72-76 kb (RiPKSr4) on the 95 kb contig (no. 3) (detailed analysis in Chapter 6).

At least thirteen ORFs were found and located on the 117 kb contig of BAC24P12. Significantly, given that the PKS1 and PKS5 genes were located in this region and also ketoacetyl CoA thiolase gene (KAT) coding acetoacetyl CoA synthase involved in the common biosynthesis of polyketide compounds was located at the end of this predicted segment as twenty-three genes were located on four contigs of BAC29M05.

Interestingly, there were some regions revealing an identical gene order between the both BACs. The nine ORFs (at 27 to 96 kb) on the third contig of 29M05 showed identities to the same genes on 24P12 (at 53 to 117 kb), but the regions were inverted (Figure 5.6 B). Another three ORFs (at 1 to ~53 kb) of BAC24P12 were also on the first (1 to 24 kb) and the second contig (~ 22 to 46 kb) of BAC29M05. Taken together this indicates that this is the same region of the chromosome or a closely related or duplicated part of the chromosome. In addition, there were eight genes on BAC29M05 (at 1 to 21 kb of the second and 1 to 26 kb of the third contig) that were not found in the other BAC. This shows that the BACs appear overlapping, with a common region covered by both of them.

The presence of these 5 genes on BAC29M05, which cannot be found in the same region on BAC24P12, indicated the possibility that the BACs represent information from different regions that might possible arising from unequal crossing over or gene duplication. Alternatively, an anomaly might be arisen during the preparation of the BACs or fosmids, thus allowing unconnected elements to join. However both drafts were composed of ~2 genes per 10 kb within certain parts, which likely corresponded to presence of ~1 gene per 5.7 kb in gene-rich regions of *R. idaeus.*

In summary, the drafts of BAC24P12 were composed of the twelve adjacent genes along 117 kb in length (contig no. 3) and BAC29M05 contained twenty-three genes on ~170 kb in total length (four contigs). A conclusion of predicted genes and their coding proteins that are found on the BACs are showed in Table 5.1. Some genes that provide health benefits to human are found, for instant Cycloartenol synthase(CAS); Lupeol synthase (LUP) and ATP-binding cassette transporter (ABC transporter) (detailed analysis in Chapter 6), including some genes that are important keys for plant growth such as Expansin (EXP) and Cellulose synthase (CesA).

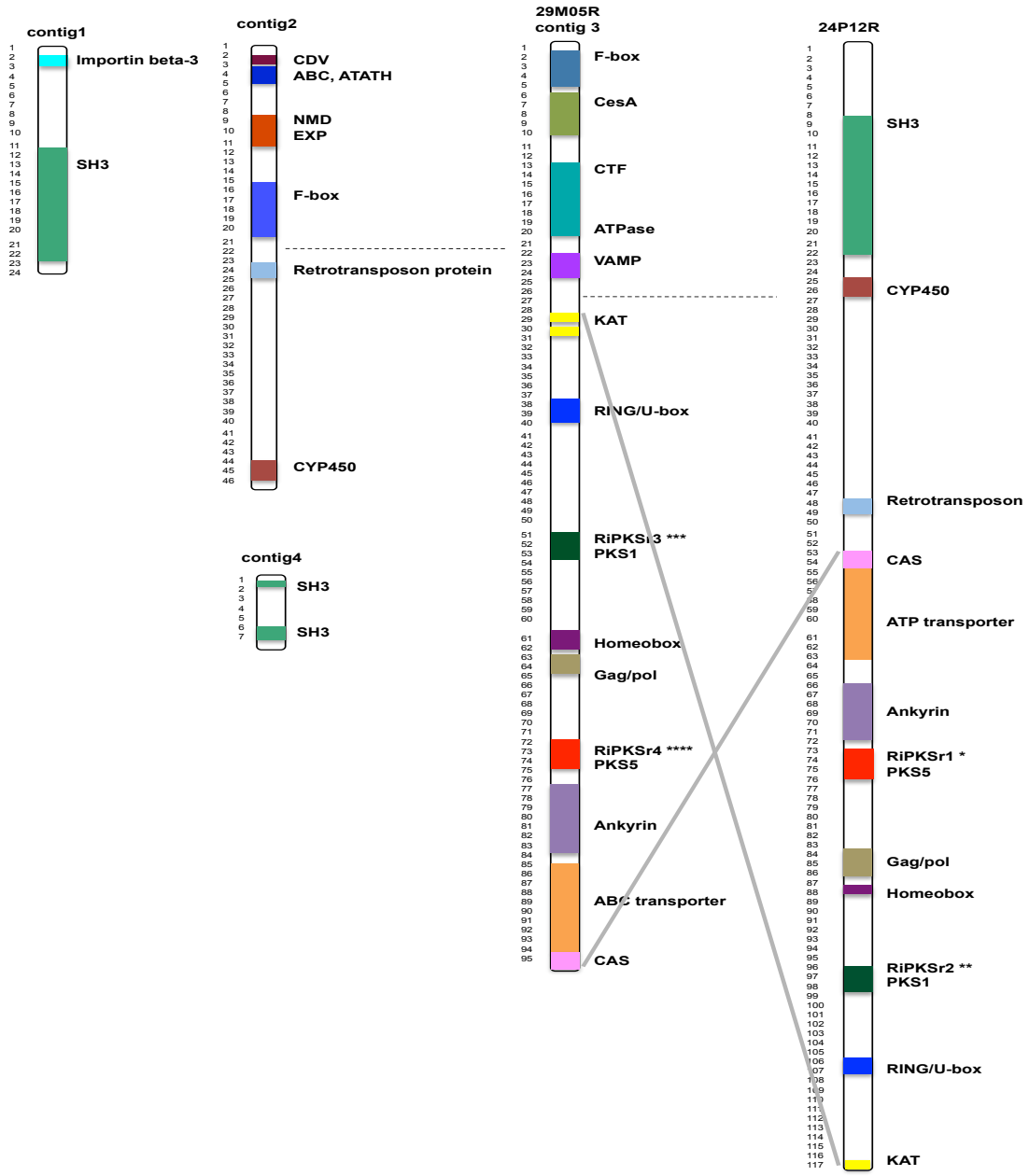


Figure 5.6 A Schema of the order of genes on BAC29M05 and BAC24P12 derived from raspberry genome cv. Glen Moy. Twelve predicted genes were respectively located on the 117 kb contig of BAC24P12 and twenty-three genes were placed on all four contigs of BAC29M05. Nine genes on the third contig of 29M05 (at 27 to 96 kb) were similar and in the same order as genes on 24P12 (at 53 to 117 kb). The two PKS-like genes (PKS1 and PKS5) were on both maps at RiPKSr1-4 regions.

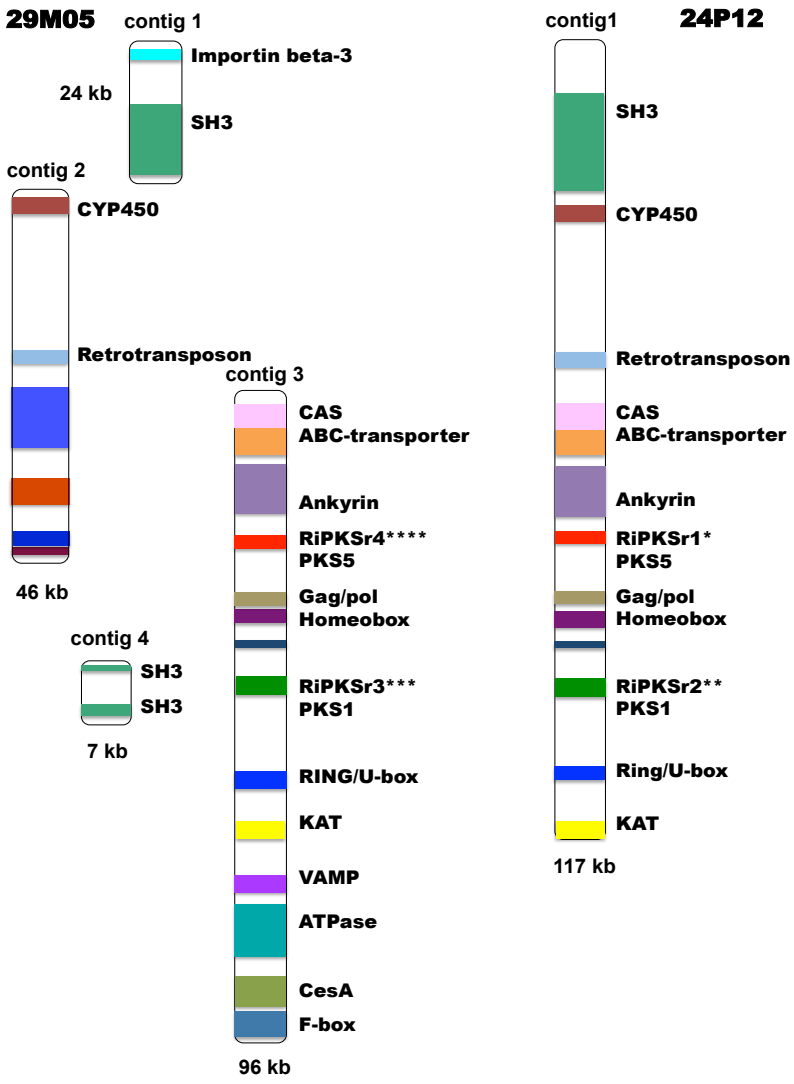


Figure 5.6 B Comparison of gene regions within an entire length of the maps derived from the two BACs : 24P12 (117 kb) and 29M05 (~170 kb). Segments of genes found in the BAC24P12 were separately on the all three contigs (no. 1, 2 and 3) of the BAC29M05. As both BACs appeared overlapping part containing nine gene regions (CAS, ABC-transporter, Ankyrin, 2 RiPKS genes, Gag/pol, Homeobox, RING/U-box and KAT).

Table 5.1 Gene sequences and coding proteins found in BAC29M05 and 24P12

	Name of gene/protein	24*	29*
1	3-ketoacyl-CoA thiolase / (KAT gene)	√	√
2	RING/U-box domain containing protein / (RING/U-box gene)	√	√
3	Aromatic polyketide synthase / (PKS gene)	√	√
4	Chalcone synthase / (CHS gene)	√	√
5	Homeobox protein / (Homeobox gene)	√	√
6	Gag-pol polyprotein / (gag/pol gene)	√	√
7	Transposon protein, retrotransposon protein / (Transposon gene, retrotransposon gene)	√	√
8	Ankyrin repeat domain-containing protein 13C / (Ankyrin gene)	√	√
9	ATP-binding cassette transporter / (ABC transporter gene)	√	√
10	Cycloartenol synthase / (CAS gene)	√	√
11	Lupeol synthase / (LUP gene)	√	√
12	Cytochrome P450 / (CYP450)	√	√
13	SH3 domain-containing protein / (SH3 gene)	√	√
14	Importin beta-3 protein / (Importin gene)	√	√
15	Nonsense-mediated mRNA decay protein / (NMD gene)		√
16	Expansin-like protein / (EXP gene)		√
17	F-box family protein / (F-box gene)		√
18	Cellulose synthase / (CesA gene)		√
19	Chromosome transmission fidelity / (CTF/CHL gene) / in yeast		√
20	AAA-type ATPase family protein / (ATPase gene)		√
21	Vesicle-associated membrane protein / (VAMP gene)		√
22	ATP-binding cassette transporter / ABC, ATATH		√
23	Nonsense-mediated mRNA decay protein / (NMD gene)		√

(24*, BAC2P12, 29*, 29M05)

5.2.6 Comparison the Ordering Genes of Raspberry

Based on the comparison the ordering PKS/CHS genes of raspberry to genes mapped on LG3 and LG7 of wild strawberry genome (*Fragaria vesca* subsp. *vesca* in NCBI databases, there are several gene families that are the same as those on the both BACs such as ATPase; ABC transporter; CAS; KAT; LUP; VAMP; F-box; U-box; ACAT; Ankyrin; EXP; CesA; Importin; homeo-box and Cell division control protein etc. One CHS gene coding flavonone isomerase 2-like and another one coding anthocyanidin 3-*O*-glucosyntransferase 5-like were mapped on LG3 but their nucleotides and amino acids differ from those of the four RiPKS regions of the both BACs (24P12 and 29M05) of red raspberry cv. Glen Moy in this thesis). Interestingly, two CHS genes coding chalcone synthase 2-like are on the same location of LG7 (LOC101298162 and LOC101298456), being 389 amino acid residues (with five amino residues different between them both) and located ~3.2 kb from each other. These two CHSs genes are next to the two genes coding actin-related protein 7-like and polygalacturonase-like with distance ~2.8 kb. RiPKSr1p of BAC24P12 (393 aa) differs from the both CHS2-like with fifty-one polypeptide residues but RiPKSr3p of BAC29M05 (391 aa) is less dissimilar with seventeen and eighteen residues (no result shown).

Moreover, comparison the ordering genes of raspberry to plants in different family such as soybean (*Glycine max*) showed that at least seven CHS-like and CHS genes mapped on Chromosome 1, 8, 9 and 11 (NCBI databases). For instance, two CHS1 genes (389 aa) (ID732575 and 100790997) are next to two CHS5 genes (388 aa) (100792581 and 100791524) and another CHS2 gene (100779649) respectively, with in total distance ~30.5 kb (at 8,462.5 to 8,493 kb) on chromosome 8. However, although there are dissimilarities in ordering genes and distances among CHS and their adjacent genes, this indicated the presence multiples PKS/CHS gene regions on the same chromosomal location and also happens in red raspberry cv. Glen Moy.

5.2.7 Construction of Mini Physical Maps

To generate the initial physical maps of both BACs for further mapping on the red raspberry chromosome, two sets of fosmid clones encompassing along the each BAC were reanalysed with several restriction enzyme digestions, and the restriction map used to determine their fingerprints (results of gel electrophoresis shown in Appendix). This led to the detection of clone overlaps and clones of the Minimal Tiling Path.

5.2.7.1 Physical Map Derived from BAC24P12

5.2.7.1.1 The Minimal Tiling Path of the Physical Map of BAC24P12

In order to start the physical map construction, an order of fosmid clones within the assembled contig of 117 kb of BAC24P12 was clarified. At least five clones (L1; J9; J2; L9 and A1) subsequently belonged to the predicted region of the BAC with ~100 kb, except the first 17 kb (Figure 5.7). Although MTP of BACs consisted of only five clones that did not cover an entire length, they represented ~90% of the map and included the regions of genes of interest, thus being enough to construct the mini physical map of this BAC.

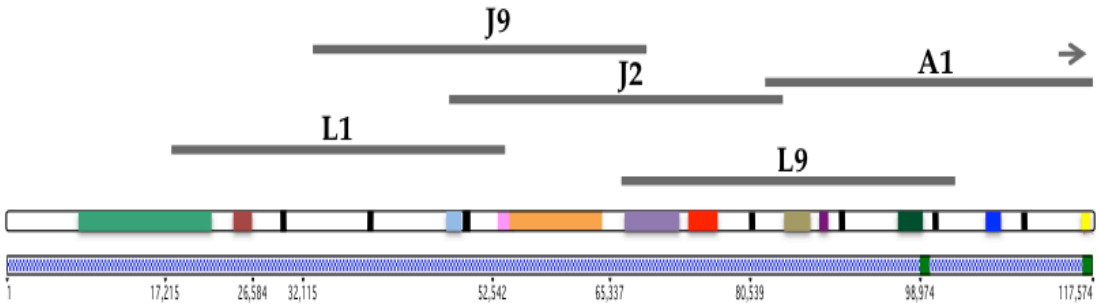


Figure 5.7 Schema of an ordering fosmid clone set of the MTP of BAC24P12. The five clones (L1; J9; J2; L9 and A1) presented approximately 90% in length (~100 kb in length) along this part at 17 to 117 kb of BAC24P12. Clones shared approximately ~ 50 to 60% in length with adjacent clones.

5.2.7.1.2 Fingerprint Analysis of the MTP for the Physical Map of BAC24P12

To analyze overlaps of fosmid clones of the MTP, the DNA was digested with several restriction endonucleases and then separated by agarose gel electrophoresis. Rare cutting enzyme that did not cut in the vector were used (Appendix A.3), N*ot* I; X*ho* I; P*me* I; M*lu* I; S*wa* I; A*fe* I; E*ag* I; S*al* I; S*ma* I and X*ma*I were used for single and double enzymes digestions. Single restriction digestions were carried out with all these enzymes but double digestions were performed with some enzymes such as S*wa* I/X*ho* I; S*wa* I/M*lu* I; S*wa* I/P*me* I and X*ho* I/M*lu* I respectively. The pattern of digested fragments was compared with the predicted using Cut Map (Sequencer software) to assist in rapidly determining their overlaps (Appendix A.4 to A.8).

Digestion with single restriction enzyme with M*lu* I and X*ho* I gave the same pattern as predicted. Digestion with M*lu* I showed 24.9 Kb of L1 and 16 Kb of L9, being similar to their expected patterns (Appendix A.4). In the X*ho* I digest pattern, three pairs of adjacent clones (L1 and J9; J9 and J2; J2 and L9) showed the same fragment length (Appendix A.5).

Followed by digestion with double enzymes such as X*ho* I/M*lu* I; S*wa* I/P*me* I and S*wa* I/X*ho* I, several clones produced similar bands (Appendix A.6 to A.8). Fragment patterns of X*ho* I/M*lu* I digestion demonstrated the order of these clones (Appendix A.6). L1 and J9 produced the same fragment of 1.4 kb, while J9 gave a fragment of 10 kb linking to J2. Subsequently J2 and L9 also showed a fragment of 7.3 kb. These fragments demonstrated the order of the adjacent clones. In addition, L9 also produced the other three fragments of approximately 2, 1.7 and 1.12 kb that were similar as the expected patterns predicted by Cut Map. Therefore this confirmed the true position of L9. The final clone, A1, gave two fragments of 2.2 kb and also may provide 21 kb, thus assuring the true position of clone A1 placed on the map. Although certain clones showed incompletely digested patterns leading to disappearing of some fragments, the results still presented possibility for being the true positive MTP of BAC24P12 (Figure 5.8).

The pattern of bands was also analyzed using Experion^{TM} Automated Electrophoresis System (BIO-RAD) but it was more difficult to identify the restriction fragment

lengths (no results shown). Together with using SynGene Gen tools (GeneSnap version 6.08.04), the result showed that patterns in many clones were either alike or the same as the size predicted by Cut Map (Appendix A.9 to A.11). However this software could not detect multiple fragments located at almost the same size on the gel correctly, particularly larger than 10 kb.

5.2.7.1.3 Determination of Overlapping Fosmid Clones

For the final analysis, a set of overlapping fosmid clones of the MTP was examined. Some pairs of adjacent fosmid clones proved their positions on the map, and therefore the positions of the clones should be correct. All fosmid clones of the MTP showed overlaps with approximately 50% (~18 kb) of an average length of fosmid clones (~36 kb). The result following this fingerprint analysis showed that the order of the MTP was L1; J9; J2; L9 and A1 respectively, with approximately 100 kb in length (since at ~17 to 117 kb).

5.2.7.1.4 Mini Physical Map of BAC24P12

Following the above results, the mini physical map of the BAC24P12 was edited. It was approximately 100 kb in length, composed of at least twelve predicted genes that were contained by a set of at least five fosmid clones. Importantly, one fosmid clone (J2) carried the RiPKSr1 region coding PKS5 gene (at 73 to 76 kb) and three clones (A1; J3 and J4) contained RiPKSr2 region coding PKS1 gene (at 96 to 99 kb). However, clone L9 contained both regions of RiPKSr1 and RiPKSr2 respectively. For the other additionally analyzed genes; CY P450, ABC transporter and CAS (described in chapter 6) were located in clone L1 and J9 respectively (Figure 5.7 and 5.8).

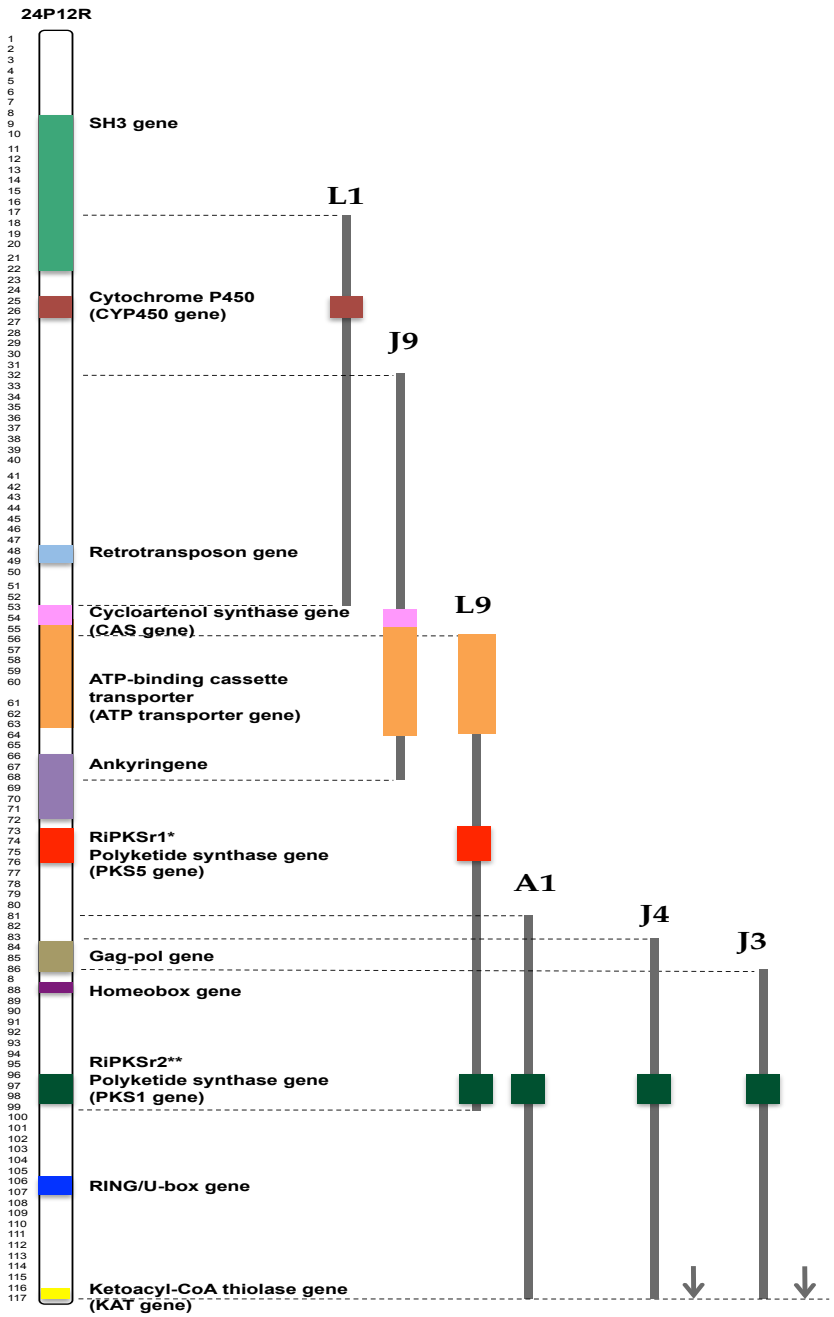


Figure 5.8 The schema of a mini constructed physical map of BAC24P12 with approximately 117 kb in length. Twelve genes were located and there were at least four overlapping fosmid clones covering 100 kb (at 17 to 117 kb). While one fosmid clone (J2) carried the RiPKSr1 region coding PKS5 gene and three clones (A1; J3 and J4) contained RiPKSr2 region coding PKS1 gene, only L9 clone carried both RiPKSr1 and RiPKSr2 regions.

5.2.7.2 Physical Map derived from BAC29M05

5.2.7.2.1 The Minimal Tiling Path of the Physical Map of 29M05

By presenting the longest assembled sequences of BAC29M05, the fosmid clones were first ordered within a contig of 96 kb. At least ten clones were in series for the MTP of this BAC. The seven clones (F6; J7; A5; D8; G10; C1 and C8) covered the ordered sequences at 7 to 88 kb. To assist a configuration of the MTP covering the entire length, the other three clones were chosen for the first 7 kb (L1 and G5) and the final 8 kb (D12) (Figure 5.9).

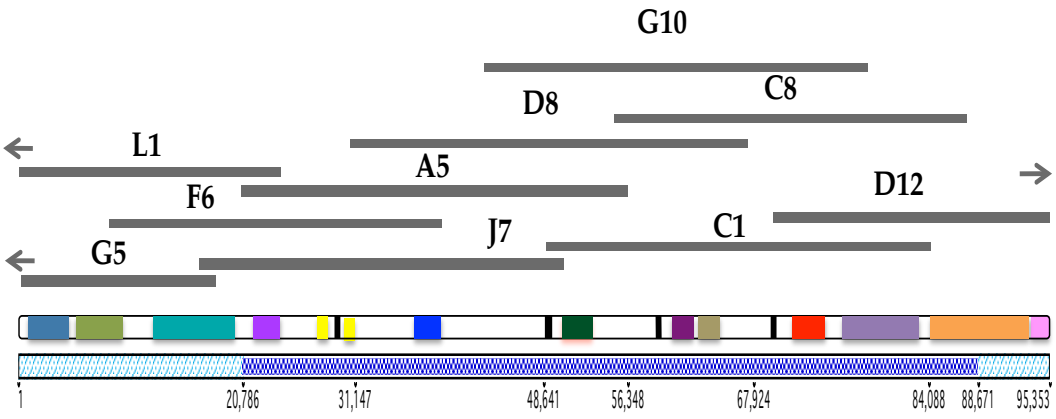


Figure 5.9 The ordering of fosmid clones of the MTP of the assembled contig (no. 3) of BAC29M05. A set of ten clones covered approximately 96 kb. Seven clones (F6; J7; A5; D8; G10; C1 and C8) were along the length at 7 to 88 kb. The other three clones (L1; G5 and D12) were added at the beginning and the end of the to complete an entire length of the map. As each clone shared approximately 70% (25 kb) with adjacent clone.

5.2.7.2.2 Fingerprint Analysis of the MTP for the Physical Map of BAC29M05

In order to determine the ordering fosmid clones of the MTP, an analysis was carried using the same set of restriction enzymes previously used for BAC24P12. The fragmented patterns of digestion also were equated to predicted sizes along the map shown in Appendix 9.18 to 9.20. The digestion with X*ho* I (Appendix A.12) showed the overlapping sequences of clone F6 and J7 with a fragment of ~4.9 Kb. J7; A5 and D8 produced the same fragment length of ~2.2 kb. Particularly, D8 presented a large fragment at ~23 Kb and ~1.7 kb that were compared to the predicted fragments. While clone C8 produced certain small fragments and also 7.3 kb that showed in clone D12 as well. Although D12 was the final clone without comparing to any clone at the end of MTP, the clone produced a fragment of~10 kb. In addition, at the beginning of the MTP, clone G5 presented a fragment of length ~ 9.9 kb, thus proving the true position of this clone.

While double enzyme digestion provided confirmation for certain clones, a M*lu* I/X*ho* I double digestion of Clone L1 showed two additional fragments at 3.3 kb and ~1.2 kb. C1 also showed more two fragments at 1.7 and 1.3 kb (Appendix A.13). Although some clones (A5; C1; G10; L1 and D12) showed different patterns from the predicted, several clones produced the correct length of fragment. Furthermore, determination with double rare enzyme, S*wa* I/P*me* I, showed fragments of 6.5 kb clarifying the pattern of the two clones L1 and J7 (Appendix A.14). Three other clones (G10; C1 and C8) produce ~10.7 kb in similarity. This assisted in proving the order of at least five clones of the MTP of the BAC29M05. While analyzing the fingerprints using SynGene Gen tools, certain fragments were same the size as predicted by Cut Map (Appendix A.15 to A.17).

5.2.7.2.3 Determination of Overlapping Fosmid Clones

The set of ten adjacent fosmid clones of the MTP of BAC29M05 showed a continuous overlaps along the map length. The seven clones of the MTP were overlapping approximately 70%(~25 kb) of fosmid length. The restriction mapping demonstrated the order of the MTP including F6; J7; A5; D8; G10; C1; C8 and D12 respectively,

with approximately 89 kb (~7 to 96 kb). Importantly, the previous Southern blot analysis used to screen fosmid clones, from the library in Chapter 4, showed positive hybridization of clones including A5 D8; G10; C1; C8 and D12. This confirmed the order of clones on the map.

5.2.7.2.4 Mini Physical Map of BAC29M05

A mini physical map of BAC29M05 was constructed that was 98 kb in physical length of this part of the red raspberry chromosome. The map consisted of at least eight ordered fosmid clones that sequentially comprised thirteen predicted genes. The four clones (A5; D8; C8 and D12) of the MTP carried either RiPKSr3 or RiPKSr4 region coding the PKS1 or PKS5 genes (at approximately 51-53 and 72-75 kb). Notably, two clones (C1 and G10) contained both RiPKSr3 and RiPKSr4 regions (Figure 5.10). In addition, ABC transporter and CAS genes were located at the end of the map (at ~85 to 96 kb) and were carried by clone D12 as the CYP450 gene was separately located on another contig of 46 kb (contig no. 2) and was carried by clone K8.

5.2.7.3 Difference of the Mini Physical Maps

There were some differences between the two maps in length and identical genes that were located on the maps. Unequal length resulted in different number of genes. All of the length of BAC29M05 is gene-rich region (with 13 genes/96 kb) while only a half of BAC24P12 is gene-rich region (with 12 genes/117 kb). Map of BAC24P12 begins at SH3; CYP450 and Retrotransposon gene and the ordering genes of CAS to KAT gene. In parallel, map of BAC29M05 starts at CAS to KAT gene and then VAMP; ATPase; CTF; CesA and F-box gene respectively. This obviously presented that both maps carry the overlapping segment of 53 kb, including difference of additional genes along opposite direction of both maps. This could be useful for connecting this part to the further segments on the chromosome or other BACs from the original library.

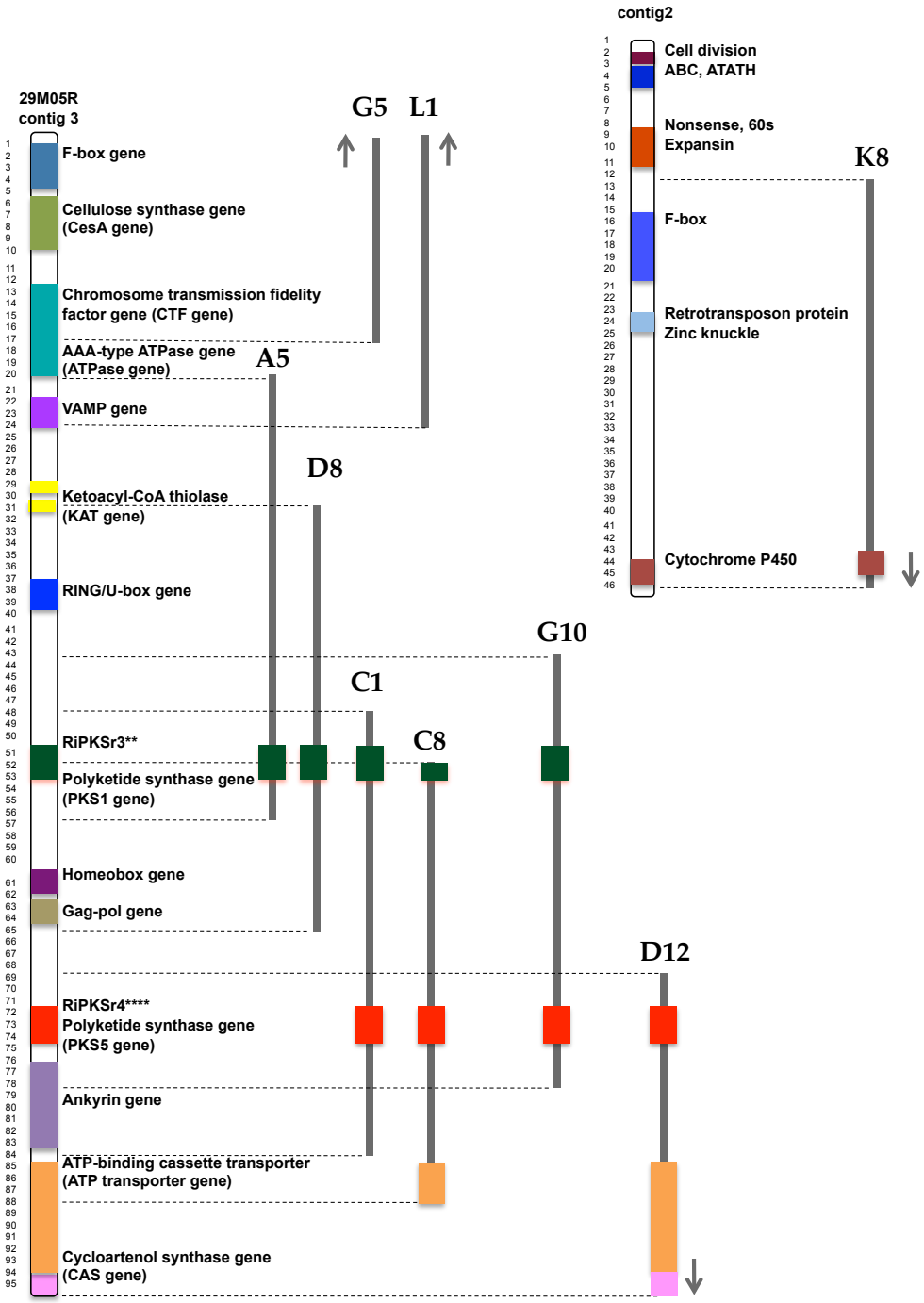


Figure 5.10 A mini constructed physical map for 98 kb of BAC29M05. The map contains thirteen predicted genes and comprises a MTP of eight fosmid clones. Six clones (A5; D8; G10; C1; C8 and D12) of the MTP carried PKS gene regions while two clones (C1 and G10) contained both RiPKSr3 and RiPKSr4 regions coding PKS1 and PKS5 genes respectively.

5.3 DISCUSSION

5.3.1 BAC Sequence Assembly

Following the initial results of 454 sequencing, the number and length of large contigs was important data for determining the quality of 454 sequencing. As greater number of large contigs leads to a higher quality and easier assembly of sequences, shorter lengths may indicate low quality of 454 sequences and there will be some problems for assembly. In this chapter, the success of the assembly of the BACs depended on both the number and size of combined sequences resulting from the previous 454 sequencing. Reviewing average numbers of ~24 thousands reads per BAC was a small amount produced by 454 FLX sequencing. Although this type of sequencing is generally considered to produce long reads of 400-800 bp, an average of read lengths in the BACs was only 273 bp per read. It seemed that the efficiency of 454 sequencing of these runs was less than average.

As an assembly of 454 sequences and paired end sequences of BAC24P12 was successful because of sufficient quality of the resulting 454 sequence. In contrast, it was not easy to combine all contigs and estimate an accuracy size of insert DNA in BAC29M05 and 31B12. Although, adding sequences of BAC paired end sequences could assist in determining an overall length of BAC24P12, it could not be done for the other two BACs (29M05 and 31P12) because of the presence of gaps. Therefore other sequencing such as primer walking is needed to solve this problem of these two BACs. While the two BACs (24P12 and 29M05) always revealed similar restriction digests, PCR, southern blotting and also initial annotation of sequences derived from their PCR amplifications (Chapter 3) and also fosmid paired-end sequences (Chapter 4). This chapter strengthened the similarities between their sequences that confirmed that they might be overlapping BACs. Annotations of longer sequences of these BACs not only revealed similarity but also proved that they were derived from red raspberry genomic DNA, presenting ~3 kb of nucleotide sequence identified as PKS genes of red raspberry (Accession no. EU862821.1) by Blastn aligning. While their assembly provided the two large contigs including ~213 kb (96 and 117 kb) of red raspberry BAC genome, they are new information of this area of the genome.

5.3.2 Fosmid Library Assistances to Assembly

Combination of the available Sanger sequences of fosmid ends against 454 sequences gave an indication of the accuracies of the assemblies of BAC24P12 although there was still up to ~30 kb missing. The assemblies of the BAC29M05 has not been possible at certain points of the contigs with some gaps because of the limitation of random distribution of fosmid clones in BAC. However, since the predicted sizes of gaps were less than the average insert size of the fosmid clones, these gaps could be closed using approaches such as primer/DNA walking. Certain next generation sequencing technologies for instance, Ion torrent sequencing with clone by clone can be used to close these gaps.

