QTL Mapping, Gene Identification and Genetic Manipulation of Glucosinolates in *Brassica rapa* L.

BY

Arvindkumar H. Hirani

A Thesis
Submitted to the Faculty of Graduate Studies at the University of Manitoba
in Partial Fulfillment of the Requirements
for the Degree of

DOCTOR OF PHILOSOPHY

Department of Plant Science
University of Manitoba
Winnipeg, Canada R3T 2N2

July 2011

ACKNOWLEDGEMENTS

I would like to express my very sincere thanks to my supervisor, Dr. Genyi Li, Department of Plant Science, University of Manitoba, for providing excellent guidance and valuable advice throughout my PhD. program. I am also grateful for his encouragement and assistance in the course of my research.

I would also like to thank my advisory committee members, Dr. Peter B. E. McVetty, Department of Plant Science, University of Manitoba and Dr. Anne C. Worley, Department of Biological Sciences, University of Manitoba for their valuable suggestions, helpful discussions and guidance for my research.

I would like to thank Dr. Fouad Daayf for giving me access to the HPLC equipment facility in his Lab. I would like to extend my thanks to Mr. Lorne Adam and Dr. Abdelbasset El Hadrami for their assistance in HPLC for glucosinolate analysis. I would like to thanks Dr. Tom Ward for valuable help at the LC-MS facility in the Manitoba Chemical Analytical Laboratory, Department of Chemistry, University of Manitoba.

My sincere thanks go to Dr. Carla Zelmer, Dr. JianFeng Geng, Dr. Gao Feng, Dr. Santosh Kumar, Ravneet Behla and Roger Watts for their technical support and helpful discussions. I appreciate all the past and present members of Dr. Li's Lab for their advice and help in molecular techniques. My thanks also go to the greenhouse staff, Ian Brown and Cathy Bay. I would like to thanks all the members of Department of Plant Science.

I am heartily grateful to the Heavenly Father, God, who gifted me this life and being always my source of strength, energy and hopes for the best. I praise and thank you my almighty Creator.

I am deeply and forever indebted to my parents, Hasmukhray and Kantaben for their love, support and encouragement throughout my entire life. I am also very grateful to my wife, Ashmita. Without her love, encouragement, understanding and support, it would not have been possible to finish this thesis. I am also thankful to my only sister

Sangita, and my brothers Vipul and Dilip. Special thanks to my friends for their help and support.

Finally, I acknowledge that this research would not have been possible without the financial assistance of Genome Canada, Genome Alberta, NSERC, the University of Manitoba Graduate Studies and the Department of Plant Science at the University of Manitoba. I express my gratitude to all of them for financial support throughout my PhD. program.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

2,3DH3MB	2,3-dihydroxy-3-methylbutyrate
2,3DH3MP	2,3-dihydroxy-3-methylpentanoate
2A2HB	2-aceto-2-hydroxybutyrate
2AL	2-acetolactate
2MP	2-Methylpropyl
2OB	2-oxobutyrate
3MOB	Methyl-2-oxobutanoate
3MOP	Methyl-2-oxopentanoate
4OH	4-Hydorxyglucobrassicin
AFLP	Amplified fragment length polymorphism
AHAS	Acetohydroxyacid synthase
ALY	Glucoalyssin
ANOVA	Analysis of variance
APCI	Atmospheric pressure chemical ionization
BAR6	Yellow sarson (genotype of *B. rapa*)
BC	Backcross
BCAT	Branched chain amino transferase
Bo-	*Brassica oleracea*
Br-	*Brassica rapa*
BU-RILs	BAR6 x USU9 recombinant inbred lines
CIM	Composite interval mapping
CYP	Cytochromes P450
DH	Doubled haploid
DHAD	Dihydroxyacid dehydratase
ESI	Electro spray ionization
FAE	Fatty acid elongation
GBNAP	Glucobrassicanapin
GERU	Glucoerucin
GNAP	Gluconapin
GSL-ALK	Glucosinolate side chain modification gene
GSL-AOP	Glucosinolate side chain modification gene
GSL-ELONG	Glucosinolate side chain elongation gene for ≥ 4C
GSL-FMO	Glucosinolate flavin monooxygenases
GSL-PRO	Glucosinolate side chain elongation gene for 3C
GSLs	Glucosinolates
GST	Glutathione *S*-transferase
HPLC	High performance liquid chromatography
IAA	Indole-3-acetic acid

IAN	Indole acetonitrile
ILE	Isoleucine
IPM	Isopropylmalate
IPMDH	Isopropylmalate dehydrogenases
IPMI	Isopropylmalate isomerases
IPMS	Isopropylmalate synthase
KARI	Ketolacid reductoisomerase
LC-MS	Liquid chromatography-mass spectrometry
LG	Linkage group
LOD	Logarithm of the odds
Mal-Der	Malate derivative
MAM	Methylthioalkylmalate synthase
MAS	Marker assisted selection
MTBO	4-methylthio-2-oxobutanoate
MTOHP	4-methylthio-2-oxoheptanoate
MTOHX	4-methylthio-2-oxohexanoate
MTOP	6 methylthio-2-oxopentanoate
NIRS	Near infrared reflectance spectroscopy
PDA	Photodiode array
PRO	Progoitrin
Pyr	Pyruvate
QTL	Quantitative trait loci
RAPD	Random amplified polymorphic DNA
RFLP	Restriction fragment length polymorphism
RI16	Chinese cabbage (genotype of *B. rapa*)
RILs	Recombinant inbred lines
RNAi	RNA interference
SCAR	Sequence characterized amplified region
SNP	Single nucleotide polymorphism
SRAP	Sequence related amplified polymorphism
SR-RILs	Sarson x RI16 recombinant inbred lines
SSR	Simple sequence repeats
ST	Sulfotransferase
STS	Sequence tagged sites
TAGSL	Total aliphatic glucosinolates
TGSL	Total glucosinolates
TIGSL	Total indole glucosinolates
UGT	Glucuronosyltransferases
USU9	Canadian spring type DH lines (genotype of *B. rapa*)
VAL	Valine
XRF	X-ray fluorescence spectroscopy

ABSTRACT

Glucosinolates are amino acid derived plant secondary metabolites with a moiety of thioglucose and a moiety of sulfonated oxime bound to various R-groups. Glucosinolates and their breakdown products are found in the order *Capparales* or *Brassicales* depending upon classification and are an important class of phytochemicals involved in direct or indirect plant-microbe, plant-insect, plant-animal and plant-human interactions. In this study, quantitative trait loci (QTL) mapping of individual leaf and seed glucosinolates was performed in *B. rapa* using two recombinant inbred line (RIL) populations, designated SR-RILs and BU-RILs. Nine gene-specific molecular markers were developed and tagged on different linkage groups on the SR-RIL genetic map and 4 gene-specific markers were integrated on the BU-RIL genetic map for glucosinolate biosynthesis genes, *GSL-ELONG*, *GSL-PRO*, *GSL-FMO$_{OX1}$*, and *GSL-AOP/ALK*. In addition, genetic manipulation of aliphatic glucosinolates through homoeologous gene replacement and/or introgression was performed using aneuploid backcrosses followed by marker assisted selection in *B. rapa*.

Identification of QTL for leaf and seed glucosinolates was performed using a genetic map of *B. rapa* from a recombinant inbred line population (SR-RILs developed from a cross of Chinese cabbage by turnip rapeseed). The genetic map was developed using a total of 1,579 molecular markers including 9 gene-specific markers for glucosinolates. Several large and small effect QTL for progoitrin (2-hydroxy-3-butenyl), gluconapin (3-butenyl), glucoalyssin (5-methylsulfinylpentyl), glucobrassicanapin (4-pentenyl), 2-methylpropyl (isobutyl) and 4-hydroxy-indole-3-yl-methyl (4-hydroxyglucobrassicin) glucosinolates in leaves and seeds were identified which

explained between 6 and 54% of the phenotype variance. Interestingly, a major QTL for 5C aliphatic glucosinolates in leaves and seeds was co-localized with a candidate *Br-GSL-ELONG* locus on linkage group A3, displayed co-segregation with the co-dominant SCAR marker BrMAM1-1. The same locus (*Br-GSL-ELONG*) was also co-localized with the QTL regulating seed and leaf gluconapin, suggesting that the gene(s) might have either multi-functional characteristics for 4C and 5C side chain aliphatic glucosinolate biosynthesis or differentiated functional properties of tandem duplicate genes at the locus.

A second recombinant inbred line population (BU-RILs) derived from a cross of yellow sarson (BAR6) and a canola type DH line (USU9) segregated for glucoerucin, gluconapin and progoitrin as major 4C aliphatic glucosinolates together with 4-hydoxyglucobrassicin indole glucosinolate. A genetic map was developed using SRAP markers and four glucosinolate gene-specific loci were integrated on the map. Phenotyping for seed glucosinolates was performed in greenhouse and field environments but only in greenhouse environments for leaf glucosinolates. Analysis of variance revealed that there were no significant differences between these two environments for seed glucoerucin, gluconapin, sum of 5C glucosinolates or total aliphatic glucosinolates. However, differences for seed progoitrin, sum of 4C, indole and total glucosinolates were observed, suggesting that biotic and abiotic stresses might have impacted the production of these glucosinolates. Four QTL for seed glucoerucin and three QTL for seed gluconapin were identified in greenhouse and field environment conditions. These QTL accounted for up to 49% of the phenotypic variance. Three QTL were identified for 4C aliphatic glucosinolates in seeds, suggesting that at least three *Br-GSL-ELONG* loci with epistatic interaction might be involved in 4C side chain aliphatic glucosinolate

biosynthesis in *B. rapa*. The results of a genetic correlation study of 4C glucosinolates suggested that glucoerucin is the initial precursor for the remaining 4C aliphatic glucosinolates, glucoraphanin, gluconapin and progoitrin.

Genetic manipulation of aliphatic glucosinolates was performed using aneuploid backcrosses followed by marker assisted selection in *B. rapa*. A resynthesized *B. napus* line was crossed and then backcrossed with three *B. rapa* genotypes, RI16, BAR6 and USU9 to develop aneuploid lines for the glucosinolate genes *GSL-ELONG*$^{-}$ and *GSL-PRO*$^{+}$ followed by marker assisted selection. Early and advanced backcross progenies displayed variable marker transmission rates ranging from 4 to 73%. In the RI16 genetic background, reduction in 5C aliphatic glucosinolates was observed in 15 (30%) of the BC_3F_2 plants and those plants were positive for *GSL-ELONG*$^{-}$ markers. The A-genome and gene-specific SCAR marker BrMAM1-1 was absent in those plants which showed a reduction in 5C aliphatic glucosinolates, suggesting that the functional allele had been replaced by a non-functional *GSL-ELONG*$^{-}$ allele from *B. oleracea*. Some plants with functional *GSL-ELONG* locus replacement had a reduction of 15 to 25 µmole/g of seed 5C aliphatic glucosinolates compared to the recurrent parent RI16. However, several advanced backcross plants with introgression of *GSL-ELONG*$^{-}$ that did not show reduction in 5C aliphatic glucosinolates were also observed. All positive *GSL-PRO*$^{+}$ advanced backcross plants in the RI16 genetic background produced the 3C aliphatic glucosinolate sinigrin, suggesting that introgression of or replacement with a functional *GSL-PRO*$^{+}$ gene from *B. oleracea* might be responsible for the production of sinigrin in *B. rapa*.

FOREWORD

This thesis is written in manuscript style as outline by the Department of Plant Science, University of Manitoba. The thesis begins with a general introduction and literature review followed by the three manuscripts that comprise the main parts of the thesis. Each manuscript consists of an abstract, introduction, materials and methods, results and discussion. The thesis ends with a general discussion, conclusion, a cited list of references and appendices. All the manuscripts were written according to the Theoretical and Applied Genetics (TAG) journal format. The manuscripts will be submitted to scientific journals for publication.

1. GENERAL INTRODUCTION

The genus *Brassica* possesses three diploid and three amphidiploid species, all of which are commercially important agricultural crop species. These species have been cultivated worldwide as sources of vegetable oil as well as for leafy, floret and root vegetables. For example, *B. napus*, *B. juncea*, *B. carinata*, *B. rapa* and *B. nigra* are all grown for edible and industrial oil production. *Brassica napus* is predominantly cultivated in Canada, China, Australia, northern Europe, whereas *B. juncea* and *B. rapa* are cultivated on the Indian subcontinent as the major edible oilseed crops (Downey 1990, FAO 2008 http://faostat.fao.org/site/339/default.aspx). Similarly, different morphs of *B. rapa* and *B. oleracea* are grown as vegetable crops. Rapeseed/canola quality is based on the fatty acid profile of the oil and low total glucosinolates in the meal. In *Brassica* vegetables, however, desirable glucosinolate profiles in plant tissues viz., leaves, flowers, roots are important quality parameters.

Glucosinolates are sulphur containing secondary metabolites biosynthesized by many plant species in the order *Capparales* (*Brassicales*). Physical tissue or cell injury leads to the breakdown of glucosinolates through the hydrolytic action of the enzyme myrosinase, resulting in the production of compounds including isothiocynates, thiocyanates and nitriles. Derivative compounds of glucosinolates have a wide range of biological functions including anti-carcinogenic properties in humans, anti-nutritional effects of seed meal in animals, insect pest repellent and fungal disease suppression (Mithen *et al.* 2000, Brader *et al.* 2006).

Glucosinolates play important role in the nutritional qualities of *Brassica* products. *Brassica* products are consumed as oil, meal and as vegetables. Rapeseed (*B. napus*, *B. juncea* and *B. rapa*) is a source of oil and has a protein-rich seed meal. Edible quality of oil, however, is adversely affected by high concentration of erucic acid. The seed meal remaining after oil extraction is rich in protein with balanced amino acids. High glucosinolates in the seed meal pose health risks to livestock (Fenwick *et al.* 1983, Griffiths *et al.* 1998). Consequently, plant breeders have nearly eliminated erucic acid from the seed oil and have dramatically reduced the level of seed glucosinolates (>100 µmole/g seed to <30 µmole/g seed) via conventional breeding, allowing the nutritious seed meal to be used as an animal feed supplement. There is, however, a significant residual content of glucosinolates in rapeseed/canola seed meal (over 10 µmole/g seed) and further reduction of the total glucosinolate content would be nutritionally beneficial (McVetty *et al.* 2009). Therefore, to produce healthy seed meal from rapeseed, it is important to genetically manipulate glucosinolate content. *Brassica* vegetables (*B. rapa* and *B. oleracea*) are highly regarded for their nutritional qualities, they are a good source of vitamin A and C, dietary soluble fibres, folic acid, essential micro nutrients and low in calories, fat and health beneficial glucosinolates such as glucoraphanin and sulforaphane. Breeding objectives for these Brassica crops include the enhancement of beneficial glucosinolates and reduction of others. It is, therefore, important to understand the genetic, biosynthetic, transportation and accumulation mechanisms for glucosinolates in *Brassica* species.

Several studies to elucidate various glucosinolate biosynthesis steps at the gene or transcriptional levels have been reported for the model plant and *Brassica* relative, *A.*

thaliana. Figure 2.5 illustrates the current understanding of the biosynthesis pathway of glucosinolates in *Brassica* and *Arabidopsis*. Several glucosinolate side chain elongation genes (*AtMAM1*, *AtMAM2* and *AtMAM3*) have been cloned and functionally characterized in *A. thaliana* (Compos de Quiros *et al.* 2000, Kroymann *et al.* 2001, Textor *et al.* 2007) and *B. oleracea* (*BoGSL-ELONG*) (Li and Quiros 2002, Zheng *et al.* 2010). Similarly, two genes (*IPMS1* and *IPMS2*) for glucosinolate side chain elongation at the propyl branch (Kraker *et al.* 2007) also have been cloned and functionally characterized in *Arabidopsis*. This work suggests that multiple functions in propyl glucosinolate and leucin amino acid biosynthesis may exist. Additionally, glucosinolate side chain modification genes, *GSL-FMO*$_{OX1-5}$ for oxygenation of methylthio- to methylsulfinyl-, *GSL-ALK/AOP* for methylsulfinyl- to alkenyl- and to hydroxyl- aliphatic glucosinolates also have been characterized in *A. thaliana* (Kliebenstein *et al.* 2001c, Li *et al.* 2008). In the amphidiploid *Brassica* species *B. napus* and *B. juncea*, QTL mapping was performed and several QTL for seed glucosinolates were identified (Magrath *et al.* 1994, Toroser *et al.* 1995, Ramchiary *et al.* 2007, Bisht *et al.* 2009). There is, however, only a single study reporting QTL mapping of glucosinolates in leaves (Lou *et al*. 2008) of *B. rapa*.

The diploid species *B. rapa* (possessing the A-genome) has several morphologically divergent subspecies i.e., *rapa, oleifera, pekinensis, chinensis, parachinesis, japonica* and *nipposinica*. Some of the subspecies of *B. rapa* have been cultivated for leafy, root or floret vegetables or as an oilseed. Additionally, *B. rapa* is diploid progenitor of two amphidiploid species (*B. napus* and *B. juncea*), which are globally cultivated for vegetable oil. Despite the importance of *B. rapa* to the world food

and oil production, there is limited genetic information available for glucosinolate biosynthesis in *B. rapa*. It is, therefore, important to understand glucosinolate biosynthesis pathway in *B. rapa* and to identify the major genes/loci involved. With this knowledge, glucosinolate content and composition may be manipulated in various *Brassica* crops.

This research was conducted under the following fundamental objectives:

1. Identification of QTL for leaf and seed aliphatic glucosinolate profile and content in *B. rapa* recombinant inbred lines derived from a cross between a Chinese cabbage (RI16, medium GSLs with 5C) and a yellow sarson (BAR6, high GSLs without 5C)
2. Identification of QTL for leaf and seed aliphatic glucosinolate profile and content in *B. rapa* recombinant inbred lines developed from a cross between a yellow sarson (BAR6, high GSLs with gluconapin 4C) and a canola type DH line (USU9, low GSLs with glucoerucin 4C)
3. Homoeologous *GSL-ELONG* gene replacement and *GSL-PRO* gene introgression for aliphatic glucosinolate profile and content manipulation in *B. rapa* using marker assisted selection in backcrosses

1. In this study, following hypotheses were tested by three research experiments: That the individual glucosinolate biosynthesis genes/loci could be mapped on specific linkage group(s) of the A-genome
2. That the glucobrassicanapin and glucoalyssin (total 5C aliphatic glucosinolates) are controlled by a single *Br-GSL-ELONG* locus in *B. rapa*

3. That glucoerucin (4C) aliphatic glucosinolate is biosynthesized by a single major locus in canola quality *B. rapa*
4. That the production of aliphatic glucosinolates in *B. rapa* can be manipulated by replacement or introgression of glucosinolate biosynthesis genes through backcrosses using marker assisted selection

2. LITERATURE REVIEW

2.1. Historical background of *Brassica* species

The crops belonging to the genus *Brassica* have been of great importance to humanity. Since ancient times, *Brassica* crops have been used for many purposes, including vegetables, oilseeds, feed, condiments, fodder, green manure and even medical treatments. Early history suggests that rapeseed has been cultivated for several thousand years with its origin in the Mediterranean region although exact time of domestication and the place of origin are still unknown. Sanskrit writings in 2000-1500 BC characterized species identified as *B. rapa* and *B. napus* as oleiferous forms and mustards, respectively. *Brassica juncea* and *B. rapa* are believed to have been crop plants in India long before the Christian era. The Greek, Roman and Chinese literature of 500-200 BC referred *B. rapa* as rapiferous forms and were also described for various medicinal properties (Downey and Röbellen 1989). In early times, rapeseed oil was used as a lamp oil, which in later centuries led gradually to its use as a valuable cooking oil.

Brassica species are diverse in terms of morphology, agronomy and quality traits. Domestication of rapeseed in Europe seems to have begun in the early Middle Ages. In 1620, *B. rapa* was first recorded in Europe by the Swiss botanist Casper Banhin (Gupta and Pratap 2007). . As a result, *Brassica* crops were adapted and cultivated in many parts of the world (Mehra 1966). Rapeseed was introduced in Canada before the Second World War (McVetty *et al.* 2009). Commercial cultivation in Canada began during the Second World War to supply lubricating oil for steamships. Canada's first *B. rapa* rapeseed cultivar, Arlo, with high erucic acid (40 to 45%) and high glucosinolate content

(>150 µmole/g seed) was developed in 1958 using selection from open pollinated populations (McVetty *et al.* 2009). Initially, *B. rapa* was the dominant cultivated species of *Brassica* in western Canada. In late 1980s, a large acreage of *B. rapa* and *B. napus* was grown in the Prairie Provinces. Subsequently, the production area of *B. rapa* declined to about 15 – 20% of its former area in 1990s. The reduction in acreage of *B. rapa* resulted from the introduction of herbicide tolerance canola, which provided the early planting and high yield advantages of *B. napus* cultivars. Currently, *B. rapa* is still grown in small areas in Canada because of its early maturity. Research efforts are underway to develop disease resistant hybrid varieties to increase yield potential of *B. rapa*. *Brassica rapa* are grown as a winter sarson crop in Asian countries such as India, Pakistan, China and Bangladesh. Vegetable forms of *B. rapa* (Chinese cabbage, turnip, pak choi, komatsuna, mizuna green and rapini) are widely cultivated in many parts of the world (Prakash and Hinata 1980, Takuno *et al*. 2007).

2.1.1. Genomic relationships in *Brassica* species

Genomic relationships between the three diploid and three amphidiploid *Brassica* species were initially established in the 1930s based on various taxonomical and cytogenetic studies (Figure 2.1) (Morinaga 1934, U 1935). Three allotetraploid *Brassica* species namely, *B. napus* (AACC, 2n=38), *B. juncea* (AABB, 2n=36) and *B. carinata* (BBCC, 2n=34) have been derived from three diploid elementary species, *B. rapa* (AA, 2n=20), *B. nigra* (BB, 2n=16) and *B. oleracea* (CC, 2n=18).

The genomic relationships of *B. napus* with *B. rapa* and *B. oleracea* have been confirmed by the resynthesis of *B. napus* from *B. rapa* x *B. oleracea* crosses (U 1935,

Downey *et al.* 1975, Olsson and Ellerstrom 1980). The close relationship between the six *Brassica* species made it feasible to incorporate a trait from one species into others to make the crops more suitable to agricultural systems. Thus, complex traits like glucosinolates can also be manipulated as required through interspecific hybridization. It has been relatively easier to make interspecific crosses among some of these six species (e.g. *B. napus* x *B. rapa*) compared to others (e.g. *B. rapa* x *B. oleracea*). Wide hybridizations are normally performed by the application of embryo rescue techniques. The most recent advances in genome sequencing technology, bioinformatics and data mining have opened an avenue for comparative analysis of ESTs, BACs, genes (families), whole chromosomes and even entire genomes to determine evolutionary relationship between these species and their ancestors (Gao *et al.* 2004, Gao *et al.* 2006, Punjabi *et al.* 2008, Mun *et al.* 2009, Qiu *et al.* 2009, Nagoaka *et al.* 2010).

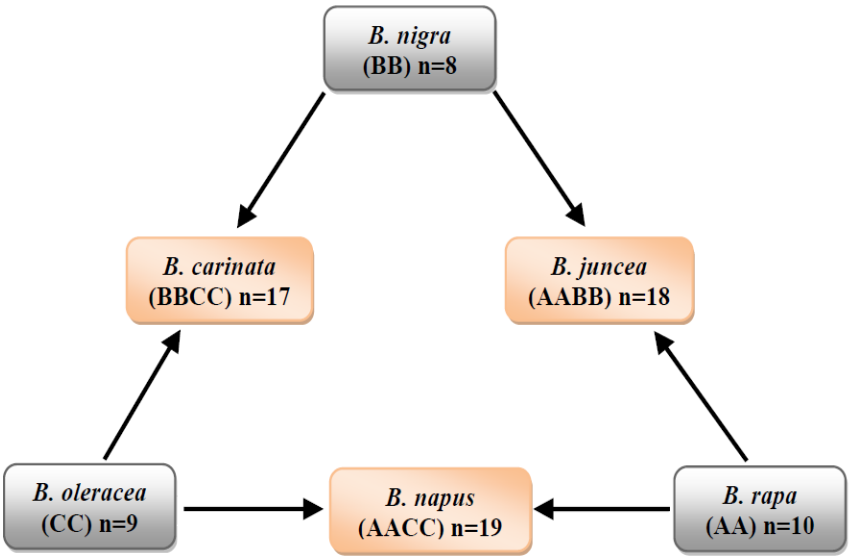


Figure 2.1 U-triangle of genomic relationship between diploid and amphidiploid *Brassica* species (U 1935).

2.1.2. Low glucosinolate rapeseed and canola

Early forms of domesticated rapeseed and their cultivars possessed a high concentration of glucosinolates (100 to 180 µmole/g) in their oil-free seed meal. The presence of glucosinolates in rapeseed had hindered the use of rapeseed meal in livestock industries due to anti-nutritional effects of its hydrolysis products in animals. As a result, in the 1970s, plant breeders searched germplasm collections for low glucosinolate contents. A Polish spring rape cultivar, Bronowski, with low glucosinolate content was discovered by The Agriculture Canada Research Station in Saskatoon (Kondra and Stefansson 1970). This sole genetic source of the low glucosinolate trait has been used to develop all the low glucosinolate cultivars in *B. napus* and *B. rapa* worldwide through conventional plant breeding. *B. napus* and *B. rapa* cultivars with low content of erucic acid and glucosinolate were developed, which ushered in a new era for *Brassica* crop production and its consumption. The world's first double low (low erucic acid and low glucosinolate content) *B. napus* and *B. rapa* cultivars, Tower and Candle, respectively, were developed by pedigree selection in the progenies of interspecific crosses in 1970s (Stefansson and Downey 1995, McVetty *et al.* 2009). In Canada, this new type of oilseed rape was designated "Canola". The term "Canola" applies to any rapeseed cultivars with erucic acid content of <2% and glucosinolates content of <30 µmol/g in oil-free seed meal. The Canola term is a registered trademark of the Canadian Canola Association. The name is derived from **Can**adian **O**il **L**ow **A**cid (Canola Council of Canada, 2010a, http://www.canola-council.org/canola_the_official_definition.aspx). Currently, most rapeseed (high erucic acid) and canola cultivars have glucosinolate levels <15 µmole/g in oil-free seed meal. The development of low erucic acid and low glucosinolate cultivars

has also been undertaken for other *Brassica* rapeseed species (e.g. *B. juncea*) and in other parts of the world for the quality improvement of their oils and seed meals.

2.2. Economic importance of *Brassica* species

The family *Brassicaceae* (syn. *Cruciferae*) is one of the crucial plant families for humans and animals and supplies several products from various plant parts. The little cruciferous weed *A. thaliana* has become an important model organism for the study of plant molecular biology, including the related crop species. The mustard family (*Brassicaceae*) is the fifth largest monophyletic angiosperm family, comprising 338 genera and about 3700 species in 25 tribes (Beilstein *et al.* 2006). The genus *Brassica* is one of the 51 genera of the tribe *Brassiceae* and includes the economically valuable crop species. *B. napus, B. rapa, B. juncea, B. carinata* and *B. nigra* are grown for edible and industrial oil as well as nutritionally valued seed meal.

Globally, rapeseed and canola oil is being utilized for human consumption, industrial applications and as a feedstock for biodiesel production. Canola oil is considered a healthy edible oil due to its high level of monounsaturated fatty acid (61%), lower level of saturated fatty acid (7%) and moderate amount of polyunsaturated fatty acid (22%) in its overall fatty acid profile (McVetty and Scarth 2002). Rapeseed that has erucic acid levels greater than 45% also has many industrial applications such as plasticizers, slip agents for fibreglass and oil for the lubrication industry. Additionally, the seed meal is a marketable source of protein rich animal feed supplement.

Rapeseed is the world's third leading oil producing crop after palm and soybean, and it contributes about 15% to the global total vegetable oil production. Canada was the

top rapeseed producing country in the world with 12.6 million MT productions in 2008 (FAO 2008). Canola/rapeseed contributes about $14 billion annually to the Canadian economy along with the generation of about 200,000 jobs throughout Canada in the areas of production, transportation, exporting, crushing and refining (Canola Council of Canada, 2010b http://www.canolacouncil.org/canadian_canola_industry.aspx). Canola/rapeseed meal is the second most popular protein feed ingredient in the world after soybean meal. Protein content of canola/rapeseed meal ranges from 36 to 39%, with a good amino acid profile for animal feeding (Newkirk *et al*. 2003). The major producers and consumers of canola/rapeseed meal are Australia, Canada, China, European Union and India. Along with oil production, *Brassica* species also produce different forms of vegetables and are the most widely cultivated vegetable crops in the world. Most of the production is consumed locally with a small amount of international trade. *B. napus* and *B. juncea* are used as vegetables in Asian countries like China, Japan and India. *B. rapa* is differentiated into seven groups viz., var. *compestris*, *pekinensis*, *chinensis*, *parachinesis*, *narinosa*, *japonica* and *rapa*. *Brassica rapa* is cultivated for leafy and root vegetables in the form of Chinese cabbage, pak choi and turnip; *B. oleracea* is cultivated for leafy and floret vegetables in various morph types such as cabbage, cauliflower, kale, collard, kohlrabi, brussels and broccoli.

2.3. Plant secondary metabolites and their functions

The sessile nature of plants requires them to produce a large numbers of defence compounds including primary and secondary metabolites. It is believed that the currently discovered plant metabolic compounds account for only about 10% of the actual compounds present naturally within the plant kingdom (Schwab 2003, Wink 2003). Plant

secondary metabolites are organic biochemical compounds produced in plants during normal growth and development. While they are not directly involved in plant growth, development or reproduction, these secondary metabolites play vital roles in plant defence mechanisms, acting for example, phytoalexins and phytoanticipins. Phytoalexins are antimicrobial defence metabolites synthesized *de novo* in response to biotic and abiotic stresses. Phytoalexins are involved in induced plant defence mechanisms including lytic enzymes, oxidizing agents, cell wall lignifications and pathogenesis-related proteins and transcript stimulation (Pedras *et al.* 2008). Phytoanticipins are low molecular weight antimicrobial compounds which are constitutively active for defence. Their production may be increased under high biotic or abiotic stresses (Pedras *et al.* 2007). Certain classes of phytoanticipins require enzymatic modification and derivation in order to become active within the defence systems of the plant. Plant secondary metabolites are broadly categorized into three groups based on their biosynthetic origin

(i) Flavonoids and allied phenolic and polyphenolic compounds

(ii) Terpenoid compounds

(iii) Nitrogen and sulphur containing alkaloid compounds

2.3.1. Glucosinolates as secondary metabolites

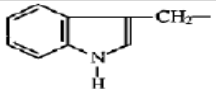
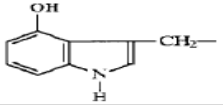
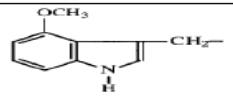
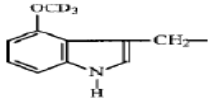
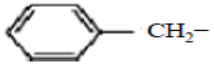
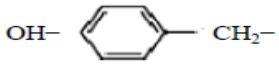
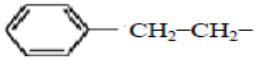
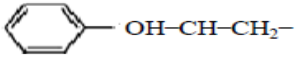
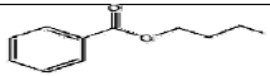
Glucosinolates are sulphur rich, nitrogen containing anionic natural products, derived from specific amino acids and their precursors (Fenwick *et al.* 1983). Glucosinolates are reported almost exclusively from the order *Brassicales* (*Capparales*), which possesses about 15 families such as *Brassica*ceae, *Capparaceae* and *Caricaceae*. Glucosinolates are also reported in a few members of the family *Euphorbiaceae*, a very

distinct family to other glucosinolate containing families (Rodman *et al.* 1996). Glucosinolates coexist with endogenous thioglucosidases called myrosinases in cruciferous plant species and activate plant defence mechanism against biotic and abiotic stresses. Tissue disruption causes systemic interactions between glucosinolates and myrosinases in the presence of moisture. The interaction produces numerous compounds with diverse biological activities (Bones and Rossiter 1996, Halkier 1999). Glucosinolates are some of the most extensively studied plant secondary metabolites; various enzymes and transcription factors involved in biosynthesis have been studied in the model plant *Arabidopsis* and to some extent in *Brassica* crops species. The broad functionality, physiochemical and genetic studies of glucosinolates have led to a model status for research on secondary metabolites (Sønderby *et al.* 2010).

2.3.2. Chemical structures of glucosinolates

Table 2.1 Trivial name and chemical formula of R side-chains of glucosinolates identified in *Brassica* species.

GSL Name	Trivial name	RF	R Side chain	Chemical Structure	Mol. Wt.#
Aliphatic 3C	Sinigrin	1.00	2-Propenyl	CH_2=CH–CH_2–	279
	Glucoibervirin	0.80	3-Methylthiopropyl	CH_3S–CH_2–CH_2–CH_2–	327
	Glucoiberin	1.07	3-Methylsulfinylpropyl	CH_3SO–CH_2–CH_2–CH_2–	343
	Glucocheirolin	1.26	3-Methylsulfonylpropyl	CH_3SO_2–CH_2–CH_2–CH_2–	179.26
	Glucoputranjivin	1.00	1-Methylethyl	CH_3–CH–CH_3	281
	Glucosisymbrin	1.32	2-hydroxy-1-methylethyl	OH–CH_2–CH–CH_3	298
Aliphatic 4C	Gluconapin	1.11	3-Butenyl	CH_2=CH–CH_2–CH_2–	293
	Progoitrin	1.09	(2R)-2-Hydroxy-3-butenyl	CH_2=CH–CH(OH)–CH_2–	309
	Epi-progoitrin	1.09	(2S)-2-Hydroxy-3-butenyl	CH_2=CH–CH(OH)–CH_2–	309
	Glucoerucin	1.04	4-Methylthiobutyl	CH_3S–CH_2–CH_2–CH_2–CH_2–	341
	Glucoraphasatin	0.40	4-Methylthio-3-butenyl	CH_3S–CH=CH–CH_2–CH_2–	340
	Glucoraphanin	1.07	4-Methylsulfinylbutyl	CH_3SO–CH_2–$(CH_2)_2$–CH_2–	357
	Glucoraphenin	0.90	4-Methylsulfinyl-3-butenyl	CH_3SO–CH=CH–CH_2–CH_2–	355
	Glucoconringiin	1.00	2-Hydroxy-2-methylpropyl	CH_3=CH(OH)–CH_2	312
Aliphatic 5C	Glucoalyssin	1.07	5-Methylsulfinylpentyl	CH_3SO–CH_2–CH_2–CH_2–CH_2–CH_2–	371
	Glucobrassicanapin	1.15	Pent-4-enyl	CH_2=CH–CH_2–CH_2–CH_2–	307
	Glucoberteroin	1.05	5-Methylthiopentyl	CH_3S–CH_2–CH_2–CH_2–CH_2–CH_2–	354
	Gluconapoleiferin	1.00	2-Hydroxy-pent-4-enyl	CH_2=CH–CH_2–CH–OH–CH_2–	323
	Glucocleomin	1.07	2-Hydroxy-2-methylbutyl	CH_3–CH_2–CH(OH)–CH_2–CH_2–	326
Aliphatic 6C	Glucolesquerellin	1.00	6-Methylthiohexyl	CH_3S–CH_2–CH_2–CH_2–CH_2–$(CH_2)_2$–	370
	Glucohesperin	1.00	6-Methylsulfinylhexyl	CH_3SO–CH_2–CH_2–CH_2–CH_2–$(CH_2)_2$–	385
Aliphatic 7C	Glucoarabishirsutain	1.00	7-Methylthioheptyl	CH_3S–CH_2–CH_2–CH_2–CH_2–$(CH_2)_3$–	384
	Glucoibarin	1.00	7-Methylsulfinylheptyl	CH_3SO–CH_2–CH_2–CH_2–CH_2–$(CH_2)_3$–	399
Aliphatic 8C	Glucoarabishirsuin	1.10	8-Methylthiooctyl	CH_3S–CH_2–CH_2–CH_2–CH_2–$(CH_2)_4$–	398
	Glucohirsutin	1.10	8-Methylsulfinyloctyl	CH_3SO–CH_2–CH_2–CH_2–CH_2–$(CH_2)_4$–	414

GSL Name	Trivial name	RF	R Side chain	Chemical Structure	Mol. Wt.[#]
Indole	Glucobrassicin	0.29	3-Indolylmethyl		368
	4-Hydroxyglucobrassicin	0.28	4-Hydroxy-3-indolylmethyl		384
	4-Methoxyglucobrassicin	0.25	4-Methoxy-3-indolylmethyl		398
	Neoglucobrassicin	0.20	N-Methoxy-3-indolylmethyl		398
Aromatic	Glucotropaeolin	0.95	Benzyl		329
	Glucosinalbin	0.50	p-Hydroxybenzyl		345
	Gluconasturtiin	0.95	2-Phenethyl		343
	Glucobarbarin	1.09	(2S)-2-Hydroxy-2-phenethyl		360
	Glucomalcomiin	0.40	3-Benzoyloxypropyl		402

Mol. Wt.[#]: molecular weight of desulfoglucosinolates, RF: response factor (Haughn *et al.* 1991, Griffiths *et al.* 2000, Brown *et al.* 2003).

2.3.3. Glucosinolates and their biological functions in agriculture and nature

Glucosinolates are a uniform group of thioglucosides with an identical core structure called β-D-glucopyranose bound to a (Z)-N-hydroximinosulfate ester by a sulphur atom with a variable R group. Approximately 120 glucosinolates differing in their R group side chains have been identified (Halkier and Gershenzon 2006). These glucosinolates are categorized into three classes based on their precursor amino acids and side chain modifications (Table 2.1). Kliebenstein *et al.* (2001a) suggested that these three classes of glucosinolates are independently biosynthesized and regulated by different sets of gene families from separate amino acids. Each class is briefly discussed below.

2.3.3.1. Aliphatic glucosinolates

Aliphatic glucosinolates are the major group of glucosinolates in *Brassica* species, contributing about 90% of the total glucosinolate content of the plant. Glucosinolates are constitutively biosynthesized *de novo* in cruciferous plants, although their degradation is highly regulated by spatial and temporal separation of glucosinolates and myrosinases within the plant based on environmental and biotic stresses (Drozdowska *et al.* 1992). Hydrolysis of glucosinolates produces a large number of biologically active compounds that have a variety of functions. The most common hydrolysis products of aliphatic glucosinolates in many cruciferous species are isothiocyanates that are formed by the rearrangement of aglycone with carbon oxime adjacent to the nitrogen at neutral pH while at acidic pH, nitriles are the predominant products (Fahey *et al.* 2001). These

unstable compounds are cyclised to a class of substances responsible for goiter in animals (Griffiths *et al.* 1998).

By contrast, sulforaphane is one of the derivatives of glucoraphanin, an aliphatic glucosinolate that has several beneficial properties for humans and animals. It is known as an inducer of phase II enzymes such as glutathione transferases and quinone reductases of the xenobiotic pathway in human prostate cells (Zhang *et al.* 1992, Faulkner *et al.* 1998). The phase II enzymes are involved in the detoxification of electrophilic carcinogens that can lead to mutations in DNA and cause different types of cancers (Mithen *et al.* 2000). Enhanced consumption of cruciferous vegetables appears to reduce the risk of cancers (Nestle 1997, Talalay 2000, Brooks *et al.* 2001). The sulforaphane content of these vegetables could be a leading factor in the reduction. Another less documented health benefit of sulforaphane is the inhibition of *Helicobacter pylori*, a pathogen of peptic ulcers and gastric cancer (Fahey *et al.* 2002). Sulforaphane also protects human retinal cells against severe oxidative stresses (Gao *et al.* 2001).

Isothiocyanates and other breakdown products of glucosinolates play important roles as repellents of certain insects and pests (Rask *et al.* 2000, Agrawal and Kurashige 2003, Barth and Jander 2006, Benderoth *et al.* 2006). Leaves of the mutant *myb28myb29* in *Arabidopsis* with low aliphatic glucosinolate content, when fed to the lepidopteran insect *Mamestra brassicae,* enhanced larval weight by 2.6 fold (Beekwilder *et al.* 2008). Glucosinolates may have specific repellent or anti-nutritional effects on specific classes of insects and microorganisms. Some *in vitro* studies demonstrated that glucosinolate degradation products, isothiocyanates and nitriles, inhibited fungal and bacterial pathogen growth (Brader *et al.* 2001, Tierens *et al.* 2001). In *Arabidopsis,* over expression of

CYP79D2 from cassava increased accumulation of isopropyl and methylpropyl aliphatic glucosinolates and transformed plants showed enhanced resistance against a bacterial soft-rot disease (Brader *et al.* 2006). Birch *et al.* (1992) reported that biotic stresses such as pest damage in *Brassica* species alters glucosinolate profiles in roots, stems, leaves and flowers. This suggests that a phytoanticipin property of glucosinolates is involved in the plant defence mechanisms of *Brassica*. Glucosinolates and their breakdown products have many biological functions, with a few compounds acting as biopesticides, biofungicides and soil fumigants, while others play roles in attraction of pollinators and provide oviposition cues to certain insects. The attraction of specialized insects could be due to the glucosinolate-sequestering phenomenon of some insects including harlequin bugs, sawflies, and some homoptera including aphids (Bridges *et al.* 2002, Mewis *et al.* 2002).

2.3.3.2. Indole glucosinolates

Indole (heterocyclic) glucosinolates in cruciferous plants (including *Arabidopsis*) are derived from tryptophan and possess variable R group side chains. The relatively high content of indole glucosinolates in the model plant *Arabidopsis* has enhanced our knowledge of the biosynthesis, transportation and functional properties of this class of glucosinolates (Petersen *et al.* 2002, Brown *et al.* 2003). Side chain modification in indole glucosinolates occurs through hydroxylations and methoxylations catalysed by several enzymes. Indole glucosinolate types and contents in different organs of the plant are strongly affected by environmental conditions. Four main indole glucosinolates have been identified in most cultivated *Brassica* species: glucobrassicin, neoglucobrassicin, 4-methoxyglucobrassicin and 4-hydroxyglucobrassicin. Similar to aliphatic glucosinolates,

breakdown products of indole glucosinolates have multiple biological functions. Indole-3-carbinol derived from glucobrassicin has potent anticarcinogenic activity (Hrncirik *et al.* 2001). The indole glucosinolate derived compound 4-methoxyglucobrassicin has strong insect deterrent activity (Kim and Jander 2007, De Vos *et al.* 2008). Osbourn (1996) reported antimicrobial activities of indole glucosinolates and their breakdown products in *Brassica* species. Several studies suggest that there is a metabolic association between indole glucosinolates and the plant hormone indole-3-acetic acid (IAA). In the consecutive reactions, indole glucosinolates are degraded into indole acetonitrile (IAN), which is then hydrolyzed by nitrilases into IAA (Figure 2.2). In clubroot infected *Brassica* roots, indole glucosinolate-based induction of IAA was observed to be responsible for gall formation. The IAA production from indole glucosinolates during gall formation is associated with a signalling cascade of IAA and cytokinin complex (Ugajin *et al.* 2003). Structural similarity data indicates that the indole alkaloid, brassinin, and possibly other cruciferous phytoalexins are derived from glucobrassicin. Studies in rapeseed, mustard and *Arabidopsis* have suggested that methyl jasmonate and wounding induce the biosynthesis of particular indole glucosinolates (Bodnaryk 1992, Brader *et al.* 2001).

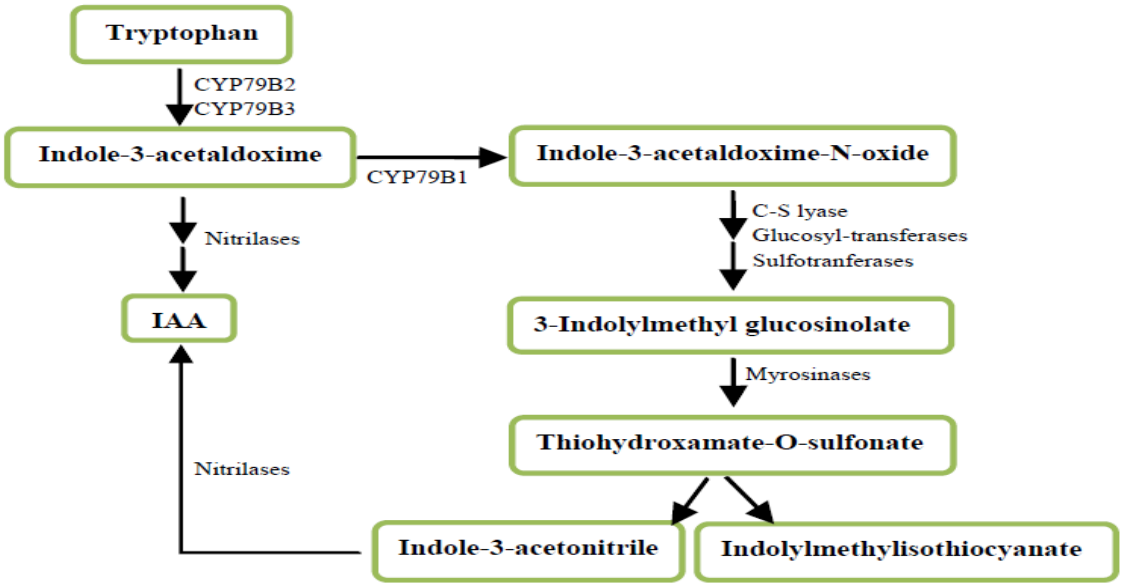


Figure 2.2 Biosynthetic pathway and breakdown products of indole glucosinolates (De Vos *et al.* 2008). In this pathway, IAA produces from precursors and derivatives of 3-indolylmethyl glucosinolate by various nitrilases.

2.3.3.3. Aromatic glucosinolates

The third class of glucosinolates in cruciferous species is aromatic or benzylic glucosinolates, derived from the aromatic parental amino acids phenylalanine and tyrosine. Very limited information is available regarding aromatic glucosinolates at qualitative or quantitative levels. Aromatic glucosinolates are biosynthesized independently from other glucosinolates, which is apparently due to involvement of different amino acid precursors in the biosynthesis of the different classes of glucosinolates (Kliebenstein *et al.* 2001a). Cloning and functional characterization of the *CYP79A* gene of *Arabidopsis* suggests that cytochrome P450-dependent monooxygenase catalyzes the reaction from phenylalanine to phenylacetaldoxime in aromatic glucosinolate biosynthesis (Wittstock and Halkier 2000). Five aromatic glucosinolates have been identified in *Brassicaceae*: glucotropaeolin, glucosinalbin, gluconasturtiin,

glucobarbarin and glucomalcomiin. The distinctive aroma and spiciness of condiment *Brassica* plant parts, such as the leaves and seeds of white (*Sinapis alba*) and black (*B. nigra*) mustards, is due to the presence of these aromatic glucosinolates (Fenwich *et al*. 1983).

2.3.4. Biosynthesis of aliphatic glucosinolates

Aliphatic glucosinolates are the most abundant class in *Brassica* species, therefore, the genetic of biosynthesis is described in more detail. Aliphatic glucosinolates are biosynthesized from five amino acids (methionine, alanine, leucine, isoleucine and valine) (Halkier and Gershenzon 2006). Biosynthesis of aliphatic glucosinolates occurs in three stages at two different locations. The first chain elongation step is catalyzed by *BCAT4* in the cytosol (Schuster *et al.* 2006), whereas development of core structures and secondary side chain modification reactions take place in the chloroplasts (Textor *et al.* 2007, Sawada *et al.* 2009). Chain elongation steps produce propyls (3C), butyls (4C), pentyls (5C), hexyls (6C), heptyls (7C) and octyls (8C) aliphatic glucosinolates in cruciferous species including *Arabidopsis*. Glucosinolate side chain modification reactions involve oxygenation, hydroxylation, alkenylation and benzoylation, which are controlled by several gene families. The pattern of glucosinolate biosynthesis varies from organ to organ within the plant; young leaves, buds, flowers and silique walls all have higher rates of glucosinolate biosynthesis than roots, old leaves and presumably seeds (Brown *et al.* 2003). Various studies also suggest that transportation of glucosinolates and their breakdown products from organ to organ via phloem occurs upon requirement to protect the plant. Seeds, however, are the most important store of total glucosinolates produced by the plants (Brudnell *et al.* 1999). Seeds contain much higher glucosinolates

concentrations than other plant parts and it is thought that leaf glucosinolates are the basis for accumulations of total glucosinolates in seeds (Klienbestein *et al.* 2001a). This suggests that long distance transportation of glucosinolates from source to sink occurs. A few reports discuss an independent pathway of glucosinolate biosynthesis in seeds, resulting in the high concentration of glucosinolate in seeds (Du and Halkier 1998, Osbourn 1996). Experimental evidence, however, is not strong enough to support a separate pathway at this time.

2.3.4.1. Parental amino acid biosynthesis and condensation

Methionine is the main precursor of aliphatic glucosinolates in *Brassica* species. The enzyme BCAT4 catalyzes the initial chain elongation reaction to produce 2-oxo acid from methionine, an analogous process to the formation of the branched chain amino acid valine to its chain-elongated homolog leucine (Figure 2.3). In *Arabidopsis*, a *bcat4* mutant showed about a 50% reduction in total aliphatic glucosinolates and at the same time increased the level of free methionine and S-methyl-methionine (Schuster *et al.* 2006). This suggests that the *BCAT4* gene produces an enzyme which is involved in the first deamination reaction. Subsequently, three consecutive reactions of transformations occur. The first is a transamination and condensation reaction with acetyl-CoA catalyzed by *GSL-ELONG* in *Brassica* species (Li and Quiros 2002). This is homologous to *MAM1* in *Arabidopsis* (Campos de Quirose *et al.* 2000, Benderoth *et al.* 2006, Textor *et al.* 2007). The same reaction occurs for 3C aliphatic glucosinolates which is controlled by isopropylmalate synthase (*IPMS1*, *IPMS2*). Isopropylmalate synthase is homologous to *MAM1* in *Arabidopsis* (Kliebenstein *et al.* 2001b, Field *et al.* 2004) and to *GSL-PRO* in *Brassica* species (Li *et al.* 2003, Gao *et al.* 2006). The second isomerisation reaction is

controlled by isopropylmalate isomerises (*IPMI*) and third reaction is oxidative decarboxylation controlled by isopropylmalate dehydrogenases (*IPM-DH*) (Figure 2.3) (Wentzell *et al.* 2007, Sawada *et al.* 2009).

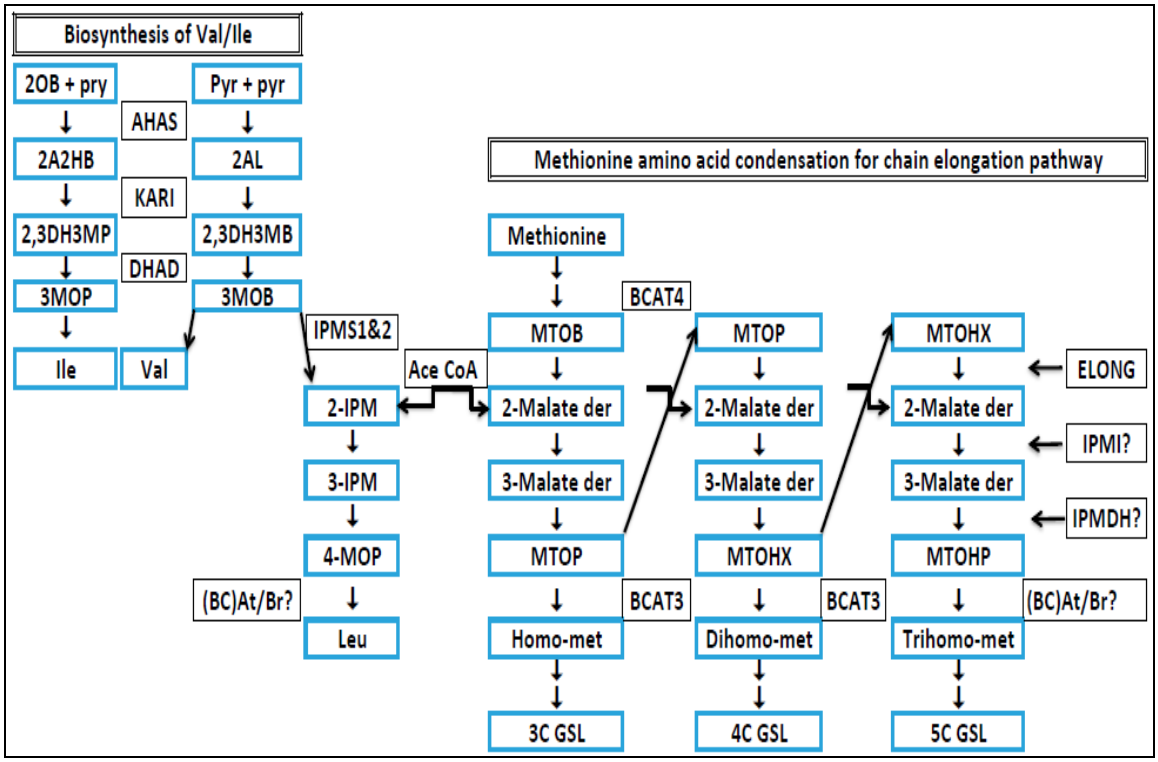


Figure 2.3 Methionine amino acid condensation pathway. All the reactions are catalyzed by *BCATs*, *ELONGs*, *IPMIs* and *IPM-DHs* gene families for 3C, 4C and 5C glucosinolates. Genes shown in gray boxes and derivative products shown in blue boxes (Kroymann *et al.* 2001, Sawada *et al.* 2009). BCAT- branched chain amino transferase, MTOB- 4-methylthio-2-oxobutanoate, MTOP- 6 methylthio-2-oxopentanoate, MTOHX- 4-methylthio-2-oxohexanoate, IPMI- isopropylmalate isomerases, IPMDH- isopropylmalate dehydrogenases, MOB- methyl-2-oxobutanoate, MOP- methyl-2-oxopentanoate, AHAS- acetohydroxyacid synthase, KARI- ketolacid reductoisomerase, DHAD- dihydroxyacid dehydratase, 2AL- 2-acetolactate, 2A2HB- 2-aceto-2-hydroxybutyrate, 2OB- 2-oxobutyrate, 2,3DH3MB- 2,3-dihydroxy-3-methylbutyrate, 2,3DH3MP- 2,3-dihydroxy-3-methylpentanoate.

These three consecutive reactions produce elongated 2-oxo acids with one or more methylene groups. These compounds are either transaminated by the BCAT enzyme to yield homo-methionine, which can enter into the core glucosinolate skeleton structure formation, or proceed through another round of chain elongation (Figure 2.3). Overall, the methionine amino acid condensation pathway produces a range of methionine derivatives such as homo-methionine, dihomo-methionine, and trihomo-methionine, which proceed to the next biosynthesis step called glucosinolate core skeleton formation (Figure 2.3).

2.3.4.2. Glucosinolate core skeleton formation

Glucosinolate core skeleton structure formation has been well characterized in *Arabidopsis*, with at least 13 enzymes and five different biochemical reactions, i.e., oxidation, oxidation with conjugation, C-S cleavage, glucosylation and sulfation (Grubb and Abel 2006, Halkier and Gershenzon 2006) involved in the formation. The precursors are catalyzed into aldoxime by cytochromes belonging to the *CYP79* gene family (Figure 2.4). At least seven *CYP79s* were identified and functionally characterized in *Arabidopsis*. The *CYP79F1* gene converts all short chain methionine derivatives, whereas *CYP79F2* gene is involved in conversions of the long chain methionine derivatives. Similarly, *CYP79B2* and *CYP79B3* catalyze tryptophan derivatives, and *CYP79A2* catalyzes phenylalanine substrates (Figure 2.4) (Zang *et al.* 2008). Subsequently, aldoximes are oxidized into either nitrile oxides or aci-nitro compounds by *CYP83A1* for methionine derivates and *CYP83B1* for tryptophan as well as phenylalanine derivates. This proceeds to a non-enzymatic conjugation to produce S-alkly-thiohydroximates. In this sulphur rich chemical pathway, the next step is C-S cleavage by C-S lyase from S-

alkly-thiohydroximate to thiohydroximic acid; C-S lyase forms an enzymatic complex with an S-donating enzyme. The *c-s lyase* mutant of *Arabidopsis* showed complete lack of aliphatic and aromatic glucosinolates in *Arabidopsis*, suggesting that this single gene family has a crucial role in skeleton processing (Mikkelsen *et al.* 2004).

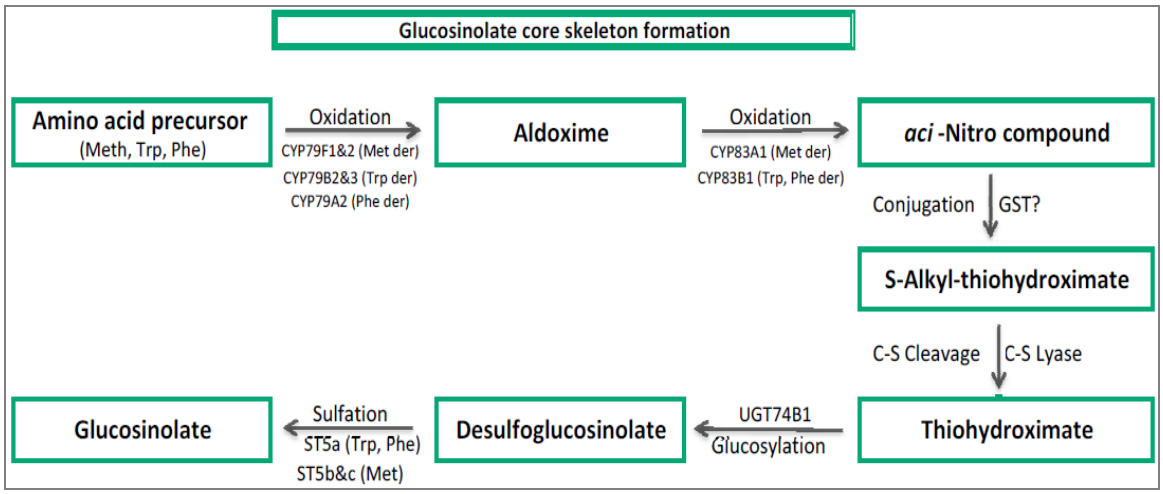


Figure 2.4 Glucosinolate core skeleton structure formation by cytochromes. Methionine amino acid precursors produce aliphatic, tryptophan produces indole and phenylalanine produces aromatic glucosinolate core structures (Grubb and Abel 2006, Halkier and Gershenzon 2006). ST- sulfotransferase, UGT- glucuronosyltransferases, GST- glutathione S-transferase.

In the glucosylation step, desulfoglucosinolate is formed by a member of the *UGT74* family. The final reaction of core skeleton formation is accomplished with sulfation of desulfoglucosinolates to produce intact glucosinolates by sulfotransferases *AtST5a*, *AtST5b* and *AtST5C* in *Arabidopsis*. Biochemical characterization of sulfotransferases in *Arabidopsis* revealed that *AtST5a* favour to sulfate phenylalanine and tryptophan derived desulfoglucosinolates, whereas *AtST5b* and *AtST5c* favour to sulfate long chain aliphatic glucosinolates (Piotrowski *et al.* 2004). In a comparative analysis

study between *Arabidopsis* and *B. rapa*, at least 12 paralogs of sulfotransferases were known to be responsible for this reaction (Zang *et al.* 2008). In glucosinolate skeleton formation reactions, the first four biosynthesis reactions take place in the chloroplast and the last reaction of sulfation occurs in the cytosol. This suggests that shuttle transporters play important roles in the entire biosynthesis process (Klein *et al.* 2006).

2.3.4.3. Side chain modification in aliphatic glucosinolates

After glucosinolate core skeleton structure formation, the core skeletons are subjected to a set of reactions known as side chain modification or secondary transformation. Side chain modifications of glucosinolates are the last crucial enzymatic reactions on intact glucosinolates before their transport to sinks or biological degradation by myrosinases occur. Hydrolysis products of individual glucosinolates are recognized based on side chain variation in R groups. A hydrolysis product of glucoraphanin has anticancer properties. The R group modifications of glucoraphanin change their chemical properties, therefore, hydrolysis products have anticarcinogenic functions. Hydrolysis products of progoitrin, however, have anti-nutritional effect in animals, which reduce the palatability of rapeseed meal.

Side chain modification begins with the oxidation of sulphur in the methylthio precursor to produce methylsulfinyl and then methylsulfonyl moieties (Figure 2.5). In *Arabidopsis*, this reaction is catalyzed by the flavin monooxygenases, *GSL-FMO*$_{OX1-5}$ located within the *GSL-OX1* locus on chromosome I. Phylogenetic analysis revealed a main group of *GSL-FMOs* for cruciferous species, which is further categorized according to subspecies, indicating that functional diversity of S-oxygenation of glucosinolates

exists (Hansen *et al.* 2007, Li *et al.* 2008). Knockout mutant and over expression studies suggested that *GSL-FMO*$_{OX1-4}$ catalyzes the 4-methylthiobutyl to 4-methylsulfinyl reaction and *GSL-FMO*$_{OX5}$ is involved in the S-oxygenation of long chain glucosinolates in *Arabidopsis* (Li *et al.* 2008). In *Brassica* vegetables, products of *GSL-FMOs* catalyses are the sources of anticancer compounds from aliphatic glucosinolates. It will be beneficial to identify these genes/loci in *Brassica* species so that they might be further used to manipulate aliphatic glucosinolates towards favourable forms.

A second round of binary side chain modification changes methylsulfinyl to alkenyl- and to hydroxyl- aliphatic glucosinolates (Figure 2.5). In *Arabidopsis* these reactions are controlled by a *GSL-ALK*/*GSL-OHP* locus that has three tandem repeats (*GSL-AOP1*, *GSL-AOP2* and *GSL-AOP3*), which encode 2-oxoglutarate-dependent dioxygenases located on chromosome IV. Functional characterization indicates that *GSL-AOP2* catalyzes the reaction to alkenyl, whereas *GSL-AOP3* controls the reaction toward hydroxyalkenyl. The function of *GSL-AOP1*, however, is not clear in *Arabidopsis*, it might be involved in both reactions (Figure 2.5) (Hall *et al.* 2001, Kliebenstein *et al.* 2001c, Mithen *et al.* 1995). The *GSL-ALK* and *GSL-OHP* are either closely linked on the same genomic region or allelic variants of a single genetic locus though they may show variable functions. In *Arabidopsis*, *GSL-OHP* catalyzes the reaction only for 3C aliphatic glucosinolate branches, whereas *GSL-ALK* is involved in 3C, 4C and 5C aliphatic glucosinolate branches. There is no clear functional information available for long chain (6C and so on) aliphatic glucosinolate branches and presumably *GSL-ALK* accomplishes these reactions in *Arabidopsis* (Kliebenstein *et al.* 2001c, Parkin *et al.* 1994). In *B. oleracea*, *GSL-ALK* was inferred by positional cloning and biochemical analysis. The

functional allele in collard and the non-functional allele (with 2 bp deletion creating a frame-shift mutation) in broccoli were confirmed. A locus or loci of *GSL-ALK* is also believed to have a role in the catalysis of methylsulfinyl to alkenyl glucosinolates (Li and Quiros 2003). Hydroxylation changes alkenyl to hydroxy aliphatic glucosinolate (in butyls, pentyls, hexyls and so on) biosynthesis branches in *Arabidopsis* and *Brassica* species; these sets of reactions are controlled by *GSL-OH* dependant on the presence of both *GSL-AOP2* and *GSL-ELONG*. In *Brassica*, the final product of this reaction in 4C glucosinolate biosynthesis is progoitrin and its hydrolytic derivative, oxazolidine-2-thione which causes goiter in animals. These compounds are major obstacles to the use of *Brassica* crops as animal feed (Fenwick *et al.* 1983).

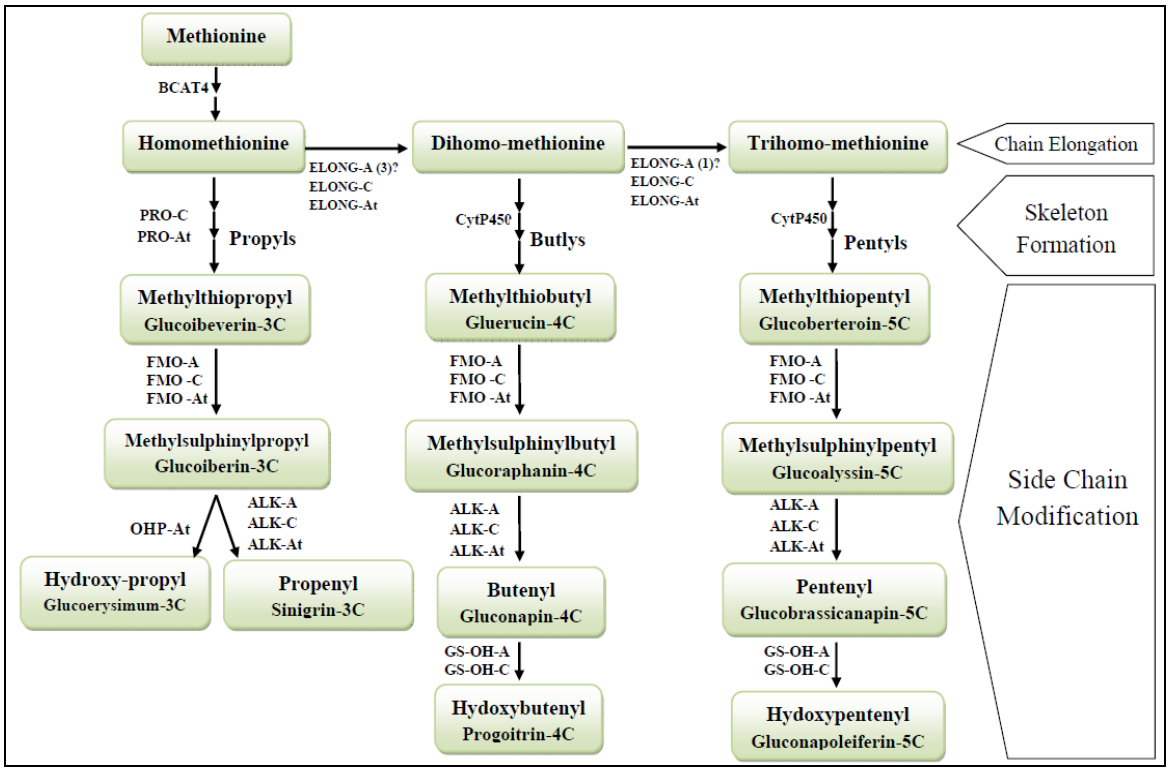


Figure 2.5 Glucosinolate core structure and side chain modification pathway for 3C, 4C and 5C aliphatic glucosinolates. In the biosynthesis steps, gene symbols ending with A indicate A genome, C for C genome and At for *A. thaliana* (Magrath *et al.* 1994, Mithen *et al.* 1995, Li and Quiros 2003, Mahmood *et al.* 2003).

2.3.5. Diversity of glucosinolates in *Brassicaceae*

Glucosinolates are united by their unique basic skeleton (β-D-glucopyranose) but glucosinolates are diverse in their origins, side chain modifications, degradations and final biological functions. In addition to structural diversity, a diversity of glucosinolates is seen between families, genera, species, subspecies and different accessions of subspecies. This diversity provides insight into glucosinolate biosynthesis at the genomic, physiological, biochemical and host-pathogen interaction levels. The natural variation of glucosinolate profiles between species or different cultivars of same species permits the investigation of the effects of QTL or genes and gene interactions. This can be utilized for advanced breeding applications like MAS, trait introgression and gene pyramiding for beneficial glucosinolates. In *Arabidopsis*, naturally occurring variations in glucosinolates were identified and quantified for 34 types of glucosinolates in the leaves of 39 ecotypes (Hogge *et al.* 1988, Reichelt *et al.* 2002). Similarly, different morphotypes of *B. rapa* possess eight different glucosinolates with gluconapin and glucobrassicanapin as predominant aliphatic glucosinolates (He *et al.* 2000). Padilla *et al.* (2007) reported 16 different glucosinolates among 116 accessions of turnip greens.

The wide range of variation in glucosinolate profiles provides the opportunity to study individual glucosinolates for their potent biological activities *in planta*. Within different forms of *B. oleracea,* 12 different glucosinolates have been detected. The beneficial glucosinolate glucoraphanin showed significant variation ranging from 44 to 274 µmole/g seed in different genotypes of broccoli (Mithen *et al.* 2000, Rangkadilok *et al.* 2002). Furthermore, variation in concentration of individual glucosinolates also exists in cultivars of the same species.

2.4. Quantitative trait loci for glucosinolates in major *Brassica* species

Glucosinolate biosynthesis in *Brassica* crops has quantitative inheritance, which is regulated by complex genetic factors and affected by environmental factors. Glucosinolates are functionally diverse and well recognized plant secondary metabolites; so they have been extensively studied in terms of QTL mapping, biosynthesis gene cloning and functional characterization in *Arabidopsis* (Kliebenstein *et al.* 2001a, Kliebenstein *et al.* 2001c, Compos de Quiros *et al.* 2000, Brown *et al.* 2003, Benderoth *et al.* 2006, Textor *et al.* 2007, Li *et al.* 2008). However, very limited genetic, biochemical and metabolomic information is available on glucosinolate biosynthesis, transport and final product utilization in *Brassica* crops including *B. rapa*. There has been a few QTL mapping studies reported for major *Brassica* crop seed glucosinolates. Uzunova *et al.* (1995) mapped four QTL for total seed glucosinolate content in a *B. napus* DH population, which accounted 61% total phenotypic variance. Similarly, Toroser *et al.* (1995), based on a RFLP linkage map, identified two larger and three small effect QTL for total aliphatic glucosinolate content using a DH population in *B. napus*. These QTL explained 70% of the total phenotypic variance. This suggests that several loci with additive or epistatic effect are involved in total seed glucosinolate biosynthesis in different genetic backgrounds. Howell *et al.* (2003) reported QTL mapping for total seed glucosinolates analyzed by X-ray fluorescence (XRF) and near-infrared reflection spectroscopy (NIRS) in two inter-varietal *B. napus* backcross populations. They identified four QTLs accounting for 76% of the phenotypic variance in the Victor x Tapidor population. These three QTL accounted for 86% of phenotypic variance in this second population. These studies, however, were limited to either total seed

glucosinolates or 3C, 4C and 5C aliphatic glucosinolates, and did not infer the genetic loci for individual aliphatic, indole or aromatic glucosinolates. Furthermore, there were no reports of publicly available molecular markers for marker assisted selection of glucosinolates. Such markers, if developed, could be used in breeding to manipulate glucosinolate profiles and contents in *Brassica* crop species.

In another amphidiploid species, *B. juncea,* several studies were conducted for QTL mapping of seed glucosinolates. Cheung *et al.* (1998) detected two QTL for 2-propenyl and three QTL for 3-butenyl glucosinolates which explained between 89% and 81% of total phenotypic variance. This QTL mapping study was carried out in a DH population derived from the F_1 of two *B. juncea* parental lines, J90-4317 (low glucosinolates) and J90-2733 (high glucosinolates). Mahmood *et al.* (2003) reported three QTL for 2-propenyl glucosinolate content which explained 78% of the phenotypic variance, while five QTL for total seed aliphatic glucosinolates explained phenotypic variance between 30% and 45%. In this study a DH population and an RFLP linkage map was used. Similarly, Ramchiary *et al.* (2007) reported six QTL for seed glucosinolate content in the F_1DH and advanced backcross DH (BC_4DH) of *B. juncea*. Some of the large effect QTL in advanced backcross (BC_4DH) of *B. juncea* were fine mapped using a candidate gene approach and comparative sequence analyses of *Arabidopsis* and *B. oleracea* (Bisht *et al.* 2009). The results suggested that epistasis and additive effects of glucosinolate genes in different genetic backgrounds in *B. juncea* exist. This study, however, could not explain the homoeologous effects of genes/loci from the A- and B-genomes on the individual or total seed glucosinolate content.

In *B. oleracea*, *BoGSL-ELONG* a side chain elongation gene was cloned based on the *Arabidopsis* sequence information, and functionally characterized using an RNA interference (RNAi) approach. The results suggested that *BoGSL-ELONG* is involved in 4C and 5C aliphatic glucosinolate biosynthesis in *Brassica* species. The RNAi lines displayed an increased level of propyl glucosinolates suggesting that the precursor homo-methionine concentration enhances the activity of 3C aliphatic glucosinolate biosynthesis in *B. napus* (Li and Quiros 2002, Zheng *et al.* 2010). A natural mutation in *BoGSL-ELONG* resulting in the failure of excision of the third intron and thus producing a long cDNA fragment has been identified in a white cauliflower genotype (*B. oleracea*) lacking 4C and 5C aliphatic glucosinolates (Li and Quiros 2002). A molecular marker for this mutation would be useful in *Brassica* breeding programs for modification of glucosinolate profiles. Additionally, a gene *BoGSL-PRO* which control propyl glucosinolate biosynthesis in *B. oleracea* was sequenced using comparative analysis of the *MAM* (*methylthioalkylmalate synthase*) gene family in *Arabidopsis* (Li *et al.* 2003, Gao *et al.* 2006).

A glucosinolate side chain modification gene, *BoGSL-ALK*, was cloned using a positional cloning approach based on a closely linked SRAP marker in *B. oleracea* (Li and Quiros 2003). Functional characterization of *BoGSL-ALK* by overexpression in *Arabidopsis* and RNA interference (RNAi) in *B. napus* suggests that *BoGSL-ALK* is involved in catalyzation of either sulfinylbutyl to butenyl or hydoxybutenyl with high functional redundancy (Li and Quiros 2003, Zheng *et al.* unpublished). Interestingly, a natural frame shift mutation of 2 bp deletions was identified in broccoli, which

accumulates sulfinylbutyl glucosinolate by ceasing downstream biosynthesis of other 4C aliphatic glucosinolates.

In *B. rapa*, a single QTL mapping study for leaf glucosinolates has been reported, although it is one of the widely distributed *Brassica* species for oil and vegetable production. Lou *et al.* (2008) identified six QTL for leaf total aliphatic glucosinolate content, three QTL for total leaf indole glucosinolate conten and three QTL for leaf aromatic glucosinolates in two DH populations of *B. rapa* using an AFLP and SSR based linkage map. There was no information regarding QTL for seed glucosinolates. Glucosinolate content varies greatly between leaves and seeds (Brown *et al.* 2003). As well, there is variation in the expression patterns of the genetic loci underpinning glucosinolate production in leaves and seeds (Kliebenstein *et al.* 2001b).

2.5. Glucosinolate identification and quantification approaches

Early analysis of glucosinolates began with detection of glucosinolates and possible hydrolysis products by paper and thin-layer chromatography. The paper chromatography was applied in combination with high voltage electrophoresis, but it had many complications and low yield (Greer 1962). Danielak and Borkowski (1969) analyzed glucosinolates from seeds of 150 different cruciferous species using thin-layer chromatography. Since then, numerous techniques have been employed for quantification of total glucosinolate content with various modifications including steam distillation and titration of isothiocyanates, ELISA, sulfate-release assay, UV spectroscopy and gas chromatography of isothiocyanates. Near infrared reflectance spectroscopy (NIRS) is one of the widely used techniques for seed total glucosinolates quantification, which detects

N−H, C−H and O−H groups of total glucosinolates. NIRS is a preferred technique because it can simultaneously quantify oil and protein along with total glucosinolates in canola/rapeseed (Velasco and Becker 1998). Individual intact glucosinolates can be determined using techniques such as reverse phase HPLC-MS, thermospray LC with tandem MS in the two most common interfaces (ESI or APCI), capillary GC-MS and GC-MS-MS.

Desulfoglucosinolates usually are analyzed by reverse phase HPLC or by X-ray fluorescence spectroscopy (XRF). The reverse-phase HPLC analytical approach has been widely used for quantification of individual intact or desulfo- glucosinolates. The technique was developed in 1984 with UV based detection of either intact glucosinolates or an on-column enzymatic desulfation from plant extracts (Spinks *et al.* 1984). The photodiode array (PDA) with UV detector can distinguish spectra of aliphatic from indole and aromatic glucosinolates; the indole and aromatic glucosinolates spectra end with a shoulder. This widely applicable method for glucosinolate separation is yet subject to difficulties in interpretation of results because of differences in the time and enzymatic activity for the desulfation reaction, pH effects and mobile phase solvents with an appropriate gradient. Desulfoglucosinolates also have been analyzed by the determination of the sulfur content of the seeds using X-ray fluorescence spectroscopy (XRF) (Schnug and Haneklaus 1990). The hydrolysis products of glucosinolates, isothiocyanates, nitriles, thiocyanates and benzenedithiol, have been analyzed using techniques including GC or GC-MS and HPLC with or without fluorescent labelling (Kiddle *et al.* 2001).

2.6. Molecular markers and their applications for different traits in *Brassica rapa*

Molecular markers are efficient, reliable, time saving and cost effective tools that may enhance the capacity of conventional breeding for improvement in agronomy, quality and yield related traits of crop species without adverse effects. Morphological traits such as petal color, leaf shape *etc* were used as markers in classical breeding, where significant time and effort was required to refine crosses. There have been many practical difficulties with the use of morphological markers, including:

(i) a paucity of suitable markers and associations with agriculturally important traits (Ranade *et al.* 2001),

(ii) undesirable pleiotropic effects of many morphological markers on plant phenotypes (Ranade *et al.* 2001),

(iii) high linkage drag (Ranade *et al.* 2001), and

(iv) trait of interest easily can be lost in a breeding cycle if there is no strong linkage between marker and traits (Ranade *et al.* 2001).

Advancements in molecular biology tools and techniques have overcome some of the difficulties of classical breeding. Different types of DNA molecular markers (hybridization based e.g. RFLP; PCR based e.g. SSR, RAPD, SCAR, and SRAP) have been used for gene/QTL mapping, cloning, genetic map construction and marker assisted selection in plant breeding. Most recently, the conversion of various molecular markers (RFLP, RAPD, SRAP, AFLP, SSR, SNP etc.) to simple PCR based SCAR markers for

marker assisted selection has overcome the difficulties of other markers. It is feasible and cost effective to use SCAR markers for marker assisted selection of populations.

In *B. rapa*, many important trait-related molecular markers either have been converted to SCAR markers or have been directly utilized in marker assisted selection. Rahman *et al.* (2007) developed SRAP markers for the seed coat color gene *Br1* in *B. rapa* and converted them to SCAR markers for use in MAS for yellow seed coat color. In addition, Rahman *et al.* (2008) developed SNP markers for the *FAE1.1* gene in the A-genome and SCAR markers for *FAE1.2* gene in the C-genome, both of which are involved in erucic acid biosynthesis. These markers have been used in MAS in the HEAR (High Erucic Acid Rapeseed) breeding programs at University of Manitoba for the development of high erucic acid rapeseed. Globally, many other molecular markers have been developed and utilized in MAS breeding for agronomic, disease resistance, seed yield and quality related traits in *Brassica* species. Some important traits and their molecular markers are listed in Table 2.2, which have been utilized for trait improvement through MAS.

Table 2.2 Some of the important traits mapped and molecular markers developed in *B. rapa* for marker assisted selection in breeding.

Genes	Traits	Types of marker	Mapping populations	References
BrFLC2	Flowering time	RFLP, SSR	BC_3S_1	Schranz *et al.* (2002)
ht1-ht5	Heat tolerance	AFLP, RAPD	RILs	Yu *et al.* (2003)
VFR2	Flowering time	RAPD, RFLP	BC_3S_1	Kole *et al.* (2001)
Dwf2	Dwarf	RFLP	F_2, BC_2	Muangprom and Osborn (2004)
Crr1, Crr2	Clubroot resistance	SSR	F_2	Suwabe *et al.* (2003)
Crr3	Clubroot resistance	STS	F_3	Hirai *et al.* (2004)
CRB	Clubroot resistance	SCAR	F_2	Piao *et al.* (2004)
Eru	Erucic acid content	RFLP	F_3	Teutonico and Osborn (1994)
Fad2	Oleic acid content	SCAR	F_2	Tanhuanpää *et al.* (1996)
Fad3	Linolenic acid content	RAPD	F_2	Tanhuanpää *et al.* (1995)
TTG1	Hairiness and seed coat color	SRAP-SCAR	DH lines	Zhang *et al.* (2009)
Ac1	White rust resistance	RFLP	RILs	Kole *et al.* (2002)

2.7. Homoeology between the A and C genomes

Genome homoeology has been characterized in *Brassica* species by comparative analyses of the genetic and physical maps of *Arabidopsis* with genetic maps of *Brassica* species (Osborn *et al*. 1997, Lan *et al*. 2000, Parkin *et al*. 2002, Lukens *et al*. 2003, Parkin *et al*. 2005). These studies indicate that each genomic region has had multiple events of polyploidization and chromosome rearrangements in the *Brassicaceae* lineage after the evolutionary divergence from *Arabidopsis* approximately 14.5 to 20.4 million years ago (MYA) (Yang *et al*. 1999, Parkin *et al*. 2002). In the *Brassicaceae*, the *B. nigra* (B) genome separated from the *B. rapa*/*B. oleracea* (AC) genome lineages about 7.9 MYA (Yang *et al*. 1999, Lysak *et al*. 2005).

There are high levels of homoeology among the A- and C-genomes of *B. rapa, B. oleracea* and *B. napus* (Parkin *et al*. 2005, Punjabi *et al*. 2008). Parkin *et al*. (2003) reported stretches of collinearity on the linkage groups N1 with N11, N2 with N12 and N3 with N13 of the A- and C-genomes, respectively. Similarly, Osborn *et al*. (2003) reported reciprocal interstitial translocations of homoeologous regions of linkage groups N7 and N16, and their effects on genome rearrangements and seed yield in *B. napus*. This suggests that inter-genomic translocations and rearrangements have taken place during the evolutionary divergence of *B. oleracea* and *B. rapa* from a polyploid ancestor (Sharpe *et al*. 1995). As a result of genomic synteny, there have been several reports of homoeologous recombination between the A- and C-genome of *Brassica* species (Udall *et al*. 2004, Leflon *et al*. 2006). Cytogenetic and molecular data revealed that small and large collinear genomic regions between the A- and C-genomes of *Brassica* species allow homoeologous recombination-based trait introgression to enhance genetic variability.

2.8. Inter or intra specific breeding for trait improvement in *Brassica*

Marker assisted selection in plant breeding is well supported by the availability of molecular maps developed using various marker systems in different mapping populations. The use of molecular markers has facilitated introgression of important traits through intra or interspecific as well as inter-generic crosses. Recently, an intertribal partial hybrid between *B. rapa* and *Isatis indigotica* has been developed and genome rearrangement in this wide cross was analyzed (Tu *et al.* 2009). This type of wide hybridization remains a useful approach to enhancing the gene pool and genetic variability in cultivated crops.

In addition to gene introgression from allied species into the adapted crop species, interspecific and intergeneric crosses such as *B. rapa* x *B. oleracea*, *B. napus* x *B .nigra*, *Diplotaxis erucoides* x *B. nigra* have been carried out for the development of monosomic alien chromosome addition lines (McGrath and Quiros 1990, Chen *et al.* 1992, Cheng *et al.* 1994). These chromosome addition lines proved to be useful for gene identification as well as for investigations of chromosome homoeology between species. Disomic alien chromosome addition lines have also been reported for specific chromosomes in some wide crosses e.g. *B. napus* x *S. alba* (Wang *et al.* 2005) and *B. napus* x *R. sativus* (Budahn *et al.* 2008). The C-genome of *B. oleracea* was the first *Brassica* genome dissected through the use of chromosome addition lines. Several researchers used molecular tools such as FISH, GISH, RFLP and other markers, and cytogenetic stocks, such as chromosome addition lines, for exploring the evolution of *Brassica* genomes and their organization, and homoeologous relationships (McGrath and Quiros 1990, Hu and Quiros 1991, Wang *et al.* 2005, Budahn *et al.* 2008). Cytological investigation revealed

that some of the alien chromosomes remained unpaired as a univalent during metaphase I of meiosis. In other cases, pairing and formation of trivalents or multivalents were formed, e.g. in *B. rapa* x *B. oleracea* (Chen *et al.* 1992, Hasterok *et al.* 2005), *B. napus* x *Crambe abyssinica* (Wang *et al.* 2006), *B. rapa* x *B. oxyrrhina* (Srinivasan *et al.* 1998) addition lines. These allosyndetical associations between alien and recipient chromosomes may be one of the important tools for generation of genetic variability through insertion, deletion, rearrangement or replacement of genomic region(s) for improved germplasm development. Improvement of traits through these approaches has been considered non-GMO (Genetically Modified Organism), which does not require regulatory approval for commercialization.