Unfortunately, the least successful assembly is for BAC31B12. Although the initial resulting 454 sequences showed that there were certain large contigs, the efficiency was less. Due to more than half the contigs were short (~4 kb or less) and not enough overlap, this resulted in difficulties in connecting them together. Whilst adding BAC paired end sequences could not sufficiently assist for estimating size of DNA insert length. Again, combination with fosmid clones to link between some contigs was not possible to complete an assembly of this BAC. These seemed that 454 sequencing data might be low efficient to combine an entire sequence of this BAC. Selecting and sequencing the nucleotides of a number of clones that covering the entire length of BAC31B12 could solve this failure. However by the reason, if the original extracted DNA of this BAC, which prepared from red raspberry genome, was contaminated with chloroplast or mitochondrial DNA, in additional there could have to be recombination in the BAC, these may result in its sequence not matching with any homologous sequence in nuclear genome (Alkan *et al.,* 2011). This may be a main reason why BAC31P12 did not present any useful sequence information.

5.3.3 Continuity of Genetic Segments within the BAC Sequences

Mini physical map construction using restriction map-based fingerprinting succeeded for both BACs (29M05 and 24P12), presenting approximately 100 kb in length that derived from red raspberry genome (257 Mb). All four regions of RiPKSr1-4 coding either PKS1 or PKS5 genes including other additional genes were placed on the map drafts. Although they could present only a small part of the genome, finding certain interesting genes was useful. Additionally, the order of genes on both maps explained the reason for the presence of an overlap of the both BACs in previous Chapters (with 53 kb overlaps).

The presence of ~1 or 2 genes per 10 kb within certain parts on the two maps correlated to the presence of ~1 gene per 5.7 kb found in the gene-rich regions of *R. idaeus*. Both the PKS gene regions were within the sequence of this gene-rich region, and this indicates that these could be transmitted together. One or both of them may play an important role in the production of anthocyanin compounds associated with this region. Being co-localised on this gene-rich region of 53 kb, means that several other genes would tend to be inherited together. Similar ordering genes within both BACs indicated being the same region of the chromosome or a closely related or duplicated part of the chromosome. Although the original ordering of the genes was opposite (Figure 5.8 A), this might be resulted from the beginning of assemblies with the different ends of resulting 454 sequences of each BAC.

In addition, presence the homologous segments of BAC24P12 in all three contigs of BAC29M05 seemed that the continuity of adjacent genes on the two BACs is similar (Figure 5.8 B), the region of four additional genes (F-box; NMD/EXP; ABC/ATATH and CDV) within the BAC29M05 is the main difference. Due to the assembly of the BAC29M05 sequences could not be completed and with at least two gaps are still obstacles for assembling the entire sequences of this BAC. The further connection of these contigs could prove the true sequences of BAC29M05 including reveal the continuity of the both BACs on this chromosomal element.

However, there were some differences between the two BACs and it was possible that segments of these genes have changed due to crossing over during cell division

leading to gene recombination event. It is possible therefore that the two BACs are a pairing of homologous chromosome. This would explain the differences between them and would be advantageous for further studying both copies of the genes in parallel as well as comparison of these two maps.

5.3.4 Restriction Mapping Fingerprinting of the MTP

Sequencing of the end-paired fosmid was an essential component for confirming the physical maps. Several fosmid clones of the MTP covered almost the entire length of the map constructed from BAC24P12, and they carried genes of interest. Although the map derived from BAC29M05 was constructed from the longest part alone rather than all of the length, it was the most important element since it covered the overlapping part with the other BAC and also contained the PKS gene regions. However the MTP perfectly covered the length of its map.

Although restriction digest analysis may be unreliable on its own, using Cut Map to predict restriction sites and hence overlaps led to greater precision. From the Cut Map nucleotides on the resulting 454 assembled sequences of the BACs, if the assembly was definitely correct, fingerprints could be confirmed. Moreover the expectation in both in number and length of fragments, together with the presence of certain fragments in several clones, supported a more confident use of the restriction map based fingerprint.

High Information Content Fingerprinting (HICF) is considered a powerful tool for either large insert DNA or numerous clones (Fleury, *et al*., 2010). On the other hand, choosing suitable restriction enzymes, using agarose gel electrophoresis, together with predicting the fragmented pattern of DNA fingerprints could be used for manual analyzes for the smaller DNA size (e.g. Fosmid clones) for up to eighteen fosmid clones of MTP per BAC in this thesis. Rare cutting restriction enzymes produced 1-6 bands per clone (~1 kb up to smaller than 13 kb) and were suitable for the length of fosmid clones used (~36 to 40 kb). This therefore allowed the rapid identification of cloned fingerprints, and together was used to resolve problems particularly of both

smaller and larger fragments using agarose gel electrophoresis. However digestion using rare cutting restriction enzymes (8-bp) alone is appropriate for BAC fingerprints but it may be not suitable for fosmid cloned fingerprints. Thus digesting together with a second enzyme (8-bp or 6-bp) increases the clarity of results for smaller clone fingerprinting.

Nevertheless the two constructed physical maps are initial drafts that help to confirm the order of clones of the MTP. Importantly, fragmented digestions of some parts of the maps were dissimilar between electrophoresis and the predictions from Cut Map. Validation of the maps using a clone-by-clone approach is required for the next step.

5.4 CONCLUSION

Assemblies of the two informative genetic sources of red raspberry in this part (BAC clones and the new fosmid libraries) were success. Using next generation sequencing technologies, 454 sequencing associated with the classical Sanger sequencing, could assist a completion of nucleotide sequences within the part that is expected to encode PKSs in red raspberry cv. Glen Moy genome. The association provided the certain new information. Although one BAC failed, the assembled sequences of the other two BAC clones could be available sources for specifically PKS gene study as well as any genes included in these assemblies. Moreover, the physical maps of the two BACs (24P12 and 29M05) were a summary of the regions containing two of the PKS genes and also could be connected to the genetic map previously constructed. The maps revealed new information within the BACs including the proximity of the PKS1 and PKS5 genes involved in the biosynthesis of anthocyanins and thirteen other predicted genes along their lengths. In parallel, fosmid clones of the MTP of the maps were also produced as sources for detailed studies focusing on each gene separately. These could be the beginning of molecular applications such as clone-by-clone sequencing leading to information of the genes contained within the maps, and other manipulations.

CHAPTER 6

Identification and Analysis of Open Reading Frames from Red Raspberry cv. Glen Moy : Polyketide Synthase Genes

6.1 INTRODUCTION

Open reading frames (ORFs) are the regions of the informative nucleotide sequences indicating the functional aspect of genes as they are translated as proteins that carry out functions within the cell (Russell, 2006). Within a given sequence, ORFs can be identified using a variety of software that looks for key features of genes, such as promoter regions; transcription and translation start signals, polyadenylation signals, including analysis for codon bias. While the ORFs can give information as either the nucleotide sequence or the peptide sequence, this Chapter aimed to provide the detailed analysis of open reading frames of PKS/CHS genes at RiPKSr1-4 regions on the physical map derived from assembled sequences of each BACs (from Chapter 5). The two BACs (24P12 and 29M05) were studied on the basis that they contained the information encoding PKS/CHS genes that may be important for characteristic traits of interested. However, other genes presenting within the BACs are also of interest and may provide additional useful information.

6.1.1 General Gene Structure

A typical gene structure in eukaryotes is composed of three main regions including the promoter; coding region and terminator (Lewin, 2004 and Patthy, 2000). This begins with the promoter region followed by non-coding region called *5′ untranslated region or 5′ UTR*, and then an initial exon starting with the first codon (ATG) of translation. In between the first and the final exon, there is a set of introns and exons the number and size of which varies depending on the individual gene. Thereafter followed by the terminating exon, ending with the stop codon (TAA, TGA or TAG). This is ultimately followed by another non-coding region called the *3′ UTR,* and then final terminator polyadenylation (polyA) signal. Schema of eukaryotic gene structure is showed in Figure 6.1.

While the promoter region is the regulator for the level of gene expression, the coding region contains the genetic information for producing a specific protein. To form the mRNA strand leading to protein production, the RNA segments transcribed from the exons are ligated and a process known as splicing discards the introns. The order of

amino acids in a protein is dictated by the nucleotide sequence of the coding region that is presented in the spliced mRNA sequences (Patthy, 2000).

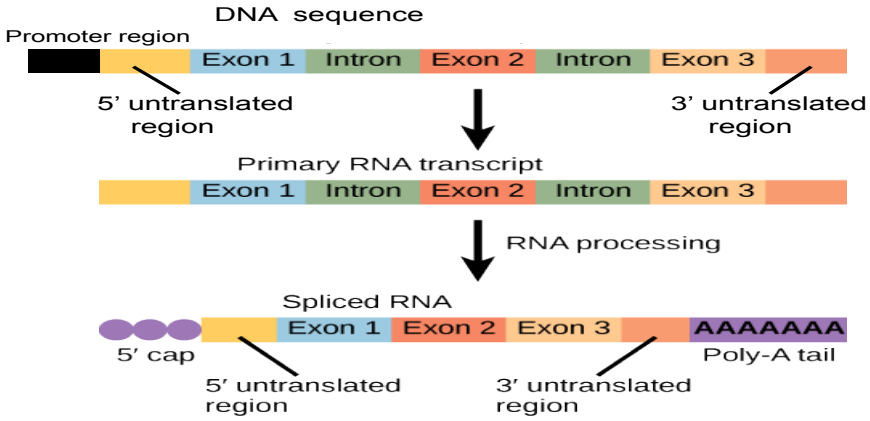


Figure 6.1 Diagram of a general gene structure in eukaryotes composted of three main regions including the promoter; coding region and terminator, begins with the promoter region followed by *5′ UTR*, start codon (ATG) of translation on the first exons; intron; stop codon, *3′ UTR* and polyadenylation (polyA) signal. The exons are ligated and introns are discarded during the process of RNA transcription. (Modified from http://cnx.org/content/m44524/latest/?collection=col11448/latest).

6.1.1.1 PKS Gene Structure

For type III polyketide synthases, the common gene structure consists of just two exons interrupted by an intron. Almost all PKS genes studied in several plants contain only one intron at a conserved position (Durbin *et al*., 2000 and Zheng *et al*., 2001), except for certain plant PKS genes that contain three introns (Ma *et al*., 2009 and Zhang *et al*., 2011). However, the order of the second exons of PKS genes does not

change as shown by comparison of homologous genes in different species. One main difference between PKS genes in diverse species is the size of the untranslated regions (UTRs), both the 5'-UTR and the 3'-UTR. Several studies suggest that the differences in the 5'-UTR influence the choice of alternative transcription initiation sites, while the 3'-UTR has an effect on polyadenylation sites for post-transcriptional processing (Rothnie, 1996), affecting stability, transport and translation of RNA (Flores-Sanchez *et al*., 2010).

6.1.2 Genetic Conservation and Variation

In organisms, genetic variation can occur at each pair of nucleotides, segments of DNA or along an entire chromosome, resulting in modification of conserved genes over the time. Whereas some genetic variation has no consequence, others can profoundly influence characteristic traits in an organism. On the other hand, genetic conservation is an essential fundamental of genetic study and is important not only to relationships between organisms but also in providing information about evolution and biodiversity relating to characteristic traits (Hartl and Jones, 2006). Generally, genes inherit continuities of structures and functions from their ancestors. Genes within the same family or superfamily have common ancestry, resulting in all members share basic structure and function (Patthy, 2000). Although genes are more distantly related, they still can be recognized their similarities.

Following the combination of the two genetic sources in Chapter 5, the assemblies of BAC sequences have formed a valuable source for investigating the PKS genes. To obtain the additional information focusing on the organization and DNA sequence of PKS/CHS genes within the BACs, this chapter concentrates on the PKS gene regions to investigate the nucleotide and amino acid sequence variation which could delivery understanding about PKS/CHS polymorphic variants in the Europe Red Raspberry *R. idaeus* (cv. Glen Moy).

6.1.3 The Aim of the Chapter

This chapter aims to identify the open reading frames of PKS/CHS genes (at RiPKSr1-4 regions) within the assembled BAC sequences derived from the genome of the Europe red raspberry *R. idaeus* and to determine the gene structure, DNA and peptide sequence variation of the genes and also other open reading frames will be analyzed.

6.2 RESULTS

According to nucleotide sequence assembly in Chapter 5, the sequence regions of the two BACs (24P12 and 29M05) containing the raspberry PKS/CHS gene clusters were investigated. An annotation of BAC sequences was carried out, analyzed and ORFs was identified. Moreover, gene structure and DNA sequence variations that may result in the presence of polymorphic alleles were investigated. Certain software was used for analysis shown in Table 6.1, such as GeneMark.hmm-E; BLAST; FGENESH and TSSP, Sequencer sequence analysis and Sequences viewer CLC.

Table 6.1 Softwares used for analysis PKS genes in BACs.

Software program	Function/Useful
GeneMark.hmm-E	Gene and protein prediction program for prokaryote and eukaryotic genomes to find Open Reading Frame and gene start site and stop site
BLAST	Basic Local Alignment Search Tool for comparing amino-acid sequences of different proteins or the nucleotides of DNA sequences and with a library or database of sequences
FGENESH *	Gene structure prediction (multiple genes, both chains) and identify 11% more correct gene models than GENEMARK
TSSP *	Prediction of PLANT Promoters
Sequencer sequence analysis	DNA sequence assembly and analysis tools for next-generation or traditional (Sanger or Capillary Electrophoresis) DNA data sets
Sequences viewer CLC	An informational tool for DNA, RNA, and protein sequence analysis based on comparison the sequences

* (Softberry test online)

6.2.1 Investigation of Open Reading Frames

To identify the ORFs within assemblies/physical maps of the both BACs, nucleotide sequences were analyzed using the gene prediction program for Eukaryotes in finding a gene start site which also shows protein-coding potential (GeneMark.hmm-E* and GeneMark-E* (http://exon.gatech.edu/eukhmm.cgi)) and allowed putative ORFs and the regions containing the predicted ORFs were used to interrogate protein sequence databases using BLASTp and BLASTx.

6.2.1.1 PKS Gene Region Identification

According to the BLASTx alignment, it identified four regions of PKS-like genes of both BACs and represented as RiPKSr1-4. While the first two regions (RiPKSr1 and RiPKSr2) were on BAC24P12, another two one (RiPKSr3 and RiPKSr4) were on BAC29M05. An assembly of BAC24P12 contains two PKS-related gene regions at 73-75 kb (RiPKSr1) and 96-99 kb (RiPKSr2). The BAC29M05 sequence carries the two PKS genes regions at 51-53 kb (RiPKSr3) and 72-76 kb (RiPKSr4) on the 96 kb (contig no.3). On both BACs, the two PKS regions were more than ~20 kb distant from one another on the same BAC (Figure 6.2 A).

In parallel to the above approach, gene finding using FGENESH software (Softberry test online) revealed that the first two sequence regions (RiPKSr1 and RiPKSr2) of BAC24P12 contained ORFs of 393 and 341 deduced amino acid whereas the other two ORFs within RiPKSr3 and RiPKSr4 on BAC29M05 encoded proteins of 391 and 239 amino acids. It is known that PKS genes normally encode proteins of around 391 amino acid residues. This suggests presence at least two PKS genes in both of the two BACs. Inverting the sequence of contig no.3 of BAC29M05 revealed a similar order of genes between this and BAC24P12, indicating the same region being investigated (Figure 6.2 B). Furthermore, as the regions of RiPKSr2 and RiPKSr3 were initially annotated as PKS1 genes and RiPKSr1 and RiPKSr4 were identical as PKS5 genes, the relationship in form of relative positions on these two BACs showed PKS1 genes are next to PKS5 gene regions.

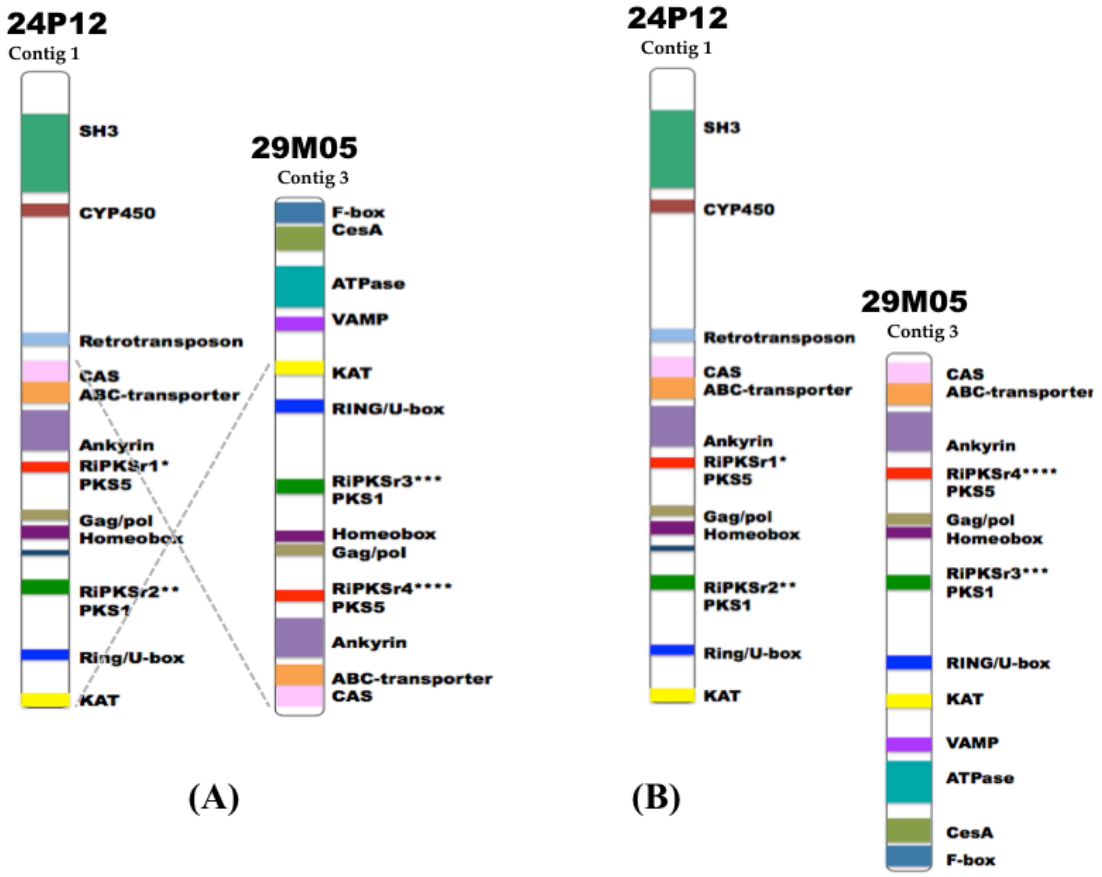


Figure 6.2 Schema of the four PKS gene regions (RiPKSr1-4) on the BACs (24P12 and 29M05). The two regions (RiPKSr1 and RiPKSr2) were at 73-75 kb and 96-99 kb on BAC24P12, and the other two (RiPKSr3 and RiPKSr4) were at 51-53 kb and 72-76 kb on BAC29M05 respectively. **A,** The BAC sequence assemblies appeared to contain two PKS genes regions. **B**, Comparison with inverting order genes of BAC29M05 and those genes of BAC24P12.

6.2.1.2 Comparisons of Protein Sequences

BLASTp identified the 4 sequence regions as encoding proteins that shared significant similarity to PKS/CHS proteins. Further comparative analysis at the amino acid level was carried out to examine this similarity in more detail. The first region, RiPKSr1p encoded a protein of 393 amino acids that was 89% identical to PKS5 of cv. Royalty (Accession no. ABV54603.1), together with to CHS5 of cv. Meeker (Accession no. AAM90650.1/AF400565.1). The second region, RiPKSr2p encoded a protein of 341 amino acids that was 91% identical over this length to PKS1 protein of cv. Royalty. Although the ORF of RiPKSr2p contained both a start and stop codon, fifty amino acids were missing. The third region, RiPKSr3p encoded a protein of 391 amino acids, was 100% identical to PKS1 of cv. Royalty (Accession no. AF292367.1). It also was 99% identical to CHS6 of cv. Meeker (Accession no. AAM90652.1). The last sequence, RiPKSr4p consisted of 239 amino acids, and was 84% identical to PKS5 cv. Royalty over this length, whilst missing both the start and stop codon and the first sixty-two amino acids of the ORF. This information is summarized in Table 6.2

Table 6.2 Summary of PKS regions and their of encoding proteins on the two BACs.

BAC Region/ORF	Length of amino acids	Best Protein Matches	Accession no	% Identity of amino acids
BAC24P12 RiPKSr1p	393	PKS5 cv. Royalty CHS5 of cv. Meeker	ABV54603.1 AAM90650.1	89%(350/393) 89%(349/393)
BAC24P12 RiPKSr2p	341	PKS1 cv. Royalty	AF292367.1	91%(304/333) 90%(300/333)
BAC29M05 RiPKSr3p	391	PKS1 cv. Royalty CHS6 cv. Meeker	AF292367.1 AAM90652.1	100%(391/391) 99%(387/391)
BAC29M05 RiPKSr4p	239	CHS6 cv. Meeker PKS5 cv. Royalty	AAM90652.1 ABV54603.1	85%(274/324) 84%(273/324)

6.2.1.3 Comparison of Four PKS Proteins from cv. Glen Moy

A comparison of predicted PKS proteins showed several different residues among the four regions. The main differences arose within the first 62 and the last 90 residues. Although RiPKSr2p had lost 34 amino acids and RiPKSr4p was missing 154 residues (Figure 6.3), two residues increased in RiPKSr1p. However these sequences could be confirmed by repeating PCR amplification of either genomic DNA of samples, BAC clones or fosmid clone containing these PKS gen regions (in Chapter 5).

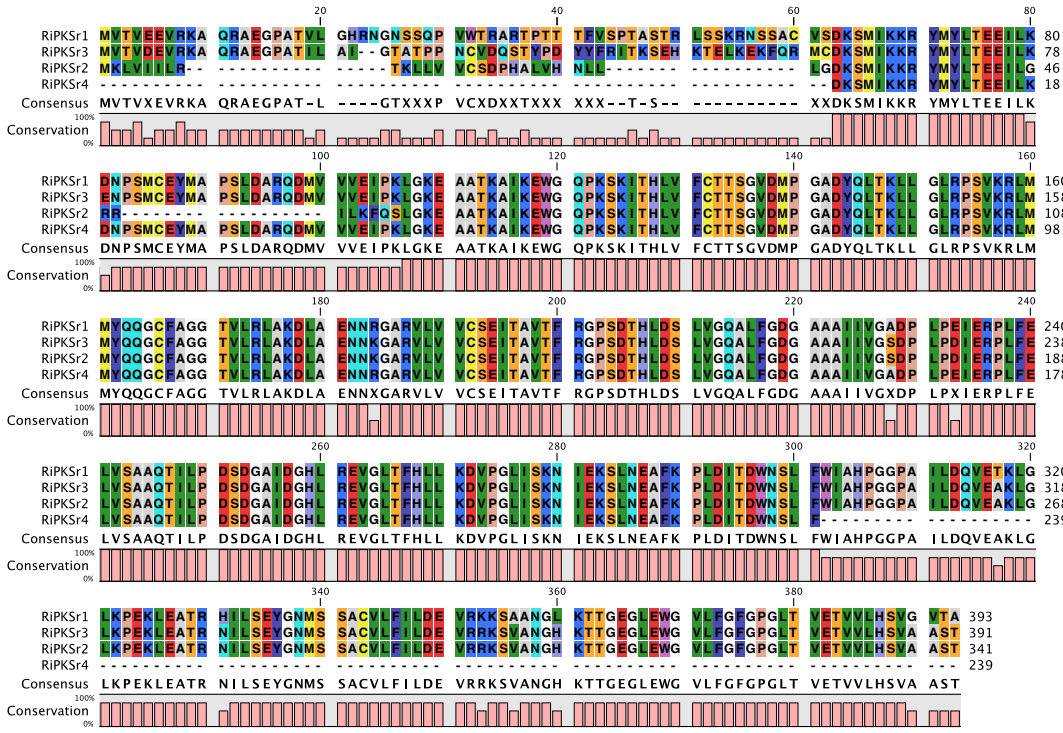


Figure 6.3 Comparison of amino acid sequence of ORFs of PKS regions (RiPKSr1-4) on the two BACs (29M05 and 24P12). The main difference was the first 62 amino acid residues. While RiPKSr2p and RiPKSr4p have been lost 34 and 154 residues respectively.

A pairwise comparison of all 4 proteins is shown in Table 6.3. Despite being shorter, remaining amino acid sequences of RiPKSr4p were 100% identical to RiPKSr1p. It seemed likely from this sequence analysis that RiPKSr1 and RiPKSr4 region should be the same gene but their origin from different BACs. This indicated that these two BACs contained DNA that may have arisen or reduced from different homologous chromosomes.

Although amino acid sequence of RiPKSr2p (341 aa) was shorter than RiPKSr3p (391 aa), with fifty amino acids in different, its sequence was still identical to RiPKSr3p with 91%(304/333). As its difference from RiPKSr1p was 88% (292/332), as almost the same level as RiPKSr3p (87%, 341/392). This demonstrated that both RiPKSr2 and RiPKSr3 region might be the same gene. However, common PKS/CHS genes compost of ~391 amino acid residues (NCBI databases), RiPKSr1p (393 aa) and RiPKSr3p (391 aa) may have typical PKS/CHS functions, these would be examined in further analysis.

Table 6.3 Pairwise comparisons of all four PKS proteins on the two BACs.

BAC	**BAC24P12**		**BAC29M05**	
ORF/Region	**RiPKSr1p (393)**	**RiPKSr2p (341)**	**RiPKSr3p (391)**	**RiPKSr4p (239)**
BAC24P12 **RiPKSr1p (393)**	-	292/332 (88%)	341/392 (87%)	239239 (100%)
RiPKSr2p (341)	-	-	304/333 (91%)	209/239 (87%)
BAC29M05 **RiPKSr3p (391)**	-	-	-	235/239 (98%)
RiPKSr4p (239)	-	-	-	-

% Identity

6.2.1.4 Detailed Comparison of Predicted PKS Proteins from cv. Glen Moy

The initial comparison identified the ORF of RiPKSr1 region (on BAC24P12) as 89% identical to PKS5 gene of cv. Royalty while the ORF of RiPKSr3 region (on 29M05) was 100% PKS1 gene. A detailed comparative analysis of amino acid residues of both PKS proteins revealed fifty-two different residues distributed over the ORFs. Main difference was at positions of 21 to 60 downstream of the amino-terminal methionine (M), of RiPKSr1p from BAC24P12 (Figure 6.4).

Comparison to other PKS protein sequences revealed that RiPKSr1p contained forty-two and fifty-two different amino acid from PKS1; PKS5; PKS6; CHS5; CHS6 and CHS11 from other species. On the other hand, RiPKSr3p showed less difference with only thirteen to fourteen residues different from other PKSs (Figure 6.5).

The 391 amino acids of RiPKSr3p includes all the active site residues (Cys164, Phe254, His303 and Asn336) (Kassim, 2009 and Jez *et al*., 2001). The amino acid substitutions in comparison to cv. Royalty and Meeker using Sequencer viewer 6 are shown in Figure 6.6 A. Comparison to CHS6 from cv. Meeker showed four different amino acid residues at positions 226; 231; 329 and 354 but completely similar to PKS1 cv. Royalty.

Certain amino acids, (Pro138; Gly163; Gly167; Leu214; Asp217; Gly262; Pro304; Gly305; Gly306; Gly335; Gly374; Pro375 and Gly376) that are important in shaping the active site of CHS (Radhakrishnan, *et al*., 2009), are found in both of the predicted PKSs of cv. Glen Moy. The amino acids those were required for coumaroyl CoA binding (Ser133; Glu192; Thr194; Thr197 and Ser338) are also conserved (Appendix A.18).

Comparison of RiPKSr1p (393 aa) to cv. Royalty and Meeker are shown in Figure 6.6 B, carried two additional residues (Arg23 and Asn24) and several different residues. Interestingly, Blast-searched alignments did not showed identification of RiPKSr1p to PKS5 of red raspberry cv. Glen Moy (with 767 nucleotides and 67 aa, Accession no. GU117608.1) in GeneBank databases. However, comparison of the sequences using

Sequence viewer analysis showed longer sequences of both RiPKSr1p (393 aa) and RiPKSr4p (239 aa) than PKS5 sequence submitted in GeneBank (Figure 6.6 C).

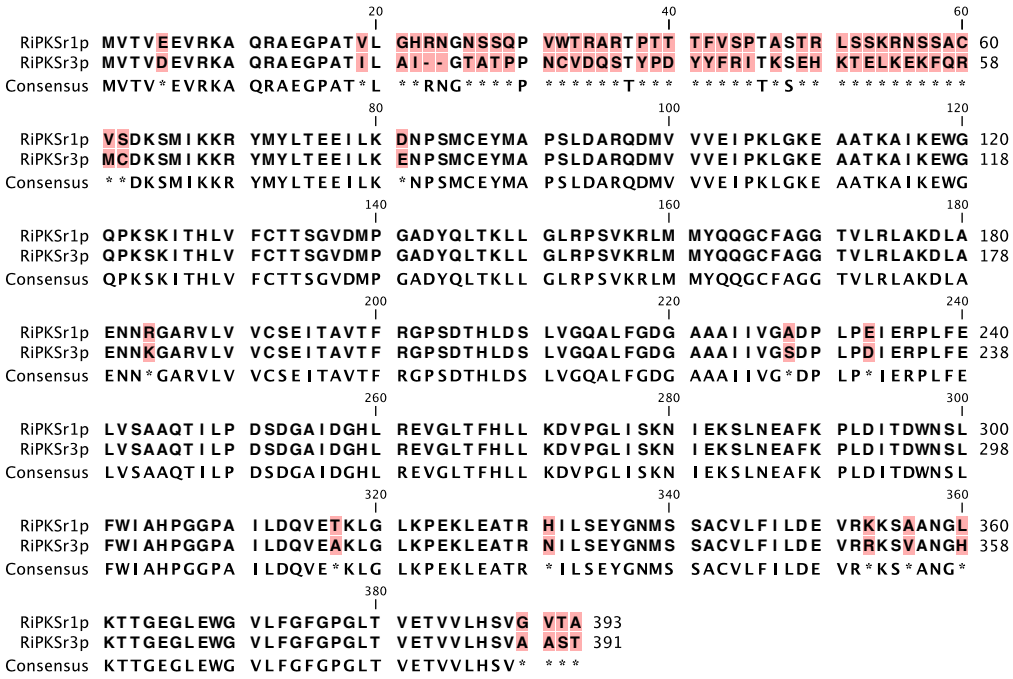


Figure 6.4 Comparison of amino acid sequences in the ORFs of the two regions at RiPKSr1 (BAC24P12) and RiPKSr3 (BAC29M05) derived from raspberry cv. Glen Moy. Differences from each other are shown in pink shading.

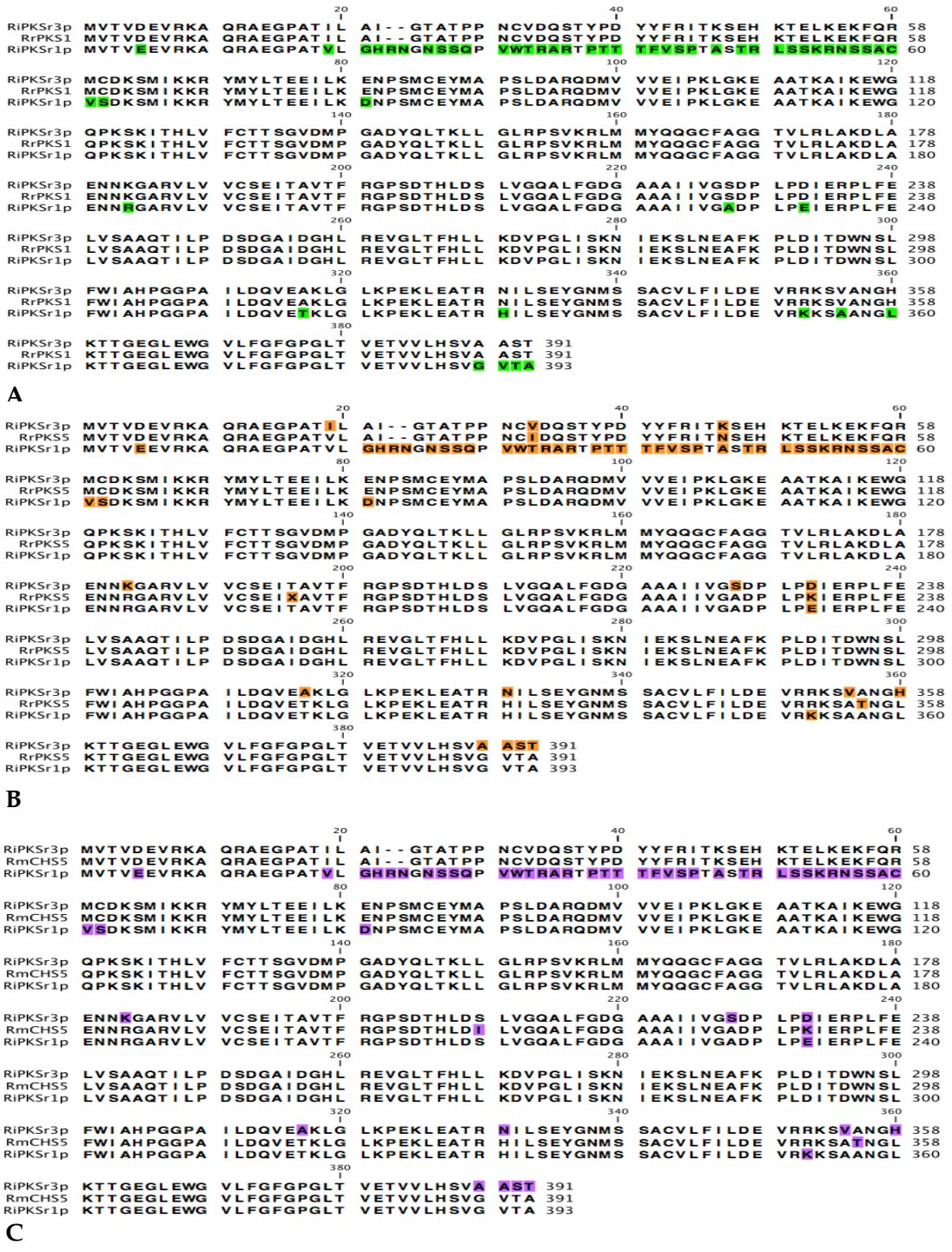


Figure 6.5 Comparison of amino acid sequences in the ORFs of the two regions at RiPKSr1p (BAC24P12) and RiPKSr3p (BAC29M05) derived from raspberry cv. Glen Moy to **A**, PKS1, and **B**, PKS5 proteins of cv. Royalty and **C**, CHS5 proteins of cv. Meeker. Differences are shown in coloured shading.

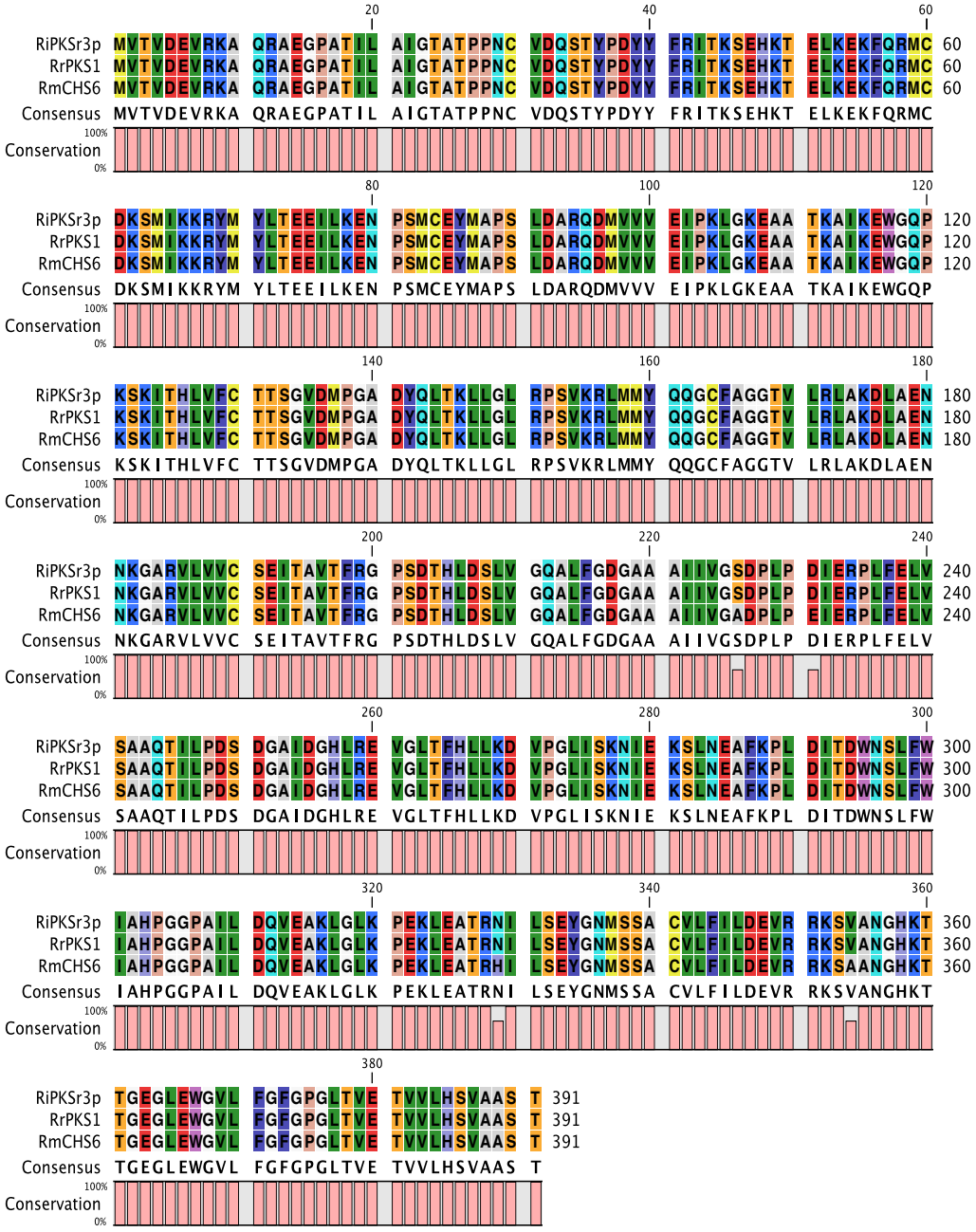


Figure 6.6 A Comparison at amino acid sequence level of RiPKSr3p (BAC29M05) to PKS1 of cv. Royalty and CHS6 of cv. Meeker using Sequencer CLC (viewer 6). RiPKSr3p (at 29-48 kb of contig no 3) of cv. Glen Moy containing 391 deduced amino acids was completely identical to PKS1 cv. Royalty but showed four different amino acid residues at 226; 231; 329 and 354 from CHS6 cv. Meeker.

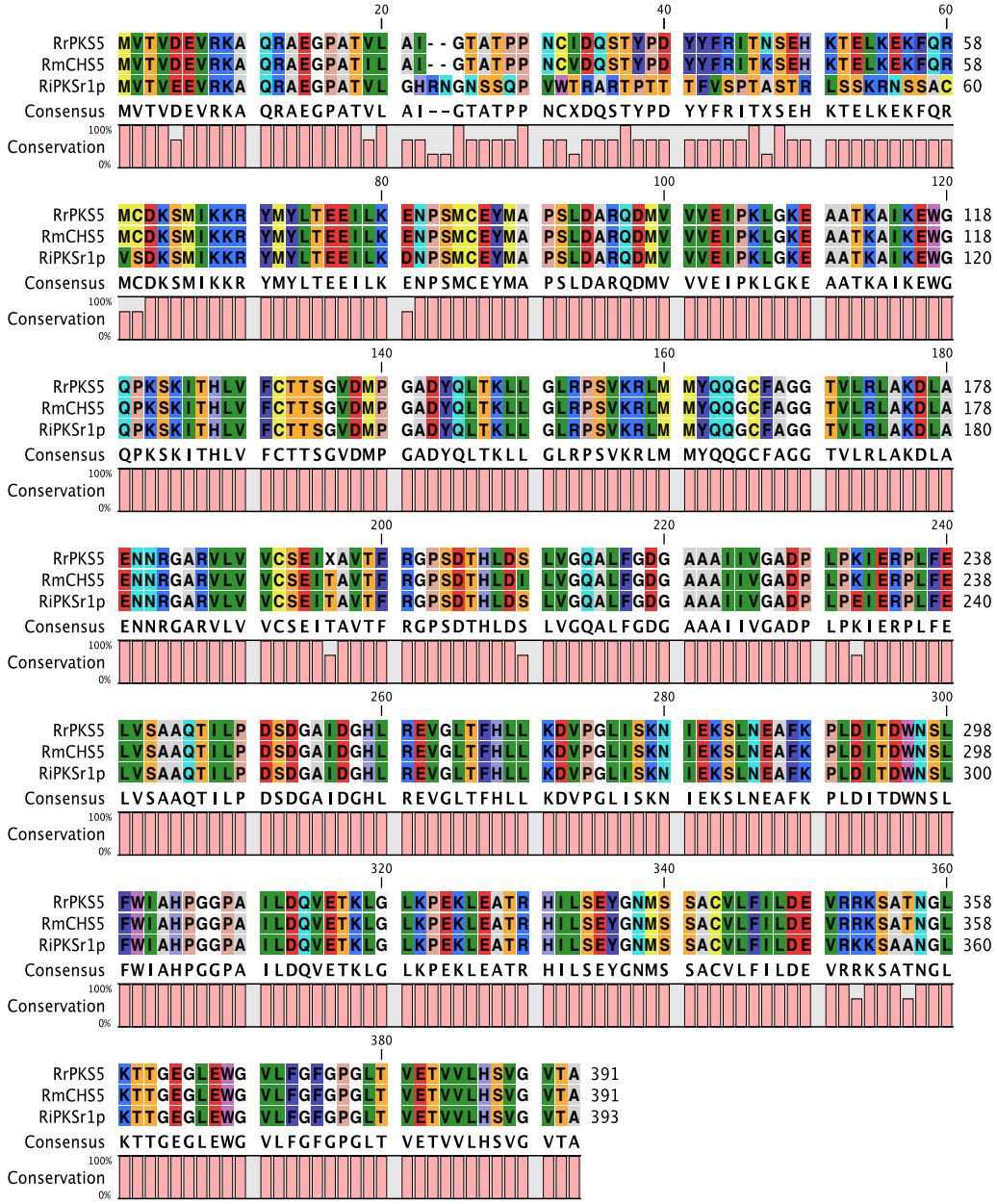


Figure 6.6 B Comparison at amino acid sequence level of RiPKSr1p (BAC24P12) to PKS5 and CHS5 of cv. Royalty and CHS11 of cv. Meeker using Sequencer CLC (viewer 6). RiPKSr1p (at 73-79 kb) of cv. Glen Moy encoding 393 amino acids carried two additional residues (Arg23 and Asn24) and several different residues.

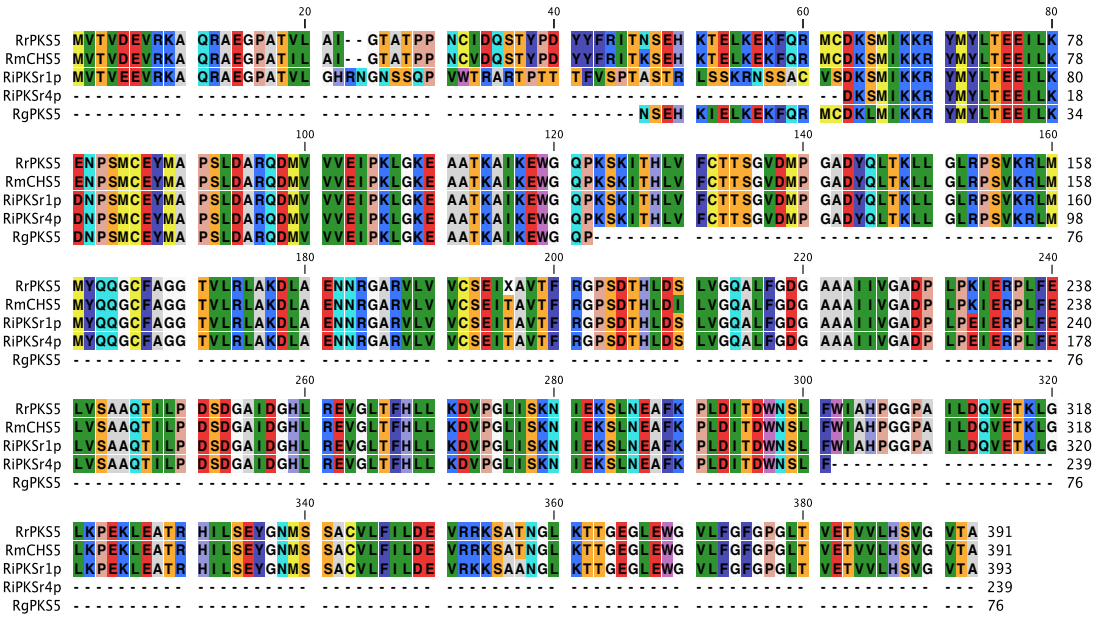
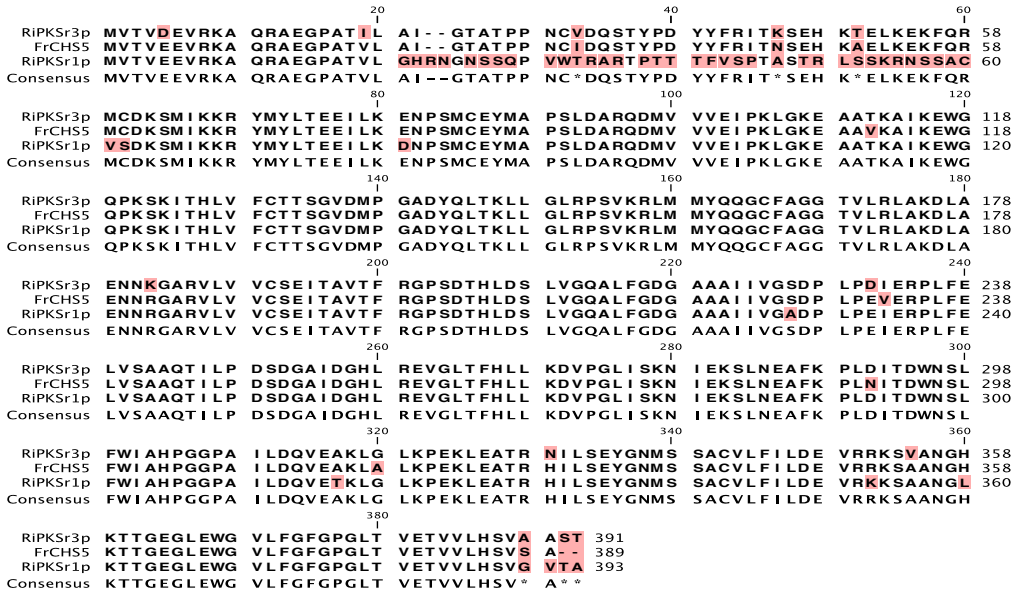


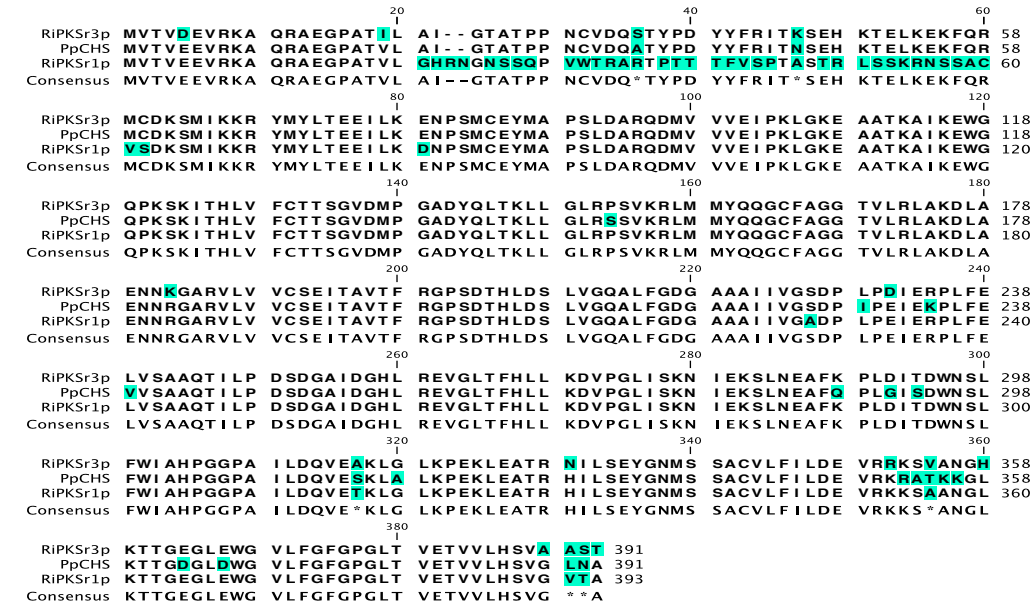
Figure 6.6 C Comparison of amino acid sequences of RiPKSr1p and RiPKSr4p to PKS5 of cv. Glen Moy (RgPKS5) submitted in GenBank databases, PKS5 of cv. Royalty (RrPKS5) and CHS5 of cv. Meeker (RmCHS5) using Sequence viewer CLC software. Both RiPKSr1p (393 aa) and RiPKSr4p (239 aa) showed longer sequences than of PKS5 submitted in GeneBank (76 aa).

6.2.1.5 Comparison of PKS Protein Sequence of cv. Glen Moy to Related Fruits

To compare the predicted PKS protein sequences of cv. Glen Moy to related fruits in the same family, commercially important fruits such as strawberry (*Fragaria vesca* subsp. Vesca, FrCHS) and prune (*Prunus persical*, PpCHS), were chosen to compare at the amino acid sequence level. Sixteen and twenty-nine amino acid residues of RiPKSr3p in BAC29M05 differed from FrCHS5 and PpCHS respectively. RiPKSr1p of BAC24P12 revealed more differences with forty-six to fifty-five residues from the two related fruits (Figure 6.7). As they are more distantly related, the comparison showed higher dissimilarities than within the same genus. However key residues, particular at active sites were conserved as previously indicated (described in 6.2.1.4).



A



B

Figure 6.7 Comparison of Glen Moy PKS amino acid residues to two related fruits in the same family. **A**, RiPKSr3p of BAC29M05 and RiPKSr1p of BAC24P12 compared to FrCHS5 of strawberry (*Fragaria vesca*). **B**, RiPKSr3p and RiPKSr1p compared to PpCHS of prune (*Prunus persical*). Differences are shown in coloured shading.

6.2.2 Analysis of PKS Nucleotide Sequences

At the nucleotide level, the RiPKSr1 region from BAC24P12 was 97% identical to PKS5 from cv. Royalty and 89% identical to CHS5 cv. Meeker. The coding sequence in the first exon showed a six base pair nucleotide insertion (at nucleotide position 184) but the second exon was still 998 bp in length. The intron was 538 bp in length 155 bp longer than the PKS1 gene from cv. Royalty (Appendix A.19). This resulted in the two additional amino acids (Arg23 and Asn24).

As the RiPKSr3 region from BAC29M05 indicated 100% identity at the nucleotide sequence level to the PKS1 gene of cv. Royalty. The genomic sequence contained two exons of 178 bp and 998 bp and an intron of 382 bp which is 1 bp less than PKS1 cv Royalty) (Appendix A.20).

6.2.2.1 Variations in Intronic Sequences

Comparison of the two introns indicated differences in the nucleotide sequences of RiPKSr1 and RiPKSr3 regions and other PKS genes. The intron of RiPKSr1 was 97% (526/540 aa) identical to that from PKS5 cv. Glen Moy (Accession no. GU117608.1) and 93% (513/551 aa) identical to PKS5 cv. Royalty (Accession no. EF694718.1) but did not align with CHS5 of cv. Meeker (Appendix A.20).

The intron in RiPKSr3 region was 100% identical (382/382 nucleotides) to the intron sequence of PKS1 cv. Royalty (Appendix A.21 and A.22). This strongly supports that hypothesis that RiPKSr3 sequence is PKS1.

6.2.2.2 Regulatory Regions and Variations

Regulatory motifs of PKS gene were successfully identified in the two BAC PKS genes. Coding regions of PKS in both BACs were also homologous at both 5' and 3' ends of those genes (Appendix A.23 and A.24).

A TATAA box motif (TATAAATA) was located at 107-115 bp upstream of the ATG translation start codon, in RiPKSr1 (BAC24P12) but two TATAA boxes were found at

143-150 and 431-439 bp upstream in RiPKSr3 (BAC29M05). In cv. Royalty, this motif was located at 108-114 bp (Kasim, 2009). Another sequence, CACGTG, motifs of bHLH reported to be involved in the light-inducibility of CHS (Martin, 1993), were found in two places, at 171-176 and 223-229 bp upstream of the ATG in RiPKSr1 and at 197-203 and 241-246 bp in RiPKSr3. While this motif was early on at 115-120 bp in Royalty (Kasim, 2009). The final hexameric sequence, AATAAT, motif recognized by RNA cleavage enzymes binding for polyadenylation, was at 965-971 bp downstream of stop codon TGA in RiPKSr1 but located in two places at 169-174 and 1106-1112 bp downstream in RiPKSr3. This same motif was found at 231-236 bp downstream in cv. Royalty (Kasim, 2009).

6.2.3 Variation in Other Genes

In addition to the PKS regions, other gene sequences were identified in the two BACs. An investigation of these regions may give information on genetic variation between homologous chromosomes and cultivars that may be useful in interpreting the PKS gene information.

6.2.3.1 Identification of the Other Genes

The three genes chosen to examine were Cytochrome P450 (CYP450); Cycloartenal synthase gene (CAS) and ATP-binding cassette transporter (ABC). The gene regions were on the two contigs (46 and 96 kb) of BAC29M05 and another set was on the contig of 117 kb of BAC24P12.

6.2.3.1.1 Cytochrome P450

The predicted region of Cytochrome P450 gene sequence (RiCYP450r1) was at 22-28 kb on BAC24P12 assembly, the ORF of RiCYP450r1p coding 482 amino acids was identified as 93%(482/485) to CYP450 of *F. vesca* subsp. *Vesca* gene. BLAST alignment revealed thirty-one dissimilar amino acids while three amino acids missed (no result shown).

One other region of Cytochrome P450 gene (RiCYP450r2) was located at 42-46 kb of the 46 kb contig on BAC29M05. The predicted ORF (RiCYP450r2p) contained 454 amino acids (1362 bp) consisting of both start and stop codons. A BLAST search identified it as a CYP450 superfamily member showing 87%(453/485 aa) homology with a *Fragaria vesca* subsp. *Vesca* gene (Acession no. XP004296643). Comparison showed thirty-two different amino acid residues while the same amount loosed.

Comparison of amino acid sequences between the two regions (RiCYP450r1p and RiCYP450r2p) with *F. vesca* subsp. *Vesca* was shown in Figure 6.8. Both regions showed twenty-five amino acid residues differencing from *F. vesca* but RiCYP450r2p was more dissimilar with twenty-nine missing residues.

Figure 6.8 Comparison of amino acid sequences of the two regions, RiCYP450r1p (on BAC24P12) and RiCYP4502rp (on BAC29M05) respectively, and *Fragaria vesca* subsp. *Vesca* gene. Differences from others are shown in pink shading. RiCYP450r1p showed twenty-five amino acid residues differencing from *F. vesca* while twenty-nine residues loosed from RiCYP450r2p.

6.2.3.1.2 ABC Transporter

On the 96 kb contig of BAC29M05, the ATP-transporter gene (ABC, RiABCr2p) was located at 85-94 kb. The ORF encoded 846 amino acids whilst being identified as 73%(577/794 aa) identical to an *Arabidopsis* gene (Accession no. 002885611.1). It also was 79%(641/811) and 83%(604/726) identical to hypothetical protein PRUPE sll0005-like of *Prunus persica* and an uncharacterized gene of *Fragaria vesca* subsp. *Vesca* gene respectively. As another ATP-transporter region (RiABCr1) of BAC24P12 was at ~55-64 kb. The ORF of 626 translated amino acids was identified as 80% (602/804) of protein ABC1, mitochondrial precursor, putative of *Ricinus communis* (Accession no. XP002518128.1). It also showed 83%(528/726) and 86%(496/811) identity to uncharacterized genes of *Fragaria* and hypothetical protein PRUPE sll0005-like of Prunus whilst being 80%(462/794) identity with *Arabidopsis* gene.

Furthermore, an alignment of amino acid sequences between both BACs was shown in Figure 6.9, and comparisons to *Arabidopsis* and *Ricinus.* The results revealed many different amino acid residues while RiABCr1p sequence was shorter than RiABCr2p with proximately two hundred and twenty amino acid residues.

In summary, amino acid sequences of both ORFs were identical to the plant genes and comparisons of RiABCr1p and RiABCr2p (Figure 6.10 and 11). Differences between the ORFs of raspberry were dissimilarity at positions of the first one hundred residues and loss of one to two hundreds amino acid residues along its sequence while adding proximately fifty to one hundreds residues at the end of the ORFs.

Figure 6.9 An alignment of amino acids of ORFs of ABC gene regions (RiABCr1p, on BAC12P12 and RiABCr2p on BAC29M05). Approximately forty-four amino acid residues were different between the two ORFs while approximately two hundred and twenty residues of RiABCr1p missed.

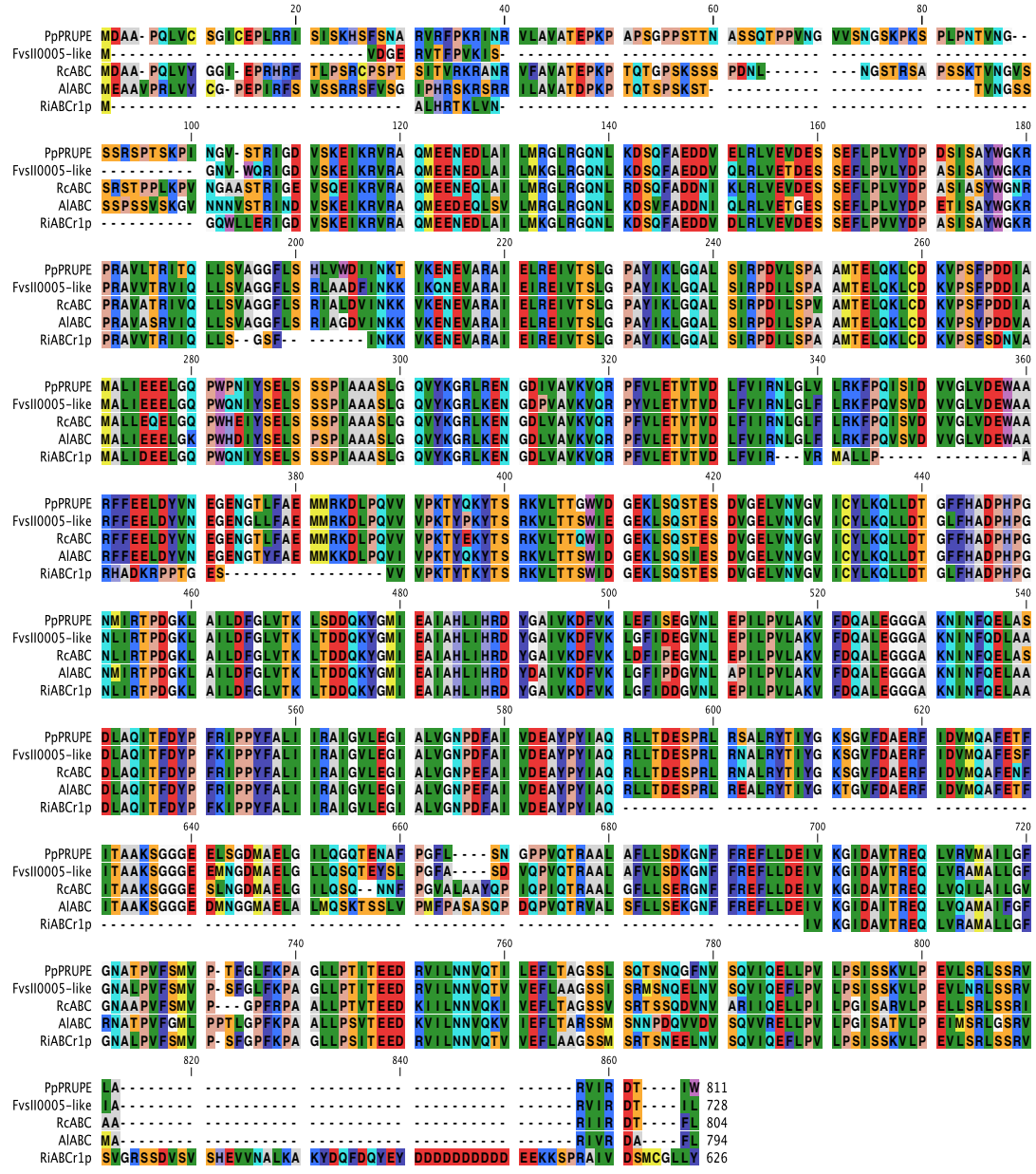


Figure 6.10 Alignment of ORF of RiABCr1p (the 626 amino acids, on BAC24P12) to plants; hypothetical protein PRUPE sll0005-like of *Prunus persica* (PpPRUPE, 811 aa) and uncharacterized protein sll0005-like of *Fragaria vesca* (Fvsll0005-like, 728 aa), protein ABC1, mitochondrial precursor, putative of *Ricinus comunis* (RcABC1, 804 aa) and *Arabidopsis* (AlABC, 794 aa), showed difference since having dissimilar and additional amino acid residues while many others were missed.

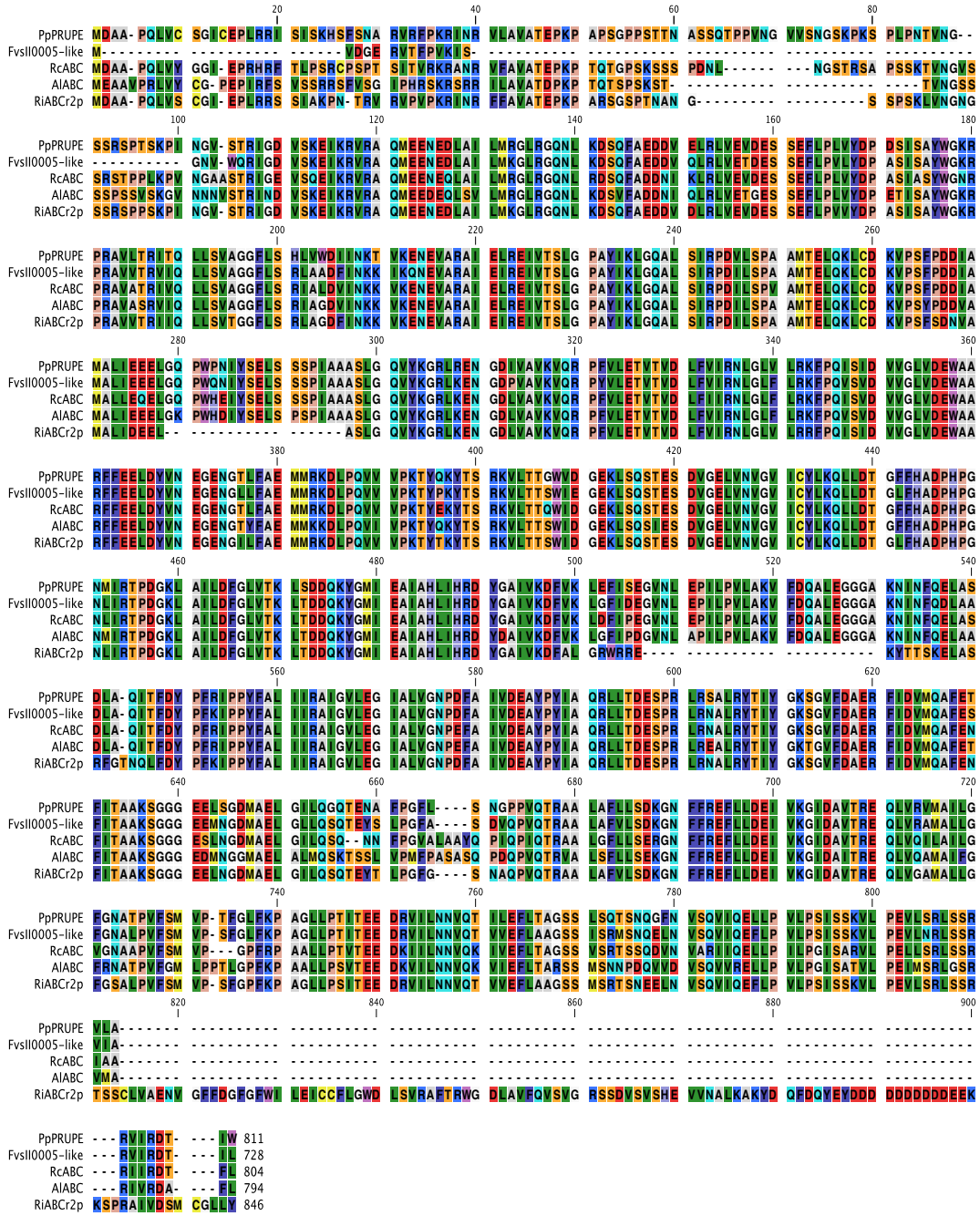


Figure 6.11 Alignment of ORF of RiABCr2p (the 846 amino acids, on BAC29M05) to plants; hypothetical protein PRUPE sll0005-like of *Prunus persica* (PRUPE(Pp), 811) and uncharacterized protein sll0005-like of *Fragaria vesca* gene (sll0005-like Fv, 728 aa), protein ABC1, mitochondrial precursor, putative of *Ricinus communis* (Rc, 804 aa) and ABC of *Arabidopsis* (Al-ABC, 794 aa), showed difference of among additional and missing amino acid residues.

6.2.3.1.3 Cycloartenal Synthase

Another region on the 96 kb contig of BAC29M05, Cycloartenal synthases gene (CAS, RiCASr2) was at 94-96 kb, encoding a protein of 213 amino acids. As one other sequence of CAS gene was at 50-54 kb of BAC24P12 (RiCASr1), containing 167 amino acid. Generally, since CAS gene composts of proximately 455-790 peptide residues, the CAS regions of both BAC (167 to 231) could present only a short length of sequences. However an alignment of the two regions presented similarity of amino acids whilst showing difference resulted from loosen residues.

An alignment between RiCASr1p and RiCASr2p showed forty-four different amino acids, in Figure 6.12. The RiCASr2p was identified as 61% (87/142 of 732 aa) of CAS *Glycine max* (Accession no. XP003556307.1) as RiCASr1p was identical to the same genes but less percentages (no result shown).

Moreover, this gene was highly similar to the lupeol synthase gene (LUP) having several medicinal properties including anti-inflammatory action. However, RiCASr2p was identified only as 66% with 89/135 out of 769 aa) of LUP *Fragaria vesca* subsp. *Vesca* (XP004308480.1); 60%(86/143 of 754 aa) of *Vitis vinifera* (XP004308480.1) and 62%(73/118 of 758 aa) of *Olea europaea.* (BAA86930) (no result shown).

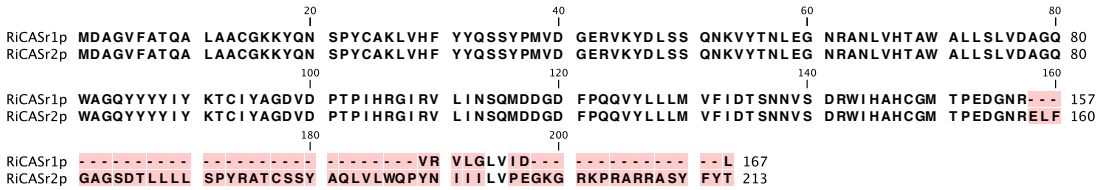


Figure 6.12 An alignment of amino acid sequences of Cycloartenal synthases gene of RiCASr1 (BAC24P24) and RiCASr2 (BAC29M05). The alignment presents forty-four different amino acid residues between both genes. CAS genes contain ~ 455-790 peptide residues but the CAS regions of both BACs could present only a short length of sequences (with 167 to 231 amino acid residues).

In summary, comparisons of both nucleotide and peptide sequences of the regions of the three gene presented additional information of the two BACs. Although each gene showed similarities of their sequences, they also demonstrated significant differences in both BACs. Together with the identification of the same gene sequence from each BAC as different proteins in the GenBank databases could support dissimilarities of genetic information between the both BACs.

6.2.3.2 PCR Amplification of the Other Genes

To validate the sequence of these genes and check for their presence in various cultivars, extracted genomic DNA of the two cultivars (Glen Moy and Latham) was amplified by PCR using primers designed to amplify the three gene sequences (ABC transporter, CAS and CYP450 (Table 2.3 B)). They were successfully amplified from both species as expected sizes that CYP450 showed fragments shorter than 700 bp while ABC and CAS were ~700 and 800 bp (Figure 6.13). The primers were also used for amplifying these expected gene sequences on both BAC (29M05 and 24P12) as they showed ~700 to 800 bp in length (no result shown). Following sequencing and BLAST searches, the amplified nucleotides were identified as the expected genes.

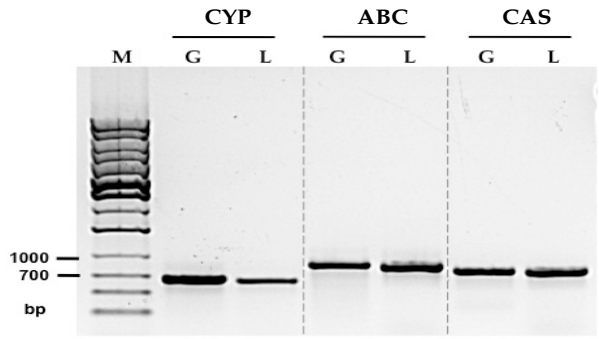


Figure 6.13 An amplification of genomic DNA of red raspberry cv. Glen Moy (GM) compared cv. Latham (L) using CYP450, ABC and CAS gene primers. The PCR Reaction were denatured at 94°C for 5 minutes and amplified by denaturing at 94°C for 1 minute, annealing at 64°C for 1 minute and extension at 72°C compared to 1 kb Molecular standard.

6.3 DISCUSSION

Analysis of an entire sequence of genes of interest the both BACs in this chapter gives much more information needed. Based on nucleotide sequence assemblies of the BACs (24P12 and 29M05), an annotation, identification and analysis demonstrated variation of PKS/CHS genes to investigate and compare their detailed information, which successfully revealed information of uncharacterized sequences of those BACs.

6.3.1 Allelic Variants of the PKS Genes

Several PKS gene regions were identified, with the same order and distance along at least ~96 kb of both BACs. This indicated that they might be individual members of a paired homologous chromosome. This was supported by the presence of the other identified ORFs, in the same order on these chromosomes (Figure 6.2).

While diploid organisms such as human, animals and plants have paired homologous chromosomes resulting in two similar or distinct copies of each gene, an individual member of each paired chromosome is inherited from the female parent, and another comes from male parental chromosome. An allele is a form of a gene, located at a specific locus on homologous chromosomes. It is one of several possible alternative forms of the gene. When the two genes are alleles of each other, they both contribute to controlling the same characteristic trait. Whilst containing two dissimilar alleles resulting in different expression of the characteristic trait, an organism is called a heterozygote, as the presence of two identical alleles is called a homozygote.

Pairs of dissimilar alleles can result in different visible characteristics of organisms, such as skin pigmentation in animals, and colours of flowers in plants. Following variations through the evolutionary process, a wild type is the “normal version” while alternative alleles are considered as new changes, modifications or mutations. Detailed analysis of deduced amino acid sequences that were translated from the nucleotides of both regions of each PKS gene obviously identified to polymorphism alleles of the genes in cv. Glen Moy.

6.3.2 A Cluster of Two PKS Genes on Each Chromosome

Finding two PKS regions on each BAC suggested that there might be two functional PKS genes on this locus of the chromosome. On the other hand, they might be duplicated genes or multiple regions of the PKS gene on this chromosome segment. In addition, finding twice and being far ~20 kb from another one may be near enough to be transmitted together from their parental chromosomes and to the next generations. In other words, these genes are linked and likely to be coinherited.

6.3.3 Specie Differences

6.3.3.1 PKS Gene Organization

The gene organization of type III polyketide synthase does not appear to be different in this cultivar. The order and number of exons and intron of PKS cv. Glen Moy still were conserved with two exons and an intervening intron that like in several plants (Durbin *et al*., 2000 and Zheng *et al*., 2001).

6.3.3.2 PKS Protein Variation

Comparison to two related fruits confirmed relationship between the two ORFs with PKS genes of plants in the same family. Although they presented more dissimilar, among active sites of PKS family still were conserved. However there are certain differences among PKS genes of the cultivar. In exons, all amino acids were conserved at the active sites, although the residues were moved ahead. While intron sequences were highly maintained at 97%-100% homology. Interestingly, the main differences of these genes were at the beginning of ORFs, the position at 1-62 of the first exon. Although active site for PKS family are at 65 to 375 (described in 6.2.1.4), changes of among amino acid before these positions may result in malfunction for anthocyanins production in plants. By another reason of loosing several residues and either the start or stop codon, they became pseudogenes lead to be malfunctioning.