Similar to agronomic, disease resistance and yield related traits, seed quality traits such as glucosinolates can be genetically manipulated using interspecific hybridization followed by marker assisted selection for introgression or replacement of a native gene with the allied gene. Natural mutations for glucosinolate biosynthesis genes have been identified in accessions of *B. oleracea* (Li and Quiros 2002, 2003) and molecular markers have been developed. These molecular markers have been employed for the manipulation of glucosinolate profiles in *Brassica* through interspecific hybridization and marker assisted selection. Niu (2008) attempted to replace the functional *GSL-ALK* gene of *B. rapa* by the null allele from *B. oleracea* (broccoli) using a gene specific SCAR marker. However, introgression of the *GSL-ALK* null allele or replacement of a single locus with small effect did not change the glucosinolate profile of the *B. rapa* in this study. This suggests that multiple loci with functional redundancy play important roles in

glucosinolate biosynthesis in *Brassica* species. This approach has met with very little or no success. This might be due to many reasons, such as:

(i) duplicated or triplicated genomic regions may mask the effect of the single locus being replaced for a quantitative traits like glucosinolate profile and concentration

(ii) lack of similarity of gene and spacer sequences between alien and host chromosomes in monosomic or disomic alien chromosome addition lines

(iii) presence of active homoeologous recombination regulator genes during meiosis

(iv) directional exchange of genetic materials in trivalent formations during meiosis because of distinct chromosome behaviour

(v) host genome chromosome numbers and amount of homology between host and alien chromosomes

On the other hand, several traits in *Brassica* species have been improved through introgression of functional genes from allied species through interspecific or inter-generic crosses such as *B. rapa* x *B. oleracea* and *B. rapa* x *B. oxyrrhina* (Chen *et al.* 1994, Srinivasan *et al.* 1998).

3. IDENTIFICATION OF QTL AND CANDIDATE GENES FOR LEAF AND SEED ALIPHATIC GLUCOSINOLATES IN *Brassica rapa* L.

Jianfeng Geng, Arvind H. Hirani, Jiefu Zhang, Carla D. Zelmer, Peter B. E. McVetty,

Fouad Daayf and Genyi Li

Department of Plant Science, University of Manitoba

Winnipeg, Manitoba Canada R3T 2N2

The article will be submitted to Theoretical and Applied Genetics journal for publication

This chapter will be submitted to TAG as Geng *et al.* (2010). In this chapter, Arvind Hirani is the second equal contributing author for this study. Major work such as phenotype evaluation of a replicate for leaf and seed in greenhouse grown RILs and a replicate from field grown RILs were performed by Arvind. Nine glucosinolate gene specific molecular markers were developed and tagged, and statistical analysis was performed by Arvind. Dr. Jianfeng Geng prepared a genetic map and evaluated RILs for a replicate of leaf and seed glucosinolates from greenhouse grown RILs.

3. Identification of QTL and candidate genes for leaf and seed aliphatic glucosinolates in *Brassica rapa* L.

3.1. Abstract

Glucosinolates are sulphur- and nitrogen-rich plant secondary metabolites and are biosynthesized in plant species belonging to the order *Capparales* or *Brassicales*. Using a recombinant inbred line (RIL) population developed from a cross of Chinese cabbage and turnip rapeseed, a genetic map in *B. rapa* was assembled with 1,579 molecular markers including 9 gene specific markers for glucosinolates. Quantitative trait loci (QTL) mapping for glucosinolates was performed with the genetic map, and these gene specific markers were used to identify the genes involved in the biosynthesis of glucosinolates. Over a dozen QTL for 2-hydroxy-3-butenyl (progoitrin), 3-butenyl (gluconapin), 2-methylpropyl (isobutyl) and 4-hydroxy-indole-3-yl-methyl (4-hydroxyglucobrassicin) glucosinolates were identified in seeds and leaves. A candidate locus for *Br-GSL-ELONG* gene on linkage group A3 was identified to co-segregate with 5C side chain aliphatic glucosinolates (glucoalyssin, glucobrassicanapin) in *B. rapa*. This locus was also co-localized with the QTL controlling seed and leaf gluconapin (4C aliphatic GSL), suggesting that the gene(s) might have either multi-functional properties for 4C and 5C side chain aliphatic glucosinolate biosynthesis or differentiated functional properties of tandem duplicate genes in the locus. Genetic correlation of 4C and 5C side chain aliphatic glucosinolates was also analyzed in the RIL mapping population and revealed that they are negatively correlated to each other.

3.2. Introduction

Glucosinolates are a sulphur-containing broad class of plant secondary metabolites with structural variability. Predominantly they are present in the economically important *Brassica* crops species. Glucosinolates are *de novo* biosynthesized and stored in plant vacuoles. Tissue damage by physical injuries, biotic or abiotic stresses, however, triggers the production of numerous volatile and non-volatile bioactive compounds through activation of myrosinase enzymes. These bioactive compounds are largely isothiocyanates, thiocyanates and nitriles. These compounds have a variety of biological functions ranging from activation of defence mechanisms against fungal and bacterial pathogens in living plants to biopesticides and biofumigants from plant residues in soil. In some cases, breakdown products act as host-plant recognition signals by specialized insects, the presence of certain glucosinolates attracting pollinators and beneficial insects. In other cases, however, glucosinolates and their breakdown products stimulate leaf surface oviposition for adapted *Brassica* pests such as cabbage and turnip root flies (Bauer *et al.* 1996, Hopkins *et al.* 1997). Glucosinolate breakdown products in seed meal are present as anti-nutritional and potent toxic compounds for mammals and poultry. On the other hand, the derivative isothiocyanate sulforaphane of glucoraphanin has been associated with reduced risk of several kinds of cancers in humans (Talalay 2000, Brooks *et al.* 2001). The varieties of biological functions of glucosinolates have led researchers to further explore the area of glucosinolate biosynthesis, biochemistry, genetics and applications in *Arabidopsis* and *Brassica* crop species.

Glucosinolate biosynthesis is regulated by at least 56 genes at multiple loci in *Arabidopsis* and *Brassica* species (Zang *et al.* 2009). Biosynthesis of aliphatic glucosinolates is accomplished in three sets of reactions, beginning with deamination of amino acid precursors, followed by elongation with acetyl-CoA and final formation of core glucosinolate structures. Subsequently, side chain modifications are mediated by sets of enzymes, producing different classes of glucosinolates with various side chains of 3C to 8C groups. The genetic bases of glucosinolates have been extensively studied in the model plant *Arabidopsis thaliana*. QTL mapping, gene cloning and their functional characterization for glucosinolate core structure formation, side chain elongation and side chain modification from methionine precursors have been well reported in *Arabidopsis* (Compos de Quiros *et al.* 2000, Kliebenstein *et al.* 2001a, Schuster *et al.* 2006, Textor *et al.* 2007, Li *et al.* 2008). Similarly, major genes for glucosinolate side chain elongations, *BoGSL-ELONG* and *BoGSL-PRO* and side chain modifications, *BoGSL-ALK/AOP3* were cloned in *B. oleracea* using *Arabidopsis* sequence information (Li and Quiros 2002, 2003) and functionally characterized using RNA interference (RNAi) of *BoGSL-ELONG* gene family in *B. napus* (Zheng *et al.* 2010).

In commercially important crop species, agronomy, yield and quality related traits are quantitative in nature and have a complex genetic basis. Identification of QTL, therefore, acts as a primary step toward dissecting the molecular basis of such complex traits. An important diploid *Brassica* species, *B. rapa* has been extensively studied for QTL mapping of several physiological, morphological and disease resistance traits such as seed coat color, flowering time, club root resistance and black rot resistance (Suwabe

et al. 2003, Lou *et al.* 2007, Soengas *et al.* 2007). There is only one study, however, focused on QTL identification of leaf glucosinolates in *B. rapa* (Lou *et al.* 2008).

Several QTL mapping studies in two amphidiploid species, *B. napus* and *B. juncea* reported large and small effect QTL on different linkage groups for total seed glucosinolates (Magrath *et al.* 1994, Toroser *et al.* 1995, Uzunova *et al.* 1995, Howell *et al.* 2003, Mahmood *et al.* 2003, Ramchiary *et al.* 2007). Magrath *et al.* (1994) reported two major loci, *Gsl-elong-A* and *Gsl-elong-C* corresponding to the A-genome and C-genomes in *B. napus*, respectively, that are involved in the production of 4C and 5C aliphatic glucosinolates. In the QTL fine mapping study, Bisht *et al.* (2009) reported four *GSL-ELONG* loci in the A-genome and two *GSL-ELONG* loci in the B-genome of *B. juncea* for glucosinolate biosynthesis along with other side chain modification loci. This suggests that the nature of glucosinolate traits within the amphidiploid *Brassica* is complex. It is, therefore, important to study the genetic regulation of glucosinolate biosynthesis in diploid *Brassica* species to determine possible loci with large and small effects on individual and total glucosinolates.

The present study focused on identification of QTL for individual aliphatic glucosinolates and the underlying genes for these large effects QTL. Major genes/loci for glucosinolate biosynthesis corresponding to significant QTL were tagged with a high-density genetic map. Furthermore, a major candidate locus of *Br-GSL-ELONG* gene on linkage group A3 was identified to co-segregate with 5C aliphatic glucosinolates in *B. rapa* and a co-dominant SCAR marker, BrMAM1-1 was developed. This marker could be useful to reduce 5C aliphatic glucosinolates in vegetable and oilseed *B. rapa* and its derivative amphidiploid species by marker assisted selection in breeding.

3.3. Materials and Methods

Plant materials and DNA isolation

A RIL (F_6) population consisting of 543 individuals were developed through consecutive selfing of an F_1. The maternal parent, a self-incompatible, brown seeded doubled haploid (DH) line 'RI16' was derived from microspore culture of an F_1 hybrid Chinese cabbage cultivar and the paternal parent, yellow sarson 'BAR6', was a self-compatible, yellow seeded pure inbred line. A total of 92 RILs (F_6) were randomly selected and used for genetic map construction and QTL analysis. Genomic DNA was isolated from young leaves of each of these 92 RILs and the two parents using a modified 2×CTAB method (Li and Quiros 2001, Sun *et al.* 2007).

Polymorphism detections and genetic map construction

DNA polymorphisms in the RIL population were identified using SRAP and SSR molecular markers. A total 207 primer pairs were used for SRAP marker detection. The PCR reactions were performed with four different color labelled primers (FAM-blue, VIC-green, NET-yellow PET-red) and unlabelled reverse primers as described previously by Sun *et al.* (2007). After PCR, 2.5 µl pooled products labelled with the four colors were added into 7.5 µl formamide mixed with 500-LIZ standard (Applied Biosystems, CA, USA) and denatured products were analyzed with the ABI 3100 Genetic Analyzer (Applied Biosystems, CA, USA). In addition, sequence information of some of the previously used SSR primers was obtained (Suwabe *et al.* 2002, Lowe *et al.* 2004, Piquemal *et al.* 2005). A total 130 SSR primer pairs were designed using NCBI sequence database of *B. rapa* (NCBI http://www.ncbi.nlm.nih.gov/nucleotide/) and a common 19

bp M13 sequence (5'-CACGACGTTGTAAAACGAC-3') was added to the 5' end of all the forward SSR primers in order to separate PCR products in the ABI 3100 Genetic Analyzer. An M13 primer labelled with one of the four different colors was added in the PCR reactions. All PCR reactions were performed in a 10 µl volume containing 7.0 µl dH_2O, 1 µl 10x PCR buffer (100mM Tris-HCl, 500mM KCl, 0.1% Triton X-100, pH-9.3), 0.3 µl 50 mM $MgCl_2$, 0.15 µl 25mM dNTPs, 0.5U Taq polymerases, 0.05 µl 10 µM forward primer, 0.15 µl 10 µM reverse primer, 0.10 µl 10 µM M13 labelled primer and 2.0 µl template DNA. The PCR running program was as follows: 94°C for 5 min; 5 cycles with 94°C for 50 s, 56°C for 50 s, 72°C for 1 min, with a 0.8°C decrease in annealing temperature at each cycle; then 25 cycles with 94°C for 50 s, 51°C for 50 s, 72°C for 1 min followed by 72°C for 7 min. All the PCR products were separated in the ABI 3100 Genetic analyzer as described in SRAP procedure.

The profiles of all SRAP and SSR markers were analyzed with GenScan (Applied Biosystems, CA, USA). The ABI files were converted to gel-like images and then all possible polymorphic loci were scored by the Genographer software (Genographer http://hordeum.oscs.montana.edu/software/genographer/index.html), and further confirmed by visualization and abundance threshold settings. Polymorphic SRAP loci were scored as dominant markers. All markers in the RIL population were examined by χ^2 test for goodness of fit to the expected segregating ratios (1:1). Linkage analysis was conducted with Join map® 3.0 (Van Ooijen and Voorrips 2001) at LOD values of 6 to 10 and all markers were assembled into 10 linkage groups.

Gene specific SCAR and SNP marker development and tagging

Homologs of *GSL-ELONG*, *GSL-PRO*, *GSL-FMO$_{OX1}$* and *GSL-AOP/ALK* genes in *B. rapa* (the A-genome) were found at NCBI and *Brassica* genome sequence databases and used to design primers to find polymorphisms between the two parents. DNA sequences within or very closely flanking regions of *GSL-ELONG* gene were used to blast on the A genome database (Brassica genome gateway http://brassica.bbsrc.ac.uk) and to find EST or BAC end sequences. Two ESTs (BR043724, BR003821) and two BAC end sequences (KBrB010E08F and KBrH121C04F) were identified, which displayed high colinearity with *GSL-ELONG* of *A. thaliana* and *B. oleracea*. Primers were designed based on the BAC and EST sequences. The PCR products amplified with 10 primer pairs and DNA from the two parents were sequenced to find INDELs or SNPs. Sequencing of PCR products was performed using the standard ABI sequencing protocol V3.1 (ABI, CA, USA). A 16 bp deletion detected in the BrMAM1-1 (KBrH121C04) locus was used to develop a SCAR marker (Table 3.1, Appendix i). A pair of primers (BrMAM1-1F, 5'-GTTTCCCTGCGTCATCAGA-3' and BrMAM1-1R, 5'-CTAAGCTCTTCGCATAGCTA-3') was used for the BrMAM1-1 SCAR marker. The PCR running program was 94°C, 1 min; 55°C, 1 min; 72°C, 1.5 min for 35 cycles. For two BrMAM3 loci, PCR was performed using two pairs of primers (CelongF5, 5'-CCACGAGCAAACACATATA-3' and CelongR5, 5'- CTTGGTGTGGTGCAGAAGA-3' for CE5; AelongF1, 5'- TTGAGGCCAGGTACCCGGA-3' and AelongR1, 5'-CAACACTTATGTATATATA-3' for AE1). Amplification condition was the same as previously described. All the PCR products were analyzed on 1.2% agarose gels. Sequence information for *GSL-PRO* loci from *B. oleracea* BAC sequences (B21F5,

AC152123.1 and B59J16, EU568372) was used to design primers. An SNP and one 12 bp deletion were detected for *GSL-PRO* loci corresponding to the C-genome BAC, B59J16 (Appendix ii). The 12 bp INDEL SCAR markers, CPRO3B_1 was scored by pair of primers (CproF3b, 5'- TGGAGTTCCAAGTTTGCATG-3' and CproR3b, 5'-GCCACTATAGTTAGAGAAC-3'). The PCR conditions were as described previously. The SNP (SNP-PRO1B) was detected using the ABI SNaPshot protocol and an SNP primer (SNP-PRO1B, 5'-TTTTTTTTTTTTGATATATAACTTGAAGAAACATA-3'). Two gene specific SCAR markers for *Br-GSL-FMO$_{OX1}$* were developed by designing primers from the BAC clone KBrS002F02 that has high collinearity with *Arabidopsis* gene *GSL-FMO$_{OX1}$*. Two *Br-GSL-FMO$_{OX1}$* gene specific SCAR markers were amplified by two pairs of primers (BrGSOX1Fa, 5'- TCGTGAAGGTCACACCGTC-3' and BrGSOX1Ra, 5'- CCAGGAATATGAGCAACA-3' for FMOFaOXRa-250; BrGSOX1Fb, 5'- ACAAGTGGGAGGTCTCTGGA-3' and BrGSOX1Ra, 5'-CCAGGAATATGAGCAACA-3' for FMOFaOXRb-250) with the PCR condition described above for other SCAR markers. Additionally, two *GSL-AOP/ALK* gene specific SCAR markers were developed and mapped in the RIL population.

Leaf and seed glucosinolate extraction and quantification by HPLC

For phenotyping of glucosinolates in leaves and seeds, 92 RILs (F_6 and F_7 generations) and their parents grown in the greenhouse as well as F_6 seeds grown in the field at the University of Manitoba were used. In the greenhouse, seeds were planted in 48-cell trays and seedlings were transferred to 20 cm pots after 3 weeks. At 2 weeks after transplanting, 250 mg young leaf tissues were collected and ground in liquid nitrogen. From the same RI lines, 200 mg air dried seeds of greenhouse grown plants were used for

seed glucosinolate extraction. Similarly, 200 mg seeds from field grown RILs (F_6) were used for seed glucosinolate extraction. Glucosinolates were desulfonated and purified using a sephadex/sulfatase method as previously described by Kliebenstein *et al.* (2001b) with some modifications. All desulfoglucosinolates were eluted in 400 µl 35% methanol solution and 50 µl was injected and run in a 5 µm column (LichroCART® 250-4 RP18, Fisher Scientific, Ottawa, Canada) coupled with the Alliance® reverse phase HPLC (Waters 2695) and photodiode array detector (Waters 996) system (Waters, MA USA). The mobile phase gradient with a flow rate of 1 ml/min solvent A (HPLC grade methanol Fisher Scientific, Canada) and solvent B (distilled water) were used for all the samples. Solvent phase gradients were programmed as follows: 8 min 7:93 A/B (v/v), 4 min 15:85 A/B (v/v), 18 min 55:45 A/B (v/v), 5 min 92:8 A/B (v/v), 5 min 92:8 A/B (v/v), 5 min 1.5:98.5 A/B (v/v), 3 min 1.5:98.5 A/B (v/v) and final 4 min 0:100 A/B (v/v) with a total of 52 min running time for each run. Desulfoglucosinolates were detected at 229 nm on a Water 996 photodiode-array detector and identity of individual peaks were confirmed based on retention time and comparison of UV absorption spectra with those of standards. All the desulfoglucosinolate absorbance data were converted into micromoles per gram of seeds or leaves and adjusted with corresponding response factors determined from the purified standards (Brown *et al.* 2003).

Six major peaks from each sample were analyzed according to their retention times and relative response factors. These were progoitrin (PRO) at 6.7 min, glucoalyssin (ALY) at 12.9 min, gluconapin (GNAP) at 14.4 min, 4-hydroxy glucobrassicin (4OH) at 15.5 min, 2-methylpropyl (2MP) at 17.9 min and glucobrassicanapin (GBNAP) at 19.4 min (Figure 3.1). Total aliphatic glucosinolates represent a sum of all aliphatic

glucosinolates (PRO, ALY, GNAP, 2MP and GBNAP) and total indole glucosinolates include three indole glucosinolates (4OH, glucobrassicin and 4-methoxy-glucobrassicin) and total glucosinolates (TGSL) include all the glucosinolates in each sample. Individual and average values of F_6 and F_7 RILs for leaf and seed glucosinolate content were used for QTL mapping.

QTL mapping for leaf and seed glucosinolates

QTL mapping was performed using Windows QTL cartographer v 2.5 (Wang *et al.* 2010) and composite interval mapping (CIM) was selected. For individual glucosinolates, a threshold for a QTL was determined by 1000 fold permutations using Zmapqtl program (Basten *et al.* 1996). QTL were declared when the LOD value was higher than the threshold. When multiple peaks were found close to each other, the peaks within one-half LOD distance and with highest LOD scores were defined as QTL positions. If the distance exceeded one-half LOD, the peaks were treated as separate QTL (Flint-Garcia *et al.* 2003).

3.4. Results

High density genetic map construction

A total of 2,966 molecular markers, which include 2,496 SRAP, 436 SSR, 25 SCAR and 9 SNP markers were scored for the RIL population. The χ^2 tests revealed highly biased segregation ratios, and in total, 1,169 of 2,496 markers (39.4%) were distorted from a 1:1 segregating ratio. Among these distorted markers, there were 973 SRAP (39%), 186 SSR (42.7%), 8 SCAR (32%) and 2 SNP (22.2%). Six hundred and ninety one (59.1%) markers were biased toward the maternal parent RI16, whereas 478

(40.9%) were biased to the paternal parent yellow sarson BAR6. In the Joinmap® software, closely linked markers and markers linked with very low LOD values were eliminated during linkage map construction. The genetic map consisted of 1,579 molecular markers including 1,429 SRAP, 129 SSR, 12 SCAR and 9 SNP that were assembled into 10 linkage groups. All the linkage groups were anchored to chromosome numbers A1 to A10 using the reported SSR markers on the previous *Brassica* genetic maps (Figure 3.3). The high density genetic map covered a genetic distance of 1,612.31 cM with average marker spacing of 1.02 cM and without major gaps of over 10 cM. The most saturated linkage group with 181 markers and the lowest one with 107 markers were linkage groups A6 and A10, respectively. Among all the linkage groups, maximum length (252.5 cM) and minimum length (107.6 cM) were observed for linkage groups A9 and A5, respectively. Eight SCAR and one SNP markers for glucosinolate genes/loci were mapped on three different linkage groups, A2, A3 and A9 (Table 3.1). In case of all the mapped markers, 568 of 1,579 (35.9%) were showed segregation distortion from 1:1. A total of 253 (44.5%) markers were biased to the maternal parent RI16 and 315 (55.5%) markers were biased toward the paternal parent BAR6.

QTL for leaf and seed glucosinolates

Glucosinolate profile and content of the parents used to develop the RIL population were very diverse (Figure 3.1, Table 3.4). In yellow sarson BAR6, gluconapin predominately was biosynthesized while in Chinese cabbage RI16, progoitrin, gluconapin, glucoalyssin and glucobrassicanapin as well as small amount of 2-methylthiopropyl glucosinolate and indole glucosinolates were all detected at a relatively low level. QTL mapping was performed for leaf (L) and seed (S) progoitrin (PRO),

glucoalyssin (ALY), gluconapin (GNAP), 4-hydroxyglucobrassicin (4OH), 2-methylpropyl (2MP), glucobrassicanapin (GBNAP) and total glucosinolates (TGSL). Threshold LOD values were obtained for each glucosinolate by 1000 permutation using the Zmapqtl program (Basten *et al.* 1996). Two QTL (LPRO-1.1 and LPRO-3.1) that were detected for leaf progoitrin explained 13% and 21% total phenotypic variance, respectively. Two QTL (LALY-3.1 and LALY-3.2) that were identified for leaf glucoalyssin explained total phenotypic variance of 32%. For leaf gluconapin, four QTL were detected (LGNAP-1.1, LGNAP-3.1, LGNAP-3.2 and LGNAP-3.3) which explained 52% total phenotypic variance. Additionally, four QTL (L2MP-3.1, L2MP-3.2, L2MP-4.1 and L2MP-8.1) were detected for 2-methylpropyl and these explained total phenotypic variance of 51%. Similarly, a cluster of three QTL (LGBNAP-3.1, LGBNAP-3.2 and LGBNAP-3.3) were identified for leaf glucobrassicanapin and explained 31% total phenotypic variance. Total leaf glucosinolates was controlled by three QTL (LTGSL-3.1, -3.2 and LTGSL-8.1), which explained total phenotype variance of 38% (Table 3.5, Figure 3.4).

In addition, three QTL (SPRO-3.1, SPRO-3.2 and SPRO-7.1) for seed progoitrin, three QTL (SALY-3.1, SALY-3.2 and SALY-9.1) for seed glucoalyssin, three QTL (SGNAP-3.1, SGNAP-3.2 and SGNAP-3.3) for seed gluconapin, three QTL (S2MP-2.1, S2MP-3.1 and S2MP-9.1) for seed 2-methylpropyl glucosinolate, two QTL (SGBNAP-3.1 and SGBNAP-3.2) for seed glucobrassicanapin and five QTL (STGSL-3.1, STGSL-5.1, STGSL-7.1, STGSL-8.1 and STGSL-9.1) for seed total glucosinolates were identified and these explained total phenotypic variance between 31% and 80% (Table 3.6, Figure 3.5). For leaf and seed QTL, phenotypic variance, linkage group, flanking

markers and other information are given in Tables 3.5 and 3.6; Figures 3.4 and 3.5 respectively.

Development of gene specific SCAR and SNP markers for glucosinolates

The high sequence conservation between the *Arabidopsis* and *B. rapa* genomes provides an adequate genetic basis for their comparative analysis. In this study, glucosinolate gene specific markers were developed with known gene sequences. A total four SCAR markers for *Br-GSL-ELONG* loci were mapped on three different linkage groups, CE5 on A2, BrMAM1-1 on A3, and AE1 on A9 (Table 3.1). Similarly, two SCAR markers FMOFaOXRa-250 and FMOFbOXRa-250 were integrated on linkage group A9 for *Br-GSL-FMO$_{OX1}$* gene. The *GSL-FMO$_{OX1}$* gene is known to be involved in methylthioalkyl to methylsulphinylalkayl reactions in *Arabidopsis* (Li *et al.* 2008). An SNP (SNP-PRO1B) and a SCAR marker (CPRO3B_1) for *GSL-PRO* gene were mapped on linkage group A9 (Appendix ii). Two SCAR markers for *GSL-AOP/ALK* genes were developed and integrated on linkage groups A3 and A9, which controlled progoitrin aliphatic glucosinolate biosynthesis.

***Br-GSL-ELONG* gene specific marker co-segregated with aliphatic glucosinolates**

Five-carbon side chain aliphatic glucosinolates are biosynthesized from trihomo-methionine by a set of enzymatic reactions. Total 5C glucosinolates (sum of glucoalyssin and glucobrassicanapin) in the RILs were grouped in two classes, which corresponded to the parental 5C aliphatic glucosinolates. There was no gluconapoleiferin (a hydroxylated 5C aliphatic glucosinolate) detected in the segregated RILs. In the RIL population, 40 lines had less than 3 µmol/g seed of 5C aliphatic glucosinolates, whereas other 51 lines

had between 9 and 25 µmole/g seed of 5C aliphatic glucosinolates. The ratio of these two classes was not significantly different from 1:1 (χ^2=0.67, df=1, P>0.1), which suggested that a major Mendelian locus controlled the content of 5C aliphatic glucosinolates. Fortunately, a codominant SCAR marker located inside the *Br-GSL-ELONG* gene co-segregated with 5C aliphatic glucosinolates when the RI lines with the low content (<3 µmol/g seed) of 5C aliphatic glucosinolates was scored as one allele and the RI lines with the high content (>10 µmol/g seed) as another allele in a locus. This single locus controlling 5C aliphatic glucosinolates in *B. rapa* was mapped on linkage group A3 (Table 3.2, Figure 3.2).

Four-carbon side chain aliphatic glucosinolates composed of glucoerucin, glucoraphanin, gluconapin and progoitrin that are biosynthesized from dihomo-methionine precursors (Magrath *et al.* 1994). The RILs segregated for gluconapin and progoitrin 4C aliphatic glucosinolates. Gluconapin content in the RIL population, however, displayed continuous variation. The lowest and highest gluconapin contents were 18 µmole/g seed and 68 µmole/g seed, respectively, in the RILs. The *Br-GSL-ELONG* gene specific SCAR marker BrMAM1-1 was closely linked with the major QTL for gluconapin in leaves and seeds, 2-methylpropyl glucosinolate in leaves, glucoalyssin in seeds and glucobrassicanapin in leaves and seeds.

Correlation between 4C and 5C aliphatic glucosinolate biosynthesis

Pearson's correlation coefficient in the RILs was analyzed for 4C, 5C and total aliphatic glucosinolates in leaves and seeds. Four-carbon and 5C aliphatic glucosinolates had a negative correlation in seeds (-0.50). A positive correlation between leaf and seed for 4C (0.51) and 5C (0.53) aliphatic glucosinolates were observed (Table 3.3). Similarly,

total aliphatic glucosinolates had a positive correlation with 4C aliphatic glucosinolates in leaves (0.91) and seeds (0.72) (Table 3.3).

3.5. Discussion

In this study, a recombinant inbred (RI) population from a cross of yellow sarson BAR6 and Chinese cabbage RI16 was used for high density linkage map construction as well as QTL mapping for glucosinolates in leaves and seeds. The RIL population segregated for glucosinolates along with a number of different traits including flowering time, self-compatibility and seed coat color. The RIL populations have an advantage over DH lines in terms of high numbers of recombination events between the parental genomes, which provide detailed genetic information to study complex traits like glucosinolates. Similarly, RIL populations have a unique advantage over F_2 populations. They provide homozygous progenies that are genetically adequate materials to determine the allelic functions of a gene.

Constructing a high-density and lab-to-lab sharable genetic map is the key step for QTL mapping, gene cloning, marker assisted selection and gene pyramiding. In this study, a high-density genetic linkage map was constructed with SRAP, SSR, SCAR and SNP markers in *B. rapa*. This map consisted of 10 linkage groups and was anchored to 10 chromosomes of the A genome with reference SSR markers. The genetic map covered a total genetic length of 1,612.31 cM, with an average marker space of 1.02 cM. This is the densest genetic linkage map in *B. rapa* and the length is comparable with other reference genetic maps (Song *et al.* 1991, Ajisaka *et al.* 1995, Nozaki *et al.* 1995, Kim *et al.* 2006, Choi *et al.*2007, Geng *et al.*2007, Chen *et al.*2009, Kapoor *et al.*2009). Microsatellites, or

simple sequence repeat (SSR) with the advantages of abundance in the plant genome, codominant inheritance, locus specificity, easy detection and reproducibility, have been exploited in almost all major crops. Compared to other crop plants such as rice and soybean (McCouch *et al.* 2002, Hwang *et al.* 2009), there are few public SSR markers available in *Brassica*, although several research groups have presented a number of SSR markers using different technologies (Plieske and Struss 2001, Suwabe *et al.*2002, Tommasini *et al.* 2003, Lowe *et al.*2004, Piquemal *et al.* 2005, Yang *et al.* 2007, Suwabe *et al.*2006, Choi *et al.* 2007, Cheng *et al.* 2009). The methods have been improved from the earlier expensive technology of probe hybridization to easy development of online databases for markers and other sequence information. In our study, a large numbers of SSR sequences (including (AT)n, (CT)n, and (GA)n) were downloaded from the NCBI databases, and a large numbers of SSR primers were developed, tested and adopted to construct the genetic map. For the newly developed 130 SSR primer pairs, 99 (76.2%) primer pairs were amplified successfully, of which 71 (71.7%) were polymorphic between the two mapping parents. The 71 SSR primer pairs provided 131 clear bands scored in the RIL population, with average polymorphic bands of 1.85 per primer pair. The results suggested that this method is accessable and efficient for SSR development and genetic map construction.

In the *Brassicaceae* family, *B. rapa* is one of the most diverse species for glucosinolates. For example, 16 different glucosinolates were identified in a samples of around 110 accessions (Padilla *et al.* 2007). Fifty-eight different morphotypes of *B. rapa* including Chinese cabbage, pak choi, turnip, komatsuna, oilseed rape and mizuna possessed eight different glucosinolates (He *et al.* 2000), suggesting that variation of

glucosinolates among *B. raoa* accessions may occur. In the current study, yellow sarson BAR6 and Chinese cabbage RI16 are distinct for aliphatic glucosinolates such as progoitrin, gluconapin, glucoalyssin and glucobrassicanapin, so the RIL population segregated for all these individual glucosinolates (Figure 3.1).

One of the parents, Chinese cabbage RI16 biosynthesized about 20 µmole/g seed 5C aliphatic glucosinolates, whereas other parent yellow sarson BAR6 produced trace amount 5C aliphatic glucosinolates. The RIL population therefore segregated for 5C aliphatic glucosinolates. Quantitative glucosinolate variation in the RIL population, especially 4C and 5C side chain aliphatic glucosinolates suggesting that multiple genetic loci are involved in glucosinolate biosynthesis. At the same time, transcriptional factors might also have an impact on total glucosinolate production. In our QTL mapping study, 5C side chain aliphatic glucosinolates were identified to be regulated by a single major effect locus, *Br-GSL-ELONG* and a codominant SCAR marker BrMAM1-1 co-segregated with 5C aliphatic glucosinolates in the RILs (Figure 3.2). It suggests that a major locus *Br-GSL-ELONG* might be non-functional in the parent yellow sarson BAR6, which produces small quantity of 5C aliphatic glucosinolates.

Glucosinolate side chain elongations and modifications are multi-step reactions. Individual genes and their loci are involved in specific glucosinolate biosynthesis reactions. The glucosinolate side chain elongations are controlled by *GSL-ELONG* for 4C and long chain aliphatic glucosinolates (Compos de Quiros *et al.* 2000, Li and Quiros 2002), whereas *GSL-PRO* controls 3C aliphatic glucosinolates (Li *et al.* 2003, Gao *et al.* 2006). Glucosinolate side chain modifications are controlled by *GSL-FMO$_{OX1}$* for methylthiobutyl to methylsulfinyl glucosinolates in *Arabidopsis* (Li *et al.* 2008) and *GSL-*

AOP/ALK for methylsulfinyl to butenyl or hydroxy butenyl aliphatic glucosinolates (Kliebenstein *et al.* 2001c, Li and Quiros 2003). In this study, we mapped three *GSL-ELONG* loci, one *GSL-PRO* and *GSL-FMO$_{OX1}$* locus, and two *GSL-AOP/ALK* loci using *Arabidopsis* and *B. oleracea* sequence information (Table 3.1). *Arabidopsis* possesses high colinearity with *Brassica* species for the genes involved in glucosinolate biosynthesis (Gao *et al.* 2006, Qiu *et al.* 2009); the major difference is the greater number of gene loci in *Brassica*. In general, *Brassica* species possess more gene loci at each step of glucosinolate biosynthesis than *Arabidopsis* due to genome duplications and chromosomal rearrangements after *Brassica*'s evolutionary divergence from *Arabidopsis* (Bowers *et al.* 2003, Lysak *et al.* 2005).

In this study, we mapped three QTL (SGNAP-3.1, SGNAP-3.2 and SGNAP-3.3) for gluconapin (4C aliphatic GSL) and a large QTL for glucobrassicanapin (SGBNAP-3.2) and glucoalyssin (SALY-3.2) (5C aliphatic GSLs) on linkage group A3 in *B. rapa*. Some of these QTL are linked with the SCAR marker BrMAM1-1 of *GSL-ELONG* locus, which explained high phenotypic variance (Table 3.6).

Some of the SCAR markers for glucosinolate genes did not show linkage with QTL identified in this study. This could be the result of both parents of the RI population sharing the same allele at a QTL. A single allele will not be detected in the QTL analysis. In a previous study, Lou *et al.* (2008) reported two QTL for glucobrassicanapin on linkage group A3 and A9. Same study reported eight QTL for gluconapin on different linkage groups with two large QTL on linkage group A3 in leaves of two DH populations of *B. rapa*. In *B. napus* and *B. juncea*, most of the previously reported work emphasized the genetic control of propyls, butyls, pentyls and/or total aliphatic glucosinolates

(Toroser *et al*. 1995, Cheung *et al*. 1998, Howell *et al*. 2003, Mahmood *et al*. 2003, Quijada *et al*. 2006). Ramchiary *et al*. (2007) mapped nine loci on different linkage groups for propyls, butyls, pentyls and total aliphatic glucosinolates in *B. juncea*. Among these nine loci, a major locus on linkage group J3 corresponding to A3 of the A-genome was involved in biosynthesis of butyl(s), pentyl(s) and total aliphatic glucosinolates. The same *GSL-ELONG* locus on linkage group A3 in this study was identified as a major locus involved in biosynthesis of gluconapin, glucoalyssin and glucobrassicanapin. Bisht *et al*. (2009) fine mapped the major locus on linkage group J3, which was identified in the previous study by Ramchiary *et al.* (2007). The fine mapping study revealed the presence of two *GSL-ELONG* loci and one *Myb28* locus on linkage group J3 in *B. juncea*. They suggested that one of these two *GSL-ELONG* loci on linkage group J3 was involved in 4C or 5C aliphatic glucosinolate biosynthesis, whereas *Myb28* is a positive regulator of the basal level of aliphatic glucosinolate production in *B. juncea*.

In conclusion, aliphatic glucosinolates in leaves and seeds are regulated by multiple loci with different interactions of which several were mapped in this study. A major QTL for glucoalyssin and glucobrassicanapin of leaf and seed was co-localized with the *GSL-ELONG* locus. This study revealed single *GSL-ELONG* locus regulates 5C aliphatic glucosinolate biosynthesis in *B. rapa*. The function of this specific *Br-GSL-ELONG* locus could be confirmed through the introduction of this locus into *B. oleracea*, such as broccoli and cauliflower, in which 5C side chain aliphatic glucosinolates are not detectable. Alternatively, resynthesized *B. napus* lines can be developed through a cross between RI16 (*B. rapa*) and broccoli or cauliflower (*B. oleracea*). Glucosinolate profiles

of these lines would determine function of the *GSL-ELONG* locus of *B. rapa* in resynthesized *B. napus*.

Table 3.1 SCAR and SNP markers integrated on different linkage groups for glucosinolate biosynthesis loci.

Genes	Name	LG	Genebank#	Primers 5'-3'
GSL-ELONG	AE1	A9	DX890412 (KBrH018C06)	F-TTGAGGCCAGGTACCCGGA R-CAACACTTATGTATATATA
GSL-ELONG	BrMAM1-1	A3	CT013275 (KBrH121C04)	F- GTTTCCCTGCGTCATCAGA R- CTAAGCTCTTCGCATAGCTA
GSL-ELONG	CE5	A2	AC149635 (B19N3)	F- CCACGAGCAAACACATATA R- CTTGGTGTGGTGCAGAAGA
GSL-FMO_{OX1}	FMOFaOXRa-250	A9	FJ376070 (KBrS002F02)	F- TCGTGAAGGTCACACCGTC R- CCAGGAATATGAGCAACA
GSL-FMO_{OX1}	FMOFbOXRa-250	A9	FJ376070 (KBrS002F02)	F-ACAAGTGGGAGGTCTCTGGA R- CCAGGAATATGAGCAACA
GSL-PRO	SNP-PRO1B	A9	EU568372 (B59J16)	F- TGGAGTTCCAAGTTTGCATG R- GCCACTATAGTTAGAGAAC
GSL-PRO	CPRO3B_1	A9	EU568372 (B59J16)	F- TGGAGTTCCAAGTTTGCATG R- GCCACTATAGTTAGAGAAC
GSL-AOP1	AOP1b-472	A9	AC232438 (KBrB002P01)	F- TTGAATATCCAGTGTAAGGTT R-TTCCATCATTTACTTTCTCAGCAG
GSL-AOP3	AOP3aFR-397	A3	AC232438 (KBrB002P01)	F- GCGTAAACATGATGGAATTT R- ATGTGTTTATTGTAACCGGG

Table 3.2 Genotype of BrMAM1-1 marker and 5C aliphatic glucosinolate content in the RIL population of *B. rapa*.

BrMAM1-1	AA	AB	BB
Genotypes	++ (51)	+− (3)	−− (40)
GH-Leaf ∑5C	>1.5 (44) 7*	1-2.5 (2) 1*	<2.0 (38) 2*
GH-Seed ∑5C	>9 (49) 2*	5-16 (3)	0-3 (38) 2*
Field-Seed ∑5C	>9 (45) 6*	5-16 (3)	0-3 (37) 3*

* Missing data for leaf and seed glucosinolate phenotypes

Table 3.3 Pearson correlation coefficients of leaf (L) and seed (S) sum of 4C (∑4C), sum of 5C (∑5C) and total aliphatic glucosinolates (TAGSL) in the RIL population.

	L∑4C	**L∑5C**	**LTAGSL**	**S∑4C**	**S∑5C**	**STAGSL**
L∑4C	1					
L∑5C	-0.14 ^{NS}	1				
LTAGSL	0.91***	0.26 ^{NS}	1			
S∑4C	0.51***	-0.41***	0.32*	1		
S∑5C	-0.53***	0.53***	-0.29*	-0.50***	1	
STAGSL	0.15 ^{NS}	-0.04 ^{NS}	0.13 ^{NS}	0.72***	0.24 ^{NS}	1

NS, P>0.01; * P<0.01; ** P<0.001; *** p<0.0001; df= 91. L∑4C- leaf sum of 4C aliphatic glucosinolates, L∑5C- leaf sum of 5C aliphatic glucosinolates, LTAGSL- leaf total aliphatic glucosinolates, S∑4C- seed sum of 4C aliphatic glucosinolates, S∑5C- seed sum of 5C aliphatic glucosinolates, STAGSL- seed total aliphatic glucosinolates

Table 3.4 Statistical parameters of RIL populations and their parents BAR6 and RI16 (average of F_6 and F_7 generations) for seed glucosinolate content (in µmole/g seeds).

Parameters	**PRO 4C**	**ALY 5C**	**GNAP 4C**	**2MP 3C**	**GBNAP 5C**	**Sum of 4C**	**Sum of 5C**	**TAGSL**
RILs means	1.7	1.1	44.5	0.7	8.0	46.3	9.1	56.1
Range	0.0-15.4	0.0-8.7	18.1-68.4	0.0-2.7	0.0-22.2	19.1-70.9	0.2-24.1	20.4-71.9
Variance	6.3	2.2	110.7	0.3	50.5	119.4	57.7	96.8
P1: BAR6	0.2	0.0	60.6	0.1	0.6	60.9	0.6	61.6
P2: RI16	1.5	2.3	35.7	0.3	11.4	37.2	13.7	51.2
Difference	-1.3	-2.3	24.9	-0.2	-10.8	23.7	-13.1	10.4

PRO- progoitrin, ALY- glucoalyssin, GNAP- gluconapin, 2MP- 2-methylpropyl, GBNAP- Glucobrassicanapin, Sum of 4C- sum of 4C aliphatic glucosinolates, Sum of 5C- sum of 5C aliphatic glucosinolates, TAGSL- total aliphatic glucosinolates

Table 3.5 Summary of QTL identified for average leaf glucosinolates from F_6 and F_7 RIL populations in *B. rapa*.