Although several amino acids had been lost, remaining residues still were highly conserved in PKS1 and PKS5. Therefore there should be two copies of PKS1 genes and another two of PKS5 in this locus on the chromosome. This revealed to the order of PKS1 and PKS5 genes by comparing of a paired homologous chromosome in raspberry cv. Glen Moy.

6.3.3.3 Nucleotide Analysis and Variation

Although the regulatory motifs of PKS/CHS (PKS1; 2 and 3) of cv. Glen Moy have been reported in Kassim (2009), some of the same motifs were found twice in this thesis, as another difference was longer length of both 5' UTR and 3' UTR sequences. Conservation of amino acid residues at these positions can lead to form of the typical PKS/CHS reaction while variations might result in non-chalcone type products.

While Kasim (2009) also revealed the presence of one polymorphic allele of PKS1 gene with 391 amino acid residues in this cultivar, this chapter not only confirmed the presence of the completed nucleotide sequences of the PKS1 gene but also reported sequence polymorphism of another allele leading to a number of dissimilar and missing amino acid residues by analysing based on the resulting 454 sequences. As a remarkable variation of both nucleotides and amino acids was significantly observed, the PKS gene regions have become more important sources in being useful for allele mining approaches of polymorphic alleles of the PKS genes in raspberry. Together with, the information can be applied for gene manipulation to study in influences of the alleles on gene expression.

At the final analysis, the identification of other three genes supported being a more benefited source of the BACs. However there still are certain genes involving to biosynthetic partway of flavonoids as well as catalytic disease prevention (presented in Chapter 5) that could be genes of interest in future study such as KAT gene coding an enzyme in synthetic pathways of flavonoid compounds but it could not analyse all of them in this thesis. As it seemed that the two BACs might be a paired homologous chromosome, this could be benefited much for studying in both two copies of genes in parallel.

However, although four PKS/CHS genes regions have been found in this thesis, only two genes may be possible allelic variations while the other two are pseudogene genes. Confirmation of BAC sequencing and assembly must be done by further sequencing of the clone of MTP of each BAC, and this would explain the large degree of variation, deletions and apparent null alleles seen for numerous genes.

6.4. CONCLUSION

The investigation of variation of organization and structure of PKS/CHS genes in the BACs revealed uncharacterized information in this segmental chromosome of cv. Gen Moy genome. The investigation of PKS genes found on both BACs indicated that the BACs might be an individual of a paired homologous chromosome or the same chromosome. This chapter also presented a longer sequence of PKS5, presence of polymorphic alleles of PKS1 and PKS5 in BAC assemblies/physical maps, and the detailed annotation of the other three genes as these all information could be basic for additional studies in *R. idaeus* cv. Glen Moy to better understanding in red raspberry genome.

CHAPTER 7

General Discussions and Conclusion

7.1 GENERAL DISCUSSION

Analysis of the three BACs derived from the genome of red raspberry cv. Glen Moy provided additional detailed information on this area of the genome. The details of the BACs were obtained by systematic construction of fosmid sub-libraries, assemblies of their sequences, annotation and investigation of the PKS genes including generation of a physical map draft of the BACs. Significantly, finding the overlap of BAC29M05 and BAC24P12 is the main reason of similarities of data along the lines described in the various chapters whilst containing a complete PKS1 gene sequence, currently identified as Naringenin chalcone-PKS by Zheng (2014) in NCBI. In this chapter, the main findings of this thesis will be summarized and discussed in relation to the work of others, and future work will be proposed.

7.1.1 Details of BACs

An initial analysis of three BAC clones from the library derived from red raspberry cv. Glen Moy provided general characteristics of those clones in Chapter 3. The two BACs (29M05 and 24P12) were similar in all analyses including the presence of the PKS/CHS gene sequences as the similarity between BAC29M05 and BAC24P12 was explained by a 53 kb overlap described in Chapter 5 and 6. As BAC31B12 appeared different based on evidence from certain analyses, except for the presence of PKS-related sequences that were similar to the first two BACs in Chapter 3 alone. There might be the probability of PCR artifacts and it needs to be explored in further analysis.

There was a set of PKS genes (PKS1; PKS2; PKS3 and PKS5) sharing high identity with CHS-type PKS (~82-91% identity with amino acid sequence) in BAC29M05 and BAC24P12. A sequence derived from a PCR product amplified using a primer for PKS1 gene aligned likely with the PKS5 gene confirming high similarity between PKS1 and PKS5 as both code for enzyme-CHS for anthocyanin synthesis (Zhang, 2001). Interestingly, BLASTp alignment showed that these sequences shared with leucoanthocyanidin dioxygenase (LDOX) (with lesser than 40% amino acid sequence identity), which is an alternative name anthocyanidin synthase (ANS). As the gene is

an essential for proanthocyanidin synthesis (Figure 1.2), its sequences share homology to flavonol synthase (FLS) gene at 49 to 87% (Pelletier *et al*., 1997). Due to these genes subsequently are in the same synthetic pathway, this is evident of presence high similar sequence of genes, which are in adjacent pathways including complexity resulting in diversity of the genes in this pathway as well as gene family.

The same patterns of hybridized fragments that detected by Southern blot analysis using the probe generated from PKS1 gene sequence suggested the existence of two PKS-like genes in both BAC29M05 and 24P12. Subsequently these were identified as PKS1 and PKS5 gene sequences in Chapter 6. This indicated that there are no obvious polymorphisms by hybridization in the two PKS genes in this part of the red raspberry genome. Interestingly, the analysis also demonstrated that the hybridizing bands, of approximately 7 and 11 kb in length with *Hind*III and 2 and 4 kb with *Eco*RI, in particular 2 kb from these double enzymes digestion, could detect *Rubus* PKS1 and PKS5 genes containing both the start and stop codon (Zhenge, 2001; Kumar and Ellis, 2003). The patterns could be useful for rapidly determining polymorphic alleles of PKS1 and PKS5 genes by RFLP technique for further applications.

7.1.2 Fosmid Sub-library Construction and Utilization

Construction of a new fosmid library derived from the three BACs was successfully completed (Chapter 4). Screening of fosmid clones in the library provided clones containing the PKS/CHS genes that were identified in Chapter 3 and confirmed the presence of these gene sequences discussed in Chapter 5. From information provided with the end-pair fosmid sequencing, several clones contained other genes of interest and will be useful for clone-by-clone analysis to better understanding in red raspberry genome.

It was observed that the new fosmid libraries of red raspberry cv. Glen Moy are a new beneficial resource for further understanding of this part of the raspberry genome. End pair sequence alignments of fosmid clones across the libraries increased the quality of the information contained in each BAC from which they were derived. Although the PKS gene sequences were rapidly identified in their sub-libraries, it seemed that there

might be no PKS gene sequences in BAC31B12. This was the first indication of a difference in these three BACs. Although there were certain similarities to BAC24P12 it is possible that the BAC31B12 might derive from another chromosome or from mitochondrial or chloroplast DNA, therefore it contains only a short stretch of a similar sequences.

Random mechanical shearing provided the unbiased fosmid library clones that were distributed along the entire length of each BAC, reviewed in Chapter 5. Although in theory construction by mechanical shearing might have cut the gene sequences into two of more parts, the selected clones each contained the entire PKS gene sequences. This is an obvious benefit of fosmid library construction in assisting assembly of BAC long sequences. However, although more fosmid clones are needed to cover the remaining length of the physical maps and to close the gaps in Chapter 5, it could be done quite easily.

7.1.3 Completion of the Entire BAC Sequences

An assembly of all three BAC sequences was described in Chapter 5. In this thesis, 454 sequencing was used to complete the assembly of the BAC sequences assisted by adding BAC end sequences to estimate the entire length of insert DNA and fosmid library, constructed in Chapter 4, to confirm an accuracy of BAC long sequences. There were three distinct cases of BAC assemblies: the assembly of BAC24P12 has been done but BAC29M05 still has some difficulties with 2 remaining gaps while BAC31B12 was not finished.

It can be observed that the assemblies of the BAC sequences have been done since an adequate quality of assemblies of resulting 454 sequences. The combined sequence of BAC24P12 was completed, with an entire DNA length of 117 kb. BAC29M05 could not be joined along all its length of ~170 kb, but the longest contig was 96 Kb providing the most important information of this BAC. Alignments of assembled sequences also strengthened the similarities between these two BACs (24P12 and 29M05), thus making them useful information sources. Indeed, through the assistance of additional information from the fosmid library, it would be possible to easily close

the gap in BAC29M05 to lead to a complete DNA sequence. In contrast, this would be more difficult for BAC31B12 due to the existence of many short length contigs resulting from the 454 sequencing, thus leaving several gaps that need to be closed to provide the entire sequences. Although more fosmid end sequences were added, it still was impossible to complete.

The presence of two or more copies of a gene within a nucleus in diploid or polyploid organisms and repetitive sequences of eukaryotes can influence the complexity and accuracy of the assembly, depending on the divergence of the duplicated genes. This may be a reason to explain to presence of several contigs that could not be assembled in BAC29M05 and 31B12 in Chapters 5. Moreover, it is possible that these problems are due to either presence of polymorphic regions of interest or misassembles of large-scale duplications.

7.1.4 Polymorphic Alleles of PKS genes in Raspberry

From the investigation of the organization and structural variation of the PKS/CHS genes in both BACs (29M05 and 24P12) described in Chapter 6, the ORFs, nucleotide and deduced amino acid sequences of PKS1 and PKS5 of red raspberry cv. Glen Moy showed 100% identity with PKS1 and 89% identical to PKS5 of cv. Royalty, respectively at the amino acid level. Variation in the genes resulted from the loss and replacement of particular amino acid residues.

The same ordering of genes along the overlapping distance for at least ~53 kb indicated that the two BACs might be either individual members of the paired homologous chromosome or segmental/duplications of the same chromosome. The presence of multiple PKS genes at gene-rich region might be inherited genes from their parents or result from a direct duplication itself. However the presence of two or more CHS genes at the same location within a distance of ~30 kb can found in certain other plants. For example, at least five CHS genes are found together in soybean (*Glycien max*) (NCBI database), thus possibly explaining its high isoflavone content and anti-oxidant property (Cseke *et al.,* 2006). In contrast, the second regions of the PKS genes found in red raspberry in this thesis are pseudogenes that confer no benefit

in enhancing anthocyanin production. This might explain why cv. Glen Moy has a pale-red colour whilst being compared to other cultivars (McCallum, 2009).

The presence of two distinct copies of each PKS gene, showing the detailed variation of nucleotide sequence that might lead to differences in functions, supported the hypothesis that there are two polymorphic PKS alleles in red raspberry cv. Glen Moy. These alleles might influence anthocyanin synthesis, as there still are many other external factors that could directly affect the production such as temperature in different growing seasons. A profile of PKS gene expression is needed to better understand this. However the existence of polymorphic alleles of the PKS genes in red raspberry cv. Glen Moy is one outcome of this thesis, as it could be applied to allele variation and allele mining in other cultivars including related plants.

7.1.5 Initially Partial Physical Maps

The new information of the segment of chromosome that carries the PKS genes of red raspberry is presented in Chapter 5. The first physical map drafts of the two BACs (24P12 and 29M05) were constructed basing on the detailed annotation of the entire BAC assemblies in Chapter 6. It reviewed the locations of the PKS genes and other genes belonging to each BAC, presenting approximately 100 kb in length that derived from red raspberry genome (257 Mb).

The detailed annotation along the entire BAC assemblies in Chapter 6 presented two first drafts of the linear order of several genes on each chromosome. Interestingly, from the information in the physical map of the two BACs provided, the mapped locations of structural genes (PKS/CHS) were in gene-rich region of the red raspberry genome. They tended to stay together from one generation to the next. Notably, there was no co-localization of any gene affecting anthocyanin synthesis or other traits on this segment of the chromosome, except for KAT (3-ketoacyl-CoA thiolase) with distance ~18-20 kb from the PKS genes.

The main difference between both maps with ~24 kb containing four different genes (F-box; NMD/EXP; ABC/ATATH and CDV) on BAC29M05 might be a result from

unequal crossing over between an individual paired-chromosome during cell division or contamination during construction of the original BAC library. Further closing of the gaps to connect all contigs of BAC29M05 could eliminate one of these possibilities, thus making accurate tools for red raspberry genomic study.

7.2 OVERALL CONCLUSIONS

According to the goal of this thesis in identifying and characterizing the PKS/CHS gene regions in the three BAC clones, which derived from the Europe red raspberry cv. Glen Moy genome, it has been achieved. The chromosomal segment of the cv. Glen Moy in this thesis contains one PKS gene (PKS1), with complete sequence (391 amino acid residues) on BAC29M05, that may cod either typical CHS or Naringenin chalcone gene that might influence to red raspberry quality focusing on anthocyanin compounds. Moreover, it also has the other three regions of PKS genes that may be pseudogenes. As all informative sources generated in this thesis, including the BAC assemblies, fosmid sub-libraries and physical map drafts provide better information, however, after additional confirming their sequences, they will be available tools for useful studying in red raspberry genome and for different applications.

7.3 FUTURE WORK

There are still several interesting areas that are a challenge for the future:

- Although the analysis of BAC31B12 seems to have failed, the presence of certain similar sequences to BAC24P12 remains unresolved. Screening fosmid clones by restriction digestion fragment patterns including sequencing can be used to solve the uncertainties with this BAC.
- Based on the physical map drafts, sequencing the selected fosmid clones of the MTP is needed to validation the maps, as well as obtaining more clones to close the remaining gaps.

- Identifying the ORFs, gene structure, and regulatory regions of the PKS genes can be the basis of studying PKS gene expression under different conditions, cells and growth phases.
- In addition, from the known PKS1 sequences of the two cv. (Royalty and Glen Moy), the utilization of polymorphic PKS allele-to-allele variation and allele mining is another interesting approach in red raspberry improvement.
- Finally, mapping all sequences of red raspberry to a physical map of the entire genome is the most challenging for future research.

REFERENCES

Abe, I. and Morita, H. (2010). Structure and function of the chalcone synthase superfamily of plant type III polyketide synthases. *Journal of Natural Product Reports.,* 27:809-838.

Abe, I., Watanabe, T. and Noguchi, H. (2005). Chalcone synthase superfamily of type III polyketide synthases from rhubarb (*Rheum palmatum*). *Proceedings of the Japan Academy, Series B.*, 81:434-440.

Aharoni, A. and Vorst, O. (2001). DNA microarrays for functional plant genomics. *Journal of Plant Molecular Biology.,* 48:99-118.

Ahmad, R., Parfitt, D. E., Fass, J., Ogundiwin, E., Dhingra, A., Gradzie, T. M., Dawei Lin, D., Joshi, N. A., Martinez-Garcia. P. J. and Crisosto, C. H. (2011). Whole genome sequencing of peach (*Prunus persica* L.) for SNP identification and selection. *Journal of BMC Genomics*.,12:569.

Akada, S., Kung, S. D. and Dube, S. K. (1993). Nucleotide Sequence and Putative Regulatory Elements of Gene 2 of the Soybean *(Glycine max)* Chalcone Synthase Multigene Family. *Journal of Plant Physiology.,*102:317-319.

Akada, S and Dube, S. K. (1995). Organization of soybean chalcone synthase gene clusters and characterization of a new member of the family. *Journal of Plant Molecular Biology.,* 29(2):189-99.

Amemiya, C.T., Zhong, T.P., Silverman, G.A., Fishman, M.C. and Zon, L. I. (1999). Zebrafish YAC, BAC, and PAC genomic libraries. *Journal of Methods Cell Biology*., 60:235-258.

Amemiya, C. T., Amores, A., Ota, T., Mueller, G., Garrity, D., Postlethwait, J. H. and Litman, G. W. (2001). Generation of P1 artificial chromosome library of the Southern pufferfish. *Journal of Gene.,* 272:283-289.

Amiour, N., Imbaud, S., Clément, G., Agier, N., Zivy, M., Valot, B., Rende, T. et al., (2012). The Use of Metabolomics Intregrated with Transcriptomic and Proteomic Studies for Identifying Key Steps Involved in the Control of Nitrogen Metabolism in Crops such as Maize. *Journal of Experimental Botany.,* 63(14):5017-5033.

Ammiraju, J. S. S., Yu, Y., Kudrna, M. L. D., Kim, H., Goicoechea. J. L., Katayose, Y., Matsumoto, T., Sasaki, J. W. T. and Wing, R. A. (2005). Random sheared fosmid library as a new genomic tool to accelerate complete finishing of rice (*Oryza sativa* spp. *Nipponbare*) genome sequence: sequencing of gap-specific fosmid clones uncovers new euchromatic portions of the genome. *Journal of Theoretical and Applied Genetics.,* 111:1596-1607.

Amsellem, L., Noyer, J. L., Bourgeois, T. L. and McKey, H. M. (2000). Comparison of genetic diversity of the invasive weed *Rubus alceifolius* Poir. (*Rosaceae*) in its native range and in areas of introduction, using amplified fragment length polymorphism (AFLP) markers. *Journal of Molecular Ecology.,* 9:443-455.

Anttonen, M. J. and Karjalainen, R. O. (2005). Environmental and genetic variation of phenolic compounds in red raspberry. *Journal of Food Composition and Analysis*., 18:759-769.

Ariyadasa, R. and Stein, N. (2012). Advances in BAC-Based Physical Mapping and Map Integration Strategies in Plants. *Journal of Biomedicine and Biotechnology.,* ID 184854, 1-11.

Asins, M. (2002). Present and future of quantitative trait locus analysis in plant breeding. *Journal of Plant Breeding*., 121:281-291.

Asfaw, A., Blair, M. W. and Struik, P. C. (2012). Multienvironment Quantitative Trait Loci Analysis for Photosynthate Acquisition, Accumulation, and Remobilization Traits in Common Bean Under Drought Stress. *Journal of Genes/Genomes/Genetics.,* 2:579-595.

Austin, M. B. and Noel, J. P. (2002). The chalcone synthase superfamily of type III polyketide synthases. *Journal of Natural Product Reports.,* 20:79-110.

Atila, S. P. and Agaoglu, Y. S. (2006). Developments in Raspberry Breeding. *Research Journal of Agriculture and Biological Sciences*., 2(4):171-177.

Bakoume´, C., Wickneswari, R., Rajanaidu, N., Kushairi, A., Amblard, P. and Billotte, N. (2007). Allelic diversity of natural oil palm (*Elaeis guineensis* Jacq.) populations detected by microsatellite markers: implications for conservation. *Journal of Plant Genetic Resources*: *Characterization and Utilization.,* 5(2):104-107.

Balen, B., Pavokovic´, D., Peharec, P. and Krsnik-Rasol, M. (2009). Biochemical markers of morphogenesis in long term horseradish (*Armoracia lapathifolia* Gilib.) tissue culture. *Journal of Scientia Horticulturae.,* 119:88-97.

Barrett, D. M. (1991). Questionnaire - Strawberry and Caneberry Quality Criteria Assessment. Oregon State University Department of Food Science and Technology, unpublished.

Barone, A., Chiusano, M. L., Ercolano, M. R., Giuliano, G., Grandillo, S. and Frusciante, L. (2008). Structural and Functional Genomics of Tomato. *International Journal of Plant Genomics.,* ID 820274, 1-12.

Batschauer, A., Ehmann, B. and Schafer, E. (1991). Cloning and characterization of a chalcone synthase gene from mustard and its light-dependent expression. *Journal of Plant Molecular Biology.,* 16:175-185.

Beekwilder, J., Hall, R. D. and Vos, C. H. De. R. (2005). Identification and dietary relevance of antioxidants from raspberry. *Journal of Biofactors*., 23(4):197-205.

Bentley, D. R., Deloukas, P., Dunham, A., French, L., Gregory, S. G., Humphray, S. J., Mungall, A. J., Ross, M. T., Carter, N. P., Dunham, I. *et al*. (2001). The physical

maps for sequencing human chromosomes 1, 6, 9, 10, 13, 20 and X. *Journal of Natural Science, Biology and Medicine.,* 409:942-943.

Brown, T. A. (2002). Genome, 2nd ec. Oxford, United Kingdom: Wiley-Liss, 600 pp.

Bonet, J., Girona, E. L., Sargent, D. J., Muñoz-Torres, M. C., Monfort1, A., Abbott, A. G., Arús1, P., Simpson, D. W. and Davik, J. (2009). The development and characterisation of a bacterial artificial chromosome library for *Fragaria vesca*. *Journal of BMC Research Notes*., 2:188.

Bao, S., Jiang, R., Kwan, W., Wang, B., Ma, X. and Song, Y. Q. (2011). Evaluation of next-generation sequencing software in mapping and assembly. *Journal of Human Genetic.,* 56:406-414.

Bowsher, C., Steer, M. and Tobin, A. (2008). Plant Biochemistry. Garland Science, Taylor&Francis Group, New York, USA. 446 pp.

Brennan, R and Graham, J. (2009). Improving Fruit Quality in Rubus and Ribes through Breeding. *Journal of Functional Plant Science and Biotechnology.,* 3(1):22-29.

Brown, T. A. (2002). Genome, 2nd ec. Oxford, United Kingdom: Wiley-Liss, 600 pp.

Bruce, A., Alexander, J., Julian, L., Martin, R., Keith, R., and Peter, W. (2002). Molecular Biology of the Cell (4th ed.). New York: Garland Publishing.

Cabezas, J. A., Cervera, M. T., Ruiz-García, L., Carreño, j. and Martínez-Zapater, J. M. (2006). A Genetic Analysis of Seed and Berry Weight in Grapevine. *Journal of Genome.,* 49(12):1572-1585.

Castañeda-Ovando, A., Pacheco-Hernández, M. L., Páez-Hernández, M. E., Rodríguez, J. A. and Galán-Vidal, C. (2009). Chemical studies of anthocyanins: a review. *Journal of Food Chemistry*., 113:859-871.

Cevallos-Cevallos, J. M., Reyes-De-Corcueraa, J. I., Etxeberriaa. E., Danyluk, M. D. and Rodrick, G. E. (2009). Metabolomic Analysis in Food Science: a review. *Journal of Trends in Food Science & Technology.,* 20:557-566.

Chagné, D., Krieger, C., Rassam, M., Sullivan, M., Fraser, J., André, C., Pindo, M., Troggio, M., Gardiner, S. E., Henry, R. A., Allan, A.C., McGhie, T. K. and Laing, W. A. (2012). QTL and candidate gene mapping for polyphenolic composition in apple fruit. *Journal of BMC Plant Biology.,* 12:12.

Chen, C., Zhou, P., Choi, Y. A., Huang, S. and Gmitter Jr, F. G. (2006). Mining and characterizing microsatellites from citrus ESTs. *Journal of Theoretical and Applied Genetics.,* 112:1248-1257.

Chen, S. and Harmon, A. C. (2006). Advances in plant proteomics. *Journal of Proteomics,,* 6:5504-5516.

Cheng, C. H., Chung, M. C., Liu, S. M., Chen, S. K., Kao, F. Y., Lin, S. J., Hsiao, S. H. et al (2005). A fine physical map of the rice chromosome 5. *Journal of Molecular Genetics and Genomics.,* 274:337-345.

Chen, C., Zhou, P., Choi, Y. A., Huang, S. and Gmitter, J. F. G. (2006). Mining and characterizing from citrus ESTs. *Journal of Theoretical and Applied Genetics.,* 112:1248-1257.

Chew, B. P. and Park, J. S. (2004). Carotenoid action on the immune response. *Journal of Nutrition.,* 134(1):257S-261S.

Cho, E., Seddon, J. M., Rosner, B., Willett, W. C. and Hankinson, S. E. (2004). Prospective study of intake of fruits, vegetables, vitamins, and carotenoids and risk of age-related maculopathy. *Journal of Arch Ophthalmology.,* 122(6):883-892.

Choumane, W., Ashtar, S., Valkoun, J. and Weigand, F. (1999). The use of DNA markers in the study of biodiversity in barley. The Third International *Triticeae* Symposium, IPGRI, May 4-8. Aleppo, Syria.

Claros, M. G., Bautista, R., Guerrero-Fernández, D., Benzerki, H., Seoane, P. and Fernández-Pozo, E. (2012). Why Assembling Plant Genome Sequences Is So Challenging. *Journal of Biology.,* 1:439-459.

Collard, B. C. Y. and Mackill, D. J. (2008). Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. *Journal of Philosophical Transactions of the Royal Society B*., 363:557-572.

Collard, B. C. Y., Jahufer, M. Z. Z., Brouwer, J. B. and Pang, E. C. K. (2005). An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection crop importment: *The basic concept. Journal of Euphytica.,* 142:169-196.

Collins, N. C., Tardieu, F. and Tuberosa, R. (2008). Quantitative Trait Loci and Crop Performance under Abiotic Stress: Where Do We Stand?. *Journal of Plant Physiology*., 147(2): 469-486.

Cong, B., Liu, J. and Tanksley, S. D. (2002). Natural alleles at a tomato fruit size quantitative trait locus differ by heterochronic regulatory mutations. *The Proceedings of the National Academy of Sciences of the United States of America.,* 99(21):13606-13611.

Costa, F., Van de Weg, W. E., Stella, S., Dondini, L., Pratesi, D., Musacchi, S. and Sansavini, S. (2008). Map position and functional allelic diversity of Md-Exp7, a new putative expansin gene associated with fruit softening in apple (*Malus* × *domestica* Borkh.) and pear (*Pyrus communis*) *Journal of Tree Genetics & Genomes*., 4:575-586.

Cseke, L. J., Kirakosyan, A., Kaufman, P. B., Warber, S. S., Duke, J. A. and Brielmann, H. L. (2006). Natural Products from Plants (2nd). CRC Taylor&Francis Group, New York, USA. 611 pp.

Dai, J. and Mumper, R. J. (2010). Plant Phenolics: Extraction, Analysis and Their Antioxidant and Anticancer Properties. *Journal of Molecules.,* 15:7313-7352.

Dale, A., Moore, P. P., McNicol, R. J., Sjulin, T. M. and Brumistrov, L. A. (1993). Genetic diversity of red raspberry varieties throughout the world. *Journal of American Society.,* 118:119-129.

Datema, E., Mueller, L. A., Buels, R., Giovannoni, J. J., Visser, R. G. F., Stiekema, W. J. and Ham, R. C. H. J. (2008). Comparative BAC end sequence analysis of tomato and potato reveals overrepresentation of specific gene families in potato. *Journal of BMC Plant Biology.,* 8:34.

Davis, T. M., Denoyes-Rothan, B. and Lerceteau-Kohler, E. (2007). Strawberry. In Kole, C. ed. Genomic Mapping and Molecular breeding in Plants IV: Fruit and Nuts. Springer, Berlin, 189-206.

Davis, T. M., Shields, M. E., Zhang, Q., Tombolato-Terzic, D., Bennetzen, J. L., Pantaroli, A. C., Wang, H., Yao, Q., SanMiguel, P. and Folta, K. M. (2010). An examination of targeted gene neighborhoods in strawberry. *Journal of Plant Biology*., 10:81e.

Dembinska-Kiec1, A., Mykkänen, A. O., Kiec-Wilk, B. and Mykkänen, H. (2008). Antioxidant phytochemicals against type 2 diabetes. *Journal of Nutrition*., 99(1):109-117.

Diaz, A., Fergany, M., Formisano, G., Ziarsolo, P., Blanca, J., Fei, Z., Staub, J. K., Zalapa, J. E., Cuevas, H. E., Dace, G., Oliver, M. and Boissot, N. et al., (2011). A consensus linkage map for molecular markers and Quantitative Trait Loci associated with economically important traits in melon (*Cucumis melo* L.). *Journal of BMC Plant Biology.,* 11:111.

Dirlewanger, E., Cosson, P., Tavaud, M., Aranzana, M. J., Poizat, C., Zanetto, A., Arús, P. and 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.). *Journal of Theoretical and Applied Genetics.,*105: 127-138.

Dongre, A. B., Raut, M. P., Bhandarkar, M. R. and Meshram, K. J. (2011). Identification and genetic purity testing of cotton F1 hybrid using molecular markers. *Indian Journal of Biotechnology.,* 10:301-306.

Dudley, J. W. (1993). Molecular markers in plant improvement: Manipulation of genes affecting quantitative traits. *Journal of Crop Science.,* 33:(in press).

Durbin, M. L., McCaig, B., Clegg, M. T. (2000). Molecular evolution of the chalcone synthase multigene family in the morning glory genome. *Journal of Plant Molecular Biology.,* 42:79-92.

Eckermann, C., Schröder, G., Eckermann, S., Strack, D., Schmidt, J., Schneider, B. and Schröder, J. (2003). Stilbenecarboxylate biosynthesis: A new function in the family of chalcone synthase-related proteins. *Journal of Phytochemistry.,* 62:271-286.

Ferrer, J. L., Jez, J. M., Bowman, M. E., Dixon, R. A. and Noel, J. P. (1999). Structure of chalcone synthase and the molecular basis of plant polyketide biosynthesis. *Journal of Nature Structural & Molecular Biology.,* 6:775-784.

Escarpa, A. and Gonzalez, M. C. (2000). Optimization strategy and validation of one chromatographic method as approach to determine the phenolic compounds from different sources. *Journal of Chromatography.,* A, 897: 161-170.

Espley, R. V., Hellens, R. P., Putterill, J., Stevenson, D. E., Kutty-Amma, S. and Allan, A. C. (2007). Red colouration in apple fruit is due to the activity of the MYB transcription factor, MdMYB10. *The Plant Journal.*, 49:414-427.

Farges, M. C., Minet-Quinard, R., Walrand, S., Thivat, E., Ribalta, J., Winklhofer-Roob, B., *et al* (2012). Immune status is more affected by age than by carotenoid depletion-repletion in healthy human subjects. *British Journal of Nutrition.,* 108(11):2054-65.

Fei, Z., Tang, X., Alba, R. and Giovannoni, J. (2006). Tomato Expression Database (TED): A Suite of Data Presentation and Analysis Tools. *Journal of Nucleic Acids Research.,* 34:766-770.

Fang, G., Blackmon, B. P., Henry, D. C., Staton, M. E., Saski, C. A., Hodges, S. A., Tomkins, J. P. and Luo, H. (2010). Genomic tools development for *Aquilegia*: construction of a BAC-based physical map. *Journal of BMC Genomics*., 11:621.

Feinbaum, R. L. and Ausubel, F. M. (1988). Transcriptional regulation of the *Arabidopsis thaliana* chalcone synthase gene. *Journal of Molecular Cell Biology.,* 8(5):1985.

Fernie, A. R. and Schauer, N. (2008). Metabolomics-assisted breeding: a viable option for crop improvement?. *Journal of Trends in Genetics.,* 25(1):39-78.

Ferreira, M. E. (2006). Molecular analysis of gene banks for sustainable conservation and increased use of crop genetic resources. In: Ruane, J., and Sonnino, A. eds. The role of biotechnology in exploring and protecting agricultural genetic resources. FAO, Rome, Italy. 121-127.

Feuillet, C., Leach, J. E., Rogers, J., Schnable, P. S. and Eversole, K. (2011). Crop genome sequencing: lessons and rationales. *Journal of Trends in Plant Science*., 16(2):77-88.

Fletcher, H., Hickcy, I. and Winter, P. (2006). Genetics (3rd). Taylor&Francis Group, Oxon, UK. 379 p.

Fleury, D., Luo, M. C., Dvorak, J., Ramsay, L., Gill, B. S., Anderson, O. D., You, F. M., Shoaei, Z., Deal, K. R. and Langridge, P. (2010). Physical mapping of a large plant genome using global high-information-content-fingerprinting: the distal region of the wheat ancestor *Aegilops tauschii* chromosome 3DS. *Journal of BMC Genomics.,* 11:382.

Flores-Sanchez, I. J., Linthorst, H. J. M. and Verpoorte, R. (2010). In silicio expression analysis of PKS genes isolated from *Cannabis sativa* L., *Journal of Genetics and Molecular Biology.,* 33(4):703-713.

Flores-Sanchez, I. J. and Verpoorte, R. (2009). Plant polyketide synthases: A fascinating group of enzymes. *Journal of Plant Physiology Biochemistry.,* 47:167-174.

Frazer, K. A., Elnitski, L. and Church, D. M., *et al* (2003). Cross-Species Sequence Comparisons: A Review of Methods and Available Resources. *Journal of Genome Research.,* 13:1-12.

Galbraith, D. W. (2006). DNA microarray analyses in higher plants. *Journal of OMICS research.,* 10(4):455-73.

Garcia-Mas, J., Benjak, A., Sanseverino, W., Bourgeois, M., Mir, G., González, V. M., Hénaff, E., *et al* (2012). The genome of melon (*Cucumis melo* L.). *Proceedings of the National Academy of Sciences*., 109(29):11872-11877.

Gardiner, S., Bus, V., Rusholme, R., Chagne, D. and Rikkerink. E. (2007). Apple. In Kole, C. ed. Genome Mapping and Molecular Breeding, Vol 4. Spring-Verlag, Heidelberg, 1-6.

Graham, J. and McNicol, R. J. (1995). An examination of the ability of RAPD markers to determine the relationships within and between *Rubus* species. *Journal of Theoretical and Applied Genetics.,* 90:1128-1132.

Graham, J., Hein, I., and Powell, W. (2007). Genome Mapping and Molecular Breeding in Plants, Fruits and Nuts. Volume 4, Springer-Verlag Berlin Heidelberg.

Graham, J., Smith, K., MacKenzie, K., Jorgenson, L., Hackett, C. and Powell, W. (2004). The construction of a genetic linkage map of red raspberry (*Rubus idaeus* subsp. *idaeus*) based on AFLPs, genomic-SSR and EST-SSR markers. *Journal of Theoretical and Applied Genetics.,* 109:740-749.

Graham, J., Smith, K., Woodhead, M. and Russell, J. (2002). Development and use of simple sequence repeat SSR markers in *Rubus* species. *Molecular Ecology Notes.,* 2(3):250-252.

Graham, J., Ratnaparkhe, M. B. and Powell, W. (2008). Molecular mapping and breeding of physiological traits. In: Kole C, Abbott AG (Eds) *Principles and Practices of Plant Genomics (Vol 2) Molecular Breeding*, Science Publishers, Inc., Enfield, NH, USA, pp 217-241.

Graham, J. and Jennings, S. N. (2009). Raspberry breeding. In: Jain SM, Priyadarshan M (Eds) *Breeding Tree Crops*, IBH & Science Publication Inc., Oxford, UK, pp 233-248.

Graham, J., Hackett, C. A., Smith, K., Woodhead, M., Hein, I. and McCallum, S. (2009). Mapping QTLs for developmental traits in raspberry from bud break to ripe fruit. *Journal of Theoretical and Applied Genetics.,* 118:1143-1155.

Griffiths, A. J. F., Miller, J. H,, Suzuki, D. T., Lewontin, R. C. and Gelbart, W. M. (1993). "Chapter 5". An Introduction to Genetic Analysis (5th ed.). New York: W.H. Freeman and Company.

González1, V. M., Garcia-Mas, J., Arús, P. and Puigdomènech, P. (2010). Generation of a BAC-based physical map of the melon genome. *Journal of BMC Genomics.,* 11:339.

Guo, Y. and Qiu, Y. J. (2013). Allele-specific marker development and selection efficiencies for both flavonoid 3′-hydroxylase and flavonoid 3′,5′-hydroxylase genes in soybean subgenus *soja. Journal of Theoretical and Applied Genetics.,* 126(6):1445-1455.

Ha, J., Abernathy, B., Nelson, W., Grant, D., Wu, X., Nguyen, H. T., Stacey, G., Yu, Y., Wing, R. D., Shoemaker, R. C. and Jackson, S. A. (2012). Integration of the Draft Sequence and Physical Map as a Framework for Genomic Research in Soybean (*Glycine max* (L.) Merr.) and Wild Soybean (*Glycine soja* Sieb. and Zucc.). *Journal of Genes/Genomes/Genetics.,*1(3):321-329.

Hadi, M. A. M. E., Zhang, F., Wu, F., Zhou, C. and Tao, J. (2013). Advances in Fruit Aroma Volatile Research. *Journal of Molecules.*, 18:8200-8229.

Hamilton, C. M., Frary, A., Lewis, C. and Tanksley, S. D. (1996). Plant Biology Stable transfer of intact high molecular weight DNA into plant chromosomes. *Proceedings of the National Academy of Sciences., USA* Vol. 93:9975-9979.

Han, Y., Gasic, K., Marron, B., Beever, J. E. and Korban, S. S. (2007). A BAC-based physical map of the apple genome. *Journal of Genomics*., 89:630-637.

Han, Y., Zheng, D., Vimolmangkang, S., Khan, M. A., Beever, J. E. and Korban, S. S. (2011). Integration of physical and genetic maps in apple confirms whole-genome and segmental duplications in the apple genome. *Journal of Experimental Botany*., 1-14.

Handelman, G. J. (2001). The evolving role of carotenoids in human biochemistry. *Journal of Nutrition*.,17(10):818-822.

Harada, T., Sunako, T., Wakasa, Y., Soejima, J., Satoh, T. and Niizeki, M. (2000). An allele of the 1-aminocyclopropane-1-carboxylate synthase gene (Md-ACS1) accounts for the low level of ethylene production in climacteric fruits of some apple cultivars. *Journal of Theoretical and Applied Genetics*., 101:742-746.

Harborne, J. B. and Baxter, H. (1993). Phytochemical Dictionaly A handbook of Bioactive Compounds from Plants. Taylor&Francis, Washington, DC, USA. 791 p.

Harker, C. L., Ellis, T. H. N. and Coen, E. S. (1990). Identification and genetic regulation of the chalcone synthase multigene family in pea. *Journal of Plant Cell*., 2:185-194.

Harlt, D. L. and Jones, E. W. (2006). Essential genetics : a genomics perspective 4th ed. Jones and Bartlett publishers. Printed in the USA. 600 p.

Hayashi, K., Matsumoto, S., Tsukazaki, H., Kondo, T., Kubo, N. and Hirai. M. (2010). Mapping of a novel locus regulating anthocyanin pigmentation in *Brassica rapa*. *Journal of Breeding Science.,* 60:76-80.

Hein, I., Williamson, S., Russell, J. and Powell, W. (2005). Isolation of high molecular weight DNA suitable for BAC library construction from woody perennial soft-fruit species. *Journal of Biotechniques*., 38:69-71.

Helariutta, Y., Kotilainen, M., Elomaa, P., Kalkkinen, N., Bremer, K., Teeri, T. H. and Albert, V. A. (1996). Duplication and functional divergence in the chalcone synthase gene family of *Asteraceae*: evolution with substrate change and catalytic simplication. *Proceedings of the National Academy of Sciences.,* 93:9033-9038.

Heldt, H. W. (1997). Plant Biochemistry&Molecular Biology. Oxford University Press Inc., New York, USA. 522 p.

Hemleben, V., Dressel, A., Epping, B., Lukacin, R., Martens, S. and Austin, M. (2004). Characterization and structural features of a chalcone synthase mutation in a white-flowering line of *Matthiola incana* R. Br.(*Brassicaceae*). *Journal of Plant Molecular Biology.,* 55(3):455-465.

Henikoff, S. and Comai, L. (2003). Single-nucleotide mutations for plant functional genomics. *Annual Review of Plant Biology.,* 54:375-401.

Hopwood, D. A. (1997). Genetic Contributions to Understanding Polyketide Synthases. Chemical Reviews., 97(7):2465-2497.

Hu, T. T., Pattyn, P., Bakker, E. G., Jun Cao, Cheng, J., Clark, R. M., Fahlgren, N., *et al* (2011). The *Arabidopsis lyrata* genome sequence and the basis of rapid genome size change. *Journal of Nature Genetics.,* 43:476-481.

Huang, J. X., Qu, L. J., Yang, J., Yin, H. and Gu, H. Y. (2004). A Preliminary Study on the Origin and Evolution of Chalcone Synthase (*CHS*) Gene in Angiosperms. *Journal of Acta Botanica Sinica.,* 46 (1):10-19.

Humphreys, M. O. (2003). Utilization of plant genetic resources in breeding for sustainability. *Journal of Plant Genetic Resources.,* 1:11-18.

Humphray, S. J., Scott, C. E., Clark, R., Marron, M., Bender. C., Camm, N., Davis, J., Jenks, A., *et al* (2007). A high utility integrated map of the pig genome. *Journal of Genome Biology.,* 8(7):R139.

Huxley, A., Griffiths, M., Levy, M. and Royal Horticultural Society (1992). New Royal Horticultural Social Dictionary of Gardening. 3000 pp.

Hutchison, C., A. (2007). DNA sequencing: bench to bedside and beyond. *Journal of Nucleic Acids Research*., 35(18):6227-6237.

Ikegawa, S., Mabuchi, A., Ogawa, M. and Ikeda, T. (2002). Allele-specific PCR amplification due to sequence identity between a PCR primer and an amplicon : is direct sequencing so reliable?. *Journal of Human Genetic*., 110: 606-608.

Jenkins, S. and Gibson, N. (2002). High-throughput SNP genotyping. *Journal of Comparative and Functional Genomics.,* 3:57-66.

Jennings, D. L. (1988). Raspberries and blackberries: their breeding, diseases and growth. Academic, London. 230 p.

Jiang, C., Schommer, C. K., Kim, S. Y. and Suh, D. Y. (2006). Cloning and characterization of chalcone synthase from the moss, *Physcomitrella patens*. *Journal of Phytochemistry.,* 67:2531-2540.

Jones, N., Ougham, H. and Thomas, H. (1997). Markers and mapping: We are all geneticists now. *Journal of New Phytologist.,* 137:165-177.

Junghans, H., Dalkin, K. and Dixon, R. A. (1993). Stress responses in alfalfa (*Medicago sativa* L.). Characterization and expression patterns of members of a subset of the chalcone synthase multigene family. *Journal of Plant Mo1ecular Biology.,* 22: 239-253.

Kabelka, E. A., Diers, B. W., Fehr, W. R., LeRoy, A. R., Baianu, I. C., You, T., Neece, D. J. and Nelson, R. L. (2004). Putative Alleles for Increased Yield from Soybean Plant Introductions. *Journal of Crop Science and Biotechnology.,* 44:784-791.

Kantety, R. V., Rota, M. L., Matthews, D. E. and Sorrells, M. E. (2002). Data mining for simple sequence repeats in expressed sequence tags from barley, maize, rice, sorghum and wheat. *Journal of Plant Molecular Biology.,* 48:501-510.

Kassim, A., Vandroux, M., Joshi, A., Zait, D., Hackett, C., Woodhead, M., Graham, J., and Paterson, A. (2008). Seasonal and genetic influences on production of characteristic volatiles in red raspberry (*Rubus idaeus*). In: *Proceedings Plant and Animal Genomes XVI Conference 2008*, San Diego, CA, USA, p 480.

Kassim, A. (2009). PhD thesis, SIPBS, University of Strathclyde.

Kassim, A., Poette, J., Paterson, A., Zait, D., McCallum, s., Woodhead, M., Smith, K., Hackett, C. and Graham, J. (2009). Environmental and seasonal influences on red raspberry anthocyanin antioxidant content and identification of QTL. *Journal of Molecular Nutrition & Food Research*., 53:625-634.

Kassim, A., Vandroux, M., Joshi, A., Zait, D., Hackett, C., Woodhead, M., Graham, J., Paterson, A. (2008). Seasonal and genetic influences on production of characteristic volatiles in red raspberry (*Rubus idaeus*). In: *Proceedings Plant and Animal Genomes XVI Conference 2008*, San Diego, CA, USA, 480 p.

Kelly, L. J. and Leitch, I. J. (2011). Exploring giant plant genomes with next-generation sequencing technology. *Journal of Chromosome Research.,* 19:939-953.

Kim, J. G., Shizuya, H., Jong, P. J., Birren, B. and Simon, M. (1992). Stable propagation of cosmid sized human DNA inserts in an F factor based vector. *Journal of NucleicAcids Research*., 20(5):1083-1085.

Klung, W. S. and Cumming, M. R. (2003). Published by Pearson Education, Inc., United States of America. 693 p.

Koch, M., Bishop, J. and Mitchell-Olds, T. (1999). Molecular systematics and evolution of *Arabidopsis* and *Arabis*. *Journal of Plant Biology.,* 1:529-537.

Korać, R. R. and Khambholja, K. M. (2011). Potential of herbs in skin protection from ultraviolet radiation. *Pharmacognosy Reviews*., 5(10):164-173.

Korban, S. S. and Chen, H. (1992). Biotechnology of Apples. In: Hammerschlag, F. and Litz, R. Editors, Biotechnology of Fruit Tree Crops, CAB Int., Oxford, 203-227.

Korzun, V. (2002). Molecular marker and their application in cereals breeding. *Cell Molecular Biology Letters*., 7:811-820.

Kumar, A. and Ellis, E. B. (2003). A family of polyketide synthase genes expressed in ripening *Rubus* fruits. *Journal of Phytochemistry.,* 62:513-526.

Kumar, A. and Ellis, B. E. (2001). The Phenylalanine Ammonia-Lyase Gene Family in Raspberry. Structure, Expression, and Evolution. *Journal of Plant Physiology.,* 127(1):230-239.

Kumar, G. R., Sakthivel, K., Sundaram, R. M., Neeraja, C. N., Balachandran, S. M., Shobha, R. N. S., Viraktamath, B. C. and Madhav, M. S. (2010). Allele mining in crops: Prospects and potentials. *Journal of Biotechnology Advances.,* 28:451-461.

Lacis, G., Kaufmane, E., Rashal, I., Trajkovski, V. and Iezzoni, A. F. (2008). Identification of self-incompatibility (S) alleles in Latvian and Swedish sweet cherry genetic resources collections by PCR based typing. *Journal of Euphytica*., 160:155-163.

Lande, R. and Thompson, R. (1990). Efficiency of marker-assisted selection in the improvement of quantitative traits. *Journal of Genetics.,* 124:743-756.

Latha, R., Rubia, L., Bennett, J. and Swaminathan, M. S. (2004). Swaminathan, Allele Mining for Stress Tolerance Genes in *Oryza* Species and Related Germplasm, *Journal of Molecular biotechnology.,* 27:100-108.

Lauer, U. and Seemulier, E. (2000). Physical Map of the Chromosome of the Apple Proliferation Phytoplasma. *Journal of Bacteriology*., 182(5):1415-1418.

Lee, D. W. and Gould, K. S. (2002). Anthocyanins in leaves and other vegetative organs: an introduction. In: Gould, K.S. and Lee, D.W. (eds). Anthocyanins in Leaves. *Advances in Botanical Research.,* Volume 37. London: Academic Press, pp.1-16.

Lea, P. J. and Leegood, R. C. (1999). Plant Biochemistry and Molecular Biology (2^{nd}). John Wiley&Sons, West Sussex, England. 364 p.

Lemieux, B., Aharoni, A., and Schena, M. (1998). Overview of DNA chip technology. *Journal of Molecular Breeding.,* 4:277-289.

Levaj, B., Dragovic-Uzelac, V., Delonga, K., Ganic, K. k., Banovic, M. and Kovaevic, D. B. (2010). Polyphenols and Volatiles in Fruits of Two Sour Cherry Cultivars, Some Berry Fruits and Their Jams. *Journal of Food Technology and Biotechnology.,* 48(4):538-547.

Lewin, B. (2004). Genes III : International Edition. Pearson Printice Hall. Printed in the USA. 1027 p.

Lewin, H. A., Larkin, D. M., Pontius, J., *et al* (2009). Every genome sequence needs a good map. *Journal of Genome Research.,* 19(11):1925-1928.

Lila, M A. (2011). Impact of Bioflavonoids from Berryfruits on Biomarkers of Metabolic Syndrome. *Journal of Functional Foods in Health and Disease.,* 1(2):13-24.

Liu, M., Li, X. Q. and Weber, C. (2002). Antioxidant and antiproliferative activities of raspberries. *Journal of Agricultural and Food Chemistry.,* 8;50(10):2926-2930.

Lopes, M. S., Belo, M. G., Mendoca, D., Sabino, G. F. and Macchado, D. A. (2006). Isolation and characterization of simple sequence repeat loci in *Rubus hochstetterorum* and their use in other species from the *Rosaceae* family, *Journal of Molecular Ecology Notes.,* 3:750-752.

Luo, M. C., Ma, Y, You, F. M., Anderson, O. D., KopecK, D., Êimková, H., ÊafáÊ, J., DoleÊel, J., Gill, B., McGuire, P. E. and Dvorak, J. (2010). Feasibility of physical map construction from fingerprinted bacterial artificial chromosome libraries of polyploid plant species. *Journal of BMC Genomics.,*11:122

Lyons, E. and Freeling, M. (2008). How to usefully compare homologous plant genes and chromosomes as DNA sequences. *Plant Journal*., 53(4):661-673.

Ma, L. A., Pang, X. B., Shen, H. Y., Pu, G. B., Wang, H. H., Lei, C. y., Wang, G., Li, G. F., Liu, B. Y. and Ye, H. C. (2009). A novel type III polyketide synthase encoded by a three-intron gene from *Polygonum cuspidatum*. *Journal of Planta.,* 229:457-469.

Mardis, E. R. (2008a). The impact of next-generation sequencing technology on genetics. *Journal of Trends in Genetics*., 24:133-141.

Malowicki, S. M. M., Martin, R. and Qian. M. C. (2008). Volatile Composition in Raspberry Cultivars Grown in the Pacific Northwest Determined by Stir Bar Sorptive Extraction-Gas Chromatography-Mass Spectrometry. *Journal of Agricultural and Food Chemistry.,* 56:4128-4133.

Marinova, D., Ribarova, F. and Atanassova, M. (2005). Total phenolics and total flavonoids in Bulgrian fruits and vegetables. *Journal of the University of Chemical Technology and Metallurgy.*, 40:255-260.

Mayer, K. F. X., Waugh, R., Langridge, P., Close, T. J., Wise, R. P., Graner, A., Matsumoto, T., Sato, K., Schulman, A., Muehlbauer, G. J., Stein, N., Ariyadasa, R., Schulte, D. and Poursarebani, N. (2012). A physical, genetic and functional sequence assembly of the barley genome. *Journal of Nature.,* 491(7426):711-716.

Maynard, V. O. (2001). Clone by clone by clone. *Journal of Nature.,* 409(15):216-218.

McCallum, S. (2009). PhD thesis, SIPBS, University of Strathclyde.

McCallum, S., Woodhead, M., Hackett, C. A., Kassim, A., Paterson, A. and Graham, J. (2010). Genetic and environmental affects influencing fruit colour and QTL analysis in raspberry. *Journal of Theoretical and Applied Genetics.,* 121(4):611-627.

McCallum, S., Zait, D., Paterson, A. and Graham, J. (2008). Development of a marker assisted breeding toolkit for premium sensory quality in raspberry. *Proceedings of 4th International Rosaceae Genomics Conference*,, Pucon, Chile, p 71.

McDougall, G. J., Dobson, P., Smith, P., Blake, A., and Stewart, D. (2005). Assessing potential bioavailability of raspberry Anthocyanins using an in vitro digestion system. *Journal of Agricultural and Food Chemistry.,* 53:896-5904.

McPherson, J. D., Marra, M., Hillier, L., Waterston, R. H., Chinwalla, A., Wallis, J., Sekhon, M., *et al* (2001). A physical map of the human genome. International Human Genome Mapping Consortium. 15;409 (6822):934-941.

Melounová, M., Vejl, P., Sedlák, P., Blažek, J., Zoufalá, J., Milec, Z. and Blažková, H. (2005). Alleles controlling apple skin colour and incompatibility in new Czech apple varieties with different degrees of resistance against *Venturia inaequalis* CKE. *Journal of Plant Soil and Environment.,* 51:65-73.

Meyers, B. C., Scalabrin, S. and Morgante, M. (2004). Mapping and sequencing Complex Genomics:Let's Get Physical Map. *Journal of Nature.,* 5:578-588.

Michaelson, M. J., Price, H. J., Ellison, J. R. and Johnston, J. S. (1991). Comparison of plant DNA contents determined by feulgen microspectrophotometry and laser flow cytometry. *American Journal of Botany*., 78(2):183-188.

Miller, J., R., Sy Koren, S. and Sutton, G. (2010). Assembly Algorithms for Next-Generation Sequencing Data. *Journal of Genomics.,* 95(6):315-327.

Ming, R., Hou, S., Feng, Y., Yu, Q., Dionne-Laporte, A., Saw, J. H., Senin, P., Wang, W., Ly, B. V. and Lewis, K. L. (2008). The draft genome of the transgenic tropical fruit tree papaya (*Carica papaya* Linnaeus). *Journal of Nature,,* 452:991-996.

Ming, R., Moore. P. H., Zee, F., Abbey, C. A., Ma, H. and Paterson, A. H. (2001). Construction and characterization of a papaya BAC library as a foundation for molecular dissection of a tree-fruit genome. *Journal of Theoretical and Applied Genetics.,*102:892-899.

Miyake, S., Takahashi, N., Sasaki, M., Kobayashi, S., Tsubota, K. and Ozawa. Y. (2012). Vision preservation during retinal inflammation by anthocyanin-rich bilberry extract: cellular and molecular mechanism. *Journal of Laboratory Investigation*., 92(1):102-109.

Moose, S. P. and Mumm, R. H. (2008). Molecular Plant Breeding as the Foundation for 21st Century Crop Improvement. *Journal of Plant Physiology.,* 147(3):969-977.

Morgante, M. and Salamini, F. (2003). From plant genomics to breeding practice. *Journal of Current Opinion in Biotechnology.,* 14:214-219.

Morozova, O. and Marra, M. A. (2008). Applications of next-generation sequencing technologies in functional genomics. *Journal of Genomics.,* 92(5):255-264.

Mozo, T., Dowar K. and Dunn, P., *et al* (1999). A complete BAC-based physical map of the *Arabidopsis thaliana* genome. *Journal of Nature Genetics.,* 22:271-275.

Mun, J. H., Kwon, S. J., Yang, T. J., Kim, H. S., Choi, B. S., Baek, S., Kim, J. S., Jin, M., Kim, J. A., *et al* (2008). The first generation of a BAC-based physical map of *Brassica rapa*. *Journal of BMC Genomics*., 9:280.

Nakitandwe, J., Trognitz, F. and Trognitz, B. (2007). Reliable allele detection using SNP-based PCR primers containing Locked Nucleic Acid: application in genetic mapping. *Journal of Plant Methods.,* 3:2

Nelson, W. M., Bharti, A. K., Butler, E., Wei, F. S., Fuks, G., Kim, H., Wing, R. A., Messing, J. and Soderlund, C. (2005). Whole-genome validation of high-information-content fingerprinting. *Journal of Plant Physiology.,* 139:27-38.

Newcomb, R. D., Crowhurst, R. N., Gleave, A. P., Rikkerink, E. H., Allan, A. C., Beuning, L. L., Bowen, J. H., Gera, E., Jamieson, K. R., Janssen, B. J., *et al* (2006). Analyses of Expressed Sequence Tags from Apple. *Journal of Plant Physiology.,* 141: 147-166.

Oeveren, J. V., Ruiter, M., Jesse, T., Poel, H. V. D., Tang, J., Yalcin, F., Janssen, A., Volpin, H., Stormo, H. K., Bogden, R., Eijk, M. J. T. and Prins, M. (2011). Sequence-based physical mapping of complex genomes by whole genome profiling. *Journal of Genome Research*., 21(4):618-625.

Pal, D. and Verma, P. (2013). Flavonoids: A powerful and abundant source of antioxidants. *International Journal of Pharmacy and Pharmaceutical Sciences.,* 5(3):95-98.

Pang P., Shen, G., Wu, W., Liu, X., Lin, J., Tan, F., Sun, X. and Tang, K. (2005). Characterization and expression of chalcone synthase gene from *Ginkgo biloba. Journal of Plant Science*., 168(6):1525-1531.

Pareek, C., S., Smoczynski, R. and Tretyn, A. (2011). Sequencing technologies and genome sequencing. *Journal of Applied Genetics*., 52:413-435.

Paterson, A. H. (1996a). Making genetic maps. In: A.H. Paterson (Ed.), Genome Mapping in Plants, R. G. Landes Company, San Diego, California; Academic Press, Austin, Texas, 23-39.

Paterson, A., Kassim, A., McCallum, S., Woodhead, M., Smith, K., Zait, D. and Graham, J. (2013). Environmental and seasonal influences on red raspberry flavour volatiles and identification of quantitative trait loci (QTL) and candidate genes. *Journal of Theoretical and Applied Genetics.,* 126:33-48.

Patthy, L. (2000). Protein evolution. MPG Book Ltd. publishers. Printed in Great Britain. 228 p.

Philippe, R., Paux, E., Bertin, I., Sourdille, P., Choulet, F., Laugier, C. and Šimková, H. (2013). A high density physical map of chromosome 1BL supports evolutionary studies, map-based cloning and sequencing in wheat. *Journal of Genome Biology*.,

Pierce, B. A. (2002). Genetics A Conceptual Approach. Freeman, New York, USA. 709 p.

Pontaroli, A. C., Rogers, R. L., Zhang, Q., Shields, M. E., Davis, T. M., Folta, K. M., SanMiguel, P. and Bennetzen, J. L. (2009). Gene Content and Distribution in the Nuclear Genome of *Fragaria vesca*. *Journal of The Plant Genome.,* 2:93-101.

Poursarebani, N., Ariyadasa. R., Zhou, R., Schulte, D., Steuernagel, B., Martis, M. M., Graner, A., Schweizer, P., Scholz, U., Mayer, K. and Stein, N. (2013). Conserved synteny-based anchoring of the barley genome physical map. *Journal of Functional & Integrative Genomics.,* 13:339-350.

Purgar, D. D., Duralija, B., Voa, S., Vokurka, A. and Ercisli, S. (2012). A Comparison of Fruit Chemical Characteristics of Two Wild Grown *Rubus* Species from Different Locations of Croatia. *Journal of Molecules.,* 17:10390-10398.

Quattrocchio, F., Wing, J. F., Van der Woude, K., Mol, J. N. and Koes, R. (1998). Analysis of bHLH and MYB domain proteins: Species-specific regulatory differences are caused by divergent evolution of target anthocyanin genes, *Plant Journal.,* 13:475 -488.

Radhakrishnan, E. K., Varghese, R. T. and Vasudevan, S. E. (2010). Unusual intron in the second exon of a Type III polyketide synthase gene of *Alpinia calcarata* Rosc. *Journal of Genetic and Molecular Biology.,* 33(1):141-145.

Radhakrishnan, E. K., and Soniya, E. V. (2009). Molecular analysis of type III polyketide synthase (PKS) gene family from *Zingiber officinale* Rosc. *African Journal of Plant Science.,* 3(3):44-48.

Ren, C., Lee, M., Yan, B., Ding, K., Cox, B., Romanov, M. N., Price, J. A., Dodgson, J. B. and Zhang, H. (2003). A BAC-Based Physical Map of the Chicken Genome. *Journal of Genome Research.,* 13:2754-2758.

Riaz, M. N. and Bushway, A. A. (1996). Compositional Analysis of Four Red raspberry Cultivars Grown in Main. *Journal of Food Quality.,* 19(6):457-465.

Robbins, J. A. and Sjulin, T.M. (1989). Postharvest storage characteristics and respiration rates in five cultivars of red raspberry. *Journal of Horticultural Science and Biotechnology.,* 24:980-982.

Romeo, J. T. and Dixon, R. A. (2002). Recent Advances in Phytochemistry, Phytochemistry in the Genomics and Post-Genomics Eras. Elsevier Science Ltd, printed in The Netherlands. 258 p.

Rothnie, H. M. (1996). Plant mRNA 3′-end formation. *Journal of Plant Molecular Biology.,* 32:43-61.

Russell, P. J. (2001). iGenetics. Benjamin Cummings. San Francisco, CA. 828 p.

Russell, P. J. (2006). iGenetics : A molecular Approach (2nd). Benjamin Cummings, San Francisco, CA. 842 p.

Saji, S., Umehara, Y., Antonio, B., A., Yamane, H., Tanoue, H., Baba, T., Aoki, H., Ishige, N., Wu, J., Koike, K., Matsumoto, T. and Sasaki, T. (2001). A physical map with yeast artificial chromosome (YAC) clones covering 63% of the 12 rice chromosomes. *Journal of Genome.,* 44(1):32-37.

Sang, T. (2002). Utility of Low-Copy Nuclear Gene Sequences in Plant Phylogenetics. *Critical Reviews in Biochemistry and Molecular Biology*., 37(3):121-147.

Sangh, C. and Dhananjaya, P. (2013). Allele Mining–Approaches and Utility. *Journal of Trends in Biosciences.,* 6:223-226.

Sapers, G. M., Galleta, G. J., Burgher, A. M. and Phillips, J. G. (1985). Composition and color of fruit and juice of thornless blackberry cultivars. *Journal of the American Society for Horticultural Science.,* 110:243-248.

Sargent, J. D., Fernández-Fernández, F., Rys, A., Knight, V. H., Simpson, D. W. and Tobutt, K. R. (2007). Mapping of *A*1 conferring resistance to the aphid *Amphorophora idaei* and *dw* (dwarfing habit) in red raspberry (*Rubus idaeus* L.) using AFLP and microsatellite markers, *Journal of BMC Plant Biology.,* 7:15.

Sasaki, T., Matsumoto, T., Antonio, B. A. and Nagamura, Y. (2005). From Mapping to Sequencing, Post-sequencing and Beyond. *Journal of Plant Cell Physiology.,* 46(1): 3-13.

Schadt, E. E., Turner, T. and Kasarskis, A. (2010). A window into third-generation sequencing. *Journal of Human Molecular Genetics.,* 19(2):227-240.

Schnable, P. S., Ware, D., Fulton, R. S., Stein, J. C., Wei, F., Pasternak, S., Liang, C., Zhang, J., Fulton, L. and Graves, T. A., *et al* (2009). The B73 maize genome: Complexity, diversity, and dynamics. *Journal of Science.,* 326:1112-1115.

Schröder, J. (1997). A family of plant-specific polyketide synthases: Facts and predictions. *Journal of Trends in Plant Science.,* 2:373-378.

Schröder, J. (2000). The family of chalcone synthase-related proteins: Functional diversity and evolution. In: Romeo, J. T., Ibrahim, R. K., Varin, L. and De Luca, V (eds) Evolution of Metabolic Pathways, V. 34. Pergamon Press, Amsterdam, pp 55-89.

Schulte, D., Ariyadasa, R., Shi, B., Fleury, D., Saski, C., Atkins, M., Jong, P., Wu, C., Graner, A., Langridge, P. and Stein, N. (2011). BAC library resources for map-based cloning and physical map construction in barley (*Hordeum vulgare* L.). *Journal of BMC Genomics.,* 12:247.

Schulte, D., Cai. D., Fan, M. K. J., Wang, S. and Jung, C. (2006). A complete physical map of a wild beet (*Beta procumbens*) translocation in sugar beet. *Journal of Molecular Genetics and Genomics.,* 275(5):504-511.

Seeram, N. P. (2008). Berry Fruits: Compositional Elements, Biochemical Activities, and the Impact of Their Intake on Human Health, Performance, and Disease. *Journal of Agricultural and Food Chemistry*., 56:627-629.

Seidl, M. F., Schneider, A., Govers, F. and Snel, B. (2013). A predicted functional gene network for the plant pathogen *Phytophthora infestans* as a framework for genomic biology. *Journal of BMC Genomics.,* 14:483.

Shawkey, M. D. and Hill, G. E. (2005). Carotenoids need structural colours to shine. *Biology Letters.* 1(2):121-124.

Shendure, J. and Ji, H. (2008). Next-generation DNA sequencing. *Journal of Nature Biotechnology*., 26(10):1135-1145.

Shukitt-Hale, B., Lau, F. C. and Joseph, J. A. (2008). Berry Fruit Supplementation and the Aging Brain. *Journal of Agricultural and Food Chemistry.,* 56:636-641.

Shulaev, V., Korban, S. S., Sosinski, B., Abbott, A. G., Aldwinckle, H. S., Folta, K. M., Iezzoni, A., Main, D., *et al* (2008). Multiple Models for *Rosaceae* Genomics. *Journal of Plant Physiology.,* 147:985-1003.

Silbersteina, L., Kovalskia, I., Huang, R., Anagnostoub, K., Molly. M., Jahnb, K. and Perl-Trevesa, R. (1999). Molecular variation in melon (*Cucumis melo* L.) as revealed by RFLP and RAPD markers. *Journal of Scientia Horticulturae.,* 79:101-111.

Simko, I., Haynes, K. G. and Jones, R. W. (2004). Mining data from potato pedigrees: tracking the origin of susceptibility and resistance to *Verticillium dahliae* in North American cultivars through molecular marker analysis, *Journal of Theoretical and Applied Genetics.,* 108:225-230.

Sistrunk, W. A. and Moore, J. M. (1983). Quality, In: *J.N.* Moore and *J.* Janick (ed). Methods in Fruit Breeding. Lafayette, W., Purdue Univ. Press, Indiana., 274-293.

Skovmand, B., Reynolds, M. P. and DeLacy, I. H. (2001). Mining wheat germplasm collections for yield enhancing traits, *Journal of Euphytica.,* 119:25-32.

Stafne, E. T. and Clark, J. R. (2004). Genetic relatedness among eastern North American blackberry cultivars based on pedigree analysis. *Journal of Euphytica.,* 139:95-104.

Stenesh, J. (1989). Dictionary of Biochemistry and Molecular Biology (2th ed). A Wiley-Interscience Publication. 288.

Stone, Z. S., Yasmin, T., Bagchi, M., Chatterjee, A., *et al* (2007). Berry anthocyanins as novel antioxidants in human health and disease prevention, *Journal of Molecular Nutrition & Food Research*., 51:756-683.

Stoner, G. D., Wang, L., Zikri, N., Chenc, T., Hecht, S. S., Huange, C., Sardoc, C. and Lechnerc, J. F. (2007). Cancer Prevention with Freeze-dried Berries and Berry Components. *Journal of Seminars in Cancer Biology,,* 17(5):403-410.

Strauss, S. H., Bousquet, J., Hipkins, V. D. and Hong, Y. P. (1992). Biochemical and molecular genetic markers in biosystematic studies of forest trees. *Journal of New Forests.,* 6:125-158.

Sunako, T., Sakuraba, W., Senda, M., Akada, S., Ishikawa, R., Niizeki, M. and Harada, T. (1999). An allele of the ripening-specific 1-aminocyclopropane-1-carboxylic acid synthase gene (*ACS1*) in apple fruit with a long storage life. *Journal of Plant Physiology.,* 119:1297-1303.

Sutherland, B. G., Tobutt, K. R., Marchese, A., Paternorter, G., Simson, D. W. and Sargent, D. J. (2008). A genetic linkage map of *Physocarpus,* a member of the *Spiraeoideae* (*Rosaceae*), based on RAPD, AFLP, RGA, SSR and gene specific markers. *Journal of Plant Breeding.,* 127:527-532.

Szajdek, A. and Borowska, E. J. (2008). Bioactive Compounds and Health-Promoting Properties of Berry Fruits: A Review. *Journal of Plant Foods for Human Nutrition.,* 63:147-156.

Tanksley, S. D. (1993). Mapping polygenes. *Annu Rev Genet.* 27:205-233.

Terol, J., Naranjo, M. A., Ollitrault, P. and Talon, M. (2008). Development of genomic resources for *Citrus clementina*: Characterization of three deep-coverage BAC libraries and analysis of 46,000 BAC end sequences. *Journal of BMC Genomics*., 9:42.

Tomaso, A. W. D. and Weissman, I. L. (2003). Construction and Characterization of Large-Insert Genomic Libraries (BAC and Fosmid) from the *Ascidian* (*Botryllus schlosseri*) and Initial Physical Mapping of a Histocompatibility Locus. *Marine Biotechnology.,* 5:103-115.

Varshney, R. K., Graner, A. and Sorrells, M. E. (2005). Genomics-assisted breeding for crop improvement. *Journal of Trends in Plant Science.,* 10:621-630.

Varshney, R. k., Hoisington, D. A. and Tyagi, A. K. (2006). Advances in Cereal Genomics and Applications in Crop Breeding. *Journal of* Trends in Biotechnology., 11:490-499.

Varshney, R. K. and Tuberosa, R. (2007). Genomics-Assisted Crop Improvemen in Genomics Approaches and Platforms. Published by Springer, The Netherlands. 386 p.

Wallace, T. C. (2011). Anthocyanins in Cardiovascular Disease1. *Journal of American Society for Nutrition, Advance in Nutrition*., 2:1-7.

Wallis, J. W., Aerts, J., Groenen, M. A. M., Crooijmans, R. P. M. A., Layman, D., Graves, T. A., Scheer, D. E., Kremitzki, C., *et al* (2004). A physical map of the chicken genome. *Journal of Nature.,* 432: 761-764.

Wang, H., Zhang, K. and Su, H. (2010). Identification of the S-genotypes of several sweet cherry (*Prunus avium* L.) cultivars by AS-PCR and pollination. *African Journal of Agricultural Research.,* 5(3):250-256.

Wang, M., Allefs, S., Van den Berg, R.G., Vleeshouwers, G. A. A., Van der Vossen, E. A. G. and Vosman, B. (2008). Allele mining in *Solanum*:conserved homologues of *Rpi-blb1* are identified in *Solanum stoloniferum. Journal of Theoretical and Applied Genetics.,* 116:933-943.

Wang, S. Y. and Lin, H. S. (2000). Antioxidant activity in fruits and leaves of blackberry, raspberry, and strawberry varies with cultivar and developmental stage. *Journal of Agricultural and Food Chemistry.,* 48(2):140-146.

Weber, C. and Liu, R. H. (2002). Antioxidant capacity and anticancer properties of red raspberry. *Journal of Acta Horticulturae.,* 585: 451-457.

Wenzel, G. (2006). Molecular plant breeding: achievements in green biotechnology and future perspectives. *Journal of Applied Microbiology and Biotechnology.,* 70:642-650.

Winter, P. and Kahl, G. (1995). Molecular marker technologies for plant improvement. *World Journal of Microbiology and Biotechnology*., 11:438-448.

Wolfgang, K. and Jean, K. (2008). Protection from Sunburn with b-Carotene-A Meta-analysis. *Journal of Photochemistry and Photobiology.,* 84:284-288.

Woodhead, M., Weir, A., Smith, K., McCallum, S., MacKenzie, K. and Graham, J. (2010). Functional Markers for Red Raspberry. *Journal of the American Society for Horticultural Science*., 135(5):418-427.

Wrolstad, R. E., Putnam, T. P. and Varsveld, G.W. (1970). Color quality of frozen strawberries: Effect of anthocyanin, pH, total acidity and ascorbic acid variability. *Journal of Food Science.,* 35:448-451.

Wrolstad, R. E. (1976). Color and pigment analysis in fruit products. Agricultural Experiment Station Bulletin, 624 pp.

Wu, C., Schramm, T. M., Zhou, S., Schwartz, D. C. and Talaat, A. M. (2009). Optical mapping of the *Mycobacterium avium* subspecies *paratuberculosis* genome. *Journal of BMC Genomics.,* 10:25.

Wu, X., Beecher, G. R., Holden, J. M. and Haytowitz, D. B. (2006). Concentrations of anthocyanins in common foods in the United States and estimation of normal consumption. *Journal of Agricultural and Food Chemistry.,* 54:4069-4075.
Wünsch, A. and Hormaza, J. I. (2002). Cultivar identification and genetic fingerprinting of temperate fruit tree species using DNA markers. *Journal of Euphytica.,* 125:59-67.

Xu, M., Korban, S. S., Song, J. and Jiang, J. (2002). Constructing a bacterial artificial chromosome library of the apple cultivar GoldRush, *Journal of Acta Horticulturae.,* 595:103-112.

Huang, U. X., Qu, L. J., Yang, J., Yin, H. and Gu, H. Y. (2004). A Preliminary Study on the Origin and Evolution of Chalcone Synthase (*CHS*) Gene in Angiosperms. *Journal of Acta Botanica Sinica.,* 46(1):10-19.

Xie, R., Zheng, L., He, S., Zheng1, Y., Yi, S. and Deng, L. (2011). Anthocyanin biosynthesis in fruit tree crops: Genes and their regulation. *African Journal of Biotechnology.,* 10(86):19890-19897.

Xu, P., Wang, J., Wang, J., Cui, R., Li, Y., Zhao, Z., Ji, P., Zhang, Y., Li, J. and Sun, X. (2011). Generation of the first BAC-based physical map of the common carp genome. *Journal of BMC Genomics.,*12:537.

Yang, J., Gu, H. and Yang, Z. (2004). Likelihood Analysis of the Chalcone Synthase Genes Suggests the Role of Positive Selection in Morning Glories (*Ipomoea*). *Journal of Molecular Evolution.,* 58:54-63.

Yang, j., Huang, J., Gu, H., Zhong,Y. and Yang, Z. (2013). Duplication and Adaptive Evolution of the Chalcone Synthase Genes of *Dendranthema* (*Asteraceae*). *Journal of Molecular Biology and Evolution.,* 19(10):1752-1759.

Yang, T., Yu, Y., Nah, G., Atkins, M., Lee, S., Frisch, D. A. and Wing, R. A. (2003). Construction and utility of 10-kb libraries for efficient clone- gap closure for rice genome sequencing. *Journal of Theoretical and Applied Genetics.,* 107(4):652-660.

Yi, J., Derynck, M. R., Chen, L. and Dhaubhadel, S. (2010). Differential expression of CHS7 and CHS8 genes in soybean. *Journal of Planta.,* 3:741-753.

Yu, Q., Tong, E., Skelton, R. L., Bowers, J. E., Jones, M. R., Murray, J. E., Hou, S., Guan, P., Acob, R. A., Luo, M. C., Moore, P. H., Alam, M., Paterson, A. H. and Ming, R. (2009). A physical map of the papaya genome with integrated genetic map and genome sequence. *Journal of BMC Genomics.,* 10:371.