GSL Name	QTL	LG	Position (cM)	LOD	R^2 (%)	Additive effect	Flanked markers	Closest SNP/SCAR/SSR	Distance (cM)
Progoitrin-4C	LPRO-1.1	A1	90.21	4.50	13.22	0.60	1220ccB121/1219abG261	CB10099-243	10.11
	LPRO-3.1	A3	80.21	6.11	20.72	-0.74	1214aaY563/1214bdB123	Na12E02-146	5.99
Glucoalyssin- 5C	LALY-3.1	A3	107.31	4.36	15.90	0.24	1214acB179/1214cdG233	SR058-303	2.61
	LALY-3.2	A3	121.41	4.51	16.29	-0.27	1219bdG346/1213bcG299	SR058-303	16.71
Gluconapin-4C	LGNAP-1.1	A1	18.41	4.46	11.79	1.15	1218adR170/SR021-198	SR021-198	0.39
	LGNAP-3.1	A3	29.71	4.02	10.39	-0.98	1214dbR219/SR057-353	SR057-353	4.29
	LGNAP-3.2	A3	149.81	6.33	17.35	-1.23	BrMAM1-1/1212acY508	BrMAM1-1	0.10
	LGNAP-3.3	A3	153.51	4.43	12.67	1.08	BrMAM1-1/SNP_tt6-211	BrMAM1-1	0.01
4OH-Indole	L4OH-1.1	A1	105.81	3.25	9.08	0.01	1207acG153/1220caY300	SR116-445	24.49
	L4OH-2.1	A2	124.01	3.99	12.19	-0.01	1214adY432/1214ddB504	BRAS011-249	59.04
	L4OH-2.2	A2	136.81	5.58	16.45	0.02	1220cdG305/1212adY255	BRAS011-249	71.81
	L4OH-6.1	A6	106.81	3.20	9.24	-0.01	1213baB288/1214caB511	Na12A01-143	2.11
	L4OH-10.1	A10	2.01	3.35	19.64	-0.02	1220caB410/1213bcY498	CB10524-240	40.19
2-Methylpropyl-3C	L2MP-3.1	A3	126.51	4.17	8.95	0.45	1213baR473/1220caB126	SR030-224	15.49
	L2MP-3.2	A3	151.11	5.83	13.04	0.56	1207aaG380/1214ddR397	BrMAM1-1	1.31
	L2MP-4.1	A4	79.91	4.17	9.20	0.42	1219bbY442/1219bcB521	SR057-288	12.49
	L2MP-8.1	A8	26.01	4.56	10.94	-0.57	1214acR722/1214bdR603	SR023-166	0.52
Glucobrassicanapin-5C	LGBNAP-3.1	A3	110.11	4.17	8.80	0.36	1220acR700/1220caB332	SR058-303	6.41
	LGBNAP-3.2	A3	126.51	4.04	8.41	0.36	1213baR473/1220caB126	SR030-224	15.49
	LGBNAP-3.3	A3	151.11	6.22	13.52	0.47	BrMAM1-1/1214ddR397	BrMAM1-1	0.02
	LGBNAP-4.1	A4	79.91	4.01	8.51	0.36	1219bbY442/1219bcB521	SR057-288	12.49
TGSL	LTGSL-3.1	A3	166.31	4.17	13.29	-1.45	1214aaG412/1214ddG371	SR016-316	1.21
	LTGSL-3.2	A3	172.41	3.96	16.31	-1.54	1207acR609/1214caY137	SR016-316	7.31
	LTGSL-8.1	A8	101.11	2.79	8.14	1.08	1207aaG545/SR011-374	SR011-374	1.08

R^2- Percentage phenotypic variance explained by QTL, LG- linkage group

Table 3.6 Summary of QTL identified for average seed glucosinolates from F_6 and F_7 RIL populations in *B. rapa*.

GSL Name	QTL	LG	Position (cM)	LOD	R^2 (%)	Additive effect	Flanked markers	Closest SNP/SCAR/SSR	Distance (cM)
Progoitrin-4C	SPRO-3.1	A3	68.81	5.17	15.68	1.12	Na14E02-126/Na12E02-118	Na14E02-126	0.01
	SPRO-3.2	A3	172.01	4.65	14.60	1.00	1213adY374/1207acR609	SR016-316	7.21
	SPRO-7.1	A7	124.81	4.95	15.01	-1.13	1212acY489/1219acG364	SR128-349	18.79
Glucoalyssin-5C	SALY-3.1	A3	119.41	9.91	23.67	-0.50	1220adB109/1207adR332	SR058-303	14.71
	SALY-3.2	A3	151.11	13.38	35.11	0.63	1207aaG380/1214ddR397	BrMAM1-1	1.31
	SALY-9.1	A9	205.01	5.07	10.69	-0.31	Ol10D08-228/SR047-229	Ol0D08-228	0.02
Gluconapin-4C	SGNAP-3.1	A3	137.91	5.60	17.06	-5.85	1214bcY265/SR030-224	SR030-224	4.09
	SGNAP-3.2	A3	150.81	9.26	26.10	-7.25	1220adB415/1207aaG380	BrMAM1-1	0.02
	SGNAP-3.3	A3	153.71	6.70	19.87	6.20	BrMAM1-1/SNP_tt6-211	BrMAM1-1	0.05
4OH-Indole	S4OH-1.1	A1	56.11	4.91	14.37	-0.22	1220caY609/1219abB164	SR027-372	5.29
	S4OH-4.1	A4	42.01	4.31	12.72	-0.29	1214bdB435/1213bcG243	SR054-234	8.69
	S4OH-9.1	A9	233.81	5.78	17.45	-0.24	1212abG710/1207adG287	SR11-257	3.58
2-Methylpropyl-3C	S2MP-2.1	A2	38.31	3.26	8.98	-0.12	1207adY407/1214bdY601	ALKSSR1-338	3.60
	S2MP-3.1	A3	203.11	4.29	12.34	0.14	1214baB725/1219bdB638	BRAS050-150	1.01
	S2MP-9.1	A9	159.51	3.37	9.47	0.13	1220cdG298/1214cbY277	Na12A02-300	1.25
Glucobrassicanapin-5C	SGBNAP-3.1	A3	137.91	8.38	25.48	3.61	1214bcY265/SR030-224	SR030-224	4.09
	SGBNAP-3.2	A3	150.41	25.89	54.25	5.26	BrMAM1-1/1220adB415	BrMAM1-1	0.01
TGSL	STGSL-3.1	A3	81.91	2.40	6.49	3.80	1214bdB123/1214daB391	Na12E02-146	4.29
	STGSL-5.1	A5	52.61	2.70	8.88	4.44	1219bcY306/1212aaB157	SNP_RISA006	8.42
	STGSL-7.1	A7	81.31	3.80	14.05	7.17	CB10278-243/SR057-198	CB10278-243	0.01
	STGSL-8.1	A8	139.81	2.55	8.17	4.31	1220bbG231/1219aaG554	SR011-374	37.8
	STGSL-9.1	A9	208.51	3.41	10.91	4.88	CB10373-229/1213baY579	CB10373-229	0.01

R^2- Percentage phenotypic variance explained by QTL, LG- linkage group

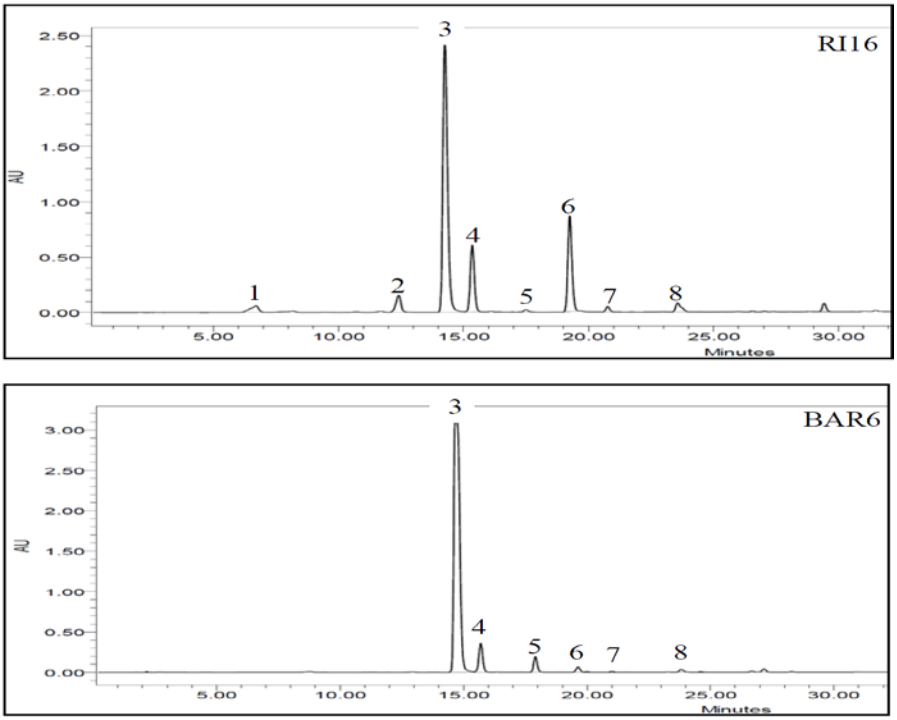


Figure 3.1 HPLC glucosinolate profiles of Chinese cabbage (RI16) and yellow sarson (BAR6) genotypes of *B. rapa*. Individual glucosinolate peaks identified by retention time and quantified by peak area. Peak #1- progoitrin; #2 - glucoalyssin; #3- gluconapin; #4- 4-hydroxyblucobrassicin; #5- 2-methylpropyl; #6- glucobrassicanapin; #7- glucobrassicin; #8- 4-methoxyglucobrassicin.

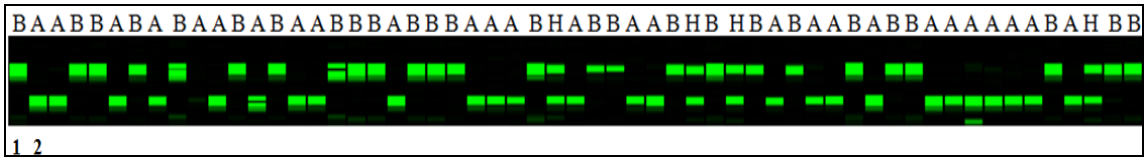


Figure 3.2 *Brassica rapa* RIL population segregating for co-dominant SCAR marker BrMAM1-1 with 5C aliphatic glucosinolates. # 1 and 2 at the bottom of the image indicate the BAR6 and RI16 parents, respectively. Polymorphism detected based on 16 bp deletion in RI16 haplotype. On top of the image, B indicates < 3.0 µmole/g seed ∑5C aliphatic glucosinolates, H indicates 5.0 to 16.0 µmole/g seed ∑5C aliphatic glucosinolates and A indicates >9.0 µmole/g seed ∑5C aliphatic glucosinolates.

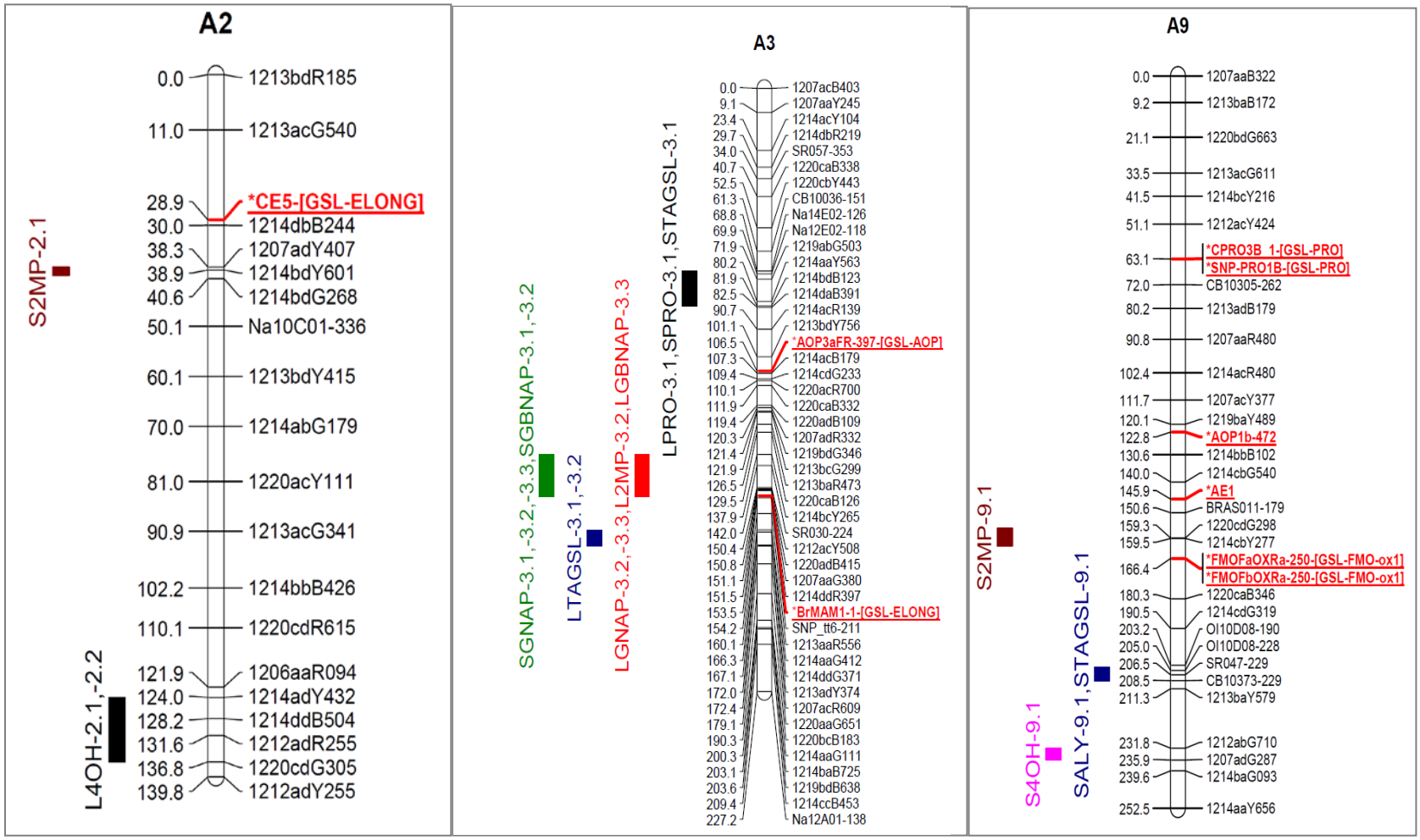


Figure 3.3 Linkage groups A2, A3 and A9 major QTL on the genetic map of *B. rapa* developed by SRAP and SSR markers. Gene specific SCAR and SNP markers are displayed in red color with underline. S- Seed GSLs, L-Leaf GSLs, N-Gluconapin, P-Progoitrin, GBN-Glucobrassicanapin, 2MP-2Methylpropyl, 4OH-4Hydroxyglucobrassicin, TAGSL-Total AGSLs

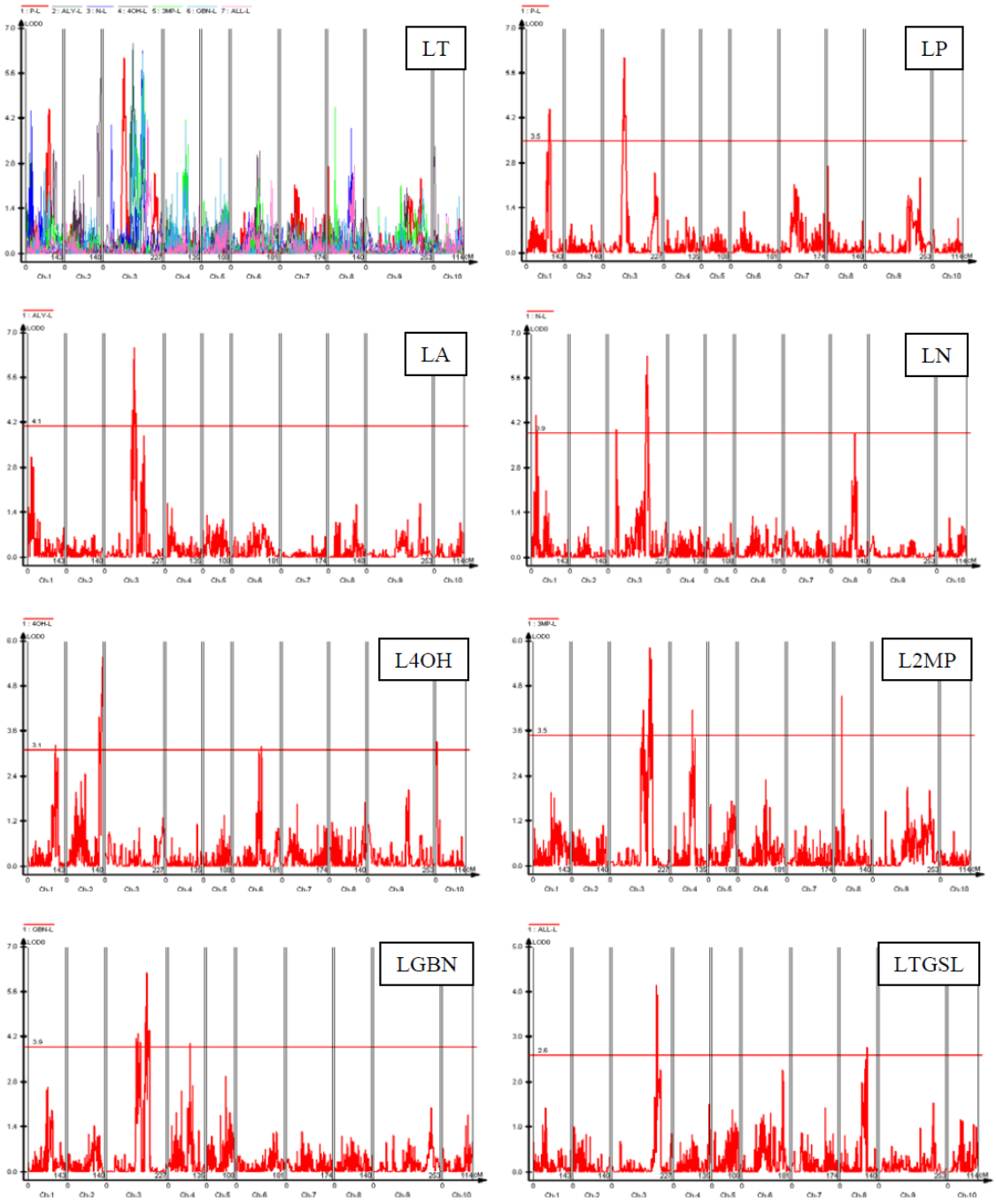


Figure 3.4 QTL map plots for glucosinolates in leaves (CIM method). LT represents all GSL components, LP for progoitrin (P), LA for glucoalyssin (ALY), LN for gluconapin (N), L4OH for indole (4OH), L2MP for 2-methylpropyl (2MP), LGBN for glucobrassicanapin (GBN), and LTGSL for total GSL (LTGSL).

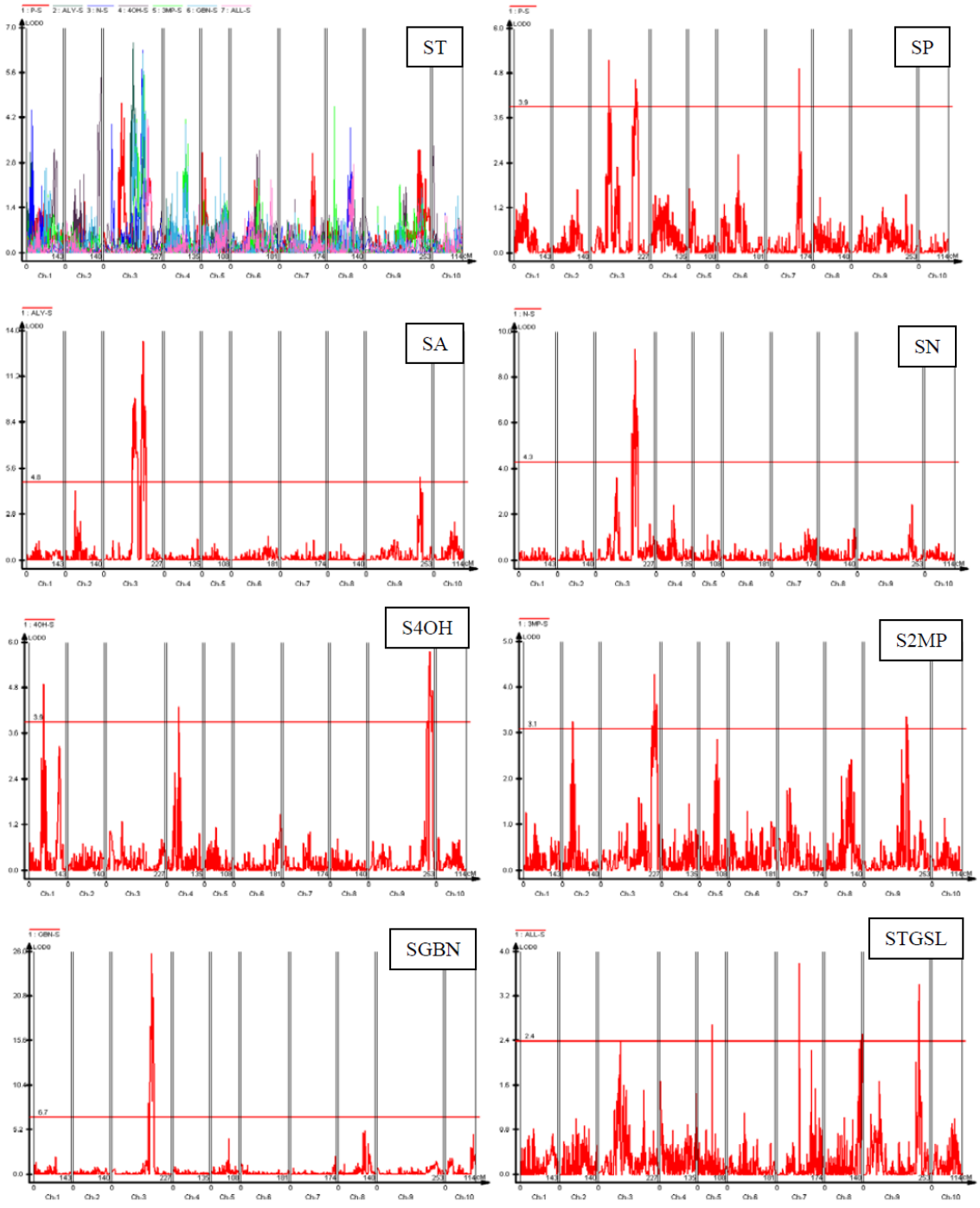


Figure 3.5 QTL map plots for glucosinolates in seeds (CIM method). ST represents all GSL components, SP for progoitrin (P), SA for glucoalyssin (ALY), SN for gluconapin (N), S4OH for indole (4OH), S2MP for 2-methylpropyl (2MP), SGBN for glucobrassicanapin (GBN), and STGSL for total GSL (STGSL).

4. IDENTIFICATION OF QTL FOR LEAF AND SEED GLUCOSINOLATES IN *Brassica rapa* L. RECOMBINANT INBRED LINES

Arvind H. Hirani, Jianfeng Geng, Carla D. Zelmer, Peter B. E. McVetty, Fouad Daayf and Genyi Li

Department of Plant Science, University of Manitoba

Winnipeg, Manitoba Canada R3T 2N2

In this chapter, Dr. Jainfeng Geng is the second author for this study. Dr. Geng developed recombinant inbred population used in this QTL mapping study.

4. Identification of QTL for leaf and seed glucosinolates in *Brassica rapa* L. recombinant inbred lines

4.1. Abstract

Glucosinolate types and contents are highly variable in *Brassica* species, and are regulated by a set of quantitative trait loci (QTL). Glucosinolate production is alos affected by different environment stimuli in plant vegetative and reproductive organs. In this study, QTL mapping was performed for leaf and seed glucosinolates in a *B. rapa* recombinant inbred line (RIL) population derived from a cross between a yellow sarson genotype (BAR6) and a canola quality DH line (USU9). The RILs segregated for glucoerucin, gluconapin and progoitrin as major 4C aliphatic glucosinolates, together with the indole glucosinolate 4-hydoxyglucobrassicin. Sequence related amplified polymorphism (SRAP) was used to develop a genetic map for QTL mapping. Four gene specific loci of *GSL-ELONG* and *GSL-PRO* were integrated on the genetic map to tag major glucosinolate biosynthesis genes. Four QTL for seed glucoerucin and three for seed gluconapin were identified in greenhouse and field environment conditions, which explained phenotypic variance of 8 to 49%. Identification of at least three QTL for 4C glucosinolates suggested that three *GSL-ELONG* loci with epistatic interaction might be involved in 4C side chain aliphatic glucosinolate biosynthesis pathway in the A-genome. Analysis of genetic correlation of 4C glucosinolates suggested that glucoerucin is the initial precursor for the remaining 4C glucosinolates. However, functional loci in the biosynthetic pathway of aliphatic glucosinolates determine concentrations of individual glucosinolates in leaves and seeds.

4.2. Introduction

The *Brassica* genus contains economically important crop species, and produces one of the most abundant plant secondary metabolites, glucosinolates. Glucosinolates are β-thioglucoside-N-hydroxysulfates, with various side chains (R) and sulfur bound to a β-D-glucopyranose moiety. Breakdown products of glucosinolates have a wide range of beneficial and harmful biological effects on flora and fauna. In the 1830s, sinigrin and sinalbin were first isolated from black and white mustard seeds, respectively. Since then, more than 120 glucosinolates have been identified which possess different chemical and functional properties (Fahey *et al.* 2002).

Physical damage to cellular integrity by biotic or abiotic stresses activates myrosinases, which cause immediate hydrolysis of glucosinolates into a variety of bioactive compounds, including isothiocyanates, thiocyanates, nitriles, oxazolidine-2-thiones and epithioalkanes. Some breakdown products have biopesticidal, biofungicidal or bionematocidal properties, which are useful to agriculture. Sulforaphane, the derivative product of glucoraphanin, possesses cancer chemo-protective properties. These breakdown products have enhanced research interest in *Brassica* crop species, especially in glucosinolate biochemistry and genetic manipulation for efficient production of favoured glucosinolates.

Glucosinolates are grouped into three classes based on their amino acids precursors. Aliphatic glucosinolates are mostly derived from methionine with small portions from leucine, isoleucine, alanine and valine (Halkier and Gershenzon 2006). Indole glucosinolates are produced from tryptophan, while aromatic glucosinolates are

derived from phenylalanine and tyrosine.

In aliphatic glucosinolates, methionine is extended by the addition of one or more methylene moieties to form different R group side chains, resulting in the production of up to 8C aliphatic glucosinolates. These chain elongation reactions are controlled by the *GSL-ELONG* gene both in *Arabidopsis* and *Brassica* species. The R groups of glucosinolates are subjected to a wide variety of secondary modifications by different enzymes. The side chain modification reactions are regulated by several sets of genes, mainly, *CytP450, GSL-FMO*$_{OX1-5}$, *GSL-ALK/AOP* or *GSL-OH*, to produce methylthioalkyl (glucoerucin), methylsulphinylalkyl (glucoraphanin), alkenyl (gluconapin) or hydroxyalkenyl (progoitrin) glucosinolates, respectively (Kliebenstein *et al.* 2001c, Chen *et al*. 2003, Li and Quiros 2003, Mithen *et al.* 2003, Li *et al.* 2008).

For R group side chain elongations and modifications, QTL mapping, followed by gene identification and functional characterization, are well described by gene specific mutant studies in *Arabidopsis* (Compos de Quiros *et al.* 2000, Kliebenstein *et al.* 2001a, Kliebenstein *et al.* 2001c, Textor *et al.* 2007, Li *et al.* 2008). In glucosinolate biosynthetic pathway, *Bo-GSL-ELONG* and *Bo-GSL-PRO* for side chain elongations and *Bo-GSL-ALK* for side chain modifications in *B. oleracea* were characterized using *Arabidopsis* sequences (Li and Quirose 2002, 2003, Gao *et al.* 2006). Limited genetic information for glucosinolate biosynthesis, however, is available for *B. rapa*, a progenitor of two commercially important amphidiploid species, *B. napus* and *B. juncea*. The A-genome holder *B. rapa* is a valuable species in the genus *Brassica*. Recently, its genome has been sequenced and is now used to understand mechanisms of polyploidy and the functional nature of duplicated or triplicated genomes (Mun *et al.* 2009). In addition, *B. rapa* has a

globally large contribution to both vegetable oil production and to vegetable crops.

Several QTL mapping studies were reported for seed glucosinolates in two amphidiploid species, *B. napus* and *B. juncea* using different mapping populations. Individual studies detected various numbers of QTL (Magrath *et al.* 1994, Toroser *et al.* 1995, Uzunova *et al.* 1995, Cheung *et al.* 1998, Howell *et al.* 2003, Mahmood *et al.* 2003, Ramchiary *et al.* 2007, Bisht *et al.* 2009). Lou *et al.* (2008) reported the lone QTL mapping study in *B. rapa* for glucosinolates in leaves and identified sixteen QTL for aliphatic glucosinolates and their ratios, three QTL for total indole glucosinolates and three QTL for total aromatic glucosinolates using two DH populations.

In this study, using a RIL population of *B. rapa* segregating for 4C aliphatic glucosinolates (glucoerucin, gluconapin and progoitrin), a SRAP marker based genetic map was constructed. QTL mapping was performed with composite interval mapping for leaf and seed gluconapin, glucoerucin, progoitrin as aliphatic glucosinolates and 4-hydoxyglucobrassicin as indole glucosinolate. Additionally, the impact of environmental conditions on glucosinolate biosynthesis and accumulation was studied in greenhouse and the field environments. Genetic correlation of individual 4C aliphatic glucosinolates was estimated and discussed in the context of glucosinolate biosynthesis.

4.3. Materials and methods

Plant materials

A yellow seeded, non-canola quality, self-compatible (high erucic acid and high total glucosinolate content) pure line of *B. rapa* from Bangladesh, yellow sarson (BAR6),

was used as the female parent in a cross with a light-brown-seeded, canola quality, self-incompatible (low erucic acid and moderate total glucosinolate content) DH line of *B. rapa* (USU9) developed at the University of Manitoba. The F_1 and subsequent generations were selfed to get homozygous RI lines. The F_7 generation of the RILs was used for genotypic evaluation and genetic map construction. The F_7 and F_8 generations of the RILs grown in greenhouse and F_7 generation of the RILs grown in the field at the University of Manitoba were used for phenotypic evaluation in this study. A total of 400 RILs were developed and 93 RILs randomly selected for this experiment, along with the two parental lines.

For greenhouse planting, the RILs and the two parental lines were seeded in a 96 cell flat containing Sunshine LA3 soil mixture. After 3 weeks, seedlings were transplanted into 15 cm pots containing a 2:2:1 mixture of soil, sand and peat. Plants were grown in a photoperiod of 16 hr light/8 hr dark in greenhouses at the University of Manitoba. Plants were uniformly fertilized biweekly with 20-20-20 (N-P-K) fertilizer at a concentration of 3.3 ml/l water.

DNA isolation and polymorphism detections by SRAP markers

Leaf tissues were collected from 93 RILs (F_7 generation) and the two parental lines and DNA extraction was performed using the CTAB (cety-trimethyl-ammonium bromide) method with some modifications as described by Li and Quiros (2002). Approximately 1.0 g of tissue was crushed in liquid nitrogen for genomic DNA extraction. Ground tissue was mixed with 8.0 ml 2x CTAB buffer (2% CTAB, 20 mM EDTA, 100 mM Tris, 1.4 M NaCl, pH-8.0) into 50 ml screw cap tubes and incubated at

65°C for 1.5 hr. After adding 8.0 ml chloroform and vigorously mixing, centrifugation was performed at 4600 rpm for 10 min. to remove plant proteins and debris. Genomic DNA was precipitated from 7.0 ml supernatant using a 0.5 volume of isopropanol with centrifugation at 4000 rpm for 1 min. DNA was washed with 70% ethanol and then the air dried pellet was dissolved in 1.5 ml dH_2O. For genotyping, five times diluted DNA was used as template to perform SRAP PCR using four color labelled forward primers (VIC-green, FAM-blue, PET-red and NET-yellow) and unlabelled reverse primers. The SRAP PCR was performed according to Li and Quiros (2001). Each 10 µl PCR reaction consisted of 7.0 µl dH_2O, 1.0 µl 10x PCR buffer, 0.3 µl 50 mM $MgCl_2$, 0.15 µl 25 mM dNTP, 0.1 µl Taq polymerases, 0.15 µl of 10 µM forward and reverse primers with 2.5 µl template DNA. The four color PCR products were pooled and DNA fragments were separated on ABI 3100 Genetic Analyzer (Applied Biosystems, CA, USA). The raw data were analyzed using GenScan software and SRAP polymorphic loci were scored using Genographer software (http://hordeum.oscs.montana.edu/genographer/).

Genetic map construction and integration of SCAR markers for glucosinolate genes

The 93 RILs and two parental lines were genotyped with 82 SRAP primer combinations. All recombination frequencies were converted into map distances using the Kosambi's mapping function. The marker segregation analysis and linkage map construction was performed using the JoinMap® 3.0 (Van Ooijen and Voorrips 2001) with LOD values between 5 and 10 to assemble 10 linkage groups.

For the mapping of glucosinolate biosynthesis genes with gene-specific loci (*GSL-ELONG* and *GSL-PRO*) in *B. rapa* (A-genome), sequence alignment was

performed with BAC end sequences from the *Brassica* genome project (http://brassica.bbsrc.ac.uk/) to design primers within genes or flanking regions. Two BAC (KBrB010E08F and KBrH121C04F) and two EST (BR043724, BR003821) sequences having high sequence similarity to those of known *GSL-ELONG* genes in *B. oleracea* and *A. thaliana* were identified. Sequencing was performed with PCR products of different primer pairs to find polymorphisms between the two parents for all possible *GSL-ELONG* loci in the A-genome. Sequencing data were analyzed and new primers were designed based on the polymorphisms. Three SCAR markers AEF3AER3-610, BrMAM1-1FRa-474 and CEF5CER5-760 were developed and integrated into linkage map for *GSL-ELONG* gene. Similarly, for the *GSL-PRO* gene, initial primers were designed based on BAC and EST sequence comparative analysis followed by sequencing PCR products to design new primers (Table 4.1). One SCAR marker, CPF4CPR2a-474, was developed for the *GSL-PRO* gene. Specific PCR conditions were used: 35 cycles with 94°C, 1 min; 55-57°C, 1 min; 72°C, 1.5 min. Three SCAR markers, AEF3AER3-610, CEF5CER5-760 and CPF4CPR2a-474 were screened on 1.2% agarose gels. A labelled M13 (CACGACGTTGTAAAACGAC) sequence was added to the 5' end of forward the primer of SCAR marker BrMAM1-1FRa-474 to score it in the ABI3100 Genetic Analyzer. Differentiation between alleles of both the parents wasa 16 bp deletion.

Comparative analysis of SRAP markers for linkage group anchoring

Comparative analysis was performed using the common SRAP primers pairs that were used in three different mapping populations to anchor linkage groups to chromosomes A1 to A10 of the genetic maps. The SRAP markers with equal (±1bp) size

bands from the same primer combinations were used as reference markers in the chromosome assignment. At least seven SRAP markers common to all populations were used to assign the linkage groups to chromosomes A1 to A10 following the recommendations of the MBGP Steering Committee (http://www.brassica.info/resource/maps/lg-assignments.php).

Leaf and seed glucosinolate extraction

Leaf and seed glucosinolates were extracted from the F_7 and F_8 generations of RILs grown in greenhouse conditions and the F_7 generation in the field. Glucosinolate extraction was performed with 200 mg air dried seeds harvested from two greenhouse grown RIL generations and a field grown RIL generation and 250 mg fresh young leaf tissues from 8 week old plants grown in greenhouse. Total seed glucosinolate isolation, purification and desulfation were performed according to Kliebenstein *et al.* (2001b) with a few modifications. Seeds were ground in liquid nitrogen, to which 2 ml boiling 70% methanol was added, and the contents were incubated at 75°C for 10 min. Samples were processed using sephadex/sulfutase desulfation reactions. Total desulfoglucosinolates were eluted into 400 µl dH_2O: methanol (70%) (1:1 v/v).

Identification and quantification of glucosinolates by LC-MS and HPLC

Seed glucosinolates from both parents were analyzed by LC-MS in order to confirm the identity of individual glucosinolates. Desulfoglucosinolates were analyzed on a Varian 212-LC binary gradient LC/MS pump equipped with Polaris^{TM} C18 A column (100mm X 2.0mm, 3 µM) (Varian Inc., CA, USA). The binary mobile phase gradient system consisted of dH_2O with 0.1% formic acid as solvent A, and acetonitrile with 0.1%

formic acid as solvent B, with a flow rate of 300 µl/min. The non-polar and polar compounds were separated using the following gradient elution program: 0.01 min 98:2 A/B, 8 min 93:7 A/B, 12 min 85:15 A/B, 30 min 45:55 A/B, 35 min 8:92 A/B, 40 min 8:92 A/B, 45 min 98.5:1.5 A/B, 48 min 98.5:1.5 A/B, and 52 min 100:0 A/B. Calibration of the system was performed by running the above mentioned mobile phase for 15 min before injecting samples. Twenty µl of desulfoglucosinolate sample was injected into the column by an HTS Pal autosampler (Varian Inc, CA, USA). The liquid flow from LC was diverted to the Varian 500 mass spectrometer LC Ion Trap (Varian Inc, CA, USA) equipped with electro spray ionization (ESI) as an interface configured for positive ionization mode. The MSD detector and interface parameters were set up as follows: needle voltage 5000 V, positive polarity capillary voltage 80 V, electron multiplier voltage 1325 V, trap damping gas flow rate 0.8 ml/min, infusion flow rate 20.0 µl/min, purge flow rate 200 µl/min, nebulizer gas pressure 50 psi, drying gas pressure 30 psi, drying gas temperature 350°C; scanning rage 110-700 m/z; scanning time 3.34 sec/scan. Nitrogen was used as the nebulizing and drying gas**.** The raw data were processed and analyzed with the Varian MS workstation software (Varian Inc, CA, USA).

Individual glucosinolate peaks identified by LC-MS was subsequently quantified by HPLC in leaves and seeds. For phenotypic evaluation of the 93 RILs and the two parental lines, desulfoglucosinolates quantification was performed. A 50 µl samples were quantified by injecting into a 5 µm column (LichroCART® 250-4 RP18, Fisher Scientific, Ottawa, Canada) coupled with a guard column (LichroCART® 250-4 RP18, 4x4 mm Fisher Scientific, Ottawa, Canada) of the Alliance® reverse phase HPLC (Waters 2695) equipped with a photodiode array detector (Waters 996, Waters, MA,

USA). Desulfoglucosinolate detection was performed by the photodiode array detector with UV light set to 229 nm. The HPLC grade methanol (A) and dH_2O (B) were used as the mobile phase gradient at a flow rate of 1 ml/min. The gradient was programmed as follows: 8 min 7:93 A/B (v/v), 4 min 15:85 A/B (v/v), 18 min 55:45 A/B (v/v), 5 min 92:8 A/B (v/v), 5 min 92:8 A/B (v/v), 5 min 1.5:98.5 A/B (v/v), 3 min 1.5:98.5 A/B (v/v) and a final 4 min 0:100 A/B (v/v) for a total of 52 min per run. Identification of individual desulfoglucosinolates was based on retention time and the UV spectra of the absorption wavelength. Abundance of each desulfoglucosinolate was converted into micromoles per gram of fresh wt. of leaves or dry wt. of seeds using the corresponding response factors. These were determined from either the purified or synthetic standards for each glucosinolate (Brown *et al.* 2003). Quantification of individual glucosinolates was carried out for glucoerucin (GERU), gluconapin (GNAP), progoitrin (PRO) (aliphatic glucosinolates) and 4-hydroxyglucobrassicin (4OH), glucobrassicin (GB) and 4-methoxyglucobrassicin (4MGB) (indole glucosinolates). Additionally, the sum of 4C ($\sum$4C), the sum of 5C ($\sum$5C), total aliphatic, total indole and total glucosinolates were estimated for leaves and seeds.

Data analysis and frequency distribution

Analysis of variance and frequency distribution for seed glucosinolate content was performed on data from two replicates (F_7 and F_8 generation of RILs) of greenhouse grown material and the F_7 generation of RILs of field grown material. Analysis of variance for seed glucoerucin, gluconapin, progoitrin, the sum of 4C glucosinolates, the sum of 5C glucosinolates, total aliphatic glucosinolates, 4-hydroxyglucobrassicin, total indole glucosinolates and total seed glucosinolates was performed. The analysis for the

randomized design used the GLM model of the SAS program (SAS Institute Inc, 1998). The mean values of the two replicates of greenhouse and the replicate values from the field were used to determine the frequency distribution of seed glucoerucin, gluconapin, progoitrin, the sum of 4C, the sum of 5C and total aliphatic glucosinolates in the RIL population. Individual values from the three replicates obtained from two different environments were used to compute Pearson correlation coefficients by SAS PROC CORR (SAS Institute Inc, 1998) for these seed glucosinolates.

QTL mapping for glucosinolates

Average data of leaf and seed glucosinolates from the two replicates in greenhouse and the independent data of seed glucosinolate from the replicate in the field were used for QTL mapping. QTL mapping was performed with Windows QTL cartographer v 2.5 (Wang *et al.* 2010) using composite interval mapping. Detection of QTL were performed at 2 cM intervals using a 10 cM window size and five background cofactors, parameters were determined by backward regression analysis. Detection of QTL was performed with genome-wide threshold values estimated from 1000 permutations of trait data across all genetic intervals (Doerge and Churchill 1996). Leaf and seed glucoerucin (GERU), gluconapin (GNAP), progoitrin (PRO), total aliphatic glucosinolates (TAGSL), 4-hydroxyglucobrassicin (4OH), total indole glucosinolates (TIGSL) and total glucosinolates (TGSL) were analyzed in this QTL mapping study.