Zafra-Stone, S., Yasmin, T., Bagchi, M., Chatterjee, A., Vinson, J. A. and Bagchi, D. (2007). Berry Anthocyanins as Novel Antioxidants in Human Health and Disease Prevention. *Journal of Molecular Nutrition & Food Research.,* 51:675-683.

Zhebentyayeva, T. N., Horn, R., Mook, J., Lecouls, A. C., Georgi, L., Swire-Clark, G., Reighard, G. L., Baird, W. V. and Abbott, A. G. (2006). A physical framework for the peach genome. *Journal of Acta Horticulturae.,* 713:83-88.

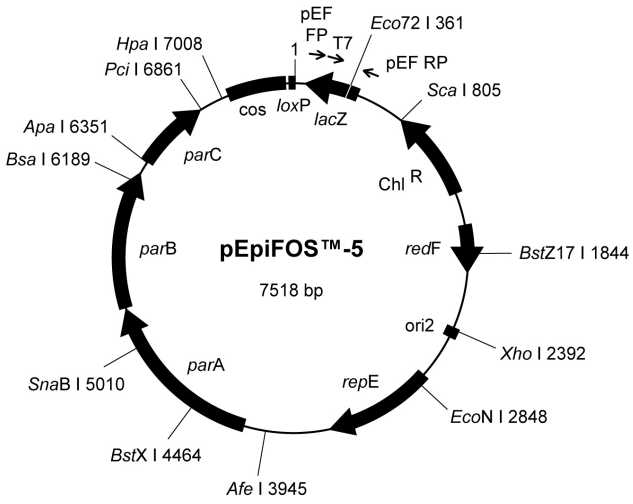
Zhebentyayeva, T. N., Swire-Clark, G., Georgi, L. L., Garay, L., Jung, S., Forrest, S., Blenda, A. V., Blackmon, B., Mook, J., Horn, R., Howad, W., Arús, P., Main, D., Tomkins, J. P., Sosinski, B., *et al* (2008). A framework physical map for peach, a model *Rosaceae* species. *Journal of Tree Genetics & Genomes.,* 4:745-756.

Zheng, D. and Hrazdina, G. (2008). Molecular and biochemical characterization of benzalacetone synthase and chalcone synthase genes and their proteins from raspberry (*Rubus idaeus* L). *Archives of Biochemistry and Biophysics.,* 470:139-145.

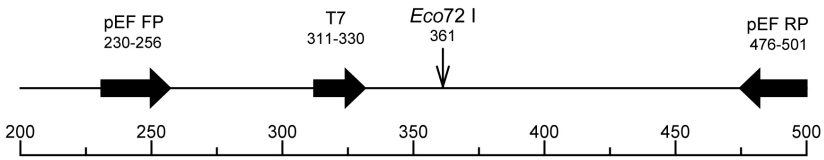
Zheng, D., Schröder, G., Schröder, J. and Hrazdina, G. (2001). Molecular and biochemical characterization of three aromatic polyketide synthase genes from *Rubus idaeus*. *Journal of Plant Molecular Biology.,* 46:1-15.

Zhou, L., Wang, L. and Peng, Z. (2011). Molecular characterization and expression analysis of chalcone synthase gene during flower development in tree peony (*Paeonia suffruticosa*). *African Journal of Biotechnology*., 10(8):1275-1284.

APPENDICES



Note: Not all restriction enzymes that cut only once are indicated above.
See the following page for further information.
Primers are not drawn to scale.

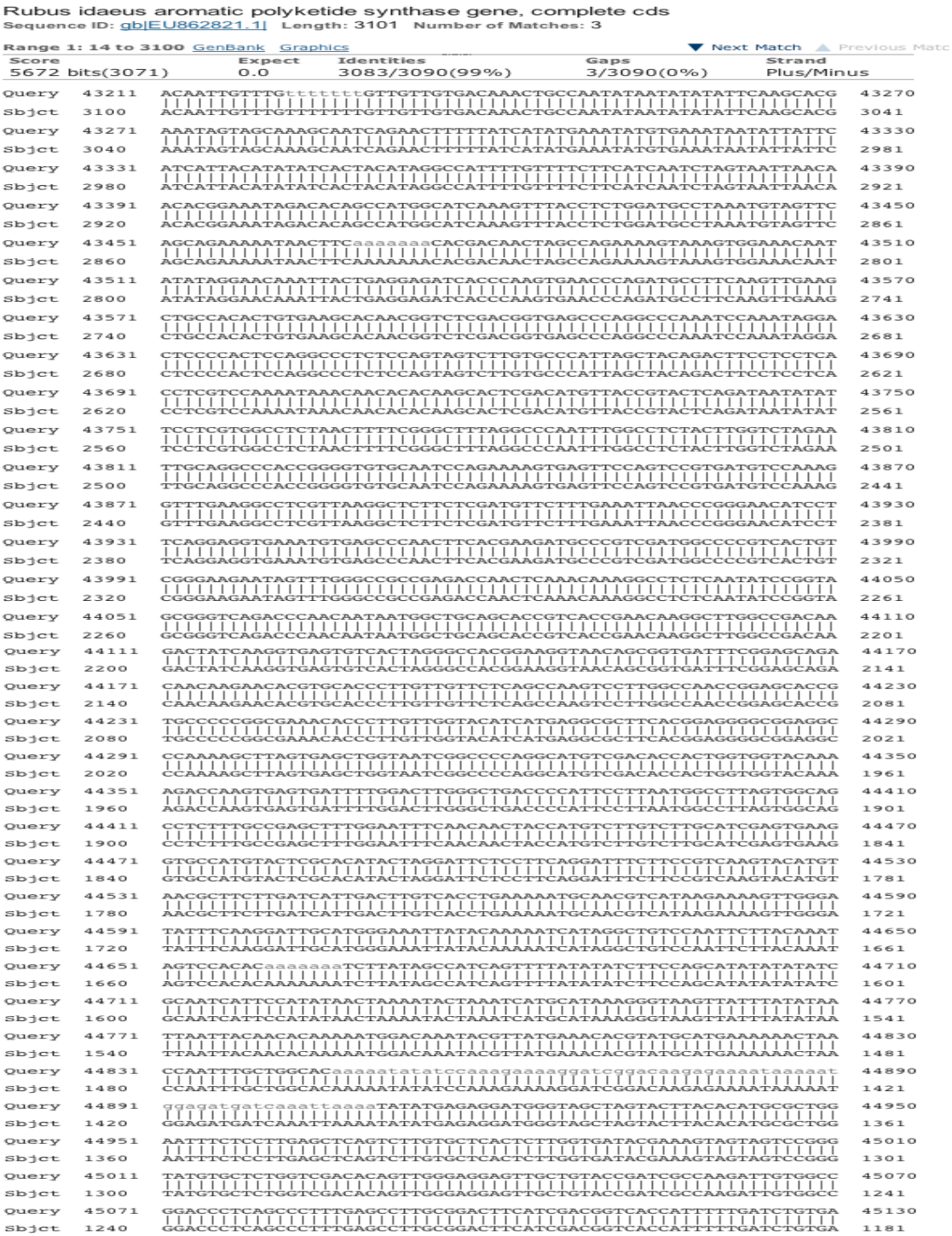


pEF FP = pCC1™ / pEpiFOS™-5 Foward Sequencing Primer
5' GGATGTGCTGCAAGGCGATTAAGTTGG 3'

pEF RP = pCC1™ / pEpiFOS™-5 Reverse Sequencing Primer
5' CTCGTATGTTGTGTGGAATTGTGAGC 3'

T7 = T7 Promoter Primer
5' TAATACGACTCACTATAGGG 3'

Appendix A.1 Structure of vector pEpiFos used for constructing sub-fosmid libraries (Invitrogen).



Appendix A.2 Blast alignment of assembled 454 sequences of BAC29M05 showing nuecleotide similarity (3083/3090) 99% to *R. idaeus* Aromatic Polyketide Synthase gene, complete cds, (PKS gene) (Accession no. EU862821.1).

Appendix A.2 Continue

Restriction Analysis of the pEpiFOS-5 Vector

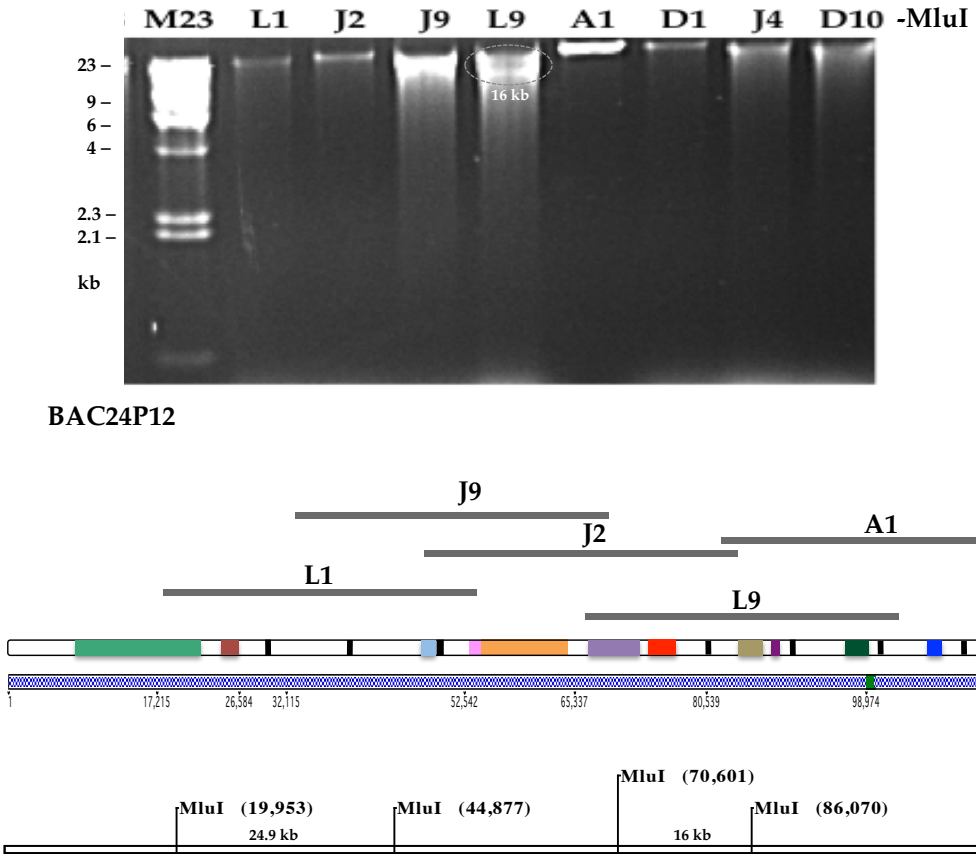
Restriction Enzymes that cut pEpiFOS-5 1 to 3 times:

Enzyme	Sites	Location	Enzyme	Sites	Location
Acc65 I	2	344, 4586	EcoR V	2	3507, 3736
Acl I	2	1133, 4978	Fsp I	3	167, 3131, 6957
Afe I	1	3945	Hind III	1	395
Afl II	2	5987, 6227	Hpa I	1	7008
Afl III	3	4352, 4526, 6861	Kpn I	2	348, 4590
Age I	3	3206, 4436, 5329	Mfe I	1	4366
Ahd I	1	6865	Msc I	2	955, 4797
Ale I	1	5922	Nae I	2	3084, 7000
Apa I	1	6351	Nar I	1	146
ApaB I	3	96, 1946, 7025	Nco I	2	917, 6566
ApaL I	1	87	Nde I	2	94, 4384
BamH I	2	353, 365	NgoM IV	2	3082, 6998
Bbs I	3	4429, 4618, 5495	Not I	2	2, 643
Bcl I	1	5177	Nru I	2	1644, 7053
BfuA I	3	390, 3393, 7160	Nsp I	3	393, 1831, 6865
Bgl I	3	651, 2550, 6999	PaeR7 I	1	2392
Bgl II	2	2525, 4592	Pci I	1	6861
Blp I	1	3858	PflF I	1	4650
Bme1580 I	3	91, 755, 6351	Pml I	1	361
BmgB I	2	4416, 7176	PpuM I	2	1728, 7237
Bmr I	3	268, 6397, 6526	Psi I	1	2501
Bpu10 I	3	1446, 3306, 4501	PspOM I	1	6347
Bsa I	1	6189	Pst I	3	387, 3404, 4945
BsaB I	2	7133, 7217	Pvu I	2	188, 5252
BsaH I	1	146	Sal I	3	377, 657, 7041
BseY I	2	5269, 6026	Sap I	2	3982, 5192
Bsm I	2	824, 1231	Sbf I	2	387, 3404
BsmB I	3	994, 1547, 3321	Sca I	1	805
BspE I	2	1222, 5146	SexA I	1	6979
BspLU11 I	1	6861	Sfi I	1	651
BspM I	3	390, 3393, 7160	Sfo I	1	147
BsrB I	3	476, 1660, 2282	SgrA I	2	4436, 5593
BsrG I	1	3159	Sim I	2	4550, 7237
BssH II	2	4843, 5387	Sma I	3	350, 651, 2872
BssS I	3	4536, 6186, 6749	SnaB I	1	5010
BstAP I	3	95, 1945, 7024	Spe I	1	6101
BstE II	1	6983	Sph I	1	393
BstX I	1	4464	Srf I	1	651
BstZ17 I	1	1844	Sse8647 I	1	1728
Bts I	2	570, 4938	Stu I	1	2553
Dra III	2	1945, 7202	Tat I	3	77, 803, 3159
Eag I	2	2, 643	Tli I	1	2392
Eco47 III	1	3945	Tth111 I	1	4650
Eco72 I	**1**	**361**	Xba I	2	371, 2571
EcoN I	1	2848	Xho I	1	2392
EcoO109 I	2	1728, 7237	Xma I	3	348, 649, 2870
EcoR I	1	332			

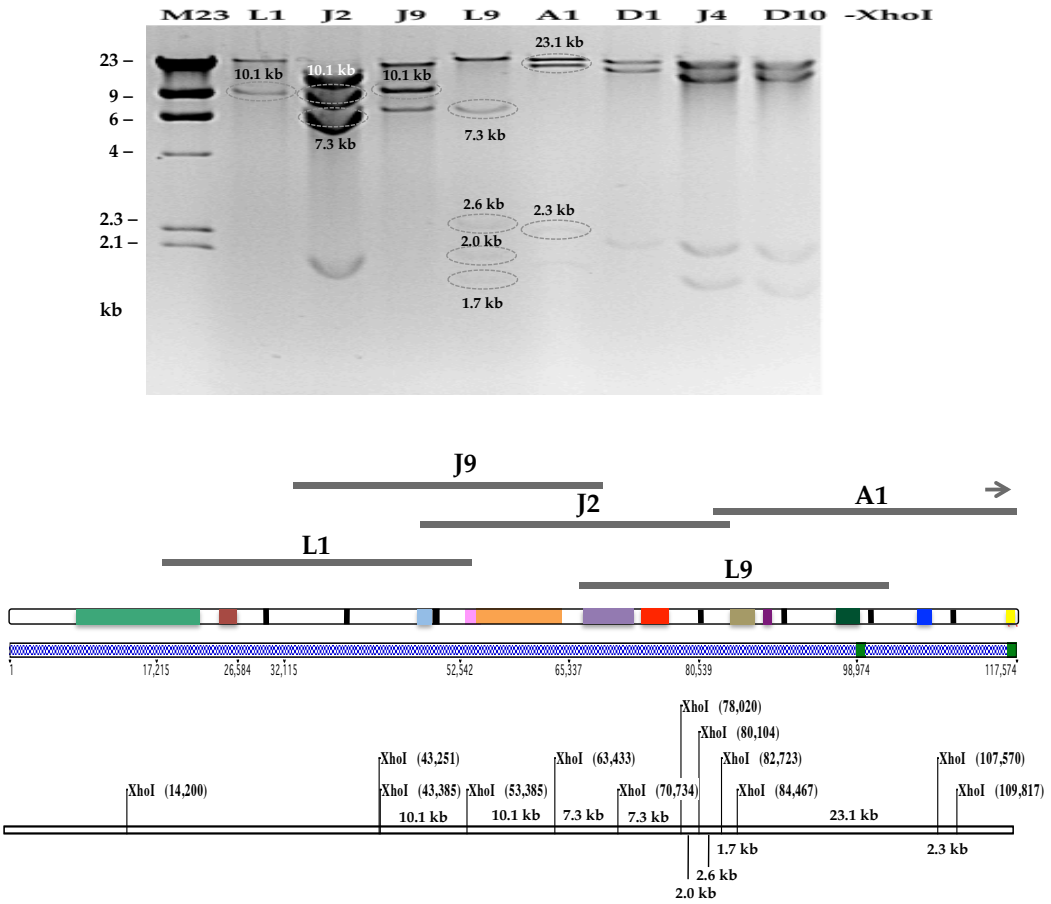
Restriction Enzymes that do not cut pEpiFOS-5:

Aat II	BciV I	Bsu36 I	Nsi I	SanD I
Asc I	BfrB I	Cla I	Pac I	Swa I
AsiS I	BsiW I	Fse I	Pme I	Xcm I
Avr II	BspD I	Mlu I	Rsr II	Tsp509 I
BbvC I	BstB I	Nhe I	Sac II	

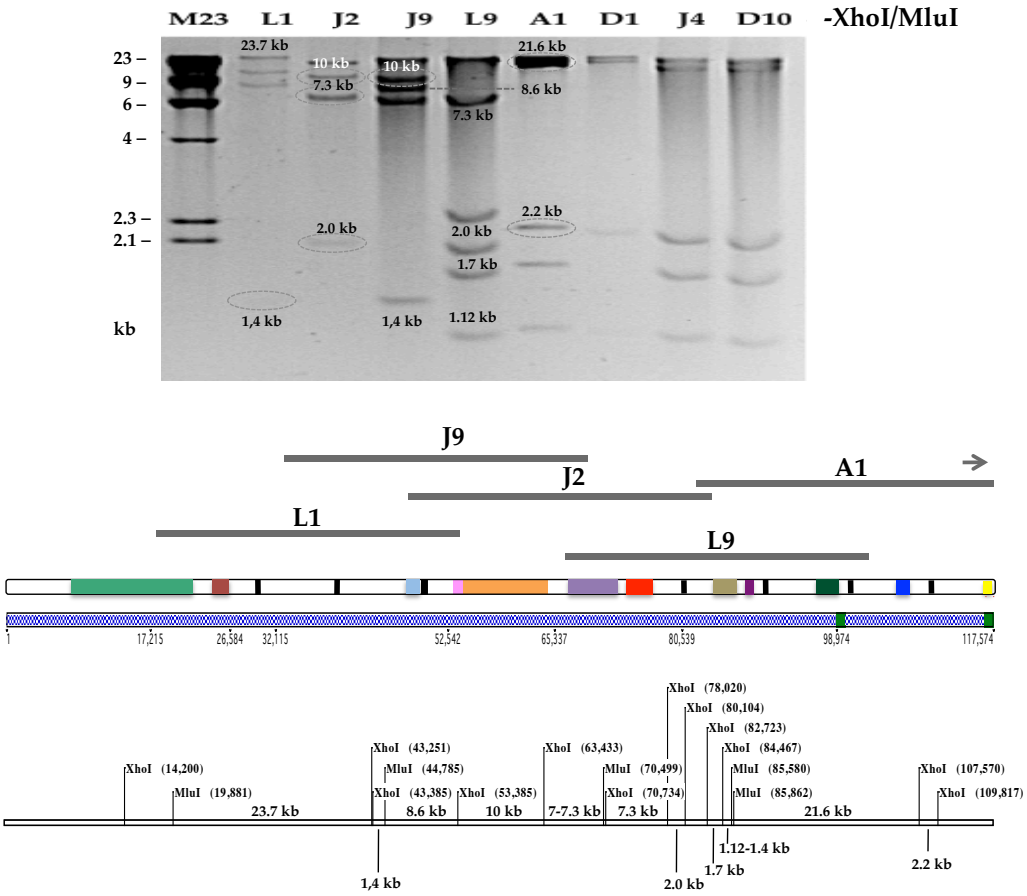
Appendix A.3 Location and number of cut sites of restriction enzymes on vector pEpiFos (Invitrogen).



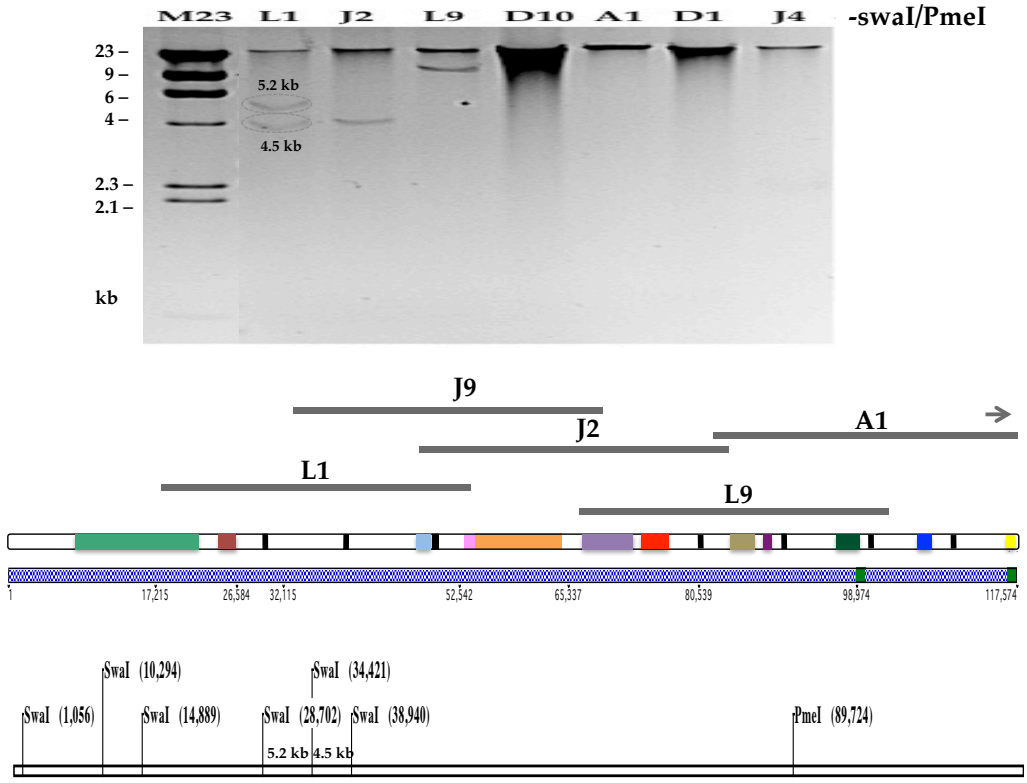
Appendix A.4 Fingerprints and overlaps analysis of eight fosmid clones derived from BAC24P12 by restriction digestion with M*lu* I.



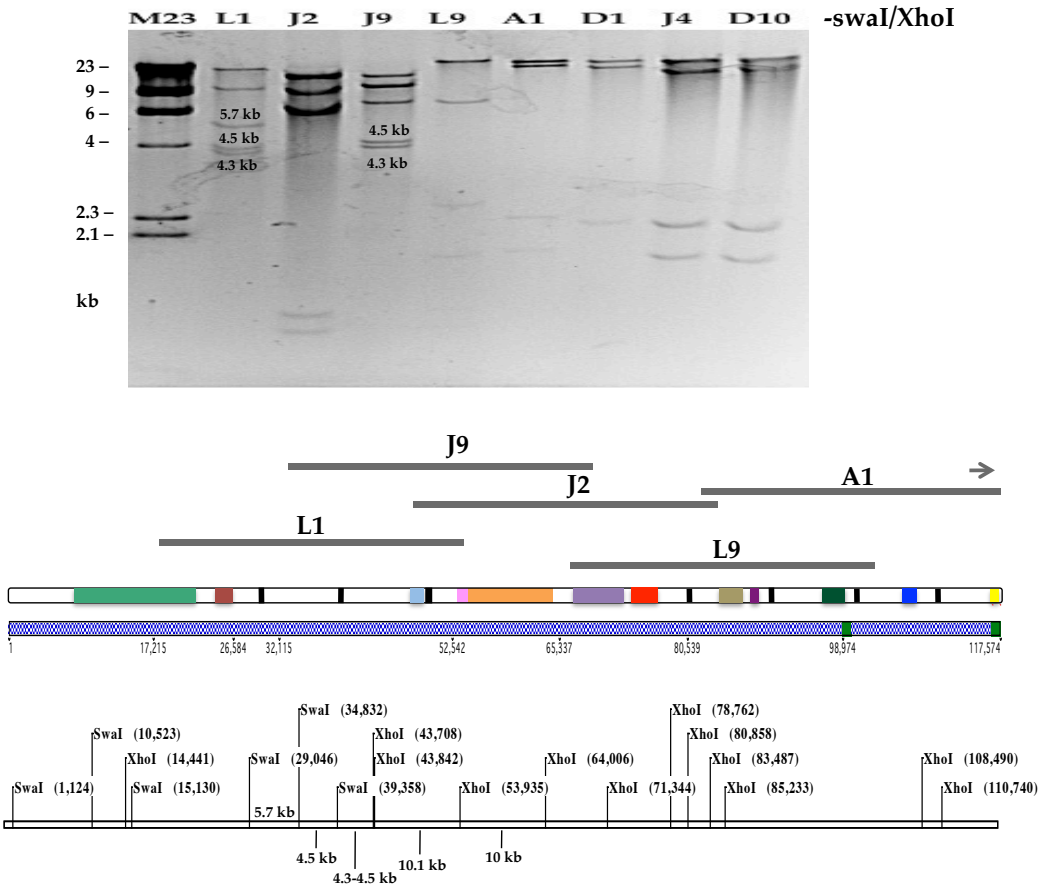
Appendix A.5 Fingerprints and overlaps analysis of eight fosmid clones derived from BAC24P12 by restriction digestion with X*ho* I.



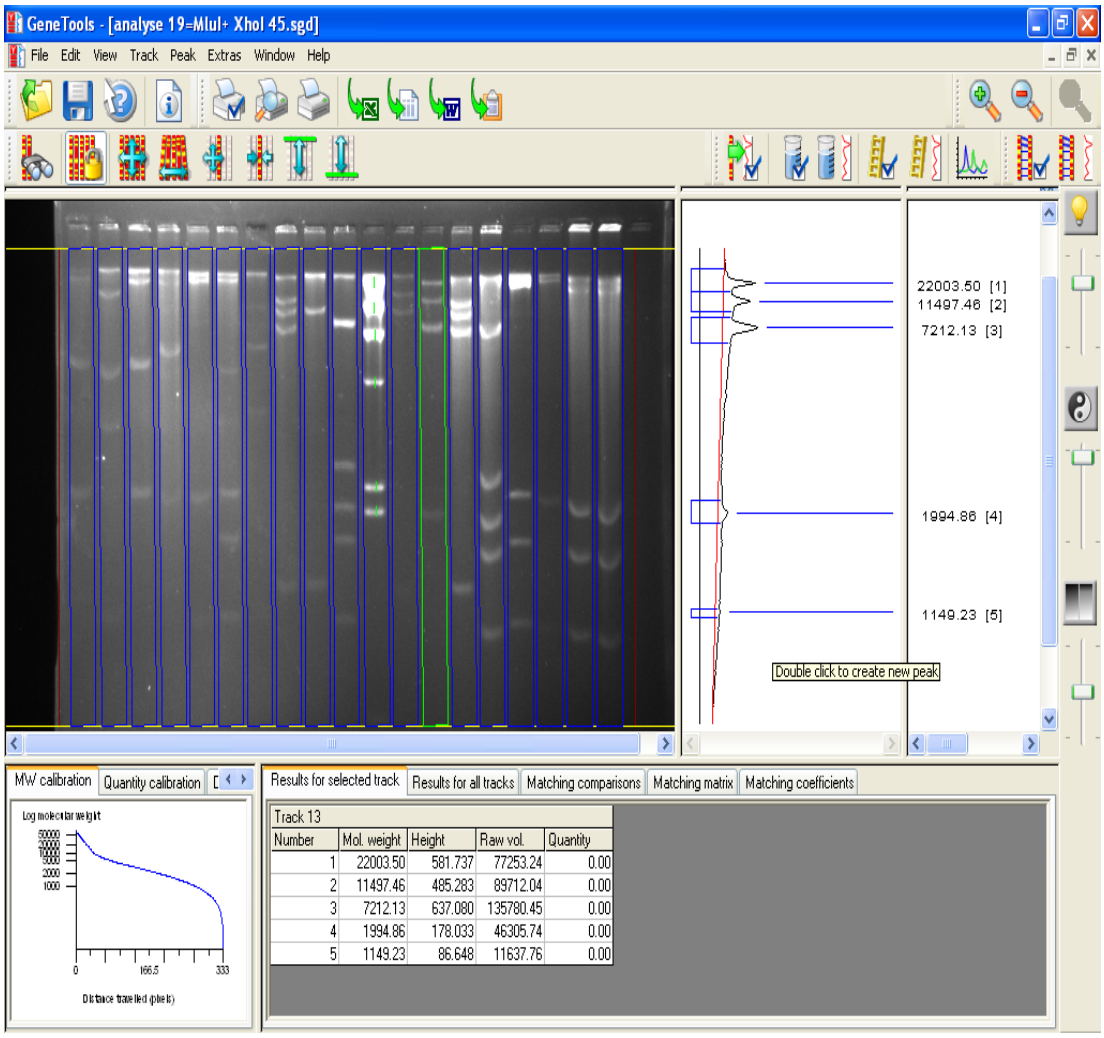
Appendix A.6 Fingerprints and overlaps analysis of eight fosmid clones derived from BAC24P12 by restriction digestion with X*ho* I/M*lu* I.



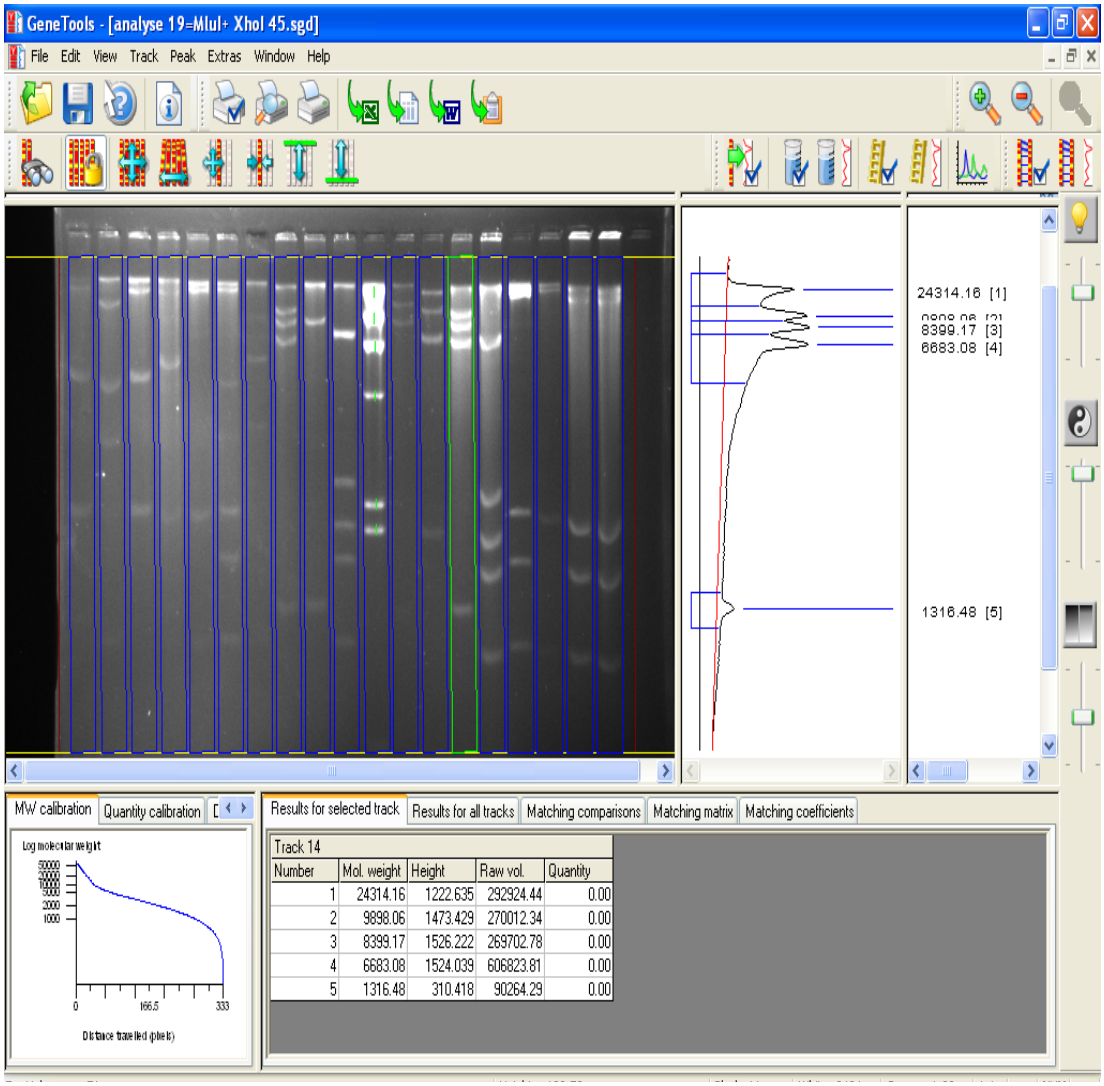
Appendix A.7 Fingerprints and overlaps analysis of fosmid clones derived from BAC24P12 by restriction digestion with S*wa* I/P*me* I.



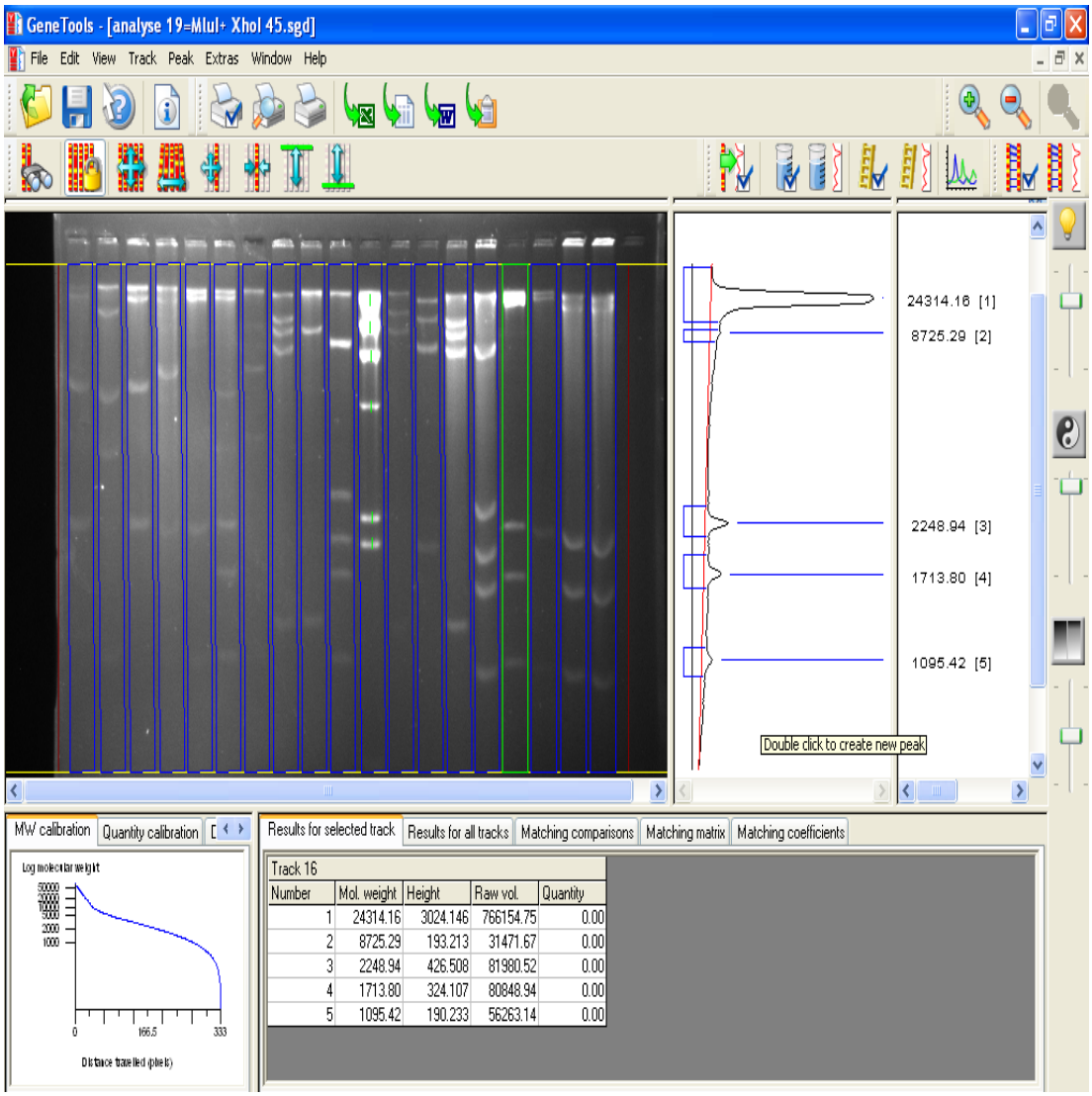
Appendix A.8 Fingerprints and overlaps analysis of fosmid clones derived from BAC24P12 by restriction digestion with S*wa* I/X*ho* I.



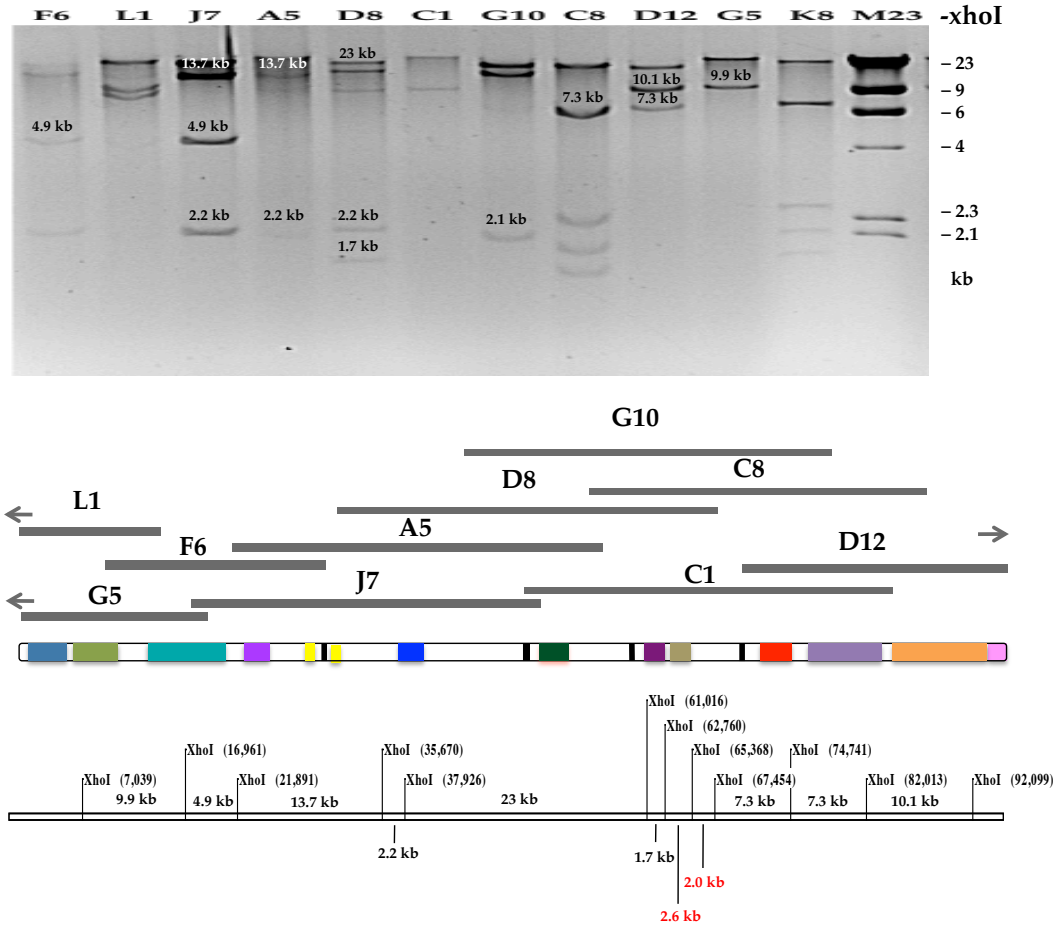
Appendix A.9 Fragment analysis of fosmid clone J2 of BAC24P12 using SynGene Gen tools. The digestion provided five fragments. At least three fragments sized by Gene Tools (10/11.4 kb), (7.3/7.2 kb), and (2.0/1.99 kb) were alike as prediction by cut map shown in Figure A.6.



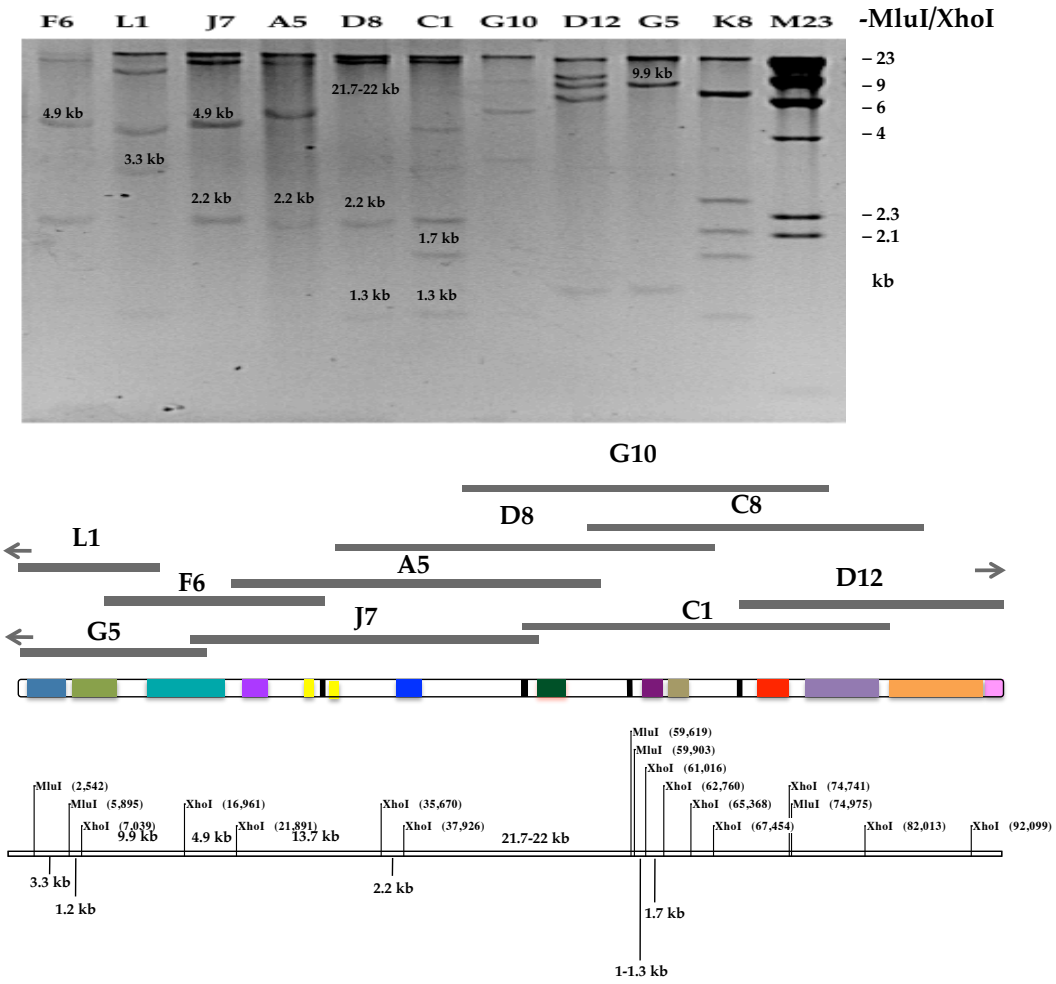
Appendix A.10 Fragment analysis of fosmid clone J9 of BAC24P12 using SynGene Gen tools. The digestion provided five fragments but three fragments sized by Gene Tools (10/9.89 kb), (8.6/8.4 kb), and (1.4/1.3 kb) were alike as prediction by cut map shown in Figure A.6.



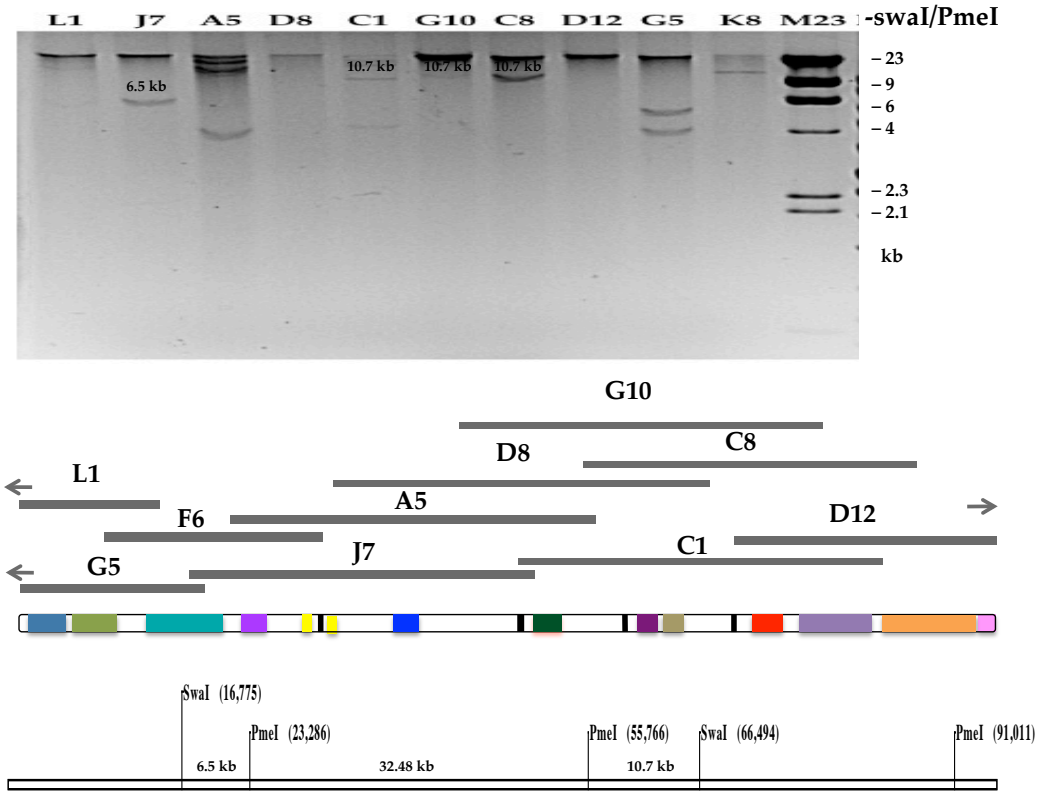
Appendix A.11 Fragment analysis of fosmid clone A1 of BAC24P12 using SynGene Gen tools. The digestion provided five fragments while two fragments sized by Gene Tools (2.2/2.24 kb) and (1.7/1.71 kb) were equal with prediction by cut map. Another size was 1.1/1.0 kb. There must be a fragment of 21 kb at the top of well but it could not separated clearly, as prediction by cut map shown in Figure A.6.



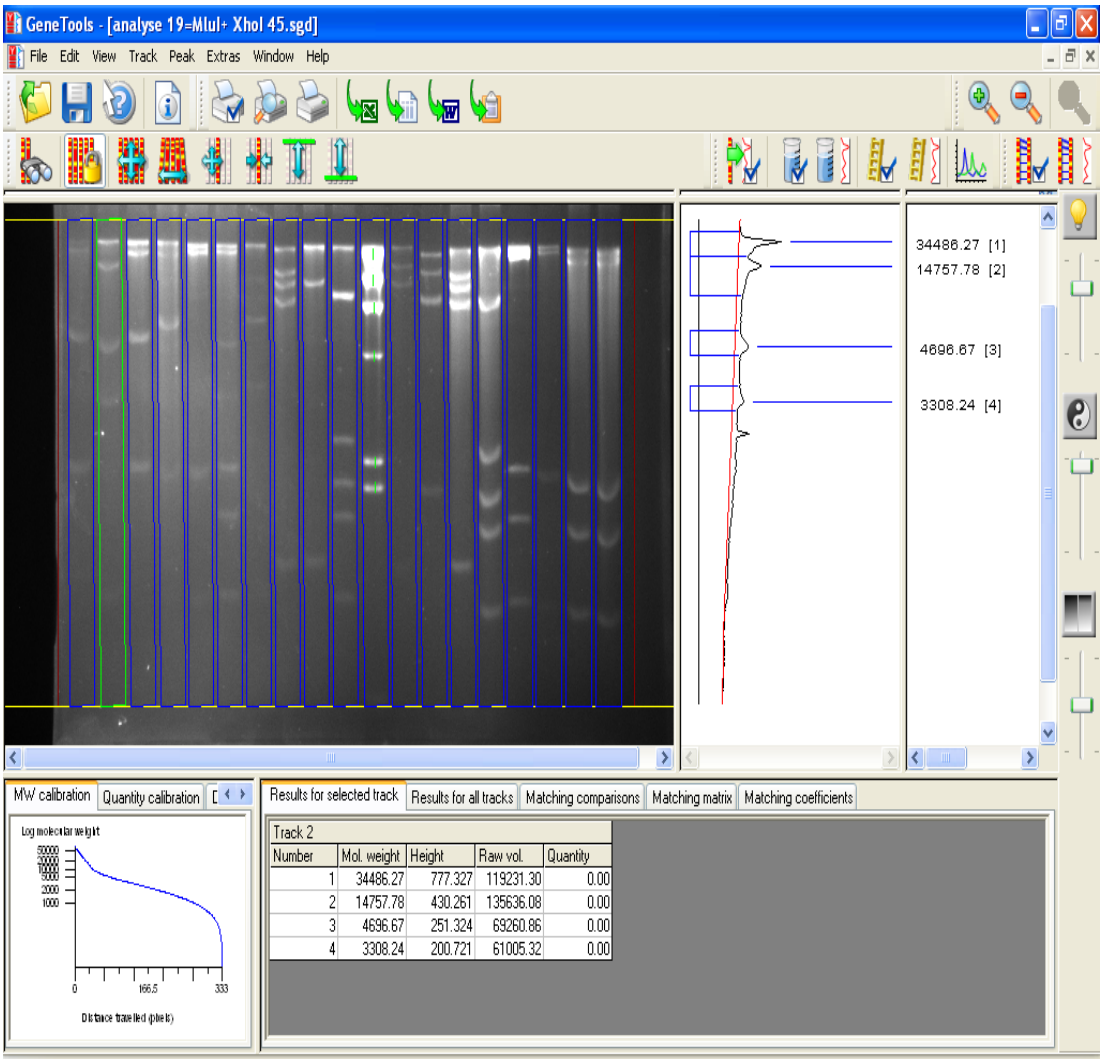
Appendix A.12 Fingerprints and overlaps analysis of eleven fosmid clones derived from BAC29M05 by restriction digestion with X*ho* I.



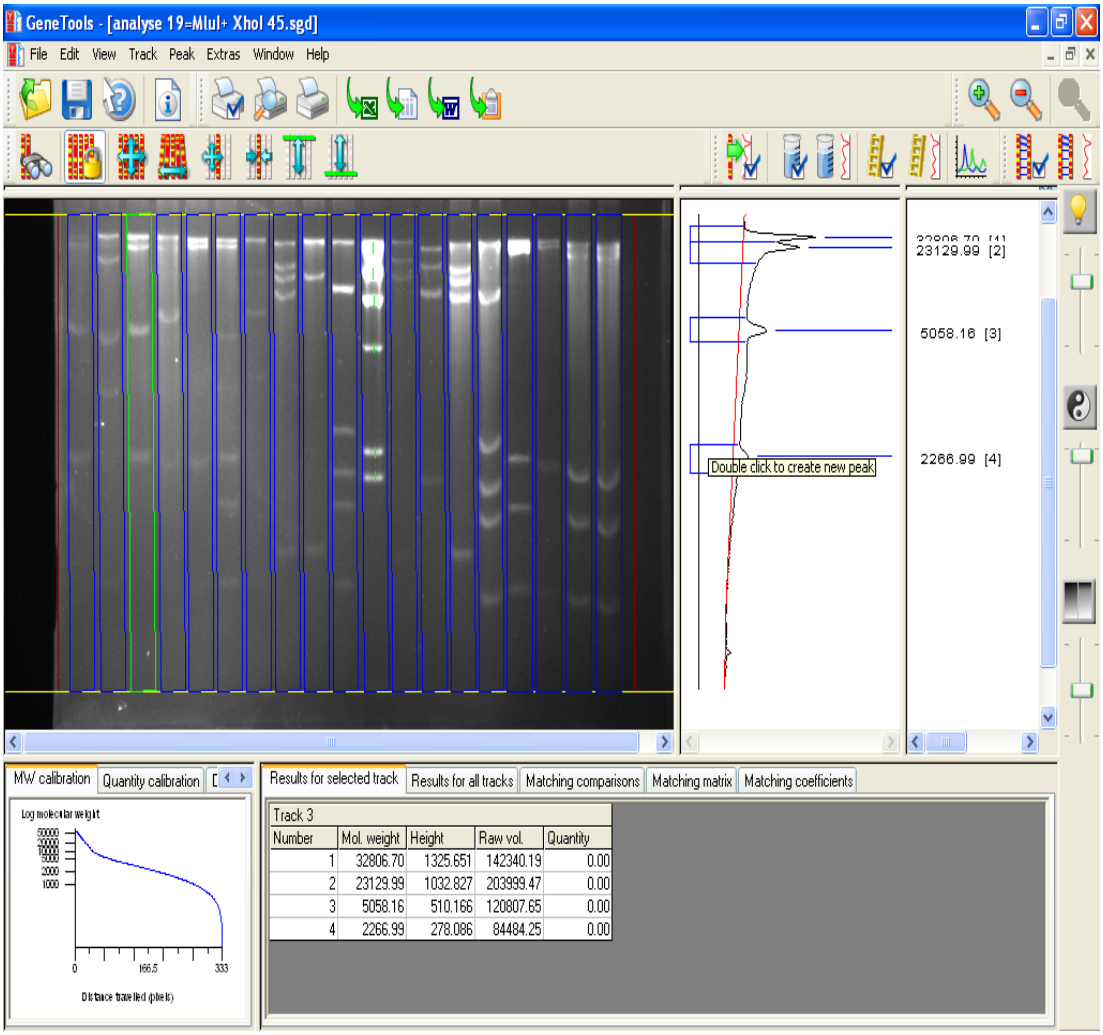
Appendix A.13 Fingerprints and overlaps analysis of ten fosmid clones derived from BAC29M05 by restriction digestion with M*lu* I/X*ho* I.



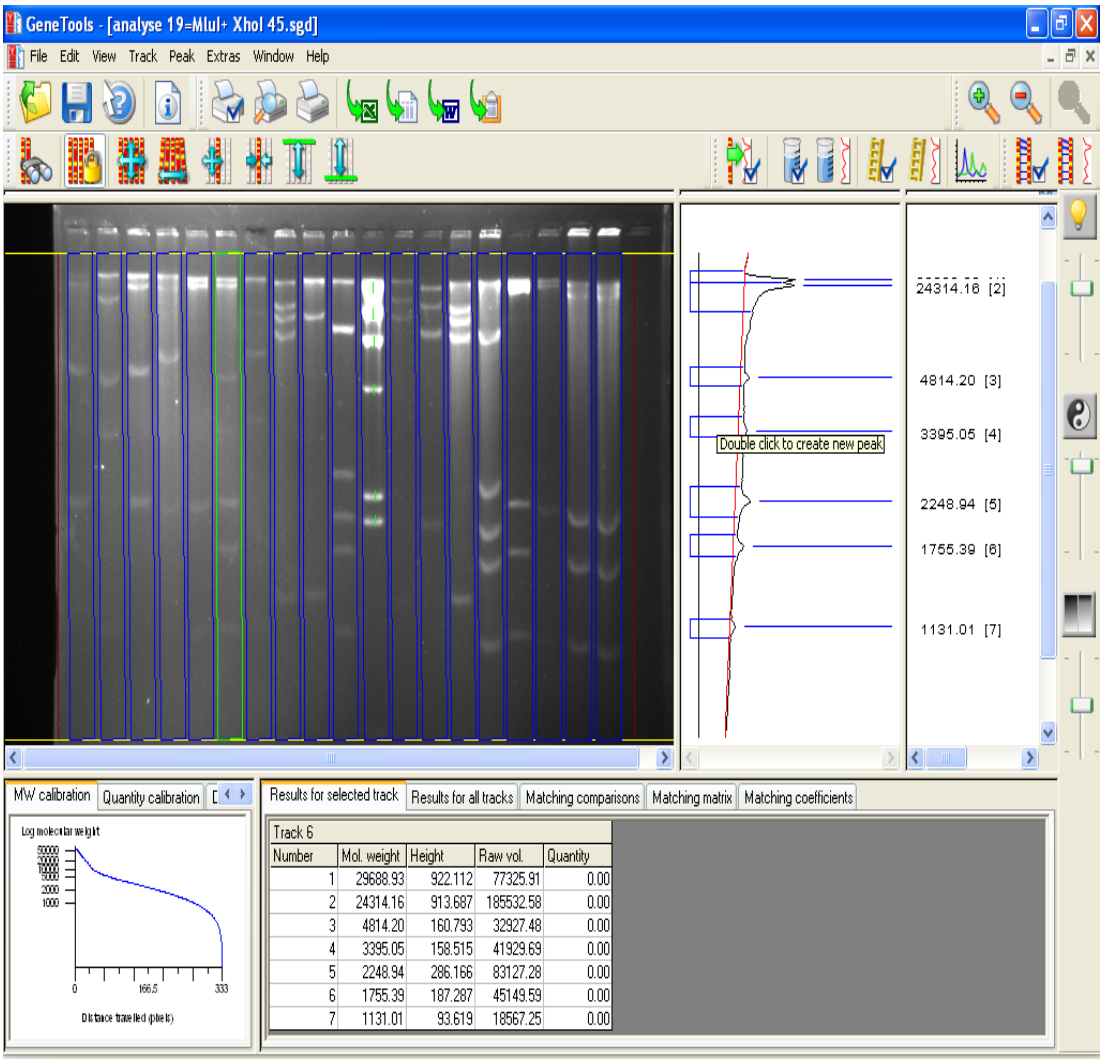
Appendix A.14 Fingerprints and overlaps analysis of ten fosmid clones derived from BAC29M05 by restriction digestion with S*wa* I/P*me* I.



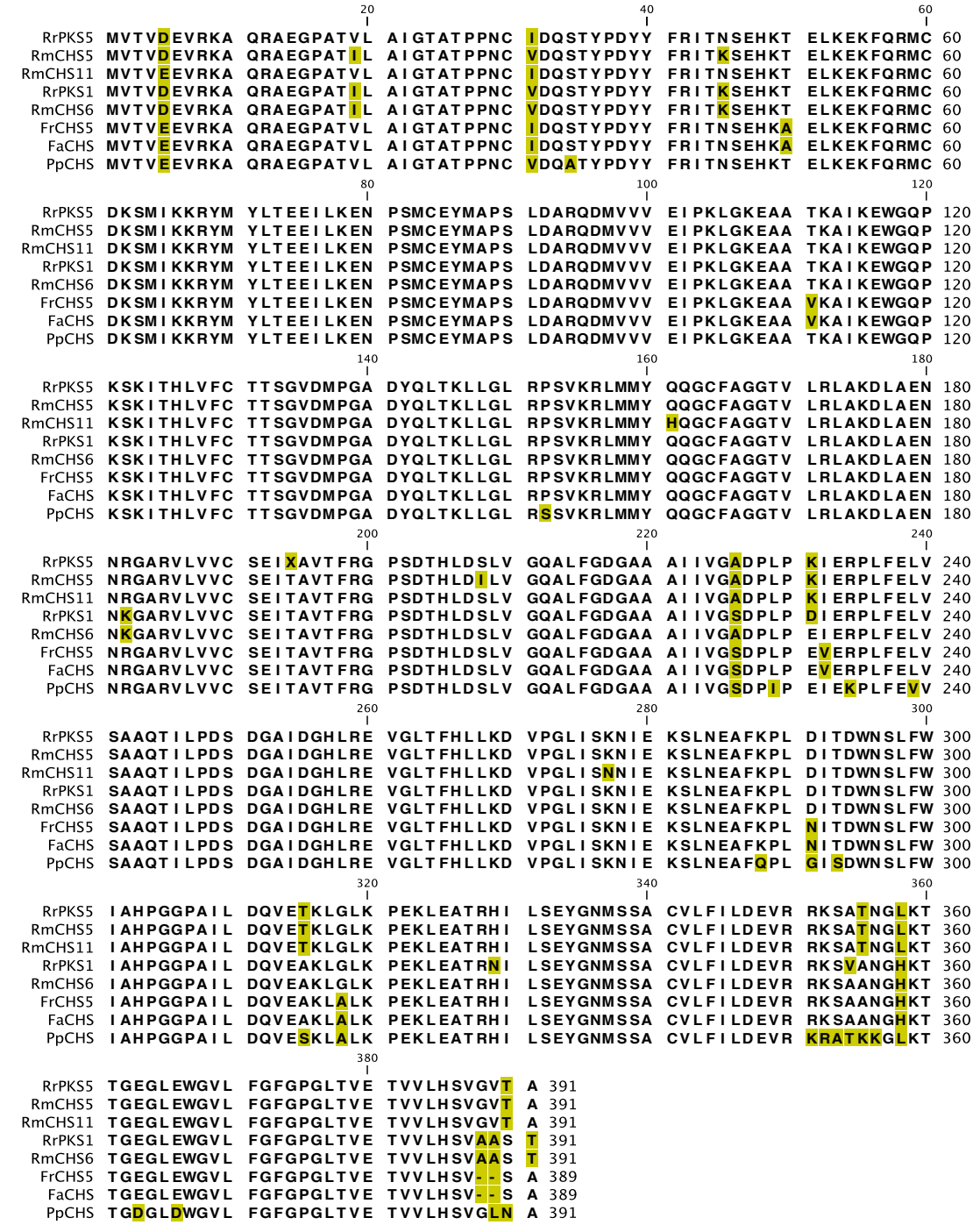
Appendix A.15 Fragment analysis of fosmid clone L1 of BAC29M05 using SynGene Gen tools. The digestion provided four to five fragments. A fragment sized by Gene Tools was equal as prediction by cut map (3.3/3.3 kb) shown in Figure A.13.



Appendix A.16 Fragment analysis of fosmid clone J7 of BAC29M05 using SynGene Gen tools. The digestion provided four fragments. Two fragment sized by Gene Tools were so equal with prediction by cut map (4.9/5.0 and 2.2/2.2 kb) shown in Figure A.13.



Appendix A.17 Fragment analysis of fosmid clone C1 of BAC29M05 using SynGene Gen tools. The digestion provided six to seven fragments. Gene Tools detected four fragments sized (1.7, 2.2, 3.3 and 4.8 kb) that were equal with prediction by cut map. Another two fragments (1.3/1.1 and 4.8/4.9) were almost same as prediction shown in Figure A.13.



Appendix A.18 Comparison to PKS/CHS protein sequences of certain fruits. Amino acids, (Pro138; Gly163; Gly167; Leu214; Asp217; Gly262; Pro304; Gly305; Gly306; Gly335; Gly374; Pro375 and Gly376), those are important in shaping the active site of CHS (Radhakrishnan *et al*., 2009), and coumaroyl CoA binding site (Ser133; Glu192; Thr194; Thr197 and Ser338) are also conserved.

LOCUS 24-RiPKS1(GM24_73-76 kb)-1182_n 1182 bp DNA linear 14-Jun-2013

DEFINITION
FEATURES Location/Qualifiers
ORIGIN

```
   1 ATGGTGACCG TCGAGGAAGT TCGCAAGGCT CAGAGGGCCG AGGGTCCGGC CACTGTCTTA
  61 GGCCATCGGA ACGGCAACTC CTCCCAACCT GTATGGACCA GAGCACGTAC CCCGACTACT
 121 ACTTTCGTAT CACCAACAGC GAGCACAAGA CTGAGCTCAA AGAGAAATTC CAGCGCATGT
 181 GTAAGTGACA AGTCAATGAT CAAGAAGCGT TACATGTACT TAACGGAAGA AATCCTGAAG
 241 GATAATCCTA GTATGTGCGA GTACATGGCA CCTTCACTCG ATGCAAGGCA AGACATGGTG
 301 GTTGTTGAAA TTCCAAAGCT CGGCAAAGAG GCTGCCACTA AGGCCATTAA GGAATGGGGT
 361 CAGCCCAAGT CCAAAATCAC CCACTTGGTC TTTTGTACCA CCAGTGGTGT CGACATGCCC
 421 GGGGCGGACT ACCAACTCAC TAAACTCTTG GGCCTCCGTC CCTCTGTCAA GCGCCTCATG
 481 ATGTATCAGC AAGGTTGCTT CGCAGGGGGC ACGGTTCTTC GGTTGGCCAA GGACTTGGCC
 541 GAGAATAACA GGGGTGCACG TGTTCTCGTT GTCTGCTCCG AAATCACCGC TGTTACCTTT
 601 CGTGGGCCTA GCGACACCCA CCTTGATAGT CTTGTGGGCC AAGCCTTGTT CGGTGACGGT
 661 GCTGCAGCTA TTATTGTTGG GGCTGACCCG TTGCCCGAGA TTGAGAGGCC CTTGTTTGAG
 721 TTGGTCTCAG CGGCCCAAAC TATTCTTCCC GACAGTGACG GGGCCATCGA TGGGCATCTT
 781 CGTGAAGTCG GGCTCACATT TCACCTCCTC AAGGATGTTC CCGGGCTGAT TTCTAAGAAC
 841 ATCGAAAAGA GCCTAAACGA GGCCTTCAAA CCTTTGGACA TCACAGATTG GAACTCACTT
 901 TTCTGGATTG CACACCCAGG TGGGCCTGCA ATTCTAGACC AAGTAGAGAC CAAATTGGGC
 961 CTAAAGCCCG AAAAGTTAGA AGCCACGAGG CACATATTAT CTGAGTACGG TAACATGTCG
1021 AGTGCTTGTG TGTTGTTTAT TTTGGACGAG GTGAGGAAGA AGTCCGCAGC TAATGGGCTC
1081 AAGACCACTG GAGAGGGTCT GGAGTGGGGA GTACTATTCG GGTTTGGGCC TGGGCTCACC
1141 GTTGAGACGG TTGTGCTTCA CAGTGTGGGT GTCACTGCTT GA
```

LOCUS 24-RiPKS1(GM24_73-76 kb)-1182_393_aa 393aa linear 14-Jun-2013

DEFINITION
FEATURES Location/Qualifiers
ORIGIN

```
  1 MVTVEEVRKA QRAEGPATVL GHRNGNSSQP VWTRARTPTT TFVSPTASTR LSSKRNSSAC
 61 VSDKSMIKKR YMYLTEEILK DNPSMCEYMA PSLDARQDMV VVEIPKLGKE AATKAIKEWG
121 QPKSKITHLV FCTTSGVDMP GADYQLTKLL GLRPSVKRLM MYQQGCFAGG TVLRLAKDLA
181 ENNRGARVLV VCSEITAVTF RGPSDTHLDS LVGQALFGDG AAAIIVGADP LPEIERPLFE
241 LVSAAQTILP DSDGAIDGHL REVGLTFHLL KDVPGLISKN IEKSLNEAFK PLDITDWNSL
301 FWIAHPGGPA ILDQVETKLG LKPEKLEATR HILSEYGNMS SACVLFILDE VRKKSAANGL
361 KTTGEGLEWG VLFGFGPGLT VETVVLHSVG VTA
```

Appendix A.19 Nucleotide sequence of RiPKSr1 region from BAC24P12 (1182 bp, at 73 to 76 kb), coding 393 amino acids, was 97% identical to PKS5 cv. Royalty and 89% to CHS5 cv. Meeker. The sequence contained two exons of 184 bp, 998 bp and an intron of 538 bp in length 155 bp longer than the PKS1 gene from Royalty.