4.4. Results

Evaluation of glucosinolates

The parents BAR6 and USU9 of the RIL population of *B. rapa* produced different types of glucosinolates. The line BAR6 produced predominantly gluconapin (~90%) with trace amount of 2-methylthiopropyl, glucoerucin, glucobrassicanapin and 4-hydroxyglucobrassicin. On the other hand, the DH line USU9 produced predominantly glucoerucin with trace amount of progoitrin, glucoraphanin, gluconapin, glucobrassicanapin and 4-hydroxyglucorassicin (Figure 4.1). Identification of individual glucosinolates in both the parents was confirmed with LC-MS and their mass (*m/z*) of daughter ion fragments corresponding to molecular weights of known desulfoglucosinolates. Each desulfoglucosinolate spectrum contained any three of four diagnostic ions corresponding to $[M+C_6H_{10}O_5]^+$, $[M+H]^+$, $[M+Na]^+$ and $[M+K]^+$ with variable intensities in both the parents. Gluconapin contained fragment ions at *m/z* 132 for $[M+C_6H_{10}O_5]^+$, *m/z* 294.1 (for $[M+H]^+$) and *m/z* 316 (for $[M+Na]^+$) in the yellow sarson BAR6 and DH line USU9. Similarly, glucoerucin contained fragment ions at *m/z* 180.3 (for $[M+C_6H_{10}O_5]^+$), *m/z* 364.2 (for $[M+Na]^+$) and *m/z* 380.3 (for $[M+K]^+$) in the yellow sarson BAR6, whereas in DH line USU9 an additional ion at *m/z* 342.1 (for $[M+H]^+$) was also observed (Figure 4.1, Appendix iii). Desulfoglucosinolate daughter ions (*m/z*) were also compared with previously reported studies (Griffiths *et al.* 2000, Zimmermann et al. 2007, Kim *et al.* 2010).

Due to qualitative and quantitative glucosinolate variations between the two parents, transgressive segregation for various glucosinolates was observed in the RILs. The transgressive segregation for aliphatic glucosinolates might be due to increased or

decreased allelic activities at different loci and their interactions. Seed glucoerucin content in the segregating RILs ranged from 0 to 51.96 µmole/g seed and was in two discrete classes corresponding to the high and low parental phenotypes. Forty five lines had less than 3.0 µmole/g seed glucoerucin, whereas the other fifty lines had greater than 10 µmole/g seed glucoerucin. Interestingly, the ratio of the two classes was not significantly different from 1:1 ratio (χ^2=0.44, df=1, P>0.1), indicating that a major locus determined the abundance of glucoerucin. Variation in the high glucoerucin group might be due to the small and cumulative effect of other loci and their interactions (Figure 4.2 A). Seed gluconapin content displayed continuous distribution in the RILs (Figure 4.1 B), and ranged from 0.73 to 88.38 µmole/g seed. A twenty two fold and one fold increase in gluconapin was observed in the RILs compared to the low parent USU9 (4 µmole/g seed) and high parent BAR6 (72 µmole/g seed), respectively. Seed progoitrin content ranged from 0 to 60.0 µmole/g seed in the segregating RILs. In addition to this, the sum of 4C (∑4C), sum of 5C (∑5C) and total seed aliphatic glucosinolates (TAGSL) displayed normal frequency distributions in the segregating RILs (Figure 4.2D, E, and F). Glucosinolates detected in the RILs were the same those observed in the parental lines. No new glucosinolates were produced by the RILs.

Genetic correlation study in 4C aliphatic glucosinolates

In this study, genetic correlation between glucoerucin, gluconapin, progoitrin, total aliphatic glucosinolates, 4-hydroxyglucobrassicin and total indole glucosinolates were analyzed in the segregating RILs. Glucoerucin had negative correlation with gluconapin (-0.83), progoitrin (-0.19) and total aliphatic glucosinolates (-0.28, Table 4.3) in seeds. Glucoerucin, however, had slightly positive correlation with 4-

hydroxyglucobrassicin (0.37) which also resulted in a positive correlation with total indole glucosinolates (0.43) in the RILs. Gluconapin had a positive correlation with total aliphatic glucosinolate (0.61), but a negative correlation with 4-hydroxyglucobrassicin (-0.56) and total indole glucosinolates (-0.61) in seeds.

Seed glucosinolate analyses for greenhouse and field environments

Glucosinolates are thought to be highly influenced by environmental factors, such as nutrient availability (especially sulfur and nitrogen), light, moisture, biotic and abiotic stresses. Analysis of variance (ANOVA) was performed for seed glucosinolate components of the RILs grown in two different environments, in the greenhouse and the field, to determine effects of phenotypic plasticity on seed glucosinolate contents. In the analysis of variance study, there was no significant difference for glucoerucin (GERU), gluconapin (GNAP), sum of 5C (Σ5C) and total aliphatic glucosinolates (TAGSL) (Table 4.2). However, a significant difference was observed in seed progoitrin between greenhouse and field replicates. This difference was also therefore found for the sum of 4C (Σ4C) aliphatic glucosinolates. Indole glucosinolates displayed significant differences between and within environments, and indicating that the production of indole glucosinolates are highly influenced by the environment.

Linkage map construction and SCAR markers tagging major glucosinolates loci

Eighty two SRAP primer combinations were used for identification of polymorphic loci. On average 3.5 useful SRAP markers were amplified per primer combination for the RIL population of *B. rapa*. A total 288 SRAP and 4 gene specific SCAR markers were used for construction of linkage maps, and the comprehensive

genetic map covered an 842.34 cM distance of the *B. rapa* genome spanning 10 linkage groups. The sizes of the linkage groups were 55.28 to 151.63 cM, where A1 and A4 were shortest linkage groups with 55.28 cM and 62.17 cM distance, respectively. On the genetic map, the average interval between two markers was 2.93 cM with a minimum of 0.015 cM and a maximum of 23.84 cM distance between adjacent markers (Figure 4.3). Individual linkage groups were assigned to chromosome number A1 to A10 based on comparative analysis of equal size SRAP markers produced from the use of the same primer combinations in the construction of two different genetic maps. Each of these maps was constructed from different mapping populations (e.g. Chapter 3, Sun *et al.* 2007). Additionally, three SCAR markers specific to *Br-GSL-ELONG* gene and a SCAR marker specific to *Br-GSL-PRO* gene were mapped on different linkage groups (Table 4.1).

QTL analysis of leaf and seed glucosinolates

QTL mapping was performed for the major 4C aliphatic glucosinolates using composite interval mapping (CIM). Three QTL on linkage groups A8, A9 and A10 for leaf glucoerucin which collectively explained 24% phenotypic variance were identified (Table 4.4). Five QTL for leaf gluconapin explaining 8 to 14% phenotypic variance were identified. One QTL for leaf progoitrin controlling 11% total phenotypic variance was detected. For leaf aliphatic and total glucosinolates, one common QTL on A9 controlling about 12% phenotypic variance was identified. Additionally, two and three QTL for leaf 4-hydroxyglucobrassicin and total indole glucosinolate, respectively, were detected which explained phenotypic variance between 11 and 28%.

For seed glucoerucin, five QTL common to both environments were identified on linkage groups A2, A3, A8, A9 and A10, explaining 4 to 45% phenotypic variance. One additional QTL for seed glucoerucin explaining 4.3% phenotypic variance appeared on A2 only in the field environment (Table 4.5, Figure 4.3). Three QTL (GH-SGNAP-3.1, GH-SGNAP-3.2, GH-SGNAP-10.1, F-SGNAP-3.1, F-SGNAP-3.2 and F-SGNAP-10.1) common to both environments were detected for seed gluconapin, which explained phenotypic variance between 7 and 40%. For seed progoitrin, a QTL (GH-SPRO-3.1) on linkage group A3 with 11% explained phenotypic variance was identified in greenhouse environment. In the field environment a single QTL (F-SPRO-10.1) on A10 was found to explain 16% of phenotypic variance. For seed total aliphatic and total glucosinolates, a pair of QTL (GH-TAGSL9.1, GH-TAGSL-9.2, F-TAGSL-9.1, F-TAGSL-9.2, GH-TGSL-9.1, GH-TGSL-9.2, F-TGSL-9.1 and F-TGSL-9.2) common to both environments were identified on linkage group A9, which explained phenotypic variance between 17 and 26%. Additionally, one QTL (GH-S4OH-9.1) in the greenhouse environment for seed 4-hydroxyglucobrassicin was detected which explained 38%of the phenotypic variance. Three QTL (F-S4OH-3.1, F-S4OH-6.1 and F-S4OH-9.1) for seed 4-hydroxyglucobrassicin were found in the field environment to explain 15 to 33% of the phenotypic variance. There were no QTL found in common for seed 4-hydroxyglucobrassicin between the two environments. Similarly, two and six QTL (GH-STIGSL-6.1, GH-STIGSL-6.2, F-STIGSL-3.1, F-STIGSL-3.2, F-STIGSL-6.1, F-STIGSL-6.2, F-STIGSL-8.1 and F-STIGSL-9.1) for seed total indole glucosinolates were identified in greenhouse and field environments, respectively, which explained phenotypic variance between 8 and 14%. There were no QTL found in common for seed

total indole glucosinolates between the two environments. QTL data for leaf and seed QTL are given in Table 4.4, table 4.5, and Figure 4.4, respectively.

4.5. Discussion

Previous studies show that many accessions of *B. rapa* predominantly biosynthesize the glucosinolates, gluconapin and glucobrassicanapin (He *et al.* 2000, Padilla *et al.* 2007). In this study, however, glucoerucin was biosynthesized as a major 4C aliphatic glucosinolate in the parental DH line USU9. The RILs segregated for 4C aliphatic glucosinolates, glucoerucin, gluconapin and progoitrin relative to the parental phenotypes. There were no new glucosinolate types identified in the segregating RILs, suggesting that *GSL-ELONG* locus corresponding to 5C aliphatic glucosinolates might be non-functional in both the parents.

Biosynthesis of aliphatic glucosinolates are regulated by the *GSL-ELONG* gene for side chain elongations (Li and Quiros 2002) with at least four loci (Bisht *et al.* 2009) in the A genome. The R group side chain modifications of aliphatic glucosinolates are controlled by the genes *GSL-FMO*$_{OX1}$ and *GSL-AOP/ALK* (Li and Quiros 2003, Mithen *et al.* 2003). The segregating RILs displayed transgressive segregation for individual seed glucosinolates. Transgressive segregation for seed progoitrin suggested that at least two highly active loci controlled progoitrin biosynthesis (Figure 4.2C). Glucoerucin is known to be the initial precursor in the biosynthesis of progoitrin. Comparision between the QTL of RIL in which produce progoitrin with those that do not suggests that a locus on the linkage group A10 corresponds to the major QTL for progoitrin (Figure 4.2C). In addition, the normal frequency distribution in the RILs for sum of 4C ($\sum$4C), sum of 5C

(∑5C) and total seed aliphatic glucosinolates (TAGSL) suggests that several genes might regulate aliphatic glucosinolate biosynthesis in *Brassica* species

The quantitative inheritance of glucosinolates is controlled by multiple genes with allelic variation at several loci, which are also influenced by various biotic and abiotic factors. Therefore, the influences of environmental factors on individual and total seed glucosinolates were studied between greenhouse and the field environments. The results of the analyses of variance performed on the three replicates suggested that glucoerucin, gluconapin, sum of 5C and total aliphatic glucosinolates did not vary significantly due to environment. Progoitrin, however, displayed significant differences among the replicates. Variation in progoitrin contents across two different environments may be related to the nature of the hydroxy compounds, which possesse variable transportation behaviour toward a sink. Another possible reason for the variation is that environmental factors and biotic stresses influence progoitrin hydrolysis activity more than for other aliphatic glucosinolates as progoitrin is biosynthesized downstream to the other 4C aliphatic glucosinolates. Similarly, indole glucosinolates, 4-hydroxyglucobrassicin and total indole glucosinolates were significantly different among the three replicates, possibly due to variable response to herbivory and defence related signalling molecules (Mikkelsen *et al.* 2004, Müller and Sieling 2006, Kim and Jander 2007) (Table 4.2). Several studies reported enhanced accumulation of indole glucosinolates in response to herbivory by root flies, lepidopteran larvae, flea beetles, sawflies and aphids (Griffiths *et al.* 1994, Rostas *et al.* 2002, Traw 2002, Müller and Sieling 2006, Kim and Jander 2007). Studies also demonstrated that defence signalling molecules such as jasmonate induce indole

glucosinolate accumulation under biotic and abiotic stresses in *Brassica* (Kliebenstein *et al.* 2002, Mikkelsen *et al.* 2004).

In the 4C aliphatic glucosinolate biosynthesis branch, significant negative correlation (-0.83) between glucoerucin and gluconapin was observed, which was supported by linked QTL on A3 and A10 with similar but opposite additive effects. This suggests that glucoerucin is an initial precursor for gluconapin in the side chain modification pathway (Table 4.5, Figure 4.4). The QTL on A2 and A8 for glucoerucin might correspond to specific side chain modification reactions, such as mono-oxygenation.

Glucoerucin, however, had a positive correlation with indole glucosinolates. A possible reason for the positive correlation is that the family of cytochrome P450 mono-oxygenases regulate conversion of amino acids to aldoximes for glucosinolates. The methionine based reactions, however, are regulated by *CYP79F1* and *CYP79F2* for aliphatic glucosinolates, and tryptophan based reactions are controlled by *CYP79B2* and *CYP79B3* for indole glucosinolates (Mikkelsen *et al.* 2000, Chen *et al.* 2003). The positive genetic correlation between aliphatic and indole glucosinolates might be due to a cluster of cytochrome P450 loci on the same region of the genome and co-segregated in the RILs. Lou *et al.* (2008) also reported that a major QTL for leaf aliphatic glucosinolates on linkage group A3 was co-localized with a QTL for indole glucosinolates. They also demonstrated a positive correlation between some aliphatic and indole compounds.

In this study, three indistinguishable QTL were detected in two different environments for seed gluconapin on linkage group A3 and A10. A QTL (GH-SGNAP-3.1 and F-SGNAP-3.1) for seed gluconapin in two environments, ~0.72 cM away from the SCAR marker BrMAM1-1FRa-474 specific to the *Br-GSL-ELONG* locus was identified. The same locus also linked with the large effect QTL on linkage group A3 for leaf and seed gluconapin in the SR-RILs (Chapter 3). Similarly, Lou *et al.* (2008) reported two homolog QTL for leaf gluconapin on A3 and A10 in two DH populations of *B. rapa*. Additionally, a QTL for seed progoitrin on A3 in greenhouse environment and a QTL on A10 in the field environment were detected, which explained 10 to 15% phenotypic variance.

Two QTL indentified for seed progoitrin potentially corresponded to the *GSL-AOP* loci in this study. The *GSL-AOP1* and *GSL-AOP3* genes presumably have at least three loci in *B. rapa*, of which two were mapped on linkage group A3 and A9 (Luo *et al.* 2008, Dr. Genyi Li, unpublished data). A QTL on A10 for leaf and seed progoitrin suggested the existence of an additional major locus. The presence of multiple *GSL-AOP* loci explained absence of glucoraphanin in the RILs and both the parents (Zang *et al.* 2008, Zang *et al.* 2009). There were discrepancies in the appearance of QTL for progoitrin, suggesting that variation in transcript levels for *GSL-AOP* genes in the glucosinolate biosynthesis pathway exists. Kliebenstein (2008) reported significant variation in transcript level, particularly for *GSL-AOP2*, *GSL-AOP3* and *GSL-OH* genes, compared to other glucosinolate biosynthesis genes in seven accessions of *Arabidopsis*. This work suggested that that tandem segmented loci possessed variable activity for glucosinolate biosynthesis.

The present study in *B. rapa* RILs suggested that multiple genetic loci regulate individual 4C aliphatic glucosinolate biosynthesis. The major *GSL-ELONG* locus for side chain elongation of 4C aliphatic glucosinolate was identified on linkage group A3 with a co-dominant SCAR marker. The common QTL in two environments can be further analysed for either conversion of SRAP markers to simple PCR based markers. Fine mapping of candidate loci and their functional characterization will advance the understanding of the complexity of 4C aliphatic glucosinolate biosynthesis in *B. rapa*. Additionally, linked SRAP or SCAR markers can be used for marker assisted selection in breeding programs to manipulate individual glucosinolates in *Brassica* species.

Table 4.1 Specific SCAR markers used for tagging glucosinolate biosynthesis genes *Br-GSL-ELONG* and *Br-GSL-PRO*.

Gene names	**Marker names**	**BAC/EST**	**LG**	**Primers (5'-3')**
GSL-ELONG	AEF3AER3-610	B19N3	A10	F- CAGTCAAATTTACCGCCTT R- GGTGGCTTTCGCGGACAC
GSL-ELONG	BrMAM1-1FRa-474	KBrH121C04	A3	F- CACGACGTTGTAAAACGAC R- CTAAGCTCTTCGCATAGCTA
GSL-ELONG	CEF5CER5-760	B19N3	A2	F- CCACGAGCAAACACATATA R- CTTGGTGTGGTGCAGAAGA
GSL-PRO	CPF4CPR2a-474	B21F5	A6	F- TGTCAGAATGATCTTGGGCT R- AGCGCAACTGGAGAAACAA

Table 4.2 Analysis of variance for seed glucosinolates in the RILs and replicates.

GSLs	Source	DF	Mean	SS	MS	CV%	F ratio
SGERU	RILs	92	18.10	83386.88	916.33	19.60	72.88.76***
	Reps	2		20.21	10.10		0.80NS
SGNAP	RILs	92	35.23	192844.57	2096.14	16.00	66.06***
	Reps	2		129.43	64.72		2.04NS
SPRO	RILs	92	6.48	33341.65	362.41	55.67	27.82***
	Reps	2		177.48	88.74		6.81**
S∑4C	RILs	92	60.43	62394.91	678.51	13.31	10.48***
	Reps	2		1684.37	842.183		13.01**
S∑5C	RILs	92	1.31	82.84	0.90	33.32	4.73***
	Reps	2		0.97	0.48		2.55NS
STAGSL	RILs	92	62.63	63619.35	691.51	10.58	15.73***
	Reps	2		340.25	169.62		2.27NS
S4OH	RILs	92	1.76	388.37	4.22	20.21	12.84***
	Reps	2		17.54	9.77		17.32***
STIGSL	RILs	92	2.73	769.23	9.23	25.45	13.52***
	Reps	2		5.89	3.04		5.91**
STGSL	RILs	92	63.45	61725.20	680.38	14.50	9.20***
	Reps	2		1103.28	566.23		7.02**

NS, P>0.01; * P<0.01; ** P<0.001; *** p<0.0001. SGNAP- seed gluconapin, SPRO- seed progoitrin, S∑4C- seed sum of 4C aliphatic glucosinolates, S∑5C- seed sum of 5C aliphatic glucosinolates, STAGSL- seed total aliphatic glucosinolates, S4OH- seed 4-hydroxyglucobrassicin, STIGSL- seed total indole glucosinolates, STGSL- seed total glucosinolates.

Table 4.3 Correlation analysis among seed glucosinolates in *B. rapa*.

	SGERU	**SGNAP**	**SPRO**	**STAGSL**	**S4OH**	**STIGSL**
SGERU	1.00					
SGNAP	-0.83***	1.00				
SPRO	-0.19*	0.04NS	1.00			
STAGSL	-0.28***	0.61***	0.55***	1.00		
S4OH	0.37***	-0.56***	-0.11NS	-0.49***	1.00	
STIGSL	0.43***	-0.61***	-0.11NS	-0.48***	0.98***	1.00

NS, P>0.01; * P<0.01; ** P<0.001; *** p<0.0001; df, 93. SGERU- seed glucoerucin, SGNAP- seed gluconapin, SPRO- seed progoitrin, STAGSL- seed total aliphatic glucosinolates, S4OH- seed 4-hydroxyglucobrassicin, STIGSL- seed total indole glucosinolates.

Table 4.4 QTL identified for glucosinolates in *B. rapa* leaves by CIM.

Trait	QTL	LG	LOD	Flanking Markers	Interval cM	Position cM	R^2 (%)	Additive Effect
LGERU	GH-LGERU8.1	A8	2.60	BUBG23BG62-353/BUEM1PM33-151	4.2	99.41	9.12	-0.59
	GH-LGERU9.1	A9	2.26	BUSA7YY235-329/BUPROPM56-329	2.8	17.14	7.87	0.56
	GH-LGERU10.1	A10	2.31	*AEF3AER3-610/BUODD3PM49-187	0.7	42.43	6.83	-0.52
LGNAP	GH-LGNAP3.1	A3	3.22	BUFADODD13-97/BUSA7TU134-374	8.4	84.30	10.41	2.45
	GH-LGNAP3.2	A3	4.89	BUFADFE44-144/BUSA7YY209-145	6.6	102.20	14.33	2.89
	GH-LGNAP7.1	A7	4.56	BUPROPM63-220/BUPROPM77-494	13.5	61.65	11.87	2.80
	GH-LGNAP8.1	A8	3.65	BUPROPM77-471/BUSA7TU126-200	0.16	111.00	9.71	-2.43
	GH-LGNAP10.1	A10	3.48	BUSA7YY235-204/BUSA7TUBR94-341	3.8	65.31	8.45	-2.22
LPRO	GH-LPRO10.1	A10	2.50	BUPROPM45-208/BUEM1PM33-503	0.33	14.85	11.15	-0.63
LTAGSL	GH-LTAGSL9.1	A9	4.41	BUPROPM56-329/BUSA7TU150-363	10.1	27.22	11.16	-2.97
L4OH	GH-L4OH8.1	A8	6.58	BUPM88BG5-523/BUFADFE29-148	5.8	64.28	28.10	0.07
	GH-L4OH8.2	A8	8.24	BUSA7TUBR60-148/BUFADFE29-186	2.9	68.02	27.66	-0.07
LTIGSL	GH-LTIGSL8.1.	A8	2.87	BUPM88BG5-523/BUFADFE29-148	5.8	64.28	14.46	-0.13
	GH-LTIGSL10.1	A10	4.25	BUEM1ODD25-344/BUSA7YY180-325	0.8	57.01	15.32	-0.14
	GH-LTIGSL10.2	A10	3.08	BUFADPM4-214/BUGA3FE01-325	2.8	58.96	11.23	0.11
LTGSL	GH-LTGSL9.1	A9	4.98	BUPROPM56-329/BUSA7TU150-363	10.1	27.22	12.03	-2.99

LGERU- leaf glucoerucin, LGNAP- leaf gluconapin, LPRO- leaf progoitrin, TAGSL- leaf total aliphatic glucosinolates, L4OH- leaf 4-hydroxyglucobrassicin, LTIGSL- leaf total indole glucosinolates, LTGSL-leaf total glucosinolates.

Table 4.5 QTL identified for glucosinolates in *B. rapa* seeds by CIM.

Trait	QTL	LG	LOD	Flanking Markers	Interval cM	Position cM	R^2 (%)	Additive Effect
SGERU	GH-SGERU2.2	A2	11.32	BUPM88BG45-188/BUFADPM73-353	1.9	40.86	4.43	16.25
	GH-SGERU3.1	A3	11.86	BUPM88CE23-375/BUEM1ODD25-429	3	28.31	44.82	-16.24
	GH-SGERU8.1	A8	14.44	BUFADFE39-131/BUSA7TU134-387	4.7	126.15	5.46	-16.20
	GH-SGERU9.1	A9	9.63	BUFADPM1-127/BUFADFE39-376	4.03	41.18	18.25	-16.25
	GH-SGERU10.1	A10	6.09	BUSA7YY180-325/BUFADFE44-112	1	58.01	17.73	-8.28
	F-SGERU2.1	A2	8.76	BUEM1PM18-298/BUFADFE39-146	2.39	21.67	4.29	-13.85
	F-SGERU2.2	A2	7.92	BUPM88BG45-188/BUFADPM73-353	1.9	40.86	4.24	13.81
	F-SGERU3.1	A3	6.78	BUPM88CE23-375/BUEM1ODD25-429	3	28.31	42.84	-13.86
	F-SGERU8.1	A8	10.32	BUFADFE39-131/BUSA7TU134-387	4.7	126.15	5.82	-14.34
	F-SGERU9.1	A9	3.37	BUFADPM1-127/BUFADFE39-376	4.03	41.18	19.31	-12.75
	F-SGERU10.1	A10	5.15	BUSA7YY180-325/BUFADFE44-112	1	58.01	15.17	-7.23
SGNAP	GH-SGNAP3.1	A3	4.63	*BrMAM1-1FRa-474/BUPROPM78-150	1.7	84.99	20.14	-12.51
	GH-SGNAP3.2	A3	5.29	BUFADODD13-97/BUEM1ODD25-429	7.5	23.80	39.97	21.89
	GH-SGNAP10.1	A10	7.62	BUFADFE44-112/BUFADODD13-384	0.26	58.03	21.17	13.45
	F-SGNAP3.1	A3	4.38	*BrMAM1-1FRa-474/BUPROPM78-150	1.7	84.99	20.53	-12.37
	F-SGNAP3.2	A3	3.16	BUFADODD13-97/BUEM1ODD25-429	7.5	23.80	7.95	7.50
	F-SGNAP10.1	A10	5.68	BUFADFE44-112/BUFADODD13-384	0.26	58.03	18.01	12.04
SPRO	GH-SPRO3.1	A3	3.19	BUPROPM78-171/ BUBG23PM117-436	3.8	43.72	11.48	4.34
	F-SPRO10.1	A10	3.51	BUODD3PM49-204/BUPROPM63-245	2.5	30.89	15.88	-5.08
STAGSL	GH-STAGSL9.1	A9	6.45	BUPROPM56-329/BUSA7TU150-363	10	27.22	22.74	-8.86
	GH-STAGSL9.2	A9	4.63	BUSA7TU150-363/BUSA7TUBR144-247	7.6	38.87	17.25	7.65
	F-STAGSL9.1	A9	6.33	BUPROPM56-329/BUSA7TU150-363	10	27.22	21.20	-8.58
	F-STAGSL9.2	A9	5.16	BUSA7TU150-363/BUSA7TUBR144-247	7.6	33.84	18.56	7.61

Trait	QTL	LG	LOD	Flanking Markers	Interval cM	Position cM	R^2 (%)	Additive Effect
S4OH	GH-S4OH9.1	A9	6.10	BUSA7TU150-363/BUFADFD03-246	5.6	28.22	38.26	1.09
	F-S4OH3.1	A3	4.23	BUBG23PM117-356/BUELONGFE03-507	2.5	67.25	14.65	0.61
	F-S4OH6.1	A6	3.82	BUSA7TUBR60-461/BUPROPM60-144	3.54	6.82	16.31	0.65
	F-S4OH9.1	A9	5.21	BUPROPM56-329/BUSA7TU150-363	10	28.22	33.06	0.97
STIGSL	GH-STIGSL6.1	A6	4.81	BUSA7TUBR60-461/BUPROPM60-144	3.54	6.82	13.54	0.72
	GH-STIGSL6.2	A6	5.17	BUSA7TUBR60-461/BUPROPM60-144	3.54	11.54	12.40	0.69
	F-STIGSL3.1	A3	4.62	BUODD3PM49-238/BUFADFE44-380	4.5	76.61	10.22	0.55
	F-STIGSL3.2	A3	5.69	BUPROPM52-424/BUELONGFE03-507	2.9	69.77	11.15	0.58
	F-STIGSL6.1	A6	3.63	BUPROPM58-270/BUEM1ODD13-198	1.7	4.13	7.79	-0.48
	F-STIGSL6.2	A6	4.79	BUSA7TUBR60-461/BUPROPM60-144	3.54	7.82	11.99	0.59
	F-STIGSL8.1	A8	3.99	BUFADFE01-431/BUSA7YY180-285	1.9	26.62	7.41	-0.47
	F-STIGSL9.1	A9	4.20	BUPROPM56-329/BUSA7TU150-363	10	27.14	8.01	0.55
STGSL	GH-STGSL9.1	A9	5.67	BUPROPM56-329/BUSA7TU150-363	10	27.22	20.81	-8.00
	GH-STGSL9.2	A9	4.49	BUSA7TU150-363/BUSA7TUBR144-247	7.6	28.22	26.05	-8.56
	F-STGSL9.1	A9	5.60	BUPROPM56-329/BUSA7TU150-363	10	27.22	19.67	-7.86
	F-STGSL9.2	A9	4.78	BUSA7TU150-363/BUSA7TUBR144-247	7.6	33.84	17.91	7.11

SGERU- seed glucoerucin, SGNAP- seed gluconapin, SPRO- seed progoitrin, TAGSL- seed total aliphatic glucosinolates, L4OH- seed 4-hydroxyglucobrassicin, LTIGSL- seed total indole glucosinolates, LTGSL- seed total glucosinolates.

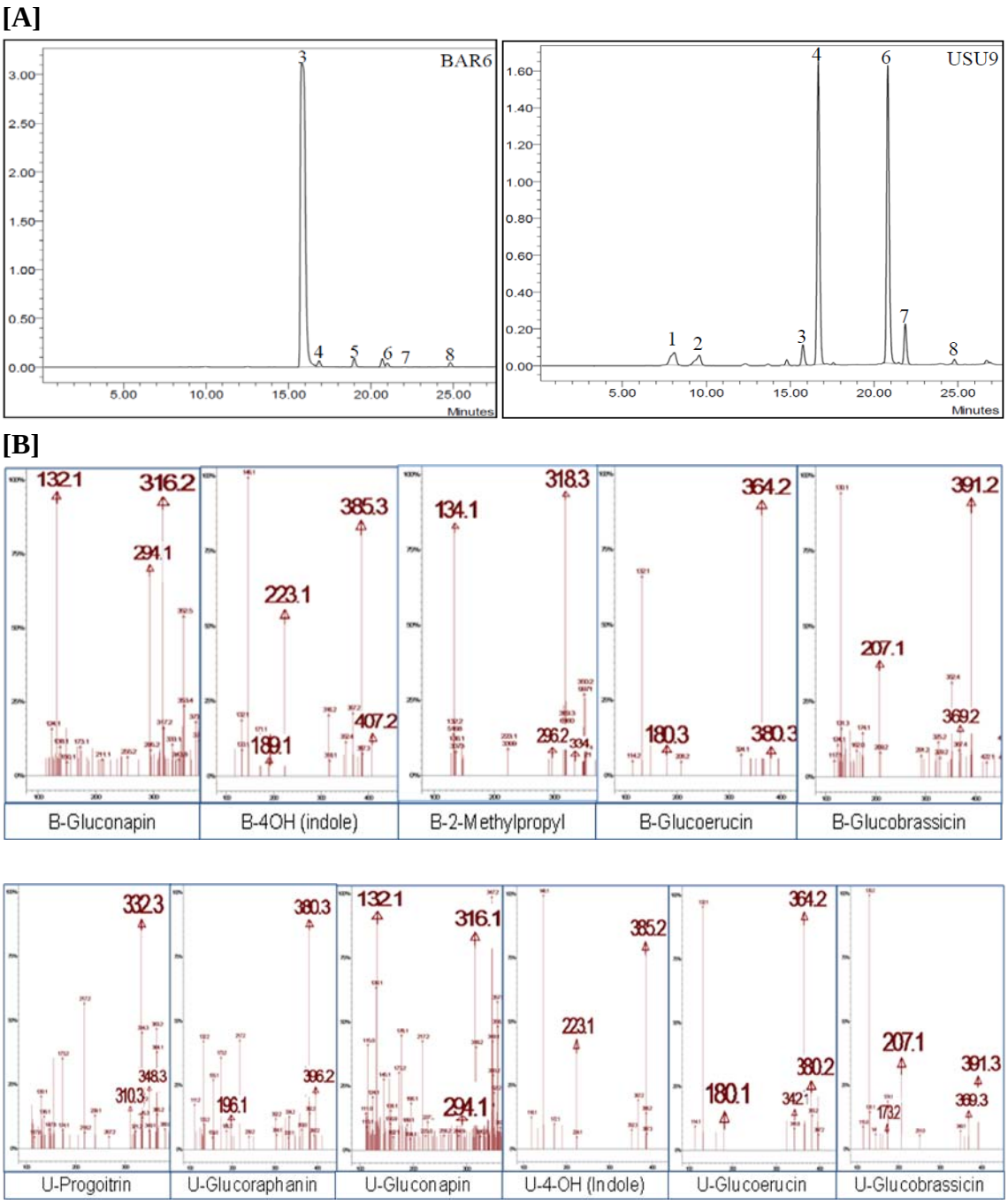
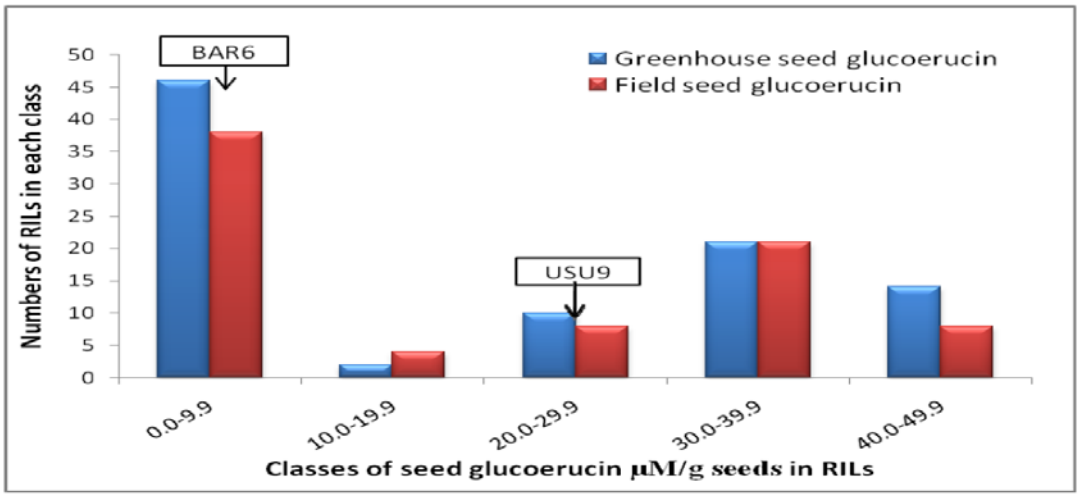
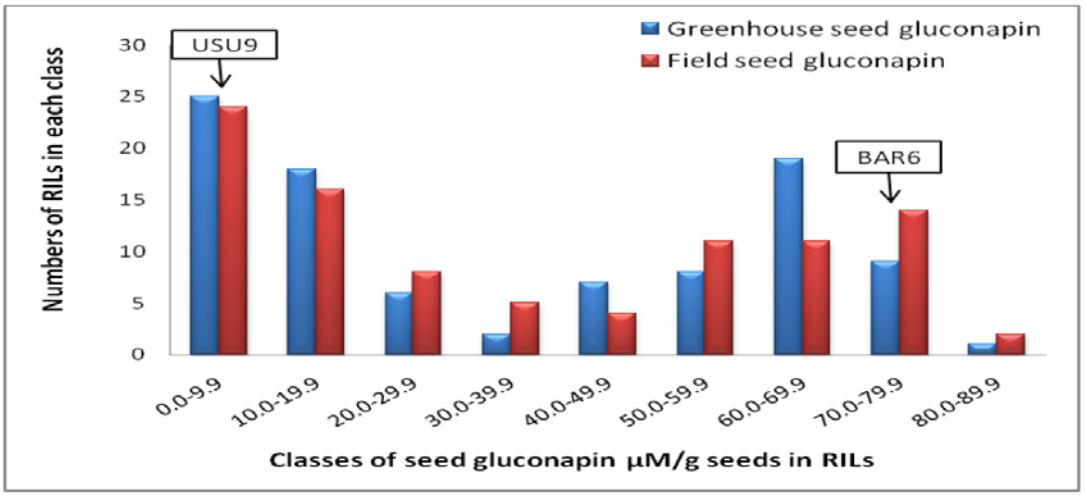


Figure 4.1 HPLC and LC-MS base identification of individual glucosinolate peaks in BAR6 and USU9. [A] HPLC profiles, peak #1 progoitrin, #2 glucoraphanine, #3 gluconapin, #4 4-hydroxyglucobrassicin, #5 2-methylpropyl, #6 glucoerucin, #7 glucobrassicin, #8 4-methoxyglucobrassicin. [B] LC-MS profiles of individual glucosinolate mass of daughter ion fragments (*m/z*) of BAR6 (B top) and USU9 (B bottom).

[A]



[B]



[C]

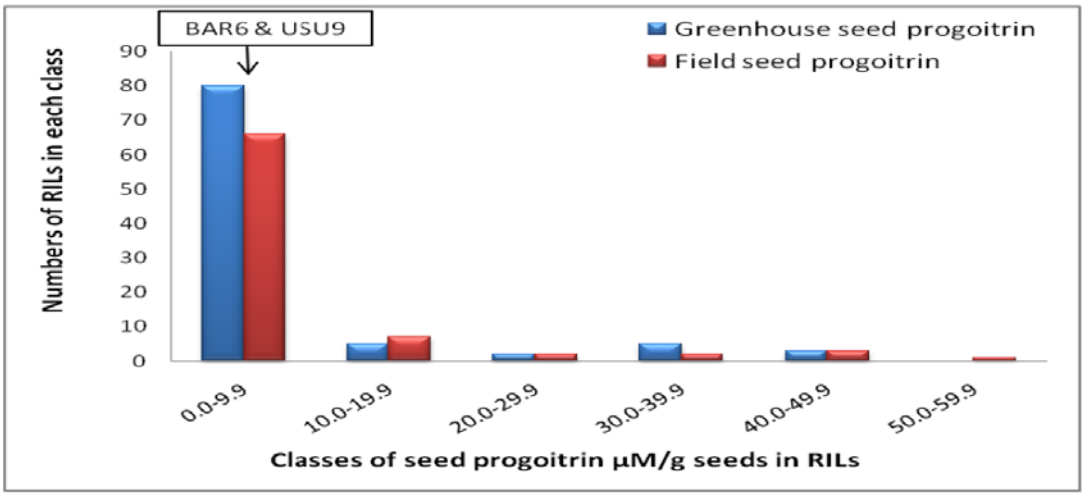


Figure 4.2 Legend on the following page

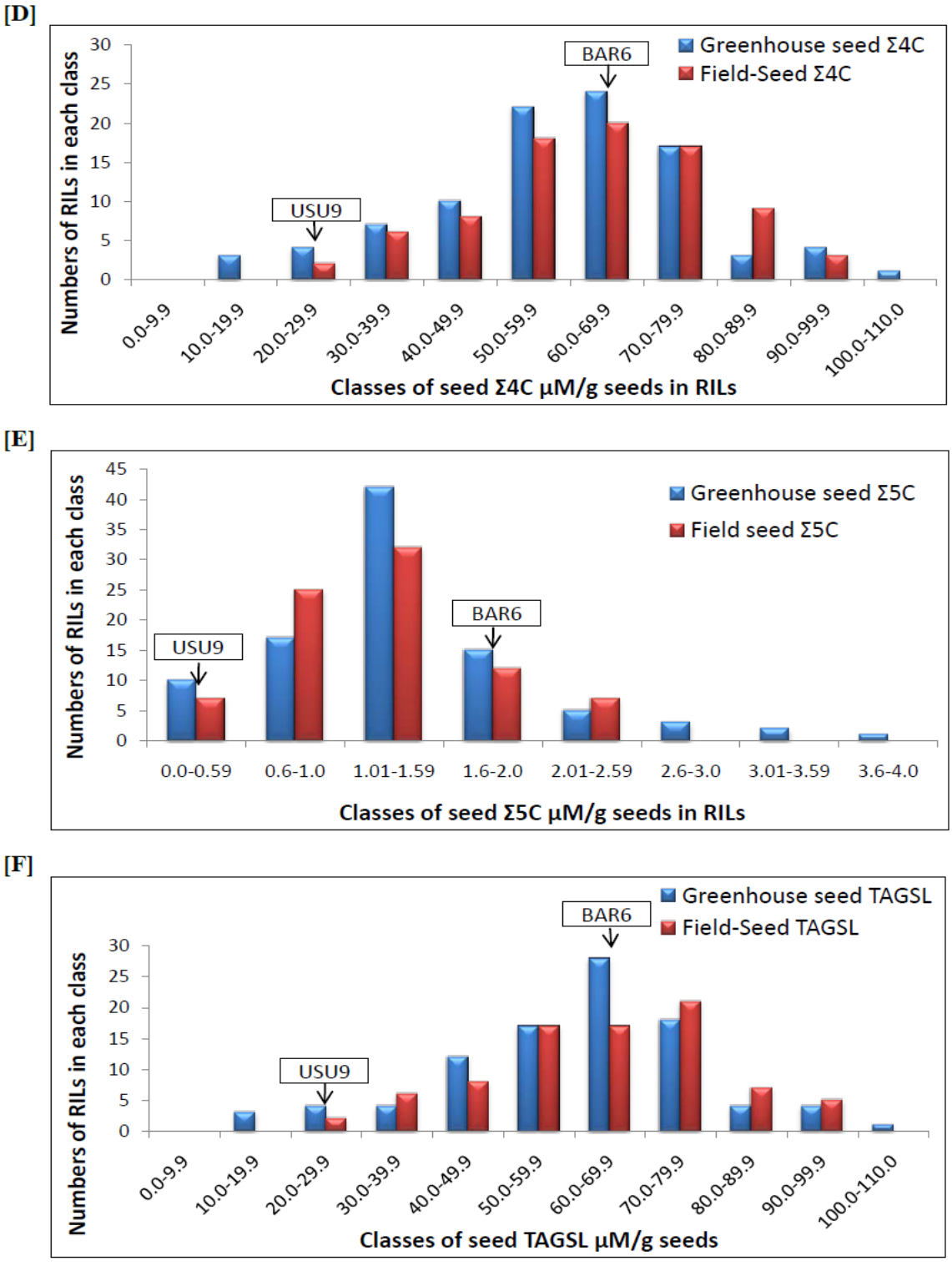


Figure 4.2 Histograms showing seed glucosinolates in the RILs of *B. rapa*. [A] glucoerucin, [B] gluconapin, [C] progoitrin, [D] sum of 4C GSL, [E] sum of 5C GSL and [F] total aliphatic GSLs.

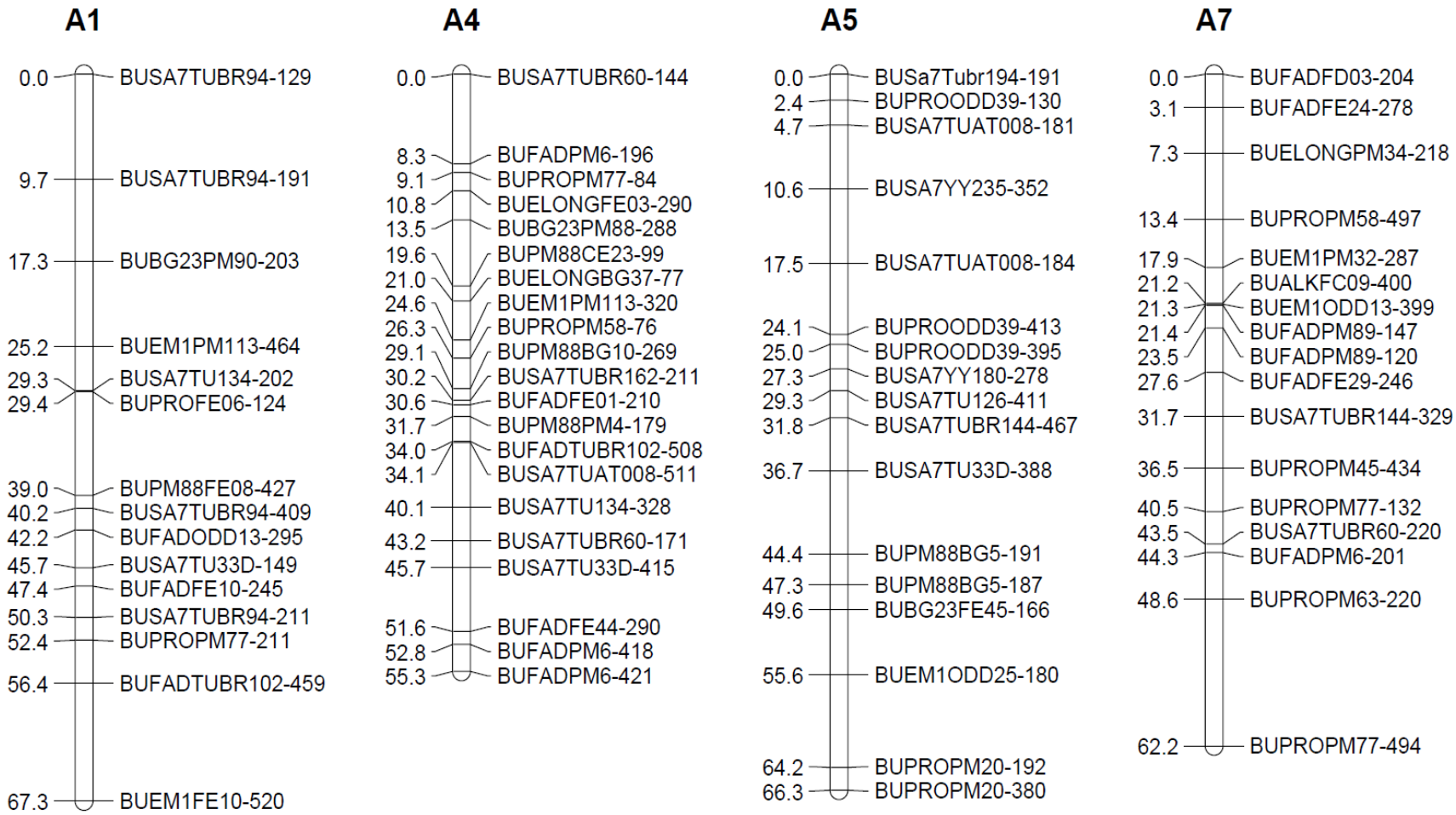


Figure 4.3 Legend on the following page

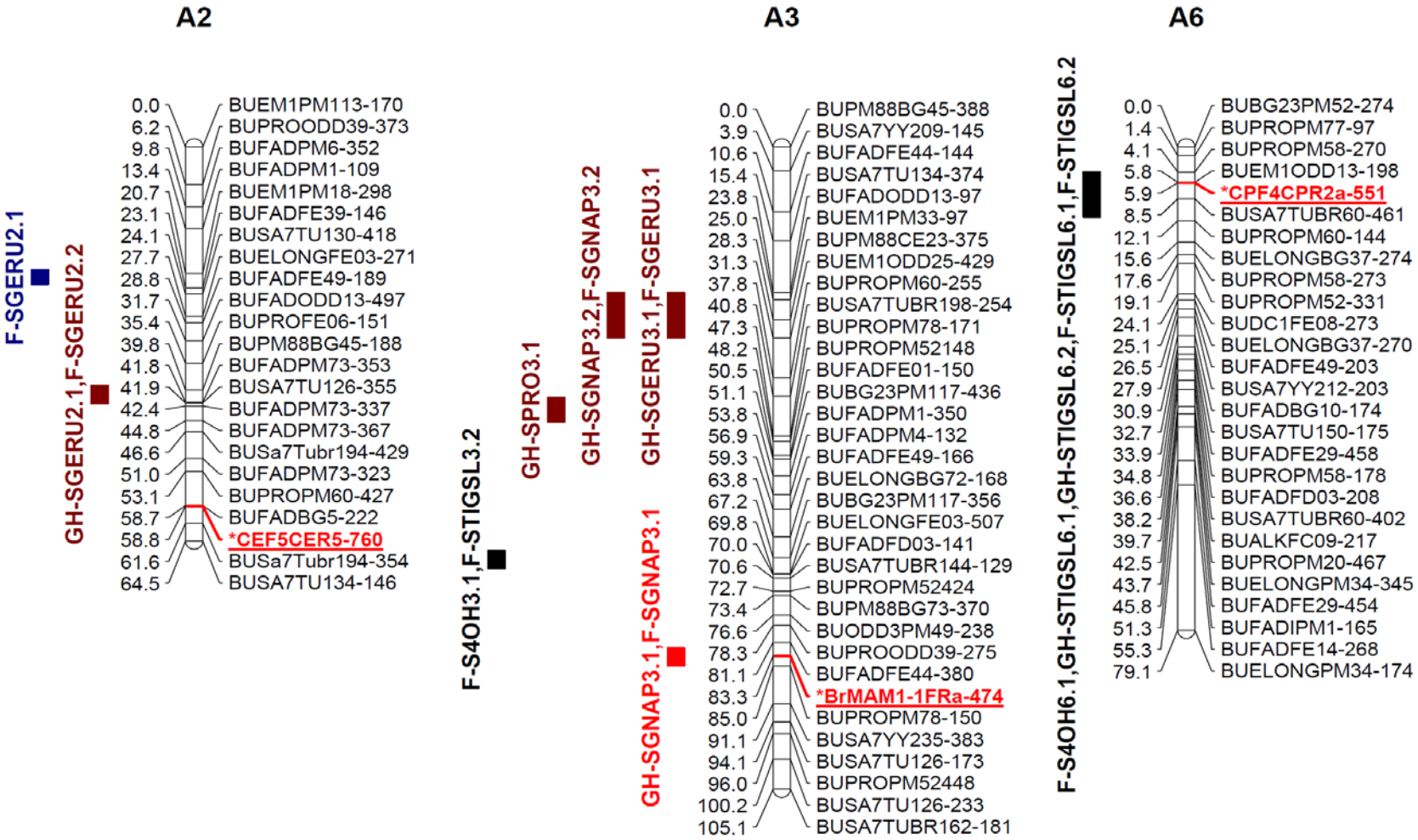


Figure 4.2 Legend on the following page

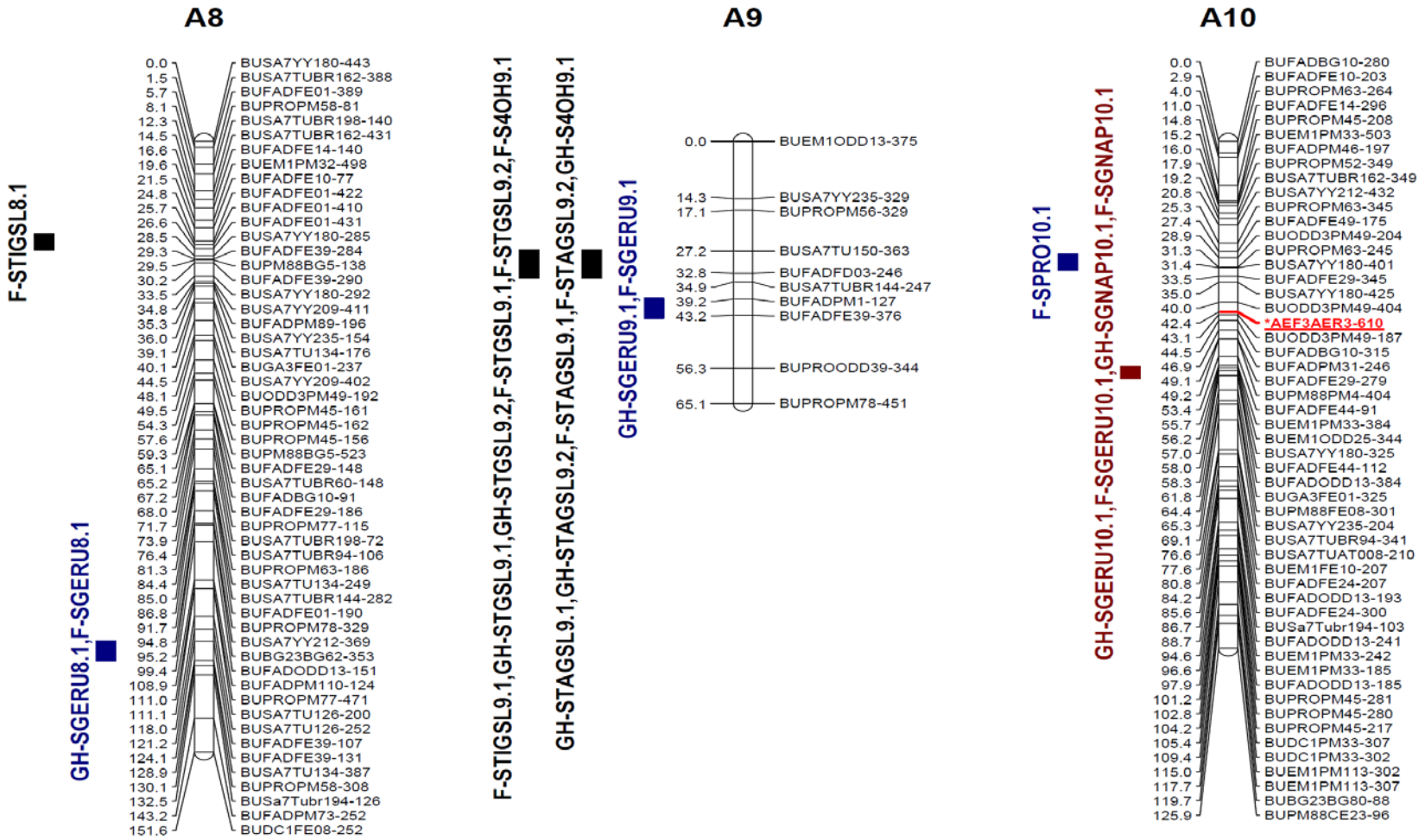


Figure 4.3 Genetic map of *B. rapa* using SRAP and the *B. rapa* RILs. Gene specific SCAR markers are indicated with asterisk, underline with bold fonts. QTL for seed glucosinolates are positioned on map by box.

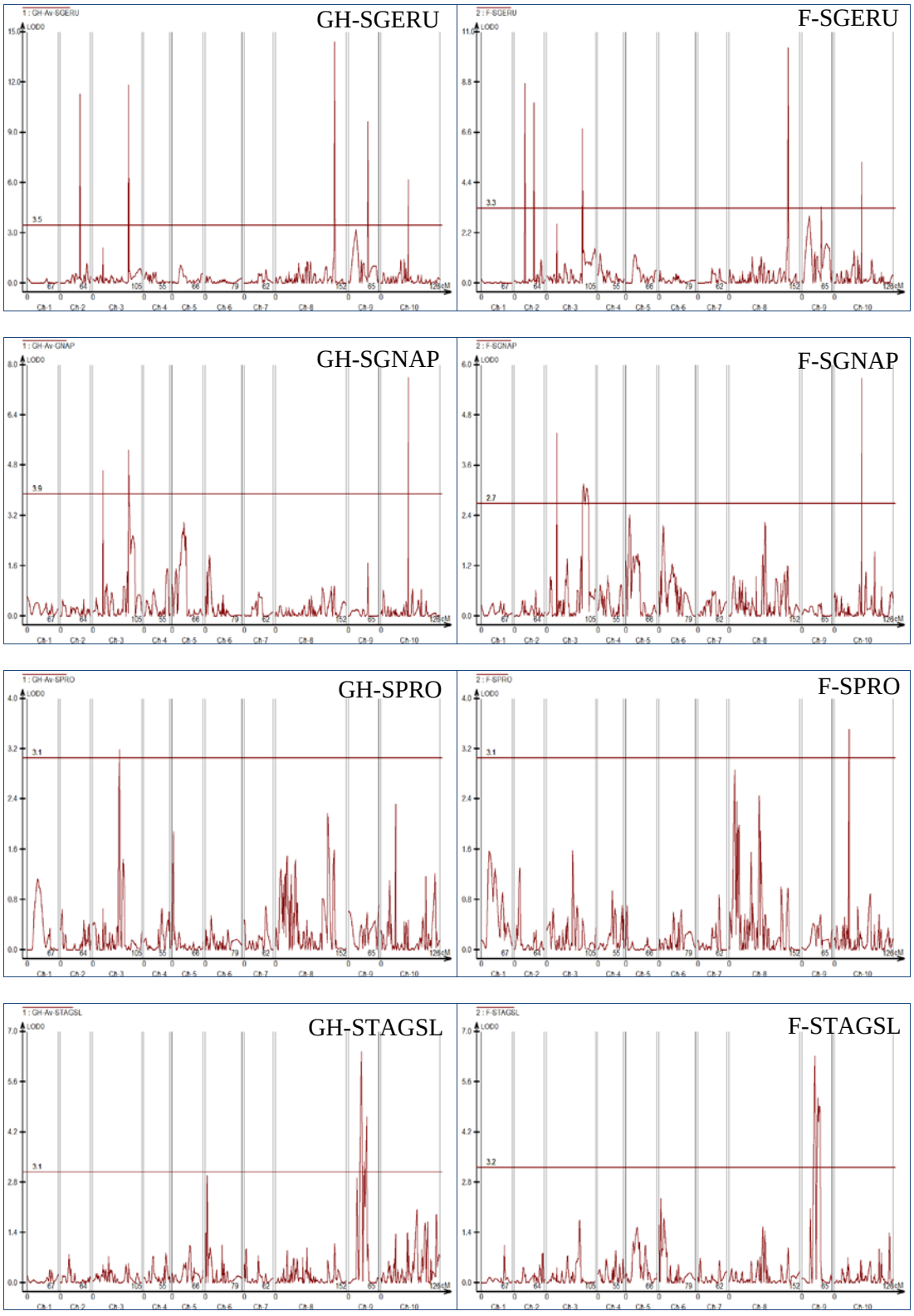


Figure 4.4 Legend on the following page

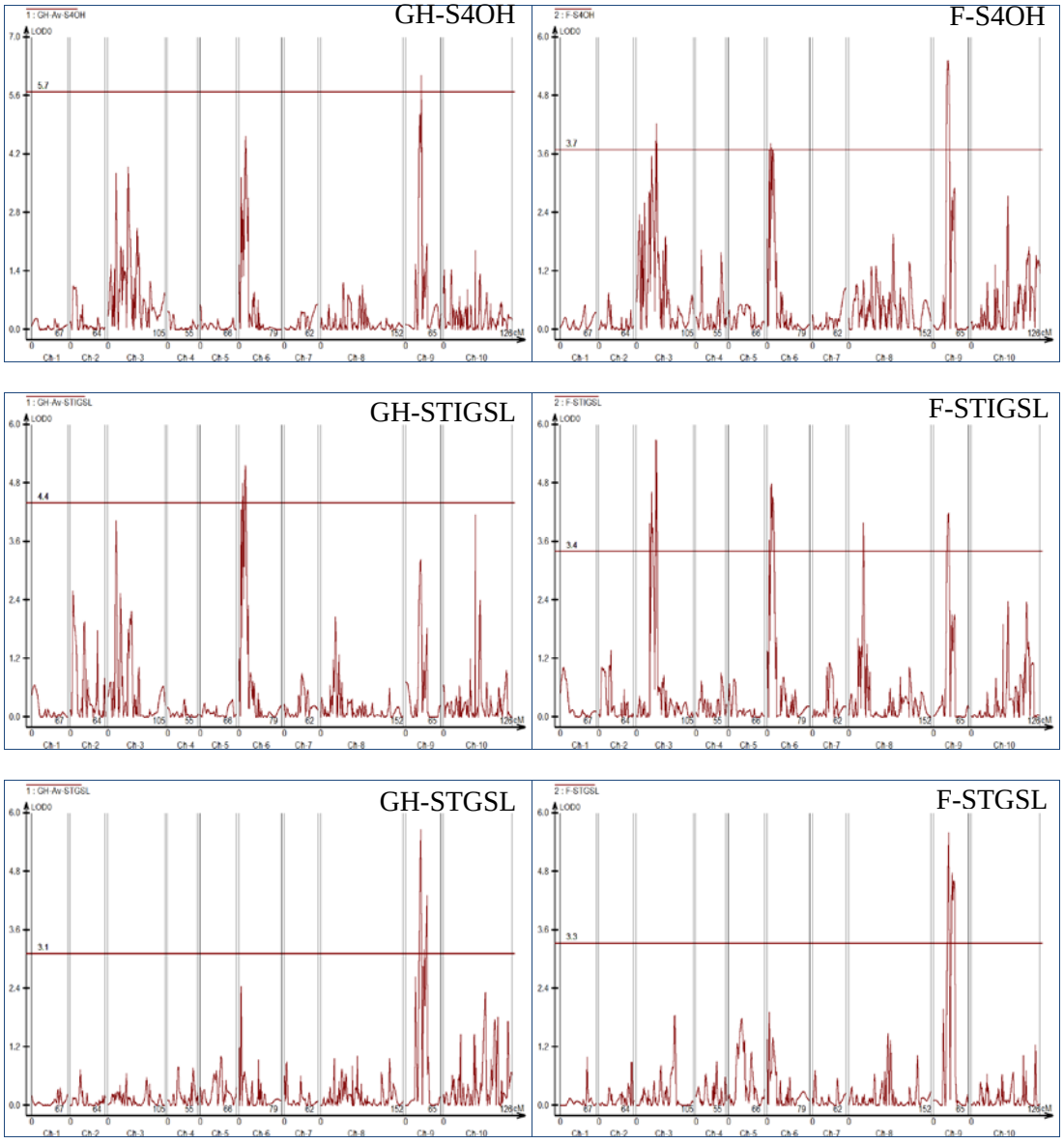


Figure 4.4 QTL plots for seed glucosinolates (CIM method). GH- Average values of two greenhouse replicates, F- Field replicate data. SGERU- seed glucoerucin, SGNAP- seed gluconapin, SPRO- seed progoitrin, STAGSL- seed total aliphatic glucosinolates, S4OH- seed 4-hydroxyglucobrassicin, STIGSL- seed total indole glucosinolates, STGSL- Total glucosinolates.

5. HOMOEOLOGOUS GENE REPLACEMENT AND INTROGRESSION FOR CHANGING ALIPHATIC GLUCOSINOLATES IN *Brassica rapa* L. BY MARKER ASSISTED BACKCROSS

Arvind H. Hirani, Carla D. Zelmer, Peter B. E. McVetty, Fouad Daayf and Genyi Li

Department of Plant Science, University of Manitoba

Winnipeg, Manitoba Canada R3T 2N2

5. HOMOEOLOGOUS GENE REPLACEMENT AND INTROGRESSION FOR CHANGING ALIPHATIC GLUCOSINOLATES IN *Brassica rapa* L. BY MARKER ASSISTED BACKCROSS

5.1. Abstract

Aliphatic glucosinolates are the predominant (~90% of total glucosinolates) sulphur-rich plant secondary metabolites in economically important *Brassica* crops. In this study, aliphatic glucosinolates were genetically manipulated through homoeologous recombination in backcross lines followed by marker assisted selection in *B. rapa*. A resynthesized *B. napus* line, from a cross between *B. rapa* and *B. oleracea*, was backcrossed with three *B. rapa* recurrent parents, RI16, BAR6 and USU9. Marker assisted selection for glucosinolate genes was utilized for each generation. Trisomic lines harbouring a chromosome with non-functional *GSL-ELONG* (*GSL-ELONG*$^{-}$) and functional *GSL-PRO* (*GSL-PRO*$^{+}$*)* genes were developed. Advanced backcross progenies (BC_3F_2) in three genetic backgrounds of *B. rapa*, RI16, BAR6 and USU9, were developed to identify homoeologous gene replacement and/or introgression. Variable marker transmission rates ranging from 4 to 73% were observed in early and advanced backcross progenies. Reduction in 5C aliphatic glucosinolates (glucoalyssin and glucobrassicanapin) was observed inBC_3F_2 progenies derived from the *B. rapa* recurrent parent RI16 that carried the *GSL-ELONG*$^{-}$ gene. . The *GSL-ELONG*$^{-}$ positive backcross progenies were also screened by the A-genome specific SCAR marker BrMAM1-1, which co-segregated with 5C aliphatic glucosinolates. The A-genome specific SCAR marker BrMAM1-1 was absent in the plants of advanced backcross progenies which showed reduction in 5C aliphatic glucosinolates. The results suggest that the functional

allele had been replaced by the non-functional *GSL-ELONG*$^{-}$ allele from *B. oleracea*. Some advanced backcross progenies (BC_3F_2) positive for the *GSL-ELONG*$^{-}$ allele did not show reduction in 5C aliphatic glucosinolates, suggesting that introgression instead of replacement of the *GSL-ELONG*$^{-}$ allele into the A-genome occurred. All the backcross plants positive for *GSL-PRO*$^{+}$ produced the 3C aliphatic glucosinolate sinigrin, suggesting that introgression of, or replacement by a functional *GSL-PRO*$^{+}$ gene from *B. oleracea* might have activated the 3C aliphatic glucosinolate biosynthesis pathway in *B. rapa*.

5.2. Introduction

In the genus *Brassica*, three diploid species, *B. rapa* L., *B. oleracea* L. and *B. nigra* L. Koch, are the evolutionary original genome donors of three amphidiploid species, namely *B. napus* (L.), *B. juncea* (L. Czern. and Coss.) and *B. carinata* (Br.). These *Brassica* species are economically important crops and are cultivated for edible oil, industrial oil and biodiesel. Since the introduction of *Brassica* species as crops in agricultural systems, many traits related to agronomy, morphology, physiology, yield, disease resistance and end product quality have been improved. Most trait improvements have been achieved through conventional breeding approaches such as intraspecific and interspecific hybridization followed by selection. The development and analysis of alien chromosome addition lines were initiated in the 1990s for dissecting *Brassica* genomes for gene mapping and genetic analyses. Chromosome addition lines have also been deployed for the introgression of suitable traits from wild or weedy relatives to the existing cultivars. Alien chromosome addition monosomic and disomic lines have been successfully developed. They have proven to be useful resources for gene identification

and chromosome homoeology studies between the genomes of *B. rapa* x *B. oleracea*, *B. napus* x *B. nigra* and *B. rapa* x *B. oxyrrhina* (McGrath and Quiros 1990, Hu and Quiros 1991, Chen *et al.* 1992, Cheng *et al.* 1994, Srinivasan *et al.* 1998). Many agronomic, disease resistance, morphological, quality traits (such as erucic acid content), flowering time, seed coat color and stem rot resistance have been successfully introgressed into cultivated *Brassica* crops through aneuploid production and subsequent backcrossing (Banga 1988, Cheng *et al.* 1994, Navabi *et al.* 2010).

Wide hybridization requires formation of bivalents or multivalents between homoeologous chromosomes for recombination, which is highly affected by homoeology of alien and host chromosomes, as well as the karyotype of the parental genomes. In wheat, wide cross-based introgression of traits were successfully achieved using the *Ph1* locus mutant since the *Ph1* locus regulates and ensures association and crossing over of homologous chromosomes in meiosis. The mutant *Ph1* gene resulted in synapsis and recombination between homoeologous chromosomes (Feldman 1966, Martinez *et al.* 2001). Using the *Ph1* mutant, leaf and stripe rust resistance genes *Lr57* and *Yr40* have been transferred from *Aegilops geniculata* into wheat (Kuraparthy *et al.* 2009). Similarly, salt-tolerance genes were successfully introgressed into wheat from *Thinopyrum bessarabicum* (King *et al.* 1997).

On the other hand, Dover and Riley (1972) reported success of wide hybridization in wheat in the presence of a functional *Ph1* locus. They suggested that several genes from *Aegilops* species suppressed the activity of the *Ph1* gene during meiosis, which resulted in successful multivalent formation and subsequent homoeologous recombination. Similar meiotic regulatory mechanisms might be involved in interspecific

or intergeneric *Brassica* crosses. In *B. napus*, a *PrBn* (*Pairing Regulator in B. napus*) gene located in the C-genome of *B. napus*, is reported to be involved in the regulation of non-homologous cross over during meiosis (Jenczewski *et al.* 2003, Nicolas *et al.* 2009). However, there is no additional experimental evidence available for the genetic regulation of meiosis in other *Brassica* species.

Glucosinolates, a class of nitrogen and sulphur containing plant secondary metabolites are found in *Brassica* species. Glucosinolates and their hydrolysis products interact in various ways with plants, microbes, insects, animals and humans. Extensive studies on glucosinolates suggest that several genes regulate glucosinolate biosynthesis in the model plant *Arabidopsis thaliana* (Compos de Quiros *et al.* 2000, Kliebenstein *et al.* 2001a, Kliebenstein *et al.* 2001c, Textor *et al.* 2007, Li *et al.* 2008). Two genes (*GSL-ELONG* and *GSL-PRO*) involved in side chain elongation, and a gene (*GSL-ALK*) involved in side chain modification were identified and characterized in *B. oleracea* (Li and Quiros 2002, 2003, Gao *et al.* 2004). Li and Quiros (2002) studied various accessions of *B. oleracea* for glucosinolate content and reported that white cauliflower produces 3C glucosinolates due to a functional *Bo-GSL-PRO* gene. However, white cauliflower lacks, of 4C aliphatic glucosinolates was due to a mutation at the splicing site of intron 3 of *Bo-GSL-ELONG*, leading to a non-functional gene in white cauliflower. In addition, they also reported that broccoli produced exclusively glucoraphanin, a 4C aliphatic glucosinolate, suggesting that broccoli has a non-functional *Bo-GSL-ALK* and a functional *Bo-GSL-ELONG*. Gene specific molecular markers were developed based on glucosinolate biosynthesis genes in *B. oleracea* for marker assisted selection to manipulate quality and quantity of aliphatic glucosinolates (Li and Quiros 2002, 2003).

The objective of this study was to manipulate aliphatic glucosinolate profiles and contents in *B. rapa* through homoeologous gene replacement or introgression from *B. oleracea* (white cauliflower) using marker assisted selection. In this study, resynthesized *B. napus* lines were developed through interspecific hybridization between *B. rapa* and *B. oleracea*. Synthetic *B. napus* was backcrossed to *B. rapa* to develop *B. rapa* x *B. oleracea* chromosome addition lines harbouring *B. oleracea* glucosinolate genes, *GSL-ELONG*$^{-}$ and *GSL-PRO*$^{+}$ through marker assisted selection. *Brassica rapa* possesses at least four loci of the *GSL-ELONG* gene involved in 4C and 5C aliphatic glucosinolate biosynthesis (Bisht *et al.* 2009). Three carbon aliphatic glucosinolates are absent, suggesting that *GSL-PRO* genes may be non-functional in *B. rapa*. Based on this knowledge, marker assisted backcross breeding was performed to replace the functional *GSL-ELONG*$^{+}$ locus/loci of *B. rapa* with the non-functional *GSL-ELONG*$^{-}$ allele from *B. oleracea* to manipulate 4C or 5C aliphatic glucosinolates in this A-genome species. Marker assisted selection also was performed for the GSL-PRO gene introgression in the *B. rapa* to produce 3C glucosinolate, sinigrin. The *GSL-ELONG*$^{-}$ gene replaced and *GSL-PRO*$^{+}$ gene introgressed advanced backcross lines were identified with gene specific markers and evaluated for glucosinolate profile and content.

5.3. Materials and Methods

Plant materials and backcross breeding scheme

A high glucosinolate content *B. rapa* line was crossed with a *B. oleracea* white cauliflower accession Snowball 76, and an embryo rescue technique was employed to produce resynthesized *B. napus* lines. One resynthesized *B. napus* line was backcrossed

with three *B. rapa* recurrent parents to produce backcross progenies. Three recurrent parents used in this research were Chinese cabbage (RI16), yellow sarson (BAR6) and a DH line (USU9). These three *B. rapa* recurrent parents had distinctive glucosinolate profiles. The genotype RI16 had gluconapin and trace amounts of progoitrin as 4C glucosinolates together with glucoalyssin and glucobrassicanapin as 5C aliphatic glucosinolates. The genotype BAR6 had a high content of gluconapin as 4C aliphatic glucosinolates, whereas the genotype USU9 had a high content of glucoerucin and trace amounts of gluconapin as 4C aliphatic glucosinolates. During recurrent backcrossing, marker assisted selection was performed in each generation with selection for genes of interest in the C-genome. In the first and second backcross cycle, recurrent parents were used as male parents. Additional reciprocal backcross was performed to enhance allosyndetical homoeologous recombination between the A-genome and alien chromosomes of *B. oleracea*. Advanced backcross progenies (BC_3F_1) were selfed to produce offspring homozygous for the introgressed or replaced major glucosinolate genes *GSL-ELONG*$^-$ and *GSL-PRO*$^+$ (Figure 5.1).

DNA extraction and C-genome specific SCAR marker assisted selection

For screening and selection, DNA was extracted from leaf tissues of 2 week old seedlings using a modified CTAB method described by Li and Quiros (2001). Instead of individual samples in Eppendorf tubes, 96 deep well plates were used. About 0.1 g leaf tissues from individual plants were placed in the wells and ground in liquid nitrogen. 500 µl 2xCTAB buffer (2%CTAB, 20mM EDTA, 100mM Tris, 1.4M NaCl, pH-8.0) was added to each well and then samples were incubated at 65°C for 1.5 hrs. 400 µl chloroform was added to each well followed by vigorous mixing and centrifugation at

6200 rpm for 10 min. Ninety microlitres of the supernatant was transferred to a 96 well plate and then DNA was precipitated with 50 µl iso-propanol followed by centrifugation at 6200 rpm for 5 min. The DNA pellet was washed with 70% ethanol, air dried and then dissolved in 80 µl dH_2O.

Marker assisted selection of non-functional *GSL-ELONG* (*GSL-ELONG*$^-$) and functional *GSL-PRO* (*GSL-PRO*$^+$) genes in backcross progenies was performed using the C-genome specific SCAR markers. The PCR reactions were conducted using the primer pair PM25+PM13 for *GSL-ELONG*$^-$ and BG89+BG54 for *GSL-PRO*$^+$ (Table 5.1). Each PCR reaction of 10 µl consisted of 7.0 µl dH_2O, 1.0 µl 10xPCR buffer, 0.3 µl 50 mM $MgCl_2$, 0.15 µl 25 mM dNTP, 0.1 µl Taq polymerases, 0.15 µl 10 µM forward and reverse primers and 2.5 µl template DNA. The PCR program was 94°C for the initial 4 min to denature genomic DNA followed by 35 cycles of 94°C for 1 min, 55°C for 1 min and 72°C for 1.5 min. For all backcross generations, the *GSL-ELONG*$^-$ and *GSL-PRO*$^+$ gene specific SCAR markers were scored on 1.2% agarose gels. One to seven plants from each line were selected using SCAR markers for the next backcross with the recurrent parents.

Leaf and seed glucosinolate extraction

Leaf glucosinolates were extracted from early (BC_1F_1) and advanced (BC_3F_1) backcross progenies to determine glucosinolate profiles. Glucosinolate analysis was performed on BC_3F_2 seeds from the plants positive for the appropriate C-genome SCAR markers. Total glucosinolate content was extracted from 200 mg air dried seeds and 250 mg fresh young leaf tissues. Leaf and seed glucosinolate purification and desulfation

were performed using sephadex/sulfutase reactions as described by Kliebenstein *et al.* (2001b) with some minor modifications. Final desulfoglucosinolates were eluted into a 1:1 (v/v) 400 µl of mixture of distilled water and 70% methanol.

Detection and quantification of aliphatic glucosinolates in backcross progenies

Leaf and seed desulfoglucosinolate separation and quantification was performed in a 5 µm column (LichroCART® 250-4 RP18, Fisher Scientific, Ottawa, Canada) coupled with the Alliance® reverse phase HPLC (Waters 2695) and photodiode array detector (Waters 996) system (Waters, MA USA). Desulfoglucosinolates were separated using mobile phase HPLC grade methanol (A) and distilled water (B) at a flow rate of 1 ml/min. Both solvents followed gradient programs as follows: 8 min 7:93 A/B (v/v), 4 min 15:85 A/B (v/v), 18 min 55:45 A/B (v/v), 5 min 92:8 A/B (v/v), 5 min 92:8 A/B (v/v), 5 min 1.5:98.5 A/B (v/v), 3 min 1.5:98.5 A/B (v/v) and a final 4 min 0:100 A/B (v/v) for a total running time of 52 min. Individual glucosinolates were identified according to retention times and quantity was calculated in micromoles per gram and adjusted with relative response factors as described by Brown *et al.* (2003).

Determination of replacement or introgression of genes through the A-genome specific markers

The A-genome specific SCAR markers (Table 5.1) for the glucosinolate side chain elongation gene *GSL-ELONG* were used to determine replacement with major effect locus/loci on linkage group A3. Advanced backcross progenies were screened with BrMAM1-1 and AE3 SCAR markers. These SCAR markers are dominant markers;

therefore, the absence of a band was considered evidence of putative replacement of the native allele(s) with the corresponding allele from the C-genome.

5.4. Results

Crossability of trigenomic hybrid with recurrent *B. rapa*

At the beginning of this research, a resynthesized *B. napus* line was developed by a cross between a *B. rapa* genotype (high glucosinolate content) and a *B. oleracea* white cauliflower genotype (lacking 4C and 5C aliphatic glucosinolate production). The resynthesized *B. napus* line was backcrossed with each of three *B. rapa* recurrent parents, RI16, BAR6 and USU9. Crossing of the resynthesized *B. napus* line with the three *B. rapa* recurrent parents, on average, produced four seeds per silique (2-8), and resulted in trigenomic hybrid (AAC) plants. These AAC hybrid plants were subsequently backcrossed to the three *B. rapa* recurrent parents. In the case of the second backcross (BC_2F_1), on average, three seeds per silique (0-5) were recorded. Seed sets in the BC_2 hybrids were very low, apparently due to genomic irregularities in the plants.

Plants (BC_2F_1) positive for the C-genome specific SCAR markers were selected and reciprocally backcrossed to the *B. rapa* recurrent parents to produce BC_3F_1 seeds. Molecular analysis of the BC_3F_1 plants with the SCAR markers revealed loss of these markers in backcross progenies of BAR6 and USU9; while the presence of these markers were recorded in the plants of the backcross progenies of RI16. These SCAR marker positive plants (30%) were selfed to produce homozygous progenies (BC_3F_2).

Marker transmission rates in backcrossed trisomic lines

In the BC_1F_1 generation, a total of 149 plants from three backcross populations were screened for the *GSL-ELONG*$^-$ marker; 36 plants (25%) were positive for this marker. The transmission rates of this marker ranged from 4 to 50% depending on the family (Table 5.2). Approximately four plants from each BC_1 family were backcrossed to the corresponding recurrent parents to generate BC_2F_1 progenies. In BC_2F_1 generation, 177 plants were screened for the *GSL-ELONG*$^-$ marker;55 positive plants (31%) were obtained in three backcross populations. Marker transmission rates varied from 8 to 63% in this generation (Table 5.2). One to three plants with appropriate markers from each family were reciprocally backcrossed to the corresponding recurrent parents to generate BC_3F_1 progenies. In BC_3F_1 progenies, marker transmission rates ranged between 4 and 50%. Plants with the *GSL-ELONG*$^-$ marker were selected to produce homozygous progenies. In the reciprocal backcross (BC_3F_1), the SCAR marker for the *GSL-ELONG*$^-$ gene was lost in the BAR6 and USU9 genetic background. In BC_3F_2 population, where the RI16 was the only recurrent parent, a total of 480 plants were screened and 232 marker positive plants (48%) were obtained. The rate of marker transmission in this generation was 35 to 73% (Table 5.2). However, 70% of BC_3F_2 family from the RI16 recurrent parent (data not shown) were found to lack the C-genome specific SCAR markers, apparently due to the loss of alien chromosomes with the advancement of the backcrosses. However, normal seed set and high marker transmission rates were observed in 30% of advanced backcross (BC_3F_2) progenies of RI16 recurrent parent.

Genome specific SCAR markers for aliphatic glucosinolates

The C-genome specific SCAR markers PM25+PM13 and BG89+BG54 were used to screen for *GSL-ELONG*$^{-}$ and *GSL-PRO*$^{+}$ genes, respectively, by marker assisted selection of backcross progenies (Figure 5.2). All backcross generations (BC_1F_1 to BC_3F_2) were screened with gene specific SCAR markers to obtain plants positive for one or both genes. In all BC_3F_1 progenies of BAR6 and USU9 recurrent parents, the C-genome specific markers were lost. However, in the RI16 backcross progenies, the C-genome specific markers for *GSL-ELONG*$^{-}$ and *GSL-PRO*$^{+}$ genes were detected in some plants. This suggests that homoeologous recombination events might had occurred and resulted in replacement or introgression of the chromosomal fragments of the C-genome into the A-genome.

There are three possibilities for the observed C-genome specific SCAR marker behaviour, (i) replacement of the A-genome fragment with that of a C-genome, (ii) introgression of C-genome fragment into the A-genome, and (iii) the existence of an additional chromosome in the backcross progenies (aneuploiy). The A-genome specific SCAR markers for *GSL-ELONG*$^{+}$ loci were used to screen the BC_3F_2 population of the recurrent parent RI16 to investigate the occurrence of gene replacement or introgression. Fifteen BC_3F_2 plants of this population did not show the A-genome specific marker BrMAM1-1 for *GSL-ELONG*$^{+}$ locus, suggesting that homoeologous recombination and gene replacement had occurred. On the other hand, several BC_3F_2 plants displayed both the A- and C-genome specific markers, suggesting that introgression of *GSL-ELONG*$^{-}$ in the A-genome had occurred. However, no phenotypic changes in leaf and seed glucosinolates were observed in these plants.

Of the BC_3F_1 plants from the RI16 recurrent parent, three plants were positive for the *GSL-PRO*$^+$ gene that produced the 3C aliphatic glucosinolate sinigrin. In these plants, a functional *GSL-PRO*$^+$ gene might be introgressed from a chromosome of the C-genome, which had activated biosynthesis of 3C aliphatic glucosinolate. However, the *GSL-PRO*$^+$ gene specific marker was not detected in any of the BC_3F_1 plants of BAR6 and USU9 nor was it found in 95% of the plants of the RI16 recurrent parent. As a result of this loss of the *GSL-PRO*$^+$ gene, no 3C aliphatic glucosinolates were detected in this group.

Modification of aliphatic glucosinolates in backcross progenies

Fifty BC_3F_2 plants that were positive for *GSL-ELONG*$^-$ markers were analyzed for glucosinolate contents and profiles. A total 15 plants (30%) were found to have 5C aliphatic glucosinolates by five times lower than their recurrent parent RI16. These plants did not show the A-genome specific marker BrMAM1-1 corresponding to *GSL-ELONG*$^+$ locus responsible for the biosynthesis of 5C aliphatic glucosinolates (Figure 5.3 and 5.4). The recurrent parent Chinese cabbage RI16 produced 25 µmole/g seed of 5C aliphatic glucosinolates, whereas backcross progenies of RI16 produced trace amount (<5 µmole/g seed) of 5C aliphatic glucosinolates (Figure 5.3 and 5.4). Results of molecular markers (A and C-genome specific) and modification of glucosinolate profiles indicate possible functional gene replacement by the non-functional *GSL-ELONG*$^-$ gene from *B. oleracea*. On the other hand, 35 plants (70%) which were positive for the *GSL-ELONG*$^-$ marker did not show any changes in glucosinolate profiles, apparently due to introgression or to the presence of the marker on additional chromosome from *B. oleracea*.

In advanced backcross progenies, the *GSL-PRO*$^{+}$ marker disappeared in most of the plants (95 %); three BC_3F_2 plants with *GSL-PRO* $^{+}$ marker, however, were observed and these were analyzed for leaf glucosinolates. The *GSL-PRO*$^{+}$ gene produced a peak of the 3C aliphatic glucosinolate sinigrin in these three backcross plants of RI16, while this peak was absent in the RI16 parent (Figure 5.5). This suggests that introgression of the *GSL-PRO*$^{+}$ gene activated 3C glucosinolate biosynthesis in *B. rapa*. However, low 3C aliphatic glucosinolate content in these plants suggested the presence of multiple active *GSL-ELONG*$^{+}$ loci for 4C and 5C aliphatic glucosinolates which might have reduced the accumulation of 3C aliphatic glucosinolates.

5.5. Discussion

The diversity of glucosinolate profiles and contents in *B. rapa* was discussed in several previous reports (He *et al.* 2000, Padilla *et al.* 2007, Lou *et al.* 2008). Limited genetic information and molecular markers, however, are available for genes involved in glucosinolate biosynthesis in *B. rapa*. In this study, three different *B. rapa* recurrent parents, RI16, BAR6 and USU9 were selected based on their diversities of glucosinolate profiles and origins were selected for backcross breeding. The Chinese cabbage RI16 produces gluconapin (4C) and glucobrassicanapin (5C) along with trace amounts of progoitrin (4C) and glucoalyssin (5C) whereas, yellow sarson BAR6 produces predominantly gluconapin (4C), and a canola quality DH line USU9 possesses predominantly glucoerucin (4C). In addition to glucosinolate profile diversity, morphological, agronomic and seed quality traits of these three parents are also diverse.