LOCUS 29-RiPKS3(GM29_48-53 kb)-1176_n 1176 bp DNA linear 14-Jun-2013

DEFINITION
FEATURES Location/Qualifiers
ORIGIN

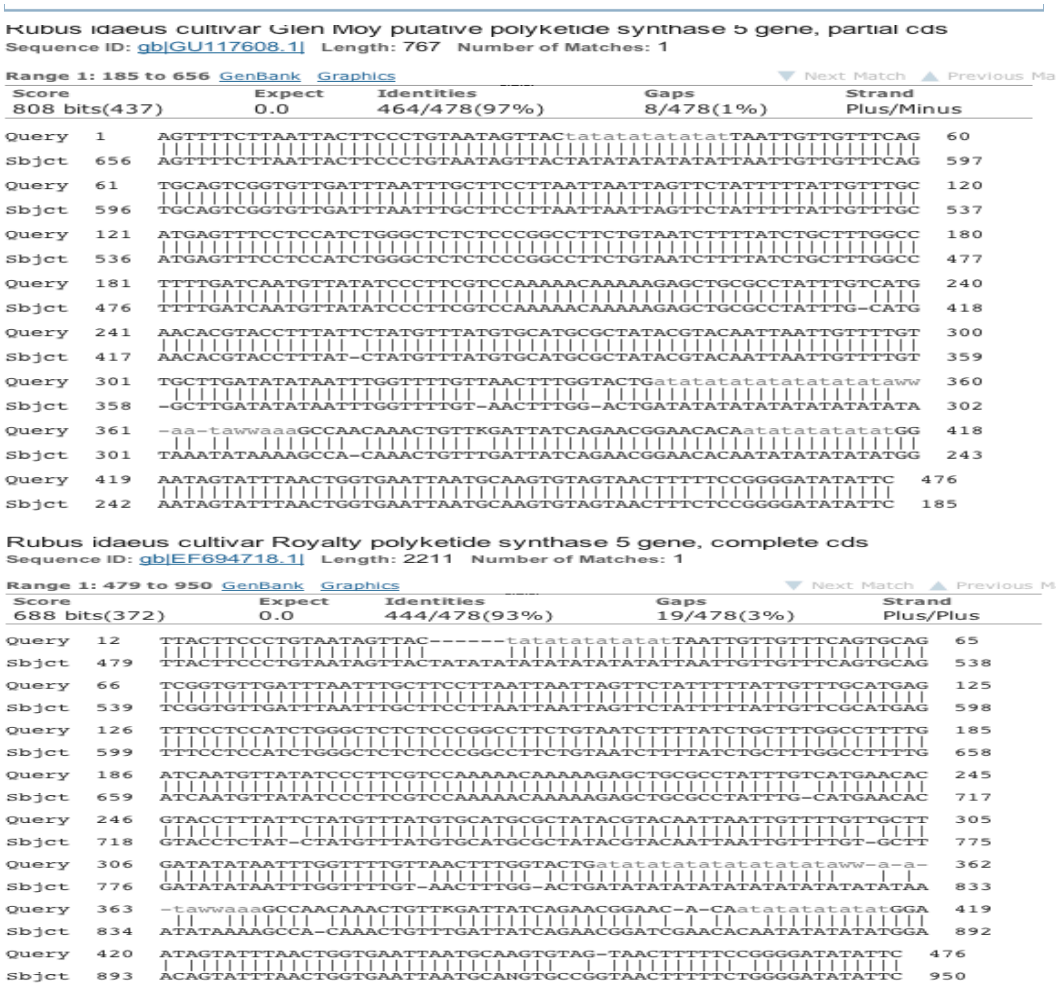
```
   1 ATGGTGACCG TCGATGAAGT CCGCAAGGCT CAAAGGGCTG AGGGTCCGGC CACAATCTTG
  61 GCGATCGGTA CAGCAACTCC TCCCAACTGT GTCGACCAGA GCACATACCC GGACTACTAC
 121 TTTCGTATCA CCAAGAGTGA GCACAAGACT GAGCTCAAGG AGAAATTCCA GCGCATGTGT
 181 GACAAGTCAA TGATCAAGAA GCGTTACATG TACTTGACGG AAGAAATCCT GAAGGAGAAT
 241 CCTAGTATGT GCGAGTACAT GGCACCTTCA CTCGATGCAA GACAAGACAT GGTAGTTGTT
 301 GAAATTCCAA AGCTCGGCAA AGAGGCTGCC ACTAAGGCCA TTAAGGAATG GGGTCAGCCC
 361 AAGTCCAAAA TCACTCACTT GGTCTTTTGT ACCACCAGTG GTGTCGACAT GCCTGGGGCC
 421 GATTACCAGC TCACTAAGCT TTTGGGCCTC CGCCCCTCCG TGAAGCGCCT CATGATGTAC
 481 CAACAAGGGT GTTTCGCCGG GGGCACGGTG CTCCGGTTGG CCAAGGACTT GGCTGAGAAC
 541 AACAAGGGTG CACGTGTTCT TGTTGTCTGC TCCGAAATCA CCGCTGTTAC CTTCCGTGGC
 601 CCTAGTGACA CTCACCTTGA TAGTCTTGTC GGCCAAGCCT TGTTCGGTGA CGGTGCTGCA
 661 GCCATTATTG TTGGGTCTGA CCCGCTACCG GATATTGAGA GGCCTTTGTT TGAGTTGGTC
 721 TCGGCGGCCC AAACTATTCT TCCCGACAGT GACGGGGCCA TCGACGGGCA TCTTCGTGAA
 781 GTTGGGCTCA CATTTCACCT CCTGAAGGAT GTTCCCGGGT TAATTTCAAA GAACATCGAG
 841 AAGAGCCTTA ACGAGGCCTT CAAACCTTTG GACATCACGG ACTGGAACTC ACTTTTCTGG
 901 ATTGCACACC CCGGTGGGCC TGCAATTCTA GACCAAGTAG AGGCCAAATT GGGCCTAAAG
 961 CCCGAAAAGT TAGAGGCCAC GAGGAATATA TTATCTGAGT ACGGTAACAT GTCGAGTGCT
1021 TGTGTGTTGT TTATTTTGGA CGAGGTGAGG AGGAAGTCTG TAGCTAATGG GCACAAGACT
1081 ACTGGAGAGG GCCTGGAGTG GGGAGTCCTA TTTGGATTTG GGCCTGGGCT CACCGTCGAG
1141 ACCGTTGTGC TTCACAGTGT GGCAGCTTCA ACTTGA
```

LOCUS 29-RiPKS3(GM29_48-53 kb)-1176_391_aa 391 aa linear 14-Jun-2013

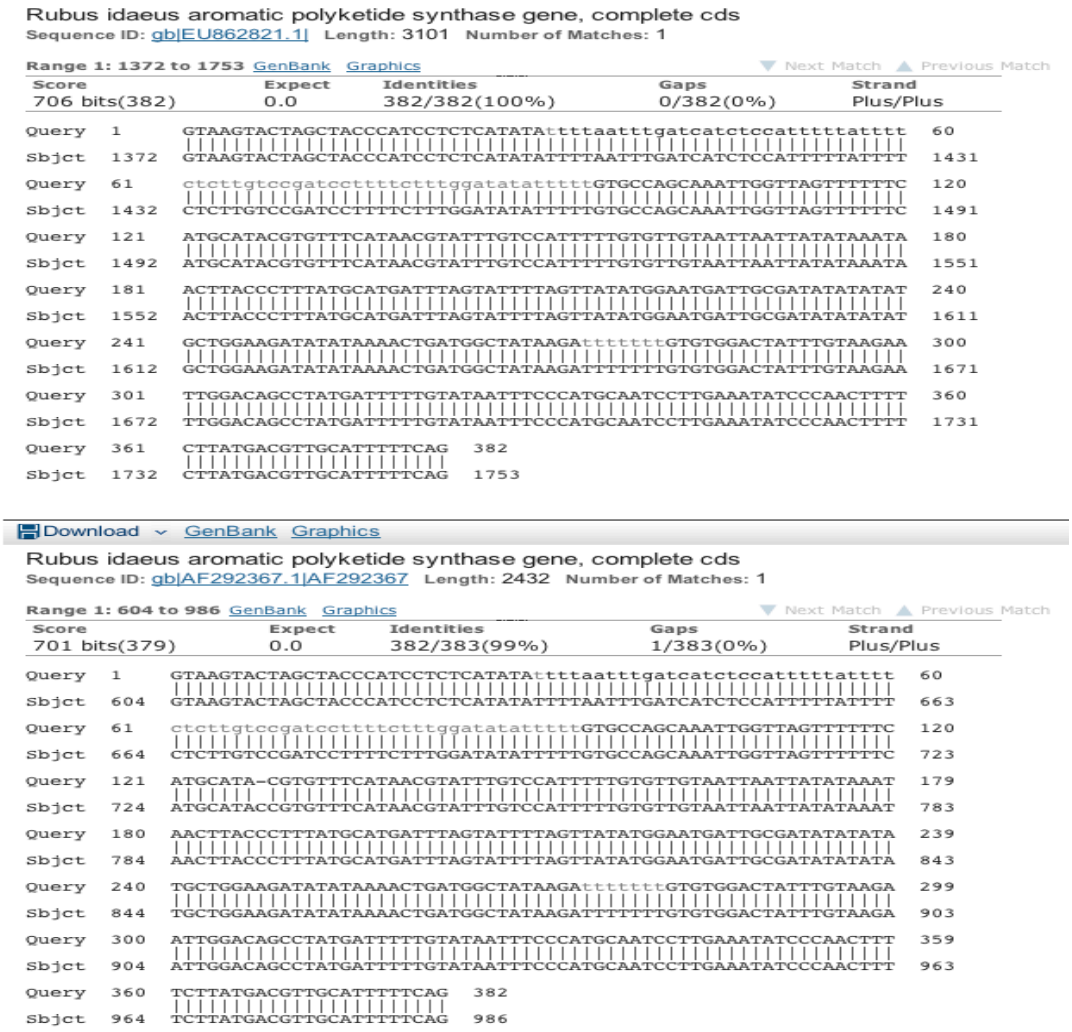
DEFINITION
FEATURES Location/Qualifiers
ORIGIN

```
  1 MVTVDEVRKA QRAEGPATIL AIGTATPPNC VDQSTYPDYY FRITKSEHKT ELKEKFQRMC
 61 DKSMIKKRYM YLTEEILKEN PSMCEYMAPS LDARQDMVVV EIPKLGKEAA TKAIKEWGQP
121 KSKITHLVFC TTSGVDMPGA DYQLTKLLGL RPSVKRLMMY QQGCFAGGTV LRLAKDLAEN
181 NKGARVLVVC SEITAVTFRG PSDTHLDSLV GQALFGDGAA AIIVGSDPLP DIERPLFELV
241 SAAQTILPDS DGAIDGHLRE VGLTFHLLKD VPGLISKNIE KSLNEAFKPL DITDWNSLFW
301 IAHPGGPAIL DQVEAKLGLK PEKLEATRNI LSEYGNMSSA CVLFILDEVR RKSVANGHKT
361 TGEGLEWGVL FGFGPGLTVE TVVLHSVAAS T
```

Appendix A.20 Nucleotide sequence of RiPKSr3 region from BAC29M05 (1176 bp), coding 391 amino acids, was 100% identical to PKS1 cv. Royalty. The sequence contained two exons of 178 bp, 998 bp and an intron of 382 bp.



Appendix A.21 Comparison of intron nucleotide sequences of RiPKSr1 region of BAC24B12 derived from red raspberry cv. Glen Moy, showing 97 % homology with intron sequence of PKS5 cv. Royalty.



Appendix A.22 Comparison of the intron nucleotide sequence of RiPKSr3 region of BAC29M05 derived from red raspberry cv. Glen Moy, showing 100% homology with intron sequence of PKS cv. Royalty.

```
LOCUS GM29-48-53 kb_sg 5970 bp   DNA  linear    UNA 05-Apr-2013
DEFINITION
FEATURES             Location/Qualifiers
ORIGIN

   1 ACGTACCCTA TATATTAAAT CCTATTGTGT GAAACGACAG GGGGCCAAGG CCACCTAAGG
  61 CCAATATACA TGGCTCCGTC CCTGACCTTC GATTAGGGCT GTCAATTCCA ACACGACCCG
 121 ATAACACGAC TCGAAACCCG CACGAAATAA AACGGGTTGA ACCCGCACGA TTAAAAAACG
 181 GGTCATTTGC GGGTCAACCC GCCAACACGA AATAAATCCG TGTAACCCGT TTATTAAACA
 241 TGTTGGGTTA GGTTATATAT AATTGGAATA ATTTTTTTGT GCCTTGGAAT AAATTTTTTG
 301 CTGTTACATT TGGATTTATC TACTTATGTT TTATTTATTA ATCATATCCA TGTATATTTA
 361 TATTTTCTTC TTATAATGTG AAGACAACAT GACAGTGAAG TTATGTAATA TTACTTTCAA
 421 GATTTTAATA TAGTTTTTTA GTTAAATGAA TCACAATGTT AACACGTTAA ATGGTGTAAA
 481 TGGGTCATTA AACTCAACAC GACACGACCT AATTATTAAA TGGGTTAGGC GGGTTGTAAA
 541 CGGGTTACCC GTTTAATAAA TGGGTTGGGT TTGGGTTTAC ATTTTTGACA CGATTATTAA
 601 ATGGGTTGGG TTTAGGTTTA CGTTTTTGAC ACGATAAAAA CTTTAACACG ACCCGAACCC
 661 AACACGACAC GACCCATTGA CAGCCCTACC TTCGATCTCG ACCAATCCCA GATCTCAACA
 721 CCCCCTACTA GAAGGCCACA TCGAACCAAC CCAGATTTGA TTCCCCAAAC CCACAACACC
 781 ATAGCCCAAA CTCCATTGCC CTAGCCACCG TAGTAGAGCC ACCAGAGTAA CCTTGGTGGG
 841 AGGCGATGGT TACTCTGGCA AGGTGGTTAC TCTGGTGGTA GAAAAATCAG TAGAGTAACC
 901 ACCATAGGGC GGTGGTTACT CTTGGATTGT TTTTTGTTTT TGTTTTTGGT CAAGGTGAAT
 961 TGGATTGTTT AGATAGAGGT TGTTACTCTT GTGAATTTTG ATGAACCTTA ATTAATTATT
1021 TGGCCCGCCG GTAAAGGGAG TACGTTATAG TTTTTAAAAA ATAATTATGA ATATTCCCAT
1081 TATTTATTAT AGACATAGAG AATGTAATGC ATGCAATTGC ACATACTGAA TGGCTCGATC
1141 GATCAAGAGT GTTCTGTAAT ATCTAATTCT TTGGGTACGG TATTAATTGG CTCATTATCC
1201 CCTTTGATGC ATCGATCAGT TGCATTCTAT TCATGATAAT ATAGATACTA TCCAGCTAGG
1261 ATCGATCTAG TACTTTTAAT CCTACGTTTA ACAAAAAATTW AAATTGATTA GAATAAAACA
1321 AAATGATTAG TGTATTATAT CGTCAATATG AAGAAATTAT GAAAGTTTGG CTTGCTATTC
1381 AAATCTCTTT AGGGGTCACA AACTCATTCT ACAATTAATA AGGCACGTAT ATATGGTGCA
1441 AGAAGAATTA AATGGCTTTG GTGTACACTA ATTTGTTTGT TGCTCGTTTA ACAAAAATAA
1501 TAATAATATA TAACAACGAA GATAAGAACA GTGAGCTTGG GTCTAGCTAG AATTTGAAAG
1561 AATCAATCAG CTACTCAAAA CTGTTTGTTC ATTTACCTAA TATATATCGA AACAACATAC
1621 AAGCACATGA TGTTAGTACT GGTGATTAAC CTTGATAGGA AGCGACATAA TATATGAACC
1681 ATAAAACGTA TACATTATGG AATGAAAAAG GAWAGGATGA GGATACGTTG AATGAGAATG
1741 TAGGTAGCTA AAGATCGAAT TTCTCAAAAT AAAATGCTAG TAAAGAAACA AATTAAGGCA
1801 AACATATTTA GTGTATTTAA TTATGATCTA TCAAGGTACA CATATACGCA TGTAATTGCA
1861 ATGTAGGTAG GCAAAATTTA CACATACTGG GTAGGTATGT AGAATTACGA CGTACCGTAC
1921 GTACACCTAG CTACATTATC AATATATATA TATATATATA TGGAAAATTG GCAACCTAAT
1981 TAATTAGCTC CACACATTAA TCACAATAAG TAAGTTGCCG TAAAAGTAAG TTGCGTTGAA
2041 TTCGATCGAT CCATCTTTAA TTAAGAATAA GTTGGATGCA TGCATGCATG GTGTGTTAGT
2101 AGCTAGTAGC TAGCTAGCTA GCTACTACGG TGGAACGTGG TAGTTCTACT AAGCCTATCA
2161 TCTTCAAGGG ACCAGAATTA AGTAGGAGCC TTGGAGGACC CGAATAATAT CAAATCTATA
2221 TATTAATTCC AAATTAGAGT TGGTAACAGG TCGATCGGTA ATTCGGTATG TTAAGTATTA
2281 CATACATATA TCTCAATTAC CATAGTCCAT ATATATATAT TGCCAACTCA TGCATCCCTA
2341 CATGCATGGA CTGTCGCTTC CTTCTTCTTT TTTGGCTAAT TGGACTGTCG CTTCCTATTA
2401 ATAATATATC CCATATATAT TCCTAAACCA CCCCAAAATG AAACTGGTAA TAATTCTCCG
2461 CACCAAATTA TTGGTAGTCT GCTCTGACCC TCACGCCCTT GTTCATAACC TACTATTGGT
2521 TTAAAAAAAA ACTGTAAATA TATATAATAT GTTAGCTAGG TTGATCGTTG CATATATTTT
2581 ATAAAGTATG ATCATATTCG TAAATTAAAA AAAACCAAAA AAAAAACAMA TCTAGCAGAT
2641 AAATAAAAAT GAATTAACCM CATCTATWTW TACAGAATTG AAGATATATA ATTTAAACCA
```

Appendix A.23 Nucleotide sequence (48-53 kb) and regulartory motifs of RiPKSr3 (PKS1) in BAC29M05 derived from cv. Glen Moy, identified 100% with PKS1 cv. Royalty, carrying two exons of 178 bp, 998 bp and an intron of 382 bp, 2 TATAA boxes (TATAAATA), 2 motifs of bHLH (CACGTG) and 2 motifs of recognizing by RNA cleavage enzymes (AATAAT).

```
2701 TATGCTTAAT TGATATATAT ATATATATAT ATATACTACG AATAGAATTA ATTTCATTAA
2761 TTAGGATTGA TTATTTGTCT CCAATGTGGT CTATAAATAA ATGTGCAAGG TCCGTGTGGT
2821 ACTGTTGCTC TATCGAAAAT AAATTTTGAA TTAAAAAGAT TAAAAAAGAA GGAGGGGGTG
2881 TGGTGATTTT TCCTGCCCTT TTGTTTGGTA GGTAAGATGA GAGAACTAAC ACATTCCAAA
2941 TCATGTGGGT TTTGATATGC AACAAACACC GCACAAAAAC TATGCACGTG CCATTGGTTG
3001 GGGGTTTCTA ACAAGGCAAT TGGCAAGCAC GTGATCCCCC AGCTACCCAC CCCTCCCTTC
3061 TCAACCCTTG TTTCATCAAA TATAAATAAC AACCCATTTC TTCGTACCAT TATATATTCT
3121 AACATCTTTG CATTGCCATT GCCATTGCCA TTGTAACATC GCAAGACTTT CACAACACAA
3181 CCTGCAGCTC CATCTGTCCT TTCTATTCAT CTTTTCTTCA CAGATCAAAA ATGGTGACCG
3241 TCGATGAAGT CCGCAAGGCT CAAAGGGCTG AGGGTCCGGC CACAATCTTG GCGATCGGTA
3301 CAGCAACTCC TCCCAACTGT GTCGACCAGA GCACATACCC GGACTACTAC TTTCGTATCA
3361 CCAAGAGTGA GCACAAGACT GAGCTCAAGG AGAAATTCCA GCGCATGTGT AAGTACTAGC
3421 TACCCATCCT CTCATATATT TTAATTTGAT CATCTCCATT TTTATTTTCT CTTGTCCGAT
3481 CCTTTTCTTT GGATATATTT TTGTGCCAGC AAATTGGTTA GTTTTTTCAT GCATACGTGT
3541 TTCATAACGT ATTTGTCCAT TTTTGTGTTG TAATTAATTA TATAAATAAC TTACCCTTTA
3601 TGCATGATTT AGTATTTTAG TTATATGGAA TGATTGCGAT ATATATATGC TGGAAGATAT
3661 ATAAAACTGA TGGCTATAAG ATTTTTTTGT GTGGACTATT TGTAAGAATT GGACAGCCTA
3721 TGATTTTTGT ATAATTTCCC ATGCAATCCT TGAAATATCC CAACTTTTCT TATGACGTTG
3781 CATTTTTCAG GTGACAAGTC AATGATCAAG AAGCGTTACA TGTACTTGAC GGAAGAAATC
3841 CTGAAGGAGA ATCCTAGTAT GTGCGAGTAC ATGGCACCTT CACTCGATGC AAGACAAGAC
3901 ATGGTAGTTG TTGAAATTCC AAAGCTCGGC AAAGAGGCTG CCACTAAGGC CATTAAGGAA
3961 TGGGGTCAGC CCAAGTCCAA AATCACTCAC TTGGTCTTTT GTACCACCAG TGGTGTCGAC
4021 ATGCCTGGGG CCGATTACCA GCTCACTAAG CTTTTGGGCC TCCGCCCCTC CGTGAAGCGC
4081 CTCATGATGT ACCAACAAGG GTGTTTCGCC GGGGGCACGG TGCTCCGGTT GGCCAAGGAC
4141 TTGGCTGAGA ACAACAAGGG TGCACGTGTT CTTGTTGTCT GCTCCGAAAT CACCGCTGTT
4201 ACCTTCCGTG GCCCTAGTGA CACTCACCTT GATAGTCTTG TCGGCCAAGC CTTGTTCGGT
4261 GACGGTGCTG CAGCCATTAT TGTTGGGTCT GACCCGCTAC CGGATATTGA GAGGCCTTTG
4321 TTTGAGTTGG TCTCGGCGGC CCAAACTATT CTTCCCGACA GTGACGGGGC CATCGACGGG
4381 CATCTTCGTG AAGTTGGGCT CACATTTCAC CTCCTGAAGG ATGTTCCCGG GTTAATTTCA
4441 AAGAACATCG AGAAGAGCCT TAACGAGGCC TTCAAACCTT TGGACATCAC GGACTGGAAC
4501 TCACTTTTCT GGATTGCACA CCCCGGTGGG CCTGCAATTC TAGACCAAGT AGAGGCCAAA
4561 TTGGGCCTAA AGCCCGAAAA GTTAGAGGCC ACGAGGAATA TATTATCTGA GTACGGTAAC
4621 ATGTCGAGTG CTTGTGTGTT GTTTATTTTG GACGAGGTGA GGAGGAAGTC TGTAGCTAAT
4681 GGGCACAAGA CTACTGGAGA GGGCCTGGAG TGGGGAGTCC TATTTGGATT TGGGCCTGGG
4741 CTCACCGTCG AGACCGTTGT GCTTCACAGT GTGGCAGCTT CAACTTGAAG GCATCTGGGT
4801 TCACTTGGGT GATCTCCTCA GTAATTTGTT CCTATATATT GTTTCCACTT TACTTTTCTG
4861 GCTAGTTGTC GTGTTTTTTT GAAGTTATTT TTCTGCTGAA CTACATTTAG GCATCCAGAG
4921 GTAAACTTTG ATGCCATGGC TGTGTCTATT TCCGTGTTGT TAATTACTAG ATTGATGAAG
4981 AAAACAAAAT GGCCTATGTA GTGATATATG TAATGATGAA TAATATTATT TCACATATTT
5041 CATATGATAA AAAGTTCTGA TTGCTTTGCT ACTATTTCGT GCTTGAATAT ATATTATATT
5101 GGCAGTTTGT CACAACAACA AAAAAACAAA CAATTGTTAT GTTTTTTTTT TCACAGAAGG
5161 GTGATTCGAT CGATTCTAAT ATTCTATTAA TTCAATCACT AGTGTATTAC AAAAAAAATT
5221 AATTCAAAGA GATACATGCA TCGATTTCAA TGGAAAATTG TAAAAACAAT ATAATTTATA
5281 AAATATTATA TTAGCCACTG CATCGGMMAA AAAATTGGCC TACCTAAGAG ATGTGCTTAG
5341 AATAGAAATT TGAGACAGAT ATATGACGTA GCAAGCTAGA GAATAAAGCT CTACGTATTT
5401 CCCTTTTTAA TTATCTATGG GTCTTGTATA AGTCCCTCAG TAATGAGTCA AATTGGAACA
5461 TTAAAATTAC ACAATTTTTT TACACAAATA TATGACCATT AGATCTAATA ATTTTTAGTA
5521 GTAGTGCAAT AGATTCTAAA TATTTTTACC TAAATTTTCT ATACTAAATG ATGTGTCATA
5581 TGATATGGCT GGCCCAACAT CATCACTTTT TATTTTGTTT ATATGCTACG TGGTCAATCA
5641 CCCTTAATAT CATACATTTT ATATTCAGCA GAAAAATAAA AATAGAGATT TTTTTGGTAA
5701 AAAAAATAGA AGAAACACGG TAACTTAATT AACCTAATAT AATATTAAAA TTACTAAAAT
5761 TTCAACAGAC CCACCACGCA CAACCATCCC GTCACACAAA ATTGTAGAGA AAAAATTTCT
5821 AATAGCAAGG AACAACTATC CTCGATTTAC ATACAAAATT TTCATATGGA ACTCAAAAAA
5881 TTACTCCTTT TGTCAATAAT ATTAGTGGGT TATAGATAAA TATTAACTTC TTCATTACAA
5941 TGAAATATAG GTTTTTTTAG AAAAAAAAAT
```

Appendix A.23 Continued

```
LOCUS GM24-72-81 kb_sq f_4   5928 bp DNA   linear   UNA 06-Apr-2013
DEFINITION
FEATURES             Location/Qualifiers
ORIGIN

   1 ATCAAAACGA TCCTGCACTC TCAATTCTCA TCCAATAGGT CTCATTTATG ACACATGTCA
  61 ATCCATAATA GGTGTAAATG TTAAAAAAAA TTGACACATG TCAACCTATG ATAGGTGACA
 121 CATGTCCTAA ATGGACCTAA TATGTCTAAT TTATGACATA TGTTGACTCA TAATAAATGT
 181 ACATGTAAAA ACAAATAGTT TAGTCTATTT TTCTTATTTA TGTGATGATT CAATGAACTT
 241 CGATCAAATT TTTAAATTTA TAGACAATTC TTAATGACAC AACTATAGAT AACAAATAGG
 301 AATTTTAATT TTTATTATTA TTATTATTAT TAGTATTATT ATTATTATTA TTATTATTAG
 361 TATTATTATT ATTATTATGT AAATTCAATA AAATAAATTG TTTGAAATGC CCCTCCTGGC
 421 TTTTGGATTT CTAGCTCCGC CCGCTTCTAT CCCATAATTT TCATAGGTGG ACAAGTGAAA
 481 ATGAAAATGC CCATAAGCTT TATGTTTGTC CTCGAGTAAT GCTGTTTGTT GGGGAATTAA
 541 CTCAAACCTA CCACGTGTGC ACACTCAATT TATAATATAA ATTTGCATCT TATGTTCATC
 601 TAACTATTGT GACACTCGCC TACCACCCAA TTTATTGGTC ACACTCACAC ACCCATGCAT
 661 AGTCAAAACA AATAATCGTG TCCCTCGGAT TAAAAACAAC AAGAAAGAAA CGAAGGGTTG
 721 GATACAACTG AAGTAGCCAA AATTTGGCTC CCGTGCTAGC AAAGGTTAGC CTTTGTTTCC
 781 TATATCCATA AACGGTTGCC ACAATAAACT TAAATCTCTC TTTTAGCTAG AGTAGTTGTC
 841 TCAGTGACAC GTCCAACTTA TCAATCATAT GCATACATTA AAATTATATA TATATGAAAA
 901 TATTTTTGGT GTTTCGCAAC TTTCGCTACG TATTCACTAT ATAGCATATG TAATTTGGGT
 961 ACAAGTGTTC ATCATGAGGT TTTCATAGAG GTTAAAATTG AGAACTTCAA GAGTGTTCAA
1021 GAGCGAGAGT GACGTCCATG CATGGCTAGA TGGTTTGCAG GGTGGCAATG GTTGTCGACA
1081 GTGGCTATTA CCTATATATA TATATTAGAA ATGATTTTAA TGGATAGATA TTCAACAAAC
1141 ATGAAGAGGT AGTTACTTTA TTCAAACAAT TAAGATTTTA ATTCAATCAA AMACATGAGA
1201 GCGTTTCTCT CTTCCATACC TAGCCGCCAT GGGCTAGGTC TTCTAGTTCT TCTATGTTGT
1261 CTCTTGTCGG CGACCTCTAC GACGGTTTCT GGTGCATGGC TGCCCCAACG ACCTCGATCT
1321 ACTGTGAGTT TCCTTGATGG TGGTGACTTT GCCGAAGGAG GAGACTTTTC ACATGACTAT
1381 GGCTTTGATC TCTTTTGGTC TGGGCTGCGA AGATTTGATC CCGTTCTATC GCTTAGGAGT
1441 GGTGATGCGC GACTCAACGA CAATGGGGTT GGAGGTCTTG TTGCTTTTCT GCTCTGTTCG
1501 TTGAAGGGGG TCGGCGGTGA CAACTCGAAG CAAGGGGGTG ACGGCGATGG GTCAGCGGCC
1561 TTGTTGTTGA TCGGCGTCTT CTGCTTGCAG TATGGGTCTT TTTGGTTGGT TGTCTTTGGT
1621 TGGGTTAGGG GTGAGGCTTT TGGGCCTTAT TCTGAGATCC AATTAGAGCG GGTCATTCCA
1681 TTGGGCTGGG CTATCGGGGA GGGTGGTTTG CCACTCTCAG TTATTACTTA GGGTTTCGTC
1741 TTTGAGCCTG CAATTATGAG AATTTTCGTG TGATCATTTG ATCGAAAAAT CTTTCAGTCG
1801 CACCACAATT GAGTGCATTG GCGGCGGAAG TTTCATTTAT TTGTTTACAT GATTGGAGAC
1861 TTTTTGTCTT TTGTAGGCTC ACTTGAATAA GTGAGCGAGA ATTGTACTAA TGAGCGGGCT
1921 TGATTTTTCT TGGCCTACTT GGCCACTGAT ATAGGGCACA AATATATGAG ATTTATTGTA
1981 TGTGCCGTTC TGGCTTCGGG ATTTAATGAT ATCTCTTTTC CTATCAAAAA AACATGAAGA
2041 GGTAGATAAA ATTTCTAGTT TAATATTTGG AGGCAGTTCA TATTAATTGG TTAGAAAAGC
2101 AAAGTTCTAA AAATCGCTAG GCGGTTGTCG AACGGCGGGA AAGGGTCTAG CGCCTAAATG
2161 ATTTTTTATT TTATTTTAAT TTTATTATAC AACATATAAA TAACTAAATA AGTGTATGCT
2221 TAAAACAAAA ACTATACTTG TATGTATTGA GAAACATAAA CATGCAAAAT GGAAGGGTAC
2281 TATAATTGAT TGAAAACATA AAATTCTGAT TTTTTTTAAA AACTTGCCTA GGCGGGGATT
2341 GTTGCTAGAA AGTATGAAGT TATGAACTAT CTAATGAGTA GTGACGACGA CAAATTAATA
2401 ATATGCTCAA TAACACAAGT GTATCAAAAT TTAAGTACAG AGTTTAATAT CGCCGTACAT
2461 ACGCTAATGA AGAGTGTCCA ATTTTCTACT CCGAGTAGTT AAATAAACAA GTTCAAGTAT
2521 TCTCTCTAAA AAAAAAAATA ATCGATAGCC TGAATATCGT ATCCAACATA ATACTCGTAA
2581 TCCGAAAGTC GTATACTCGT ATCCTTGTAC TTGGGCAAAG AACAACTCAT GGTGTCAACT
2641 CTCTTGCTCT CAACTATGTG ATTAGTCATT AGTGGCTGGC CAGAATGGTA TTACTGACCA
2701 ATTCACGAGA TCCTGCACTT GACGCCATAA ACAAACTCGG AAGGTTAGGT TGATTGCACG
```

Appendix A.24 Nucleotide sequence (72-81 kb) and regulartory motifs of RiPKSr1 (PKS5) in BAC24P12 derived from red raspberry cv. Glen Moy, identified 98% homology with PKS5 cv. Royalty, carrying two exons of 184 bp, 998 bp and an intron of 534 bp, 1 TATAA box (TATAAATA), 2 motifs of bHLH (CACGTG) and 1 motif of recognizing by RNA cleavage enzymes (AATAAT).

```
2761 TACCACATAA TTACACGACT CCAAGTAAAG TAAAGAACTA AAGATAGACG TAGACACAAT
2821 TAATTTGTTT CATAATTTTA TTTTTTTTAA GAAAAAGTAA GAGGGAGAGA TGGAGAGACG
2881 TAGGTACGCA AGACATTTTC GAAAAATGAG ATGAATATAT GCGAGGGTGA GGGTGGGTTT
2941 GATCGATGGA CAAGCCTCCA CGCCAAACTG AGAAGGGCAG GCACGTGCCA CAACTCCACA
3001 AGGCATTTGG TAGTTGAATT AGGCATTGGT GAAGCACGTG AACCCCACGC TACCCTCTCT
3061 TTTCCCTCCT CTCCCCATAA TCCCATATAA ATAAAACAAC TCCCACTCAC CATTTTGTCA
3121 CACCTTCYTC YTCCTCCTCC TCCTCCTCCT CCTCCTCCTC GATCACTGCA ACACCACAAA
3181 CCTCAACTAC CTCCTCCACT TTTCGATCAA ATGGTGACCG TCGAGGAAGT TCGCAAGGCT
3241 CAGAGGGCCG AGGGTCCGGC CACTGTCTTA GGCCATCGGA ACGGCAACTC CTCCCAACCT
3301 GTATYGACCA GAGCACGTAC CCCGACTACT ACTTTCGTAT CACCAACAGC GAGCACAAGA
3361 CTGAGCTCAA AGAGAAATTC CAGCGCATGT GTAAGTACTA GCTAGCTAGC TTCTTCTGCT
3421 GCCTCTTCAT GTTTTCCTTG TCGATCGGAT TTCCAGTTTT CTTAATTACT TCCCTGTAAT
3481 AGTTACTATA TATATATATT AATTGTTGTT TCAGTGCAGT CGGTGTTGAT TTAATTTGCT
3541 TCCTTAATTA ATTAGTTCTA TTTTTATTGT TTGCATGAGT TTCCTCCATC TGGGCTCTCT
3601 CCCGGCCTTC TGTAATCTTT TATCTGCTTT GGCCTTTTGA TCAATGTTAT ATCCCTTCGT
3661 CCAAAAACAA AAAGAGCTGC GCCTATTTGT CATGAACACG TACCTTTATT CTATGTTTAT
3721 GTGCATGCGC TATACGTACA ATTAATTGTT TTGTTGCTTG ATATATAATT TGGTTTTGTT
3781 AACTTTGGTA CTGATATATA TATATATATA TAWWAATAWW AAAGCCAACA AACTGTTKGA
3841 TTATCAGAAC GGAACACAAT ATATATATAT GGAATAGTAT TTAACTGGTG AATTAATGCA
3901 AGTGTAGTAA CTTTTTCCGG GGATATATTC AGGTGACAAG TCAATGATCA AGAAGCGTTA
3961 CATGTACTTA ACGGAAGAAA TCCTGAAGGA TAATCCTAGT ATGTGCGAGT ACATGGCACC
4021 TTCACTCGAT GCAAGGCAAG ACATGGTGGT TGTTGAAATT CCAAAGCTCG GCAAAGAGGC
4081 TGCCACTAAG GCCATTAAGG AATGGGGTCA GCCCAAGTCC AAAATCACCC ACTTGGTCTT
4141 TTGTACCACC AGTGGTGTCG ACATGCCCGG GGCGGACTAC CAACTCACTA AACTCTTGGG
4201 CCTCCGTCCC TCTGTCAAGC GCCTCATGAT GTATCAGCAA GGTTGCTTCG CAGGGGGCAC
4261 GGTTCTTCGG TTGGCCAAGG ACTTGGCCGA GAATAACAGG GGTGCACGTG TTCTCGTTGT
4321 CTGCTCCGAA ATCACCGCTG TTACCTTTCG TGGGCCTAGC GACACCCACC TTGATAGTCT
4381 TGTGGGCCAA GCCTTGTTCG GTGACGGTGC TGCAGCTATT ATTGTTGGGG CTGACCCGTT
4441 GCCCGAGATT GAGAGGCCCT TGTTTGAGTT GGTCTCAGCG GCCCAAACTA TTCTTCCCGA
4501 CAGTGACGGG GCCATCGATG GGCATCTTCG TGAAGTCGGG CTCACATTTC ACCTCCTCAA
4561 GGATGTTCCC GGGCTGATTT CTAAGAACAT CGAAAAGAGC CTAAACGAGG CCTTCAAACC
4621 TTTGGACATC ACAGATTGGA ACTCACTTTT CTGGATTGCA CACCCAGGTG GGCCTGCAAT
4681 TCTAGACCAA GTAGAGACCA AATTGGGCCT AAAGCCCGAA AAGTTAGAAG CCACGAGGCA
4741 CATATTATCT GAGTACGGTA ACATGTCGAG TGCTTGTGTG TTGTTTATTT TGGACGAGGT
4801 GAGGAAGAAG TCCGCAGCTA ATGGGCTCAA GACCACTGGA GAGGGTCTGG AGTGGGGAGT
4861 ACTATTCGGG TTTGGGCCTG GGCTCACCGT TGAGACGGTT GTGCTTCACA GTGTGGGTGT
4921 CACTGCTTGA ACTTGAAATT GAAGTTGAAG GCATCTATCT ATTTGTTCTG TGGTGATCGA
4981 TTTTATCTGC TCCTATATAT TATATATGTA TGATTTGCAT CTATTAATTT ATAGCTAGGT
5041 TTGATTTTGG GAATTTGTTT TCTTAGAGGC TTGTGTGGGC ATCATGGGGT AAGCTTTGGT
5101 GCAAATTGCT GCTGTGTTTA CCTTTCATGT TGTTTATTGC ATTTCTGAAG ACAAAAGTGT
5161 AGCGACTTAT ATATATATAT ATATATATGT TTCCATTTCA TATTGTTAAA TTACTATATA
5221 TAACTAATTT GCTTTTGTTC CATAAAGTTC GTATAATTGA GCAGATCAAT TCGTTAATCA
5281 CACCATTTAT CATGCTAATT CAACTTGTCA TTCCCACTAG CCAAGCAGGA GCATTAACCT
5341 TCTGCATCCA CAGTAGTTTT TCCATCCTTT TCCATGAAGA GAGAAATTAG GGTTAATATT
5401 TAATTACAAT CATAGTTTGA GAAATTATAG AAAAATTATA TTAATGTTGC CAATTGGTTT
5461 TATTGTGACA ATCATATTAC TTTATTATAA TATCATAGCC ACATTCGCCC ATGTGTGAAA
5521 GCCCAACGGC CACATGTACT TCACGACTTC ACGTTACCCA ATATAAGTTG TCCACATATA
5581 TGTTTTGAAA ATTCGCCACA CGTGCCAGGG CGTGTGAGAA TGTAGATTGT AGAATATTGT
5641 CCTACATTGC AAAATATAAA TTTTACAAAT GTGTTTATAA GGATTTGAGT CTGTCCAACC
5701 ATTGTCAATT ACTTTTAGTT GTGACTGGTG ACACATATTA TTTTATCACT AATGAATTAA
5761 ACTTCGGTTC CCGATGATTG AGTTATGTTG CTAATTGGTA CTGTAAAACT TTTATAAAAT
5821 AATACTTTGA GTAATTGAGT TTGAGGACTT AATAAAAATT ATTATTTTAG AGGGATTATT
5881 TATTTAGCGA TAATGAATAA TTTATTAATT TGCCGATAAA TAAATATT
```

Appendix A.24 Continued

LIST OF COMMUNICATIONS

PRESENTATIONS

- Oral presentation – WASET, International Conference Program, Venice, Italy, 27-29th April (2011). Genomic Studies of the Polyketide Synthase Genes in Red Raspberry.

- Poster presentation – Plant Genome Evolution, Amsterdam, The Netherlands, 4-6th September, (2011). Genomic Studies of the Polyketide Synthase Genes in Red Raspberry.

- Poster presentation – Sixth Rosaceous Genomic Conference, Trento, Italy, 30th September-4th October (2012). Generation of BAC-based Physical Map of the Red Raspberry Genome.

- Poster presentation – Sixth Rosaceous Genomic Conference, Trento, Italy, 30th September-4th October (2012). Fosmid Library Construction of Aromatic Polyketide Synthase in Red Raspberry *Rubus idaeus*.

- Oral presentation – ATPER, Stockholm, Sweden, 1st – 2nd June (2013). Genomic Studies of the Polyketide Synthase Genes: Generation of BAC-based Physical Map of the Red Raspberry Genome.