Glucosinolates are genetically regulated quantitative traits in *Arabidopsis* and *Brassica* species. *Brassica rapa* has a high level of genome duplication, which results in more loci with various functional properties (Punjabi *et al*. 2008). Genome duplications and rearrangements in *B. rapa* have led to a very complex genetic basis for glucosinolate biosynthesis. *Brassica rapa* and *B. oleracea* diverged quite recently (approximately 7.9 MYA) from a common ancestor when compared to *B. nigra* (approximately between 16 and 18 MYA), therefore, relatively high homoeology exists between the A and C genomes (Wroblewski *et al*. 2000, Parkin *et al*. 1995, Parkin *et al*. 2005, Punjabi *et al*. 2008). High homoeology between the A and C genomes permit crossing over and recombination events in crosses of the A and C genome holding *Brassica* species.

Several studies reported that chromosomal pairing and recombination between homologous chromosomes in trigenomic (AAC) lines of *B.rapa* x *B. oleracea* normally occurs (Attia and Röbbelen 1986, 1987), however, a very low rate of homoeologous recombination was also reported (Chen *et al.* 1992, Mikkelsen *et al.* 1996, Leflon *et al.* 2006). Homoeologous recombination and alien chromosome transmission rates depend on chromosome identity and meiotic behaviour, resulting in varied alien chromosome transmission rates among trigenomic hybrids (Leflon *et al.* 2006).

In this study, varied *GSL-ELONG*$^{-}$ marker transmission rates (4 to 73%) were observed in backcross progenies. These findings were in agreement with the previous studies of *Brassica* and other crop species. Heneen and Jørgensen (2001) reported variable RAPD marker transmission frequencies in aneuploid progenies of *B. rapa* x *B. alboglabra*. Kaneko *et al.* (2003) reported variable marker transmission rates ranging from 26 to 44% in three consecutive generations of *Raphanus sativus* possessing

monosomic additional chromosomes of *B. rapa*. Similarly, Shigyo *et al.* (2003) reported 9 to 49% transmission rates of alien chromosomes in eight *Allium* monosomic addition lines. Ali *et al.* (2001) reported transmission rates from 0 to 32% for chromosome 9 and from 14 to 88% for chromosome 6 of monosomic tomato chromosome addition lines in the cultivated potato. This suggests that different alien chromosomes have very substantial transmission rates in different crop species. Alien chromosome transmission rates appear to be higher through female parents and for the larger chromosomes (Garriga-Calderé *et al.* 1998).

Functional gene-based trait introgression has been reported in many studies of crosses between *B. napus*, *B. oleracea* or *B. alboglabra* with *B. rapa* (Guo *et al.* 1990, Cheng *et al.* 1994, Heath *et al.* 1994). However, this is the first report of homoeologous recombination and gene replacement of a functional locus in the A-genome by a non-functional C-genome specific allele. Gene replacement has led to reduction of 5C aliphatic glucosinolates in advanced backcross progenies. Replacement of large or small effect loci might have variable contributions on individual and total aliphatic glucosinolate content due to the existence of multiple loci for the *GSL-ELONG* gene in the A-genome. On the other hand, some RI16 backcross progenies with increased 4C aliphatic glucosinolates were observed. In these cases total aliphatic glucosinolate content remained the same as that of the recurrent parent. The backcross progenies with random introgression of non-functional alleles did not change in aliphatic glucosinolate content.

Loss of the markers for the alien chromosome or chromosomal fragments containing the *GSL-ELONG*$^{-}$ and *GSL-PRO*$^{+}$ genes was observed in backcross progenies for. This loss might be due to the random inclusion of the alien chromosome into the

gametes during meiosis. Retention of the marker would then depend on the specific gametes involved during fertilization of backcrosses. Several studies reported loss of markers for QTL or genes in marker assisted backcross selection in crop species (Sebolt *et al.* 2000, Ramchiary *et al.* 2007, Szadkowski *et al.* 2010).

In the future research, replacement or introgression of the *GSL-ELONG*$^{-}$ locus in these backcross progenies could be confirmed by additional backcrossing. Alternatively, screening of the closely linked A-genome specific SSR or SRAP markers will determine replace region of the chromosome within the A-genome of these lines. The Chinese cabbage lines with reduced 5C aliphatic glucosinolates achieved in this research could be used for manipulation of glucosinolate profiles in *B. napus, B. juncea* and other *B. rapa* accessions.

In summary, altered glucosinolate profiles in Chinese cabbage RI16 were achieved in this study in advanced backcross progenies through homoeologous recombination, followed by marker assisted selection. The functional characterization of the *GSL-PRO*$^{+}$ gene through gene introgression by backcrosses in this study suggests that the *GSL-PRO*$^{+}$ gene has a role in the biosynthesis of the 3C glucosinolate sinigrin in *Brassica* species.

Table 5.1 List of primers used for backcross breeding and their sequence information and genome specificity.

Gene (Marker)	Primers	Sequence 5'-3'	Specificity
GSL-ELONG^{-}	PM25	CCTGTGAGACGTTAATACC	C genome
(PM25+PM13)	PM13	GAGCTTGAGTTCTATACGC	
GSL-PRO^{+}	BG89	ATTGAAAGAACGCCTCGAAA	C genome
(BG89+BG54)	BG54	GTTGATGTGGCGATGTCTA	
GSL-ELONG^{+}	Br-MAM1-1F	GTTTCCCTGCGTCATCAGA	A genome
(BrMAM1-1)	Br-MAM1-1R	CTAAGCTCTTCGCATAGCTA	
GSL-ELONG^{+}	AE3F	CAGTCAAATTTACCGCCTT	A genome
(AE3)	AE3R	GGTGGCTTTCGCGGACAC	

Table 5.2 *GSL-ELONG*$^{-}$ (PM25+PM13) marker transmission frequency in interspecific backcross progenies.

Generation	Family name	Total plants No.	Observed plants with marker	Expected plants with marker	Observed % with marker
BC_1F_1	A†	43	8	21.5	19
	B†	30	11	15	37
	C†	28	10	14	36
	D†	23	1	11.5	4
	E†	8	3	4	38
	F†	5	1	2.5	20
	G†	3	-	1.5	-
	HxR	2	1	1	50
	IxR	2	1	1	50
	JxB	5	-	2.5	-
BC_2F_1	B01xR	5	2	2.5	40
	B19xR	6	2	3	33
	B30xR	13	1	6.5	8
	C12xR	16	10	8	63
	C27xR	48	9	24	19
	H01xR	8	3	4	38
	I01xR	9	4	4.5	44
	B27xB	3	1	1.5	33
	B30xB	7	3	3.5	43
	C03xB	8	3	4	38
	C25xB	2	-	1	-
	D15xB	4	1	2	25
	B01xU	8	2	4	25
	B19xU	6	3	3	50
	B27xU	6	1	3	17
	C25xU	2	-	1	-
	C27xU	24	9	12	38
	D15xU	2	1	1	50
BC_3F_1	RxC27(28)	38	2	19	5
	RxB01(2)	6	3	3	50
	RxC12(1)	6	2	3	33
	RxC27(23)	23	1	11.5	4
	RxC27(20)	24	9	12	38
	RxB03(2)	5	1	2.5	20
	RxC27(17)	7	3	3.5	43
	RxC27(2)	2	1	1	50
	RxC27(5)	24	6	12	25
	RxC27(10)	6	2	3	33

Generation	Family name	Total plants No.	Observed plants with marker	Expected plants with marker	Observed % with marker
BC_3F_2	RxB01(2)(3)	48	20	36	42
	RxB03(2)(4)	48	18	36	38
	RxC27(20)(2)	24	15	18	63
	RxC27(28)(3)	48	25	36	52
	RxC12(1)(6)	24	17	18	73
	RxC27(17)(2)	24	10	18	42
	RxC27(5)(21)	24	12	18	50
	RxC27(2)(1)	48	26	36	54
	RxC27(10)(2)	48	23	36	48
	RxC27(20)(23)	48	17	36	35
	RxC27(20)(10)	48	23	36	48
	RxC27(20)(1)	48	26	36	54

Resynthesized *B. napus* x *B. rapa* hybrid, recurrently backcrossed to *B. rapa*; R-RI16, B-BAR6 and U-USU9 are *B. rapa* recurrent parents. † indicates line was crossed with all three recurrent parents i.e. RI16, BAR6 and USU9.

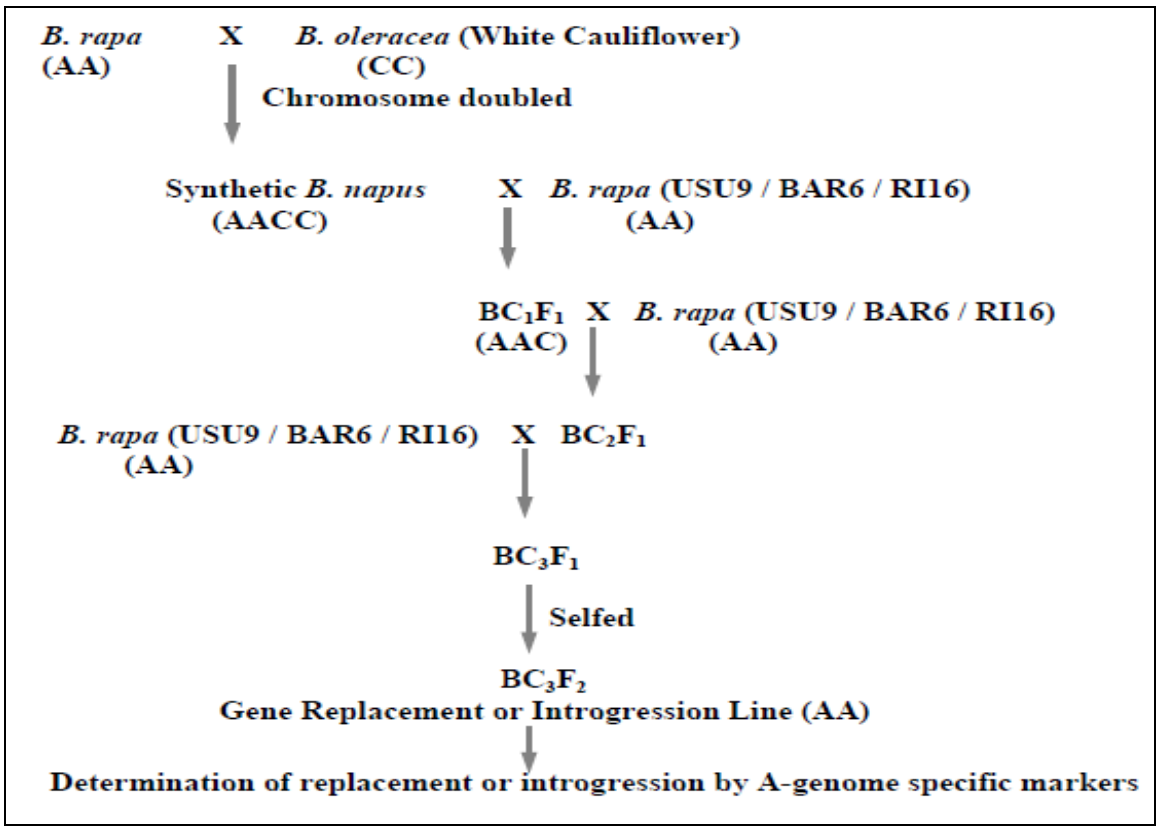


Figure 5.1 Schematic diagram for introgression or replacement of glucosinolate genes in *B. rapa* from *B. oleracea*.

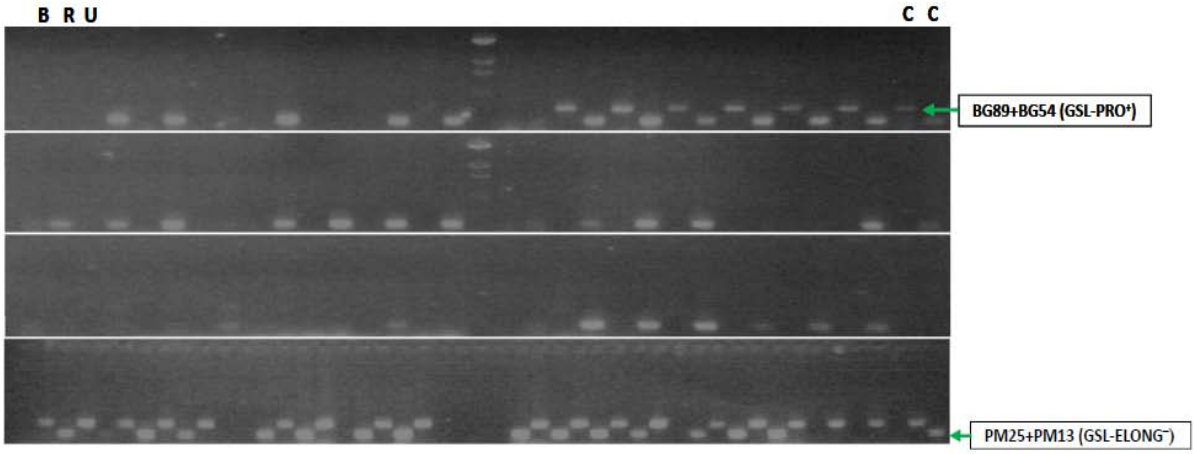


Figure 5.2 Gel image for the C-genome specific SCAR markers for *GSL-ELONG*$^{-}$ and *GSL-PRO*$^{+}$ used for selection of backcross progenies. On the top of the image, letters B, R and U represent three *B. rapa* recurrent parents BAR6, RI16, USU9, and C represents *B. oleracea* C-genome. Rest of the bands are for backcross progenies.

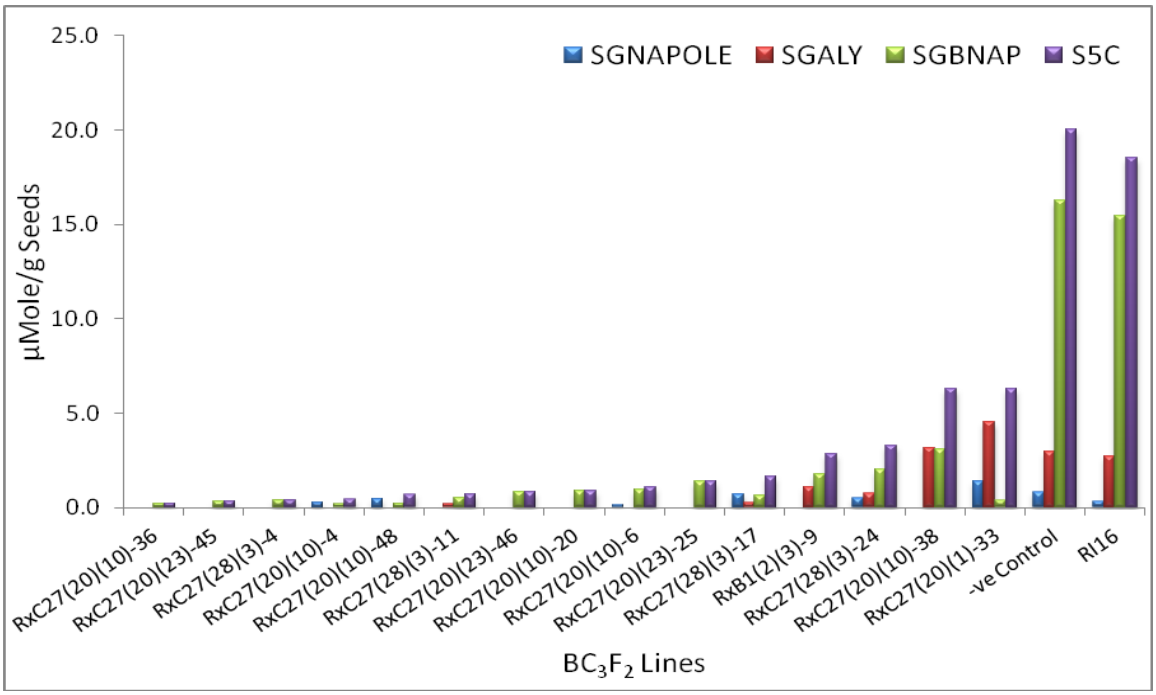


Figure 5.3 BC_3F_2 progenies of RI16 with reduction in 5C aliphatic glucosinolates in seeds. SGNAPOLE- seed gluconapoleiferin, SGALY- seed glucoalyssin, SGBNAP-seed glucobrassicanapin, and S5C- seed sum of 5C.

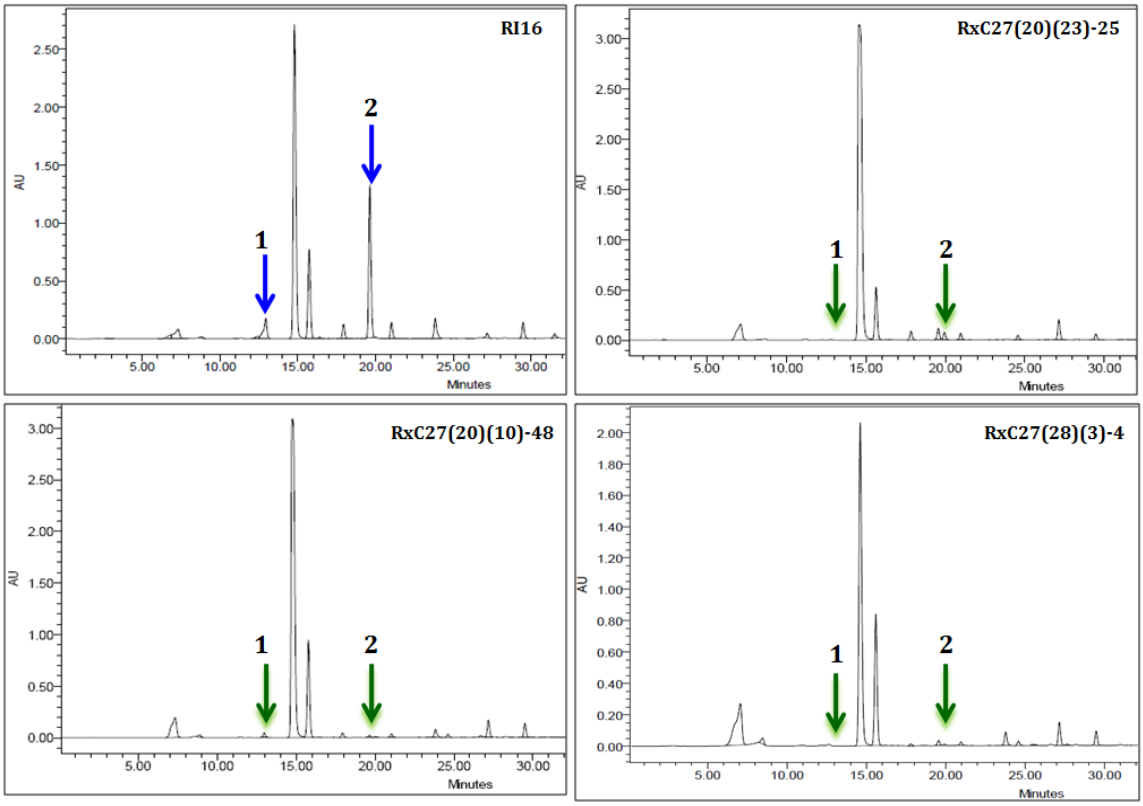


Figure 5.4 HPLC profiles of RI16 (top left) and BC_3F_2 plants of the RI16 recurrent parent (remaining three) displayed reduction of glucoalyssin and glucobrassicanapin (total 5C aliphatic glucosinolates). Arrows indicate peaks of glucoalyssin (1) and glucobrassicanapin (2).

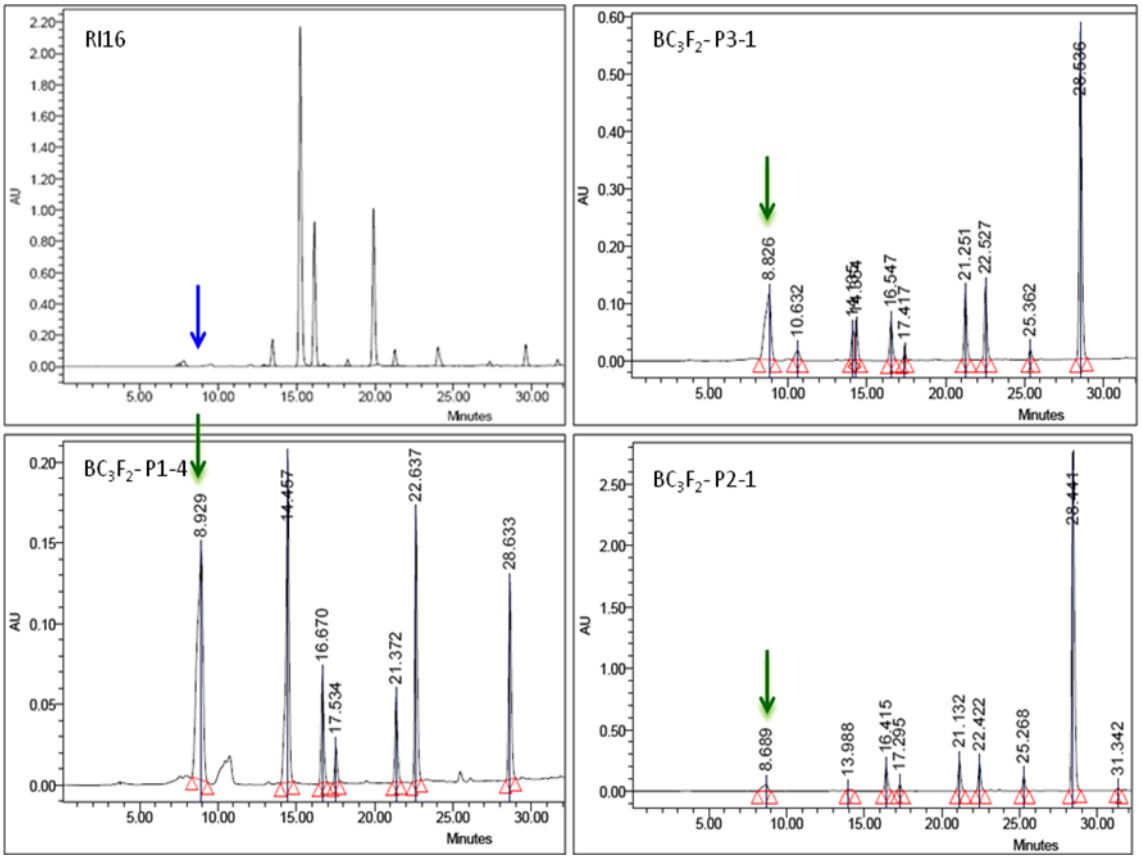


Figure 5.5 HPLC profiles of leaf glucosinolates of RI16 (top left), BC_3F_2 plants of the RI16 recurrent parent displayed a new peak of 3C glucosinolate (sinigrin). Arrows indicate the new peak of sinigrin in backcross progeneis.

6. GENERAL DISCUSSION

This thesis research was conducted to achieve the goals of QTL analysis for leaf and seed glucosinolate profile and content in two *B. rapa* RIL populations. In this study, development and integration of SCAR and SNP markers for glucosinolate biosynthesis genes in the genetic maps were used to infer the positions of large effect QTL for 4C and 5C aliphatic glucosinolates. The position of large effect QTL on linkage group A3 for 4C and 5C aliphatic glucosinolates was in agreement with previously reported studies in *B. rapa* (Lou *et al.* 2008) and *B. juncea* (Bisht *et al.* 2009). However, this is the first report of study a single genetic locus (*GSL-ELONG*) which regulates 5C aliphatic glucosinolate biosynthesis in *B. rapa*. In further support for this finding, a new approach to gene functional analysis through gene replacement or introgression accompanied by marker assisted selection was used in this study. Gene and genome specific SCAR markers were deployed for marker assisted selection to achieve glucosinolate biosynthesis gene replacement or introgression. The same *GSL-ELONG* locus discovered in the SR-RILs QTL mapping study was replaced in Chinese cabbage by homoeologous recombination in backcrosses. This resulted in the alteration of aliphatic glucosinolate profiles and contents in *B. rapa*, with 5C glucosinolate content reduced to trace amounts.

The two RIL populations developed had a common parent (BAR6), but segregated for distinct glucosinolate types (4C and 5C GSLs in SR-RILs, and glucoerucin and gluconapin in BU-RILs). QTL mapping was performed on both the RIL populations to determine large effect quantitative trait loci within the A genome. The QTL mapping study suggested that different functional genes/loci for glucosinolate biosynthesis existed in these two RIL populations.

Genetic correlation was determined for 4C and 5C aliphatic glucosinolates in SR-RILs and for glucoerucin (4C) and gluconapin (4C) aliphatic glucosinolates in BU-RILs. The genetic correlations suggested that precursor efflux and the functional characteristics of genes/loci at 4C and 5C biosynthetic branches are crucial for individual glucosinolate content. Phenotype data of the SR-RILs supported the above explanation for 4C and 5C aliphatic glucosinolate content. Presence of a single non-functional locus for 5C glucosinolate enhanced the accumulation of 4C glucosinolate in segregating SR-RILs. Similarly, *GSL-ELONG* locus replacement in backcross progenies of Chinese cabbage by marker assisted selection also enhanced 4C glucosinolate content compare to their recurrent parent. In the previous study, Zhang *et al.* (2010) also reported induction of 3C aliphatic glucosinolate in the *GSL-ELONG* gene RNAi lines of *B. napus*. These suggested that shuffling of precursors was affected by functional characteristic of locus/loci at specific biosynthesis step within the pathway.

In the SR-RIL population, QTL mapping was performed using 1,579 molecular markers including nine glucosinolate gene specific SCAR markers. Two and three replicates were used to obtain phenotypic data of leaf and seed individual glucosinolates, respectively. Three QTL were mapped on linkage group A3 for leaf and seed gluconapin, two of them from each source (LGNAP-3.2, LGNAP-3.3 for leaves and SGNAP-3.2, SGNAP-3.3 for seeds) were closely linked to the BrMAM1-1 SCAR marker. Similarly, four and two QTL were identified for leaf and seed glucobrassicanapin, respectively, and one large effect QTL from each source material (LGBNAP-3.3 for leaves and SGBNAP-3.2 for seeds) also flanked marker SCAR BrMAM1-1. This suggests that the *Br-GSL-*

ELONG locus/loci on linkage group A3 have variable functional properties or redundant function for both 4C and 5C aliphatic glucosinolate biosynthesis.

On the other hand, qualitative data for the sum of 5C aliphatic glucosinolates (glucoalyssin and glucobrassicanapin) was used to identify a major QTL involved in 5C aliphatic glucosinolate biosynthesis. The results suggest that the large effect QTL linked with the SCAR marker BrMAM1-1 on linkage group A3 is a single locus regulating 5C aliphatic glucosinolate biosynthesis in *B. rapa*. Sequence alignment of the A-genome specific *BrMAM1-1* with *A. thaliana MAM1* (*At-MAM1*) gene was performed by ClustalW2. The analysis indicated that the SCAR marker BrMAM1-1 exists within the open reading frame of the *BrMAM* gene. In *A. thaliana*, *MAM1* gene is involved in 4C and long chain (5C and larger) aliphatic glucosinolate biosynthesis.

In this study, nine molecular markers for glucosinolate biosynthesis genes were mapped on linkage groups A3, A9 and A2. Three SCAR markers, AE1, BrMAM1-1 and CE5 for *GSL-ELONG* loci were mapped; among these, BrMAM1-1 was discovered to control 5C aliphatic glucosinolates. The other two loci, corresponding to the SCAR markers AE1 and CE5, might be involved in 4C aliphatic glucosinolate biosynthesis. These SCAR markers are located within the *GSL-ELONG* gene. The absence of linkage of QTL with these SCAR markers in this study suggests a lack of allelic variation at these loci in both the parents.

Two SCAR markers, FMOFaOXRa-250 and FMOFbOXRa-250, for the *GSL-FMO_{OX1}* gene in *B. rapa* were mapped on linkage group A9. The genetic loci of *GSL-FMO_{OX1-5}* regulate the conversion of methylthiobutyl to methylsulfinylbutyl glucosinolate

in *A. thaliana* (Li *et al.* 2008) and its homologous loci might have similar functional properties in *B. rapa* and other *Brassica* species. There was no glucoerucin (methylthiobutyl) or glucoraphanin (methylsulfinylbutyl detected in the segregating SR-RILs, apparently due to the presence of highly active loci for downstream glucosinolate biosynthesis i.e. gluconapin and progoitrin. In the cultivated *Brassica* species, glucoraphanin only is reported in broccoli and purple cauliflower, both forms of *B. oleracea*. Conversely, in *B. rapa*, *B. napus* and *B. juncea*, gluconapin and progoitrin are the predominant 4C aliphatic glucosinolates with no detectable or trace amounts of glucoraphanin. The primary explanation for the lack of glucoraphanin is the presence of multiple functional *GSL-ALK/AOP* loci in the glucosinolate biosynthesis pathway in the latter three species.

In the glucosinolate side chain modification pathway, *GSL-AOP* genes (*AOP1*, *AOP2* and *AOP3*), homologs to *GSL-ALK* genes in *B. oleracea*, are involved in progoitrin biosynthesis from their precursors gluconapin and glucoraphanin. In this study, a SCAR marker, AOP1b-472 for *GSL-AOP1* on linkage group A9, and a SCAR marker AOP3aFR-397 for *GSL-AOP3* gene on linkage group A3 were mapped. The QTL analysis reported in this study is not precise enough to directly link QTL for progoitrin with candidate gene loci. A single QTL each for leaf and seed progoitrin (LPRO-3.1 and SPRO-3.1), however, were detected in the vicinity of the SCAR marker AOP3aFR-397.

In the second population, BU-RIL, QTL mapping of glucosinolates was performed using genotypic data of 288 SRAP molecular markers and four glucosinolate gene specific markers. Individual glucosinolate peaks were identified by HPLC and confirmed by LC-MS in both the parents. Glucoerucin A new peak of 4C aliphatic

glucosinolate in the canola quality DH line USU9 , which was absent in other parent, BAR6, was identified as glucoerucin. The BU-RIL population segregated for the major 4C aliphatic glucosinolates glucoerucin, gluconapin and progoitrin. There is no previous study reporting QTL analysis for both glucoerucin and glucoraphanin in *B. rapa*. Mapping of the important genetic locus/loci regulating this reaction (methylthiobutyl to methylsulfinylbutyl) would enhance our knowledge of the biosynthesis of the cancer-preventive glucosinolate compounds in *Brassica* crops. In the QTL analysis, five QTL were identified in greenhouse and field environments for seed glucoerucin which explained 4 to 45% of the phenotypic variance. Three QTL were detected for seed gluconapin in greenhouse and field environments and these explained 7 to 40% of the phenotypic variance. The single QTL (GH-SGNAP-3.1 and F-SGNAP-3.1) for gluconapin in both the environments on linkage group A3 was linked with the SCAR marker BrMAM1-1FRa-474. This QTL in BU-RIL was in agreement with the first QTL mapping study in the SR-RIL population. The common QTL to both RIL populations (SR-RIL, BU-RIL) for gluconapin infer the same *GSL-ELONG* locus on linkage group A3, which linked with the SCAR marker BrMAM1-1. Analysis of QTL in both the RIL populations (SR-RILs, BU-RILs) suggests that the *GSL-ELONG* locus on linkage group A3 is the major *MAM* gene involved in the 4C aliphatic glucosinolate biosynthesis in *B. rapa*.

Two QTL each for glucoerucin (SGERU3.1, SGERU10.1) and gluconapin (SGNAP3.2, SGNAP10.1) on linkage groups A3 and A10 at the same locations with opposite additive effects were found. This suggests that glucoerucin and gluconapin are negatively correlated to each other. Based on this negative correlation and the

biosynthesis pathway of 4C aliphatic glucosinolates, this study suggests that glucoerucin is a primary precursor for glucoraphanin, gluconapin and progoitrin. Interestingly, seed glucoerucin segregated at a 1:1 ratio in BU-RIL population. The presence of discrete high and low seed glucoerucin classes of BU-RILs in QTL analysis implies a major locus on linkage group A10. This locus is involved in glucoerucin biosynthesis with additional effects from other loci within the A-genome. The glucoerucin to glucoraphanin conversion is catalyzed by the intermediate oxygenation gene *GSL-FMO*$_{OX1}$. Several functional *GSL-AOP/ALK* loci, however, did not produce detectable levels of glucoraphanin in BU-RILs. Newly synthesized glucoraphanin appeared to be converted directly into gluconapin or progoitrin.

In addition to a major QTL for gluconapin, a QTL on linkage group A3 in the greenhouse environment and a QTL on A10 in the field environment were identified for seed progoitrin. These two QTL explained between 11 and 16% of the phenotypic variance. Furthermore, a shared QTL for seed aliphatic and seed total glucosinolates which explained 17 to 26% of the phenotypic variance was mapped. This suggested that the negative correlation between glucoerucin and gluconapin, major 4C aliphatic glucosinolates, equalized total glucosinolate content in the BU-RILs. Therefore, identification of QTL and molecular marker development for individual glucosinolates is more effective to target glucosinolate profile rather than total glucosinolate content in *B. rapa*.

Plants of *B. rapa* are highly variable for glucosinolate profile and content. However, gluconapin (4C) and glucobrassicanapin (5C) are the predominant aliphatic glucosinolates reported for the genotypes of *B. rapa*. The final portion of this research

was conducted to genetically manipulate aliphatic glucosinolate profile and content in three genotypes of *B. rapa* (RI16, BAR6 and USU9) through interspecific crosses with *B. oleracea*. Homoeologous recombination through interspecific crosses followed by marker assisted selection was performed for glucosinolate biosynthesis genes. Non-functional *GSL-ELONG*$^{-}$ and functional *GSL-PRO*$^{+}$ genes replacement and/or introgression from *B. oleracea* (white cauliflower) was attempted. In *B. oleracea* (white cauliflower), the functional characterization and comparative analysis of *GSL-ELONG* and *GSL-PRO* genes revealed that the *GSL-ELONG* is non-functional in white cauliflower. This causes a lack of 4C and 5C side chain aliphatic glucosinolates in the profile of these plants. The presence of 3C aliphatic glucosinolates indicated that the *GSL-PRO* gene is functional (Li and Quiros 2002, Gao *et al.* 2004). Based on this finding, SCAR molecular markers were used to screen interspecific backcrosses of *B. rapa* (marker assisted selection).

In the three consecutive backcross generations, variable marker transmission rates ranging from 4 to 73% were observed. Random gamete combinations might have resulted in the variable alien chromosome transmission rates. Viability of the gametes (male and female) might also not be uniform. A numbers of studies also reported variable marker transmission rates in trait introgression breeding for different crop species. Homoeologous recombinations are highly dependent on host and alien chromosome sequence homology and chromosome constitution. In this study, three *B. rapa* genotypes RI16, BAR6 and USU9 were used for backcrosses, however, RI16 seemed to be a more suitable karyotype for receipt of alien genetic materials by allosyndetical synapses than BAR6 and USU9. After reciprocal backcross and subsequent selfing, the C-genome specific markers were lost in BAR6 and USU9 backcross progenies. This indicated that

markers were lost together with additional chromosomes without homoeologous recombination events.

Fifteen BC_3F_2 progenies of the RI16 recurrent parent positive for the *GSL-ELONG*$^{-}$ markers were observed. These plants showed reduction of 5C aliphatic glucosinolates (glucoalyssin and glucobrassicanapin) relative to the parental line (RI16). In the progenies with reduced 5C glucosinolates, replacement or modification of the functional *GSL-ELONG* locus was confirmed by the disappearance of the A-genome specific SCAR marker BrMAM1-1. The same *GSL-ELONG* locus was identified as a major QTL involved in 5C aliphatic glucosinolate biosynthesis in the QTL mapping study in the SR-RIL population of *B. rapa*. This single locus correlated with the production of 20 to 25 µmole/g seed 5C aliphatic glucosinolates. Replacement of the same locus in backcross progenies of RI16 reduced 80 to 90% of 5C aliphatic glucosinolates in seeds. Two independent studies, QTL mapping in SR-RILs and *GSL-ELONG* gene replacement by marker assisted selection, confirmed that a single *GSL-ELONG* locus, located on linkage group A3 was involved in 5C aliphatic glucosinolate biosynthesis in *B. rapa*.

Three plants were also identified by marker assisted selection to be carrying the *GSL-PRO* marker. These were analyzed for leaf glucosinolates and a new small peak of sinigrin (3C) was detected. This suggested that a functional *GSL-PRO* gene introgression acted as an activator of 3C aliphatic glucosinolate biosynthesis. Sinigrin detection, however, was very low due to the presence of highly active multiple *GSL-ELONG* loci for 4C and 5C aliphatic glucosinolates. Presence of functional *GSL-ELONG* loci in the

4C and 5C glucosinolate biosynthesis pathway eliminates precursor movement towards 3C glucosinolate biosynthesis.

For future research, flanking SRAP markers to the large effect QTL for gluconapin in both the RIL populations and glucoerucin in the BU-RIL population could be used for new marker development, fine mapping of the candidate loci and functional characterization. The genomic region holding the *GSL-ELONG* locus on linkage group A3 could be cloned and functionally characterized to determine if multiple loci exist, and how they are involved in 4C and 5C aliphatic glucosinolate biosynthesis in *B. rapa*. Additionally, genetic maps constructed from the two RIL populations, SR-RILs and BU-RILs, could be integrated into a consensus map. The framework of the consensus map, with the general order of markers, could be used for more precise genetic studies of glucosinolate production and other traits in *B. rapa*. Consensus maps could be used to determine the locations of large QTLs for individual 4C aliphatic glucosinolates, particularly glucoerucin, gluconapin and progoitrin. Closely linked SSR or SNP markers could be used for future marker assisted selection or candidate loci mapping studies.

In this study, a major *GSL-ELONG* locus on linkage group A3 co-localized on the linkage map with co-dominant SCAR marker, BrMAM1-1 in the SR-RIL population. This *GSL-ELONG* locus involved in 5C aliphatic glucosinolate biosynthesis was determined by QTL mapping and confirmed by the same locus replacement in Chinese cabbage through MAS. The DNA fragment of the SCAR marker BrMAM1-1 is located within the open reading frame of the Br-MAM locus. In the sequence information of BrMAM1-1 marker, two SNPs were identified, which distinguish BAR6 and USU9 genotypes from 5C aliphatic glucosinolate producing genotype RI16 (Appendix i).

These molecular markers may also be useful tools to regulate 5C aliphatic glucosinolates in other *B. rapa* subspecies. Genetic manipulations of the *BrMAM1-1* locus through interspecific crosses followed by co-dominant marker assisted selection can be also performed in other A-genome bearing *Brassica* species such as *B. napus* and *B. juncea*. Homologous or homoeologous marker assisted selection for *BrMAM1-1* locus on linkage group A3 may help to reduce 5C aliphatic glucosinolates to the range of 15 to 25 µmole/g seed in *Brassica* species, as was seen in *B. rapa*.

7. CONCLUSIONS

Glucosinolates and their hydrolysis products have a broad range of biological functions and have been extensively studied in *Arabidopsis* and *Brassica* species. Initially glucosinolates and their hydrolysis products were studied due to their negative effects on animals. However, in the last 10-15 years, studies have also revealed human health benefits and their potential as food and plant protection agents. In the *Brassicaceae* family, *B. rapa* is an economically important crop species worldwide grown for either oil or vegetables. In addition, *B. rapa* is the diploid progenitor of two amphidiploid species, *B. napus* and *B. juncea*.

In this study, identification of QTL for glucosinolate profile and content was performed in leaves and seeds of *B. rapa* using two RIL populations, SR-RILs and BU-RILs. Nine loci were integrated on the genetic map developed from SR-RILs for the glucosinolate biosynthesis genes *GSL-ELONG*, *GSL-PRO*, *GSL-FMO*$_{OX1}$, and *GSL-AOP/ALK*. Similarly, three and one SCAR markers for *GSL-ELONG* and *GSL-PRO*, respectively, were developed and mapped on the genetic map developed from BU-RILs. Furthermore, genetic manipulation of aliphatic glucosinolate profile and content was performed through homoeologous recombination based gene replacement or introgression using marker assisted backcross selection in *B. rapa*.

In the SR-RILs, several QTL for progoitrin (2-hydroxy-3-butenyl), gluconapin (3-butenyl), glucoalyssin (5-methylsulfinylpentyl), glucobrassicanapin (4-pentenyl), 2-methylpropyl (isobutyl) and 4-hydroxy-indole-3-yl-methyl (4-hydroxyglucobrassicin) glucosinolates were identified in leaves and seeds. These QTL explained 6 to 54% of the

phenotypic variance. A candidate locus *Br-GSL-ELONG* on linkage group A3 was identified using a co-dominant SCAR marker BrMAM1-1 located within the coding region of the *BrMAM* gene. This locus was found to co-segregate with 5C aliphatic glucosinolates in *B. rapa*. Co-localization of this locus with the QTL identified for leaf and seed gluconapin indicated that this gene locus might have multi-functional properties or functional redundancy.

In the BU-RILs, a genetic map was developed using SRAP molecular markers. Glucoerucin and gluconapin were confirmed as major 4C aliphatic glucosinolates by LC-MS in both parents. The BU-RILs segregated for glucoerucin, gluconapin and progoitrin 4C aliphatic glucosinolates. Phenotype evaluation was performed in greenhouse and field environments for seed glucosinolates and in a greenhouse environment for leaf glucosinolates. Analysis of variance suggested that there were no significant differences between two environments for seed glucoerucin, gluconapin, sum of 5C and total aliphatic glucosinolates. However, differences for seed progoitrin, sum of 4C, indole and total glucosinolates were observed. This suggests that individual glucosinolates are highly influenced by environmental factors.

In the QTL mapping study for the BU-RIL population, five QTL for seed glucoerucin and three QTL for seed gluconapin which explained 4 to 45% of the phenotypic variance were identified in both environments. Two common QTL each for seed glucoerucin and seed gluconapin on linkage group A3 and A10 were identified with opposite additive effects, which was also supported by a negative correlation between them. However, the QTL for seed glucoerucin on linkage group A10 suggested a major

candidate locus due to the nature of the data. This major locus on linkage group A10 might have variable interactions with other loci that determine final glucoerucin content.

Genetic manipulation of aliphatic glucosinolate profile and content through homoeologous *GSL-ELONG* and *GSL-PRO* gene replacement or introgression was performed in three genotypes of *B. rapa*, RI16, BAR6 and USU9 interspecific crossing followed by marker assisted backcross selection. Marker transmission rates of the *GSL-ELONG*$^{-}$ gene were 4 to 73% in the backcross progenies. In advance backcross progenies of the RI16 recurrent parent, 15 (30%) plants had reduced 5C aliphatic glucosinolates. These plants were positive for the C-genome and *GSL-ELONG*$^{-}$ gene specific marker. The same plants lost the A-genome and *GSL-ELONG*$^{+}$ (*BrMAM1-1*) gene specific marker BrMAM1-1 on linkage group A3. The results verified the hypothesis that the glucobrassicanapin and glucoalyssin (total 5C aliphatic glucosinolates) are controlled by a single locus in *B.rapa*. These plants (backcross progenies of RI16 recurrent parent) had 5C aliphatic glucosinolates reduced by 15 to 25 µmole/g seed relative to their recurrent parent RI16. This supports the hypothesis that the aliphatic glucosinolate profile and content in *B. rapa* can be manipulated by replacement or introgression of glucosinolate biosynthesis genes from *B. oleracea* through backcrosses using marker assisted selection. Several other backcross progenies carrying the C-genome specific non-functional allele (*GSL-ELONG*$^{-}$) introgressed from the C-genome chromosome did not change 5C aliphatic glucosinolate profile and content in Chinese cabbage. Some advanced backcross progenies carrying a functional *GSL-PRO*$^{+}$ gene produced the 3C aliphatic glucosinolate sinigrin. This suggests that introgression of a functional *GSL-PRO*$^{+}$ gene in these plants activated the 3C aliphatic glucosinolate biosynthesis pathway.

In future research, investigation can use the SCAR marker BrMAM1-1 (along with two SNPs) in high and low 5C aliphatic glucosinolates containing genotypes and inbred lines of the A-genome containing *Brassica* species i.e. *B. rapa, B. napus* and *B. juncea*. This could determine the existence of mutations within the open reading frame of the *GSL-ELONG* locus that alters 5C aliphatic glucosinolate biosynthesis. These markers could be employed for marker assisted selection in interspecific or intergenomic crosses of the A-genome containing *Brassica* species to reduce 5C aliphatic glucosinolates up to 25 µmol in seeds.

8. LITERATURE CITED

Agrawal AA, Kurashige NS (2003) A role for isothiocyanates in plant resistance against the specialist herbivore *Pieris rapae*. J Chem Ecol 29:1403-1415

Ajisaka H, Kuginuki Y, Hida K, Enomoto S, Hirai M (1995) A linkage map of DNA markers in *Brassica campestris*. Breed Sci 45(suppl.):195–197

Ali SNH, Ramanna MS, Jacobsen E, Visser RGF (2001) Establishment of a complete series of a monosomic tomato chromosome addition lines in the cultivated potato using RFLP and GISH analyses. Theor Appl Genet 103:687-695

Attia T, Röbbelen G (1986) Cytogenetic relationship within cultivated *Brassica* analyzed in amphihaploids from the three diploid ancestors. Can J Genet Cytol 28:323-329

Attia T, Röbbelen G (1987) Digenomic triploids for an assessment of chromosome relationships in the cultivated diploid *Brassica* species. Genome 29:326-330

Banga SS (1988) C-genome chromosome substitution lines in *Brassica juncea* (L.) Coss. Genetica 77:81-84

Barth C, Jander G (2006) *Arabidopsis* myrosinases *TGG1* and *TGG2* have redundant function in glucosinolate breakdown and insect defense. Plant J 46:549-562

Basten CJ, Weir BS, Zeng ZB (1996) QTL Cartographer. North Carolina State University, Raleigh, NC

Bauer R, Birch ANE, Hopkins RJ, Griffiths DW, Simmonds MSJ, Städler E (1996) Oviposition and chemosensory stimulation of the root flies *Delia radicum* and *D. floralis* in response to plant and leaf surface extracts from resistant and susceptible *Brassica* genotypes. Entomol Exp Appl 78:61-75

Beekwilder J, Leeuwen W, Dam NM, Bertossi M, Grandi V, Mizzi L, Soloviev, Szabados L, Molthoff JW, Schipper B, Verbocht H, Vos RCH, Morandini P, Aarts MGM, Bovy A (2008) The impact of the absence of aliphatic glucosinolates on insect herbivory in *Arabidopsis*. PLoS One 3(4):e2068

Beilstein MA, Al-Shehbaz IA, Kellogg EA (2006) *Brassicaceae* phylogeny and trichome evolution. Am J Bot 93:607-619

Benderoth M, Textor S, Windsor AJ, Mitchell-Olds T, Gershenzon J, Kroymann J (2006) Positive selection driving diversification in plant secondary metabolism. Proc Natl Acad Sci USA 103(24):9118-9123

Birch ANE, Griffiths DW, Hopkins RJ, Smith WHM, McKinlay RG (1992) Glucosinolate responses of swede, kale, forage and oilseed rape to root damage by turnip root fly (*Delia floralis*) larvae. J Sci Food Agric 60:1-9

Bisht NC, Gupta V, Ramchiary N, Sodhi YS, Mukhopadhyay A, Arumugam N, Pental D, Pradhan AK (2009) Fine mapping of loci involved with glucosinolate biosynthesis in oilseed mustard (*Brassica juncea*) using genomic information from allied species. Theor Appl Genet 118:413-421

Bodnaryk RP (1992) Effects of wounding on glucosinolates in the cotyledons of oilseed rape and mustard. Phytochem 31:2671-2677

Bones AM, Rossiter JT (1996) The myrosinase-glucosinolate system, its organization and biochemistry. Physiol Plant 97:194-208

Bowers JE, Chapman BA, Rong JK, Paterson AH (2003) Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. Nature 422:433-438

Brader G, Mikkelsen MD, Halkier BA, Palva ET (2006) Altering glucosinolate profiles modulates disease resistance in plants. Plant J 46:758-767

Brader G, Tas É, Palva ET (2001) Jasmonate-dependent induction of indole glucosinolates in *Arabidopsis* by culture filtrates of the nonspecific pathogen *Erwinia carotovora*. Plant Physiol 126:849-860

Brassica genome gateway (Online database) URL. http://brassica.bbsrc.ac.uk (March 15 2007-2010)

Bridges M, Jines AME, Bones AM, Hodgson C, Cole R, Bartlet E, Wallsgrove R, Karapapa VK, Watts N, Rossiter JT (2002) Spatial organization of the glucosinolate-myrosinase system in *Brassica* specialist aphids is similar to that of the host plant. Proc Biol Sci 269(1487):187-191

Brooks J, Paton V, Vidanes G (2001) Potent induction of phase 2 enzymes in human prostate cells by sulforaphane. Cancer Epidemiol Biomarkers Prev 10:949-954

Brown PD, Tokuhisa JG, Reichelt M, Gershenzon J (2003) Variation of glucosinolate accumulation among different organs and developmental stages of *Arabidopsis thaliana*. Phytochem 62:471-481

Brudenell AJP, Griffiths H, Rossiter JT, Baker DA (1999) The phloem mobility of glucosinolates. J Exp Botany 50(335):745-756

Budahn H, Schrader O, Peterka H (2008) Development of a complete set of disomic rape-radish chromosome-addition line. Euphytica 162:117-128

Canola Council of Canada (2010a) Canola-official definition. (Online database) URL. http://www.canola-council.org/canola_the_official_definition.aspx (August 25, 2010)

Canola Council of Canada (2010b) Canola in Canada. (Online database) URL. http://www.canolacouncil.org/canadian_canola_industry.aspx (August 25, 2010)

Chen BY, Simonsen V, Lanner-Herrera C, Hennen WK (1992) A *Brassica campestris-B.alboglabra* addition line and its use for gene mapping, intergenomic gene transfer and generation of trisomics. Theor Appl Genet 84:592-599

Chen S, Glawischning E, Jørgensen K, Naur P, Jørgensen B, Olsen CE, Hansen CH, Rasmussen H, Pickett JA, Halkier BA (2003) *CYP79F1* and *CYP79F2* have distinct functions in the biosythesis of aliphatic glucosinolates in *Arabidopsis*. Plant J 33:923-937

Cheng BF, Chen BY, Heneen WK (1994) Addition of *Brassica alboglabra* Bailey chromosomes to *B. campestris* L. with special emphasis on seed colour. Heredity 73:185-189.

Cheng Y, Geng JF, Zhang J, Wang Q, Hou X (2009) The construction of a genetic linkage map of non-heading Chinese cabbage (*Brassica campestris* ssp. *chinensis* Makino). J of Genetics and Genomics 36(8):501-508

Cheung WY, Landry BS, Raney P, Rakow GFW (1998) Molecular mapping of seed quality traits in *Brassica juncea* L. Czern., and Coss. Acta Hort 459:139-147

Choi IY, Hyten DL, Matukumalli LK, Song Q, Chaky JM, Quigley CV, Chase K, Lark KG, Reiter RS, Yoon MS, Hwang EY, Yi SI, Young ND, Shoemaker RC, Van Tassell CP, Specht JE, Cregan PB (2007) A soybean transcript map: gene distribution, haplotype and single-nucleotide polymorphism analysis. Genetics 176(1):685-696

Compos de Quiros H, Magrath R, McCallum D, Kroymann J, Schnabelrauch D, Mitchell-Olds T, Mithen R (2000) α-Keto acid elongation and glucosinolate biosynthesis in *Arabidopsis thaliana.* Theor Appl Genet 101:429-437

Danielak R, Borkowski B (1969) Biologically active compounds in seeds of crucifers Part III. Chromatographical search for glucosinolates. Dissert in Pharm and Pharmacol 21:563-575

De Vos M, Kriksunov KL, Jander G (2008) Indole-3-Acetonitrile production from indole glucosinolates deters oviposition by *Pieris rapae*. Plant Physiol 146:916-926

Doerge RW, Churchill GA (1996) Permutation test for multiple loci affecting a quantitative character. Genetics 142:285-294

Dover GA, Riley R (1972) Prevention of pairing of homoeologous meiotic chromosomes of wheat by an activity of supernumerary chromosomes of *Aegilops*. Nature 24:159-161

Downey RK (1990) *Brassica* oilseed breeding – achievements and opportunities. Plant Breed 60:1165-1170

Downey RK, Röbellen G (1989) *Brassica* species. In G. Röbellen, R. K. Downey and A. Ashri, (eds) Oil crops of the world, McGraw Hill Publishing Company, New York, USA, pp. 339-362

Downey RK, Stringam GR, McGregor DI, Stefansson BR (1975) Breeding rapeseed and mustard crops. In: J.T. Harapiak (eds) Oilseed and pulse crops in western Canada. Western Cooperative Fertilizers Ltd., Calgary, Alberta, Canada, pp 157-183

Drozdowska L, Thangstad OP, Beisvaag T, Evjen K, Bones A, Iversen TH (1992) Myrosinase and myrosin cell development during embryogenesis and seed maturation. Israel J B 41:213-223

Du L, Halkier BA (1998) Biosynthesis of glucosinolates in the developing silique walls and seeds of *Sinapis alba*. Phytochem 48(7):1145-1150

Fahey JW, Haristoy X, Dolan P, Kensler T, Scholtus I, Stephenson K, Talalay P, Lozniewski A (2002) Sulforaphane inhibits extracellular, intracellular, and antibiotic-resistant strains of *Helicobacter pylori* and prevents benzo[α]pyrene-induced stomach ulcers. Proc Natl Acad Sci USA 99:7610-7615

Fahey JW, Zalcmann AT, Talalay P (2001) The chemical diversity and distribution of glucosinolates and isothiocyanates among plants. Phytochem 56:5-51

FAO (2008) FAOSTAT. (Online database) URL. http://faostat.fao.org/site/339/default.aspx (November 7, 2010)

Faulkner K, Mithen R, Williamson G (1998) Selective increase of the potential anticarcinogen 4-methylsulphinylbutyl glucosinolate in broccoli. Carcinogenesis 19:605-609

Feldman M (1966) The effect of chromosomes 5B, 5D, and 5A on chromosomal pairing in *Triticum aestivum*. Proc Natl Acad Sci USA 55:1447-1453

Fenwick GR, Heaney RK, Mullin WJ (1983) Glucosinolates and their breakdown products in food and food plants. Crit Rev Food Sci Nutr 18:123-201

Field B, Cardon G, Traka M, Botterman J, Vancammeyt G, Mithen R (2004) Glucosinolate and amino acid biosynthesis in *Arabidopsis*. Plant Physiol 135:828-839

Flint-Garcia SA, Jampatong C, Darrah LL, McMullen MD (2003) Quantitative trait locus analysis of stalk strength in four maize populations. Crop Sci 43:13-22

Gao M, Li G, Potter D, McCombie WR, Quiros CF (2006) Comparative analysis of *methylthioalkylmalate synthase* (*MAM*) gene family and flanking DNA sequences in *Brassica oleracea* and *Arabidopsis thaliana*. Plant Cell Rep 25:592-598

Gao M, Li G, Yang B, McCombie WR, Quiros CF (2004) Comparative analysis of a *Brassica* BAC clones containing several major aliphatic glucosinolate genes with its corresponding *Arabidopsis* sequence. Genome 47:666-679

Gao X, Dinkova-Kostova A, Talalay P (2001) Powerful and prolonged protection of human retinal pigment epithelial cells, keratinocytes, and mouse leukemia cells against oxidative damage: the indirect antioxidant effects of sulforaphane. Proc Natl Acad Sci USA 98:15221-15226

Garriga-Calderé F, Huigen DJ, Angrisano A, Jacobsen E, Ramanna MS (1998) Transmission of alien tomato chromosomes from BC_1 to BC_3 progenies derived from backcrossing potato (+) tomato fusion hybrids to potato: The selection of single additions for seven different tomato chromosomes. Theor Appl Genet 96:155-163

Geng JF, Zhu CS, Zhang XW, Cheng Y, Zhang YM, Hou XL (2007) A genetic linkage map of non-heading Chinese cabbage. J Amer Soc Hort Sci 132:816–823

Genographer (Online database) URL. http://hordeum.oscs.montana.edu/software/genographer/index.html (July 03, 2008)

Greer MA (1962) The isolation and identification of progoitrin from *Brassica* seed. Arch of Biochem and Biophy 99:369-371

Griffiths DW, Bain H, Deighton N, Botting NP, Robertson AAB (2000) Evaluation of liquid chromatography-atmospheric pressure chemical ionisation-mass spectrometry for the identification and quantification of desulphoglucosinolates. Phytochem Anal 11:216-225

Griffiths DW, Birch ANE, Hillman JR (1998) Anti-nutritional compounds in the *Brassicaceae* – analysis, biosynthesis, chemistry and dietary effects. J Hort Sci Biotech 73:1-18

Griffiths DW, Birch ANE, MacFarlane-Smith WH (1994) Induced changes in the indole glucosinolate content of oil-seed and forage rape (*Brassica napus*) plant in response to either turnip root fly (*Delia floralis*) larval feeding or artificial root damage. J Sci Food Agric 63:171-178

Grubb CD, Abel S (2006) Glucosinolate metabolism and its control. Trends Plant Sci 11(2):89-100

Guo ZH, Dickson MH, Hunter JE (1990) *Brassica napus* sources of resistance to black rot in crucifers and inheritance of resistance. In: Proc 6th Crucifer Genetics Workshop, Ithaca New York, USA, pp 154-155

Gupta SK, Pratap A (2007) History, origin and evolution. In: Gupta SK (eds) Advances in botanical research, rapeseed breeding. Academic Press, Elsevier, San Diego, CA, USA, pp 2-17

Halkier BA (1999) Glucosinolates. In: Ikan R (eds) Naturally occurring glycosides: chemistry, distribution and biological properties. John Wiley and Sons Ltd, London, UK, pp. 193-223.

Halkier BA, Gershenzon J (2006) Biology and biochemistry of glucosinolates. Annu Rev Plant Biol 57:303-333

Hall C, McCallum D, Prescott A, Mithen R (2001) Biochemical genetics of glucosinolate modification in *Arabidopsis* and *Brassica*. Theor Appl Genet 102:369-374

Hansen BG, Kliebenstein DJ, Halkier BA (2007) Identification of a flavin-monooxygenase as the S-oxygenating enzyme in aliphatic glucosinolate biosynthesis in *Arabidopsis*. Plant J 50:902-910

Hasterok R, Wolny E, Kulak S, Zdziechiewicz A, Maluszynska J, Heneen WK (2005) Molecular cytogenetic analysis of *Brassica rapa-Brassica oleracea* var. *alboglabra* monosomic addition lines. Theor Appl Genet 111:196-205

Haughn GW, Davin L, Giblin M, Underhill EW (1991) Biochemical genetics of plant secondary metabolites in *Arabidopsis thaliana*. Plant Physiol 97:217-226

He H, Fingerling G, Schnitzler WH (2000) Glucosinolate contents and patterns in different organs of Chinese cabbages, Chinese kale (*Brassica alboglabra* bailey)

and Choy sum (*Brassica campestris* L. spp *chinensis* Var. *Utilis* tsen et lee). J App Bot 74:21-25

Heath DW, Earle ED, Dickson MH (1994) Introgressing cold-tolerant Ogura cytoplasms from rapeseed into Pak choi and Chinese cabbage. Hort Sci 29:202-203

Heneen WK, Jørgensen RB (2001) Cytology, RAPD, and seed colour of progeny plants from *Brassica alboglabra* aneuploids and development of monosomic addition lines. Genome 44:1007-1021

Hirai M, Harada T, Kubo N, Tsukada M (2004) A novel locus for clubroot resistance in *Brassica rapa* and its linkage markers. Theor Appl Genet 108:639-643

Hogge LR, Reed DW, Underhill EW, Haughn GW (1988) HPLC separation of glucosinolates from leaves and seeds of *Arabidopsis thaliana* and their identification using thermospray liquid chromatography-mass spectrometry. J Chromatgr Sci 26:551-556

Hopkins RJ, Birch ANE, Griffiths DW, Baur R, Städler E, McKinlay RG (1997) Leaf surface compounds and the oviposition preference of the turnip root fly (*Delia floralis*): The role of glucosinolate and nonglucosinolate compounds. J Chem Ecol 23:629-643

Howell PM, Sharpe AG, Lydiate DJ (2003) Homoeologous loci control the accumulation of seed glucosinolates in oilseed rape (*Brassica napus*). Genome 46:454-460

Hrncirik K, Valusek J, Velisek J (2001) Investigation of ascorbigen as a breakdown product of glucobrassicin autolysis in *Brassica* vegetables. Eur Food Res Technol 212:576-581

Hu J, Quiros CF (1991) Molecular and cytological evidence of deletions in alien chromosomes for two monosomic addition lines of *Brassica campestris-oleracea*. Theor Appl Genet 90:258-262

Hwang TY, Sayama T, Takahashi M, Takada Y, Nakamoto Y, Funatsuki H, Hisano H, Sasamoto S, Sato S, Tabata S, Kono I, Hoshi M, Hanawa M, Yano C, Xia Z, Harada K, Kitamura H, Ishimoto M (2009) High-density integrated linkage map based on SSR markers in soybean. DNA Research 16:213-225

Jenczewski E, Eber F, Grimaud A, Huet S, Lucas MO, Monod H, Chevre AM (2003) *PrBn*, a major gene controlling homeologous pairing in oilseed rape (*Brassica napus*) haploids. Genetics 164:645-653

Kaneko Y, Yano H, Bang SW, Matsuzawa Y (2003) Genetic stability and maintenance of *Raphanus sativus* lines with an added *Brassica rapa* chromosome. Plant Breed 122:239-243

Kapoor R, Banga SS, Banga SK (2009) A microsatellite (SSR) based linkage map of *Brassica rapa*. New biotechnology 15:1-5

Kiddle G, Bennett RN, Botting NP, Davidson NE, Robertson AAB, Wallsgrove RM (2001) High-performance liquid chromatographic separation of natural and synthetic desulfoglucosinolates and their chemical validation by UV, NMR and chemical ionisatioin-MS methods. Phytochem Anal 12:226-242

Kim JH, Jander G (2007) *Myzus persicae* (green peach aphid) feeding on *Arabidopsis* induces the formation of a deterrent indole glucosinolates. Plant J 49:1008-1019

Kim JK, Chu SM, Kim SJ, Lee DJ, Lee SY, Lim SH, Ha SH, Kweon SJ, Cho HS (2010) Variation of glucosinolates in vegetable crops of *Brassica rapa* L. ssp. *pekinensis*. Food Chem 119:423-428

Kim JS, Chung TY, King GJ, Jin M, Yang T, Jin Y, Kim H, Park B (2006) A sequence tagged linkage map of *Brassica rapa*. Genetics 174:29–39

King IP, Forster BP, Law CC, Cant KA, Orford SE, Gorham J, Reader S, Miller T (1997) Introgression of salt-tolerance genes from *Thinopyrum bessarabicum* into wheat. New Physiol 137 (1):75-81

Klein M, Reichelt M, Gershenzon J, Papenbrock J (2006) The three desulfoglucosinolate sulfotransferase proteins in *Arabidopsis* have different substrate specificities and are differentially expressed. FEBS J 273(1):122-136

Kliebenstein DJ (2008) A role for gene duplication and natural variation of gene expression in the evolution of metabolism. PLoS One 3(3):e1838

Kliebenstein DJ, Figuth A, Mitchell-Olds T (2002) Genetic architechture of plastic methyl jasmonte responses in *Arabidopsis thaliana*. Genetics 161:1685-1696

Kliebenstein DJ, Gershenzon J, Mitchell-Olds T (2001a) Comparative quantitative trait loci mapping of aliphatic, indole and benzylic glucosinolate production in *Arabidopsis thaliana* leaves and seeds. Genetics 159:359-370

Kliebenstein DJ, Kroymann J, Brown P, Figuth A, Pedersen D, Gershenzon J, Mitchell-Olds T (2001b) Genetic control of natural variation in *Arabidopsis* glucosinolate accumulation. Plant Physiol 126:811-825

Kliebenstein DJ, Lambrix VM, Reichelt M, Gershenzon J, Mitchell-Olds T (2001c) Gene duplication in the diversification of secondary metabolism: tandem 2-oxoglutarate-dependent dioxygenases control glucosinolate biosynthesis in *Arabidopsis*. Plant Cell 13:681-693

Kole C, Quijada P, Michaels SD, Amasino RM, Osborn TC (2001) Evidence for homology of flowering-time gene *VFR2* from *Brassica rapa* and *FLC* from *Arabidopsis thaliana*. Theor Appl Genet 102:425-430

Kole C, Williams PH, Rimmer SR, Osborn TC (2002) Linkage mapping of genes controlling resistance to white rust (*Albugo candida*) in *Brassica rapa* (syn. *campestris*) and comparative mapping to *Brassica napus* and *Arabidopsis thaliana*. Genome 45:22-27

Kondra ZP, Stefannson BR (1970) Inheritance of major glucosinolates in rapessed (*Brassica napus*) meal. Can J Plant Sci 50:643-647

Kraker JW, Luck Katrin, Textor S, Tokuhisa JG and Gershenzon J (2007) Two *Arabidopsis* gene (*IPMS1* and *IPMS2*) encode isopropylmalate syntase, the branchpoint step in biosynthesis of leucin. Plant Physiol 143:970-986

Kroymann J, Textor S, Tokuhisa JG, Falk KL, Bartram S, Gershenzon J, Mitchell-Olds T (2001) A gene controlling variation in *Arabidopsis* glucosinolate composition is part of the methionine chain elongation pathway. Plant Physiol 127:1077-1088

Kuraparthy V, Sood S, Gill BS (2009) Molecular genetic description of the cryptic wheat-*Aegilops geniculata* introgression carrying rust resistance genes *Lr57* and *Yr40* using wheat ESTs and synteny with rice. Genome 52:1025-1036

Lan TH, DelMonte TA, Reischmann KP, Hyman J, Kowalski SP, McFerson J, Kresovich S, Paterson AH (2000) An EST-enriched comparative map of *Brassica oleracea* and *Arabidopsis thaliana*. Genome Res 10:776-788

Leflon M, Eber F, Letanneur JC, Chelysheva L, Coriton O, Huteau V, Ryder CD, Barker G, Jenezewski E, Chevre AM (2006) Pairing and recombination at meiosis of *Brassica rapa* (AA) X *Brassica napus* (AACC) hybrids. Theor Appl Genet 113:1467-1480

Li G, Gao M, Yang B, Quiros CF (2003) Gene for gene alignment between the *Brassica* and *Arabidopsis* genomes by direct transcriptome mapping. Theor Appl Genet 107:168-180

Li G, Quiros CF (2001) Sequence-related amplified polymorphism (SRAP), a new marker system based on a simple PCR reaction: its application to mapping and gene tagging in *Brassica*. Theor Appl Genet 103:455-461

Li G, Quiros CF (2002) Genetic analysis, expression and molecular characterization of *Bo-GSL-ELONG*, a major gene involved in the aliphatic glucosinolate pathway of *Brassica* species. Genetics 162:1937-1943

Li G, Quiros CF (2003) In planta side-chain glucosinolate modification in *Arabidopsis* by introduction of dioxygenase *Brassica* homolog *BoGSL-ALK*. Theor Appl Genet 106:1116-1121

Li J, Hansen BG, Ober JA, Kliebenstein DJ, Halkier BA (2008) Subclade of flavin-monooxygenases involved in aliphatic glucosinolate biosynthesis. Plant Physiol 148:1721-1733

Linkage groups assignments (Online database) URL. http://www.brassica.info/resource/maps/lg-assignments.php (April 2010)

Lou P, Zhao J, He H, Hanhart C, Del Carpio DP, Verkerk R, Custers J, Koornneef M, Bonnema G (2008) Quantitative trait loci for glucosinolate accumulation in *Brassica rapa* leaves. New Phytol 179:1017-1032

Lou P, Zhao JJ, Kim JS, Shen SX, Del Carpio DP, Song XF, Jin M, Vreugdenhil D, Wang XW, Koornneef M, Bonnema G (2007) Quantitative trait loci for flowering time and morphological traits in multiple populations of *Brassica rapa*. J Expt Bot 58:4005-4016

Lowe AJ, Moule C, Trick M, Edwards KJ (2004) Efficient large-scale development of microsatellites for marker and mapping applications in *Brassica* crop species. Theor Appl Genet 108(6):1003-1112

Lukens L, Zou F, Lydiate D, Parkin IAP, Osborn T (2003) Comparison of a *Brassica oleracea* genetic map with the genome of *Arabidopsis thaliana*. Genetics 164:359-372

Lysak MA, Koch M, Pecinka A, Schubert I (2005) Chromosome triplication found across the tribe *Brassiceae*. Genome Res 15:516-525

Magrath R, Bano F, Morgner M, Parkin I, Sharpe A, Lister C, Dean C, Turner J, Lydiate D, Mithen R (1994) Genetics of aliphatic glucosinolates. I. Side chain elongation in *Brassica napus* and *Arabidopsis thaliana*. Heridity 72:290-299

Mahmood T, Ekuere U, Yeh F, Good AG, Stringam GR (2003) Molecular mapping of seed aliphatic glucosinolates in *Brassica juncea*. Genome 46:753-760

Martinez MN, Cuñado N, Carcelén N, Romero C (2001) The *Ph1* and *Ph2* loci play different roles in the synaptic behaviour of hexaploid wheat *Triticum aestivum*. Theor Appl Genet 103:398-405

McCouch SR, Teytelman L, Xu Y, Lobos KB, Clare K, Walton M, Fu B, Maghirang R, Li Z, Xing Y, Zhang Q, Kono I, Yano M, Fjellstrom R, Declerck G, Schneider D, Cartinhour S, Ware D, Stein L (2002) Development and mapping of 2240 new SSR markers for rice (*Oryza sativa* L.) DNA Research 9:199-207

McGrath JM, Quiros CF (1990) Generation of alien chromosome addition lines from synthetic *Brassica napus*: morphology, cytology, fertility, and chromosome transmission. Genome 33:374-383

McVetty PBE, Fernando D, Li G, Tahir M, Zelmer C (2009) High-erucic acid, and low-glucosinolate rapeseed (HEAR) cultivar development in Canada. In: Hou CT, Shaw JF (eds) Biocatalysis and agricultural biotechnology. CRC, Boca Raton, FL, USA pp 43-61

McVetty PBE, Scarth R (2002) Breeding for improved oil quality in *Brassica* oilseed species. J Crop Prod 5:345-369

Mehra KL (1966) History and ethnobotany of mustard in India. Advancing Frontiers of Plant Sciences 19:51-59

Mewis IZ, Ulrich C, Schnitzler WH (2002) The role of glucosinolates and their hydrolysis products in oviposition and host-plant finding by cabbage webworm, *Hellula undalis*. Entomol Exp Appl 105:129-139

Mikkelsen MD, Hansen CH, Wittstock U, Halkier A (2000) Cytochrome P450 *CYP79B2* from *Arabidopsis* catalyzes the conversion of tryptophan to indole-3-acetaldoxime, a precursor of indole glucosinolates and indole-3-acetic acid. J Bio Chem 275 (43):33712-33717

Mikkelsen MD, Naur P, Halkier BA (2004) *Arabidopsis* mutants in the C-S lyase of glucosinolate biosynthesis establish a critical role for indole-3-acetaldoxime in auxin homeostasis. Plant J 37:770-77

Mikkelsen TR, Jensen J, Jorgensen RB (1996) Inheritance of oilseed rape (*Brassica napus*) RAPD markers in a backcross progeny with *Brassica campestris*. Theo Appl Genet 92:492-497

Mithen R, Faulkner K, Magrath R, Rose P, Williamson G, Marquez J (2003) Development of isothiocyanates-enriched broccoli, and its enhanced ability to induce phase 2 detoxification enzymes in mammalian cells. Theor Appl Genet 106:727-734

Mithen RF, Clarke J, Lister C, Dean C (1995) Genetics of aliphatic glucosinolates. III. Side chain structure of aliphatic glucosinolates in *Arabidopsis thaliana*. Heredity 74:210-215

Mithen RF, Dekker M, Verkerk R, Rabot S, Johnson IT (2000) The nutritional significance, biosynthesis and bioavailability of glucosinolates in human foods. J Sci Food Agri 80:967-984

Morinaga T (1934) Interspecific hybridization in *Brassica*. VI. Then cytology of F_1 hybrid of *B. juncea* and *B. nigra*. Cytologia 6:62-67

Muangprom A, Osborn TC (2004) Characterization of a dwarf gene in *Brassica rapa*, including the identification of a candidate gene. Theor Appl Genet 108:378-384

Müller C, Sileling N (2006) Effects of glucosinolate and myrosinase levels in *Brassica juncea* on a glucosinolate-sequestering herbivor-and vice versa. Chemoecol 16:191-201

Mun JH, Kwon SJ, Yang TJ, Seol YJ, Jin M, Kim JA, Lim MH, Kim JS, Baek S, Choi BS, Yu HJ, Kim DS, Kim N, Lim KB, Lee SI, Hahn JH, Lim YP, Bancroft I, Park BS (2009) Genome-wide comparative analysis of the *Brassica rapa* gene space reveals genome shrinkage and differential loss of duplicated genes after whole genome triplication. Genome Biol 10(10):R111

Nagoaka T, Daullah MAU, Matsumoto S, Kawasaki S, Ishikawa T, Hir H, Okazaki K (2010) Identification of QTLs that control resistance *in Brassica oleracea* and comparative analysis of clubroot resistance genes between *B. rapa* and *B. oleracea*. Theor Appl Genet 120(7):1335-1346

Navabi ZK, Strelkov SE, Good AG, Thiagarajah MR, Rahman MH (2010) *Brassica* B-genome resistance to stem rot (*Sclerotinia sclerotiorum*) in a double haploid population of *Brassica napus* x *Brassica carinata*. Can J Plant Path 32(2):237-246

NCBI (Online database) URL. http://www.ncbi.nlm.nih.gov/nucleotide/ (May 25, 2007-2010)

Nestle M (1997) Broccoli sprouts as inducers of carcinogen-detoxifying enzyme systems; clinical, dietary, and policy implications. Proc Natl Acad Sci USA 94:11149-11151

Newkirk RW, Classen HL, Scott TA, Edney MJ (2003) The availability and content of amino acids in toasted and non-toasted canola meals. Can J Anim Sci 83:131-139

Nicolas SD, Leflon M, Monod H, Eber F, Coriton O, Huteau V, Chevre AM, Jenczewski E (2009) Genetic regulation of meiotic cross-overs between related genomes in *Brassica napus* haploid and hybrids. Plant Cell 21:373-385

Niu Z (2008) Manipulation of biosynthesis of aliphatic glucosinolates in brassica crops and *Arabidopsis* through gene replacement and RNA interference. Ph.D. thesis, Depart of Plant Science, Uni of Manitoba, Winnipeg, Canada

Nozaki T, Anji M, Takahashi T, Ikehashi H (1995) Analysis of isozyme loci and their linkages in *Brassica campestris* L. Breed. Sci. 45:57-64

Olsson G, Ellerstrom S (1980) Polyploidy breeding in Europe. In: S. Tsunoda, K. Hinata, and C. Gomez-Campo (eds*) Brassica* crops and wild allies. Japan Science Society Press, Tokyo, Japan, pp 167-190

Osborn TC, Butrulle DV, Sharpe AG, Pickering KJ, Parkin IAP, Parker JS, Lydiate DJ (2003) Detection and effects of a homeologous reciprocal transposition in *Brassica napus*. Genetics 165:1569-1577

Osborn TC, Kale C, Parkin IAP, Sharpe AG, Kuiper M, Lydiate DJ, Tricj M (1997) Comparison of flowering time genes in *Brassica rapa*, *B. napus* and *Arabidopsis thaliana*. Genetics 146:1123-1129

Osbourn AE (1996) Performed antimicrobial compounds and plant defense against fungal attack. Plant Cell 8:1821-1831

Padilla G, Cartea ME, Velasco P, de Haro A, Ordas A (2007) Variation of glucosinolates in vegetable crops of *Brassica rapa*. Phytochem 68:536-545

Parkin I, Magrath R, Keith D, Sharpe A, Mithen R, Lydiate D (1994) Genetics of aliphatic glucosinolates. II. Hydroxylation of alkenyl glucosinolates in *Brassica napus*. Heredity 72:594-598

Parkin IAP, Gulden SM, Sharpe AG, Lukens L, Trick M, Osborn TC, Lydiate DJ (2005) Segmental structure of the *Brassica napus* genome based on comparative analysis with *Arabidopsis thaliana*. Genetics 171(2):765-781

Parkin IAP, Lydiate DJ, Trick M (2002) Assessing the level of collinearity between *Arabidopsis thaliana* and *Brassica napus* for *A. thaliana* chromosome 5. Genome 45:356-366

Parkin IAP, Sharpe AG, Keith DJ, Lydiate DJ (1995) Identification of the A and C genomes of the amphidiploid *Brassica napus* (oilseed rape). Genome 38:1122-1131

Parkin IAP, Sharpe AG, Lydiate DJ (2003) Patterns of genome duplication within the *Brassica napus* genome. Genome 46:291-303

Pedras MSC, Zheng QA, Sarna-Manillapalle VK (2007) The phytoalexins from *Brassicaceae*: structure, biological activity, synthesis and biosynthesis. Nat Prod Comm 2:319-330

Pedras MSC, Zheng QA, Strelkov S (2008) Metabolic changes in roots of the oilseed canola infected with the biotroph *Plasmodiophora Brassicae*: Phytoalexins and phytoanticipins. J Agric Food Chem 56:9949-9961

Petersen BL, Chen S, Hansen CH Olsen CE, Halkier BA (2002) Composition and content of glucosinolates in developing *Arabidopsis thaliana*. Planta 214:562-571

Piao ZY, Deng YQ, Choi SR, Park YJ, Lim YP (2004) SCAR and CAPS mapping of *CRb*, a gene conferring resistance to *Plasmodiophora Brassicae* in Chinese cabbage (*Brassica rapa* spp. *pekinensis*). Theor Appl Genet 108:1458-1465

Piotrowski M, Schemenewitz A, Lopukhina A, Müller A, Janowitz T, Weiler EW, Oecking C (2004) Desulfoglucosinolate sulfotransferases from *Arabidopsis thaliana* catalyze the final step in the biosynthesis of the glucosinolate core structure. J Biol Chem 279(49):50717-50725

Piquemal J, Ciniquin E, Couton F, Rondeau C, Seignoret E, Doucet I, Perret D, Villeger MJ, Vincourt P, Blanchard P (2005) Construction of an oilseed rape (*Brassica napus* L.) genetic map with SSR markers. Theor Appl Genet 111(8):1514-1523

Plieske J, Struss D (2001) Microsatellite markers for genome analysis in *Brassica*. I. Development and abundance in *Brassica* species. Theor Appl Genet 102:689-694

Prakash S, Hinata K (1980) Taxonomy, cytogenetics and origin of crop *Brassica*, a review. Opera Bot 55:1-57

Punjabi P, Jagannath A, Bisht NC, Padmaja KL, Sharma S, Gupta V, Pradhan AK, Pental D (2008) Comparative mapping of *Brassica juncea* and *Arabidopsis thaliana* using intron polymorphism (IP) markers: homoeologous relationships, diversification and evolution of the A, B and C *Brassica* genomes. BMC Genomics 9:113

Qiu D, Gao M, Li G, Quiros CF (2009) Comparative sequence analysis for *Brassica oleracea* with similar sequences in *B. rapa* and *Arabidopsis thaliana*. Plant Cell Rep 28:649-661

Quijada PA, Udall JA, Lambert B, Osborn TC (2006) Quantitative trait analysis of seed yield and other complex traits in hybrid spring rapeseed (*Brassica napus* L.): 1. Identification of genomic regions from winter germplasm. Theor Appl Genet 113:549-561

Rahman M, McVetty PBE, Li G (2007) Development of SRAP, SNP and multiplexed SCAR molecular markers for the major seed coat color gene in *Brassica rapa* L. Theor Appl Genet 115:1101-1107

Rahman M, Sun Z, McVetty PBE, Li G (2008) High throughput genome-specific and gene-specific molecular markers for erucic acid gene in *Brassica napus* (L.) for marker-assisted selection in plant breeding. Theor Appl Genet 117:895-904

Ramchiary N, Bisht NC, Gupta V, Mukhopadhyay A, Arumugam N, Sodhi YS, Pental D, Pradhan AK (2007) QTL analysis reveals context-dependent loci for seed glucosinolate trait in the oilseed *Brassica juncea*: importance of recurrent selection backcross scheme for the identification of 'true' QTL. Theor App Genet 116:77-85

Ranade SA, Farooqui N, Bhattacharya E, Verma A (2001) Gene tagging with random amplified polymorphic DNA (RAPD) markers for molecular breeding in plants. Crit Reviews in Plant Sci 20(3):251-275Rangkadilok N, Nicolas ME, Bennett RN, Premier RR, Eagling DR, Taylor PWJ (2002) Determination of sinigrin and glucoraphanin in *Brassica* species using a simple extraction method combined with ion-pair HPLC analysis. Scientia Hort 96(1):1-4

Rask L, Andreasson E, Ekbom B, Eriksson S, Pontoppidan B, Meijer J (2000) Myrosinase: gene family evolution and herbivore defense in *Brassicaceae*. Plant Mol Biol 42:93-113

Reichelt M, Brown PD, Schneider B, Oldham NJ, Stauber E, Tokuhisa J, Kliebenstein DJ, Mitchell-Olds T, Gershenzon J (2002) Benzoic acid glucosinolate esters and other glucosinolates from *Arabidopsis thaliana*. Phytochem 59:663-671

Rodman JE, Karol KG, Price RA, Sytsma KJ (1996) Molecules, morphology, and Dahlgrens expanded order *Capparales*. Syst Bot 21:289-307

Rostas M, Bennett R, Hilker M (2002) Comparative physiological responses in Chinese cabbage induced by herbivory and fungal infection. J Chem Ecol 28:2449-2463

Sawada Y, Kuwahara A, Nagano M, Narisawa T, Sakata A, Saito K, Hirai MY (2009) Omics-based approaches to methionine side chain elongation in *Arabidopsis*: characterization of gene encoding methylthioalkylmalate isomerase and methylthioalkylmalate dehydrogenase. Plant Cell Physiol 50(7):1180-1190

Schnug E, Haneklaus S (1990) Quantitative glucosinolate analysis in *Brassica* seeds by X-ray fluorescence spectroscopy. Phytochem Anal 1:40-43

Schranz ME, Quijada P, Sung S, Lukens L, Amasino R, Osborn T (2002) Characterization and effects of the replicated flowering time gene *FLC* in *Brassica rapa*. Genetics 162:1457-1468

Schuster J, Knill T, Reichelt M, Gershenzon J, Binder S (2006) *BRANCED-CHAIN AMINOTRANSFERASE4* is part of the chain elongation pathway in the biosynthesis of methionine-derived glucosinolates in *Arabidopsis*. Plant Cell 18(10):2664-2679

Schwab W (2003) Metabolome diversity: too few genes, too many metabolites? Phytochemistry 62(6):837-837

Sebolt AM, Shoemaker RC, Diers BW (2000) Analysis of a quantitative trait locus allele from wild soybean that increases seed protein concentration in soybean. Crop Sci 40:1438-1444

Sharpe AG, Parkin IAP, Keith DJ, Lydiate DJ (1995) Frequent non-reciprocal translocations in the amphidiploid genome of oilseed rape. Genome 38:1112-1121

Shigyo M, Wako T, Kojima A, Yamauchi N, Tshiro Y (2003) Transmission of alien chromosomes from selfed progenies of a complete set of *Allium* monosomic additions: The development of a reliable method for the maintenance of a monosomic addition set. Genome 46:1098-1103

Soengas P, Hand P, Vicente JG, Pole JM, Pink DAC (2007) Identification of quantitative trait loci for resistance to *Xanthomonas campestris* pv. campestris in *Brassica rapa*. Theor Appl Genet 114:637–645

Sønderby IC, Geu-Flores F, Halkier BA (2010) Biosynthesis of glucosinolates-gene discovery and beyond. Trends Plant Sci 15(5):283-290

Song KM, Suzuki JY, Slocum MK, Williams PH, Osborn TC (1991) A linkage map of *Brassica rapa* (syn. *campestris*) based on restriction fragment length polymorphism loci. Theor Appl Genet 82:296–304

Spinks EA, Sones K, Fenwick GR (1984) The quantitative analysis of glucosinolates in cruciferous vegetables, oilseed and forages using high performance liquid chromatography. Fette Seifen Anstrichmittel 86:228-231

Srinivasan K, Malathi VG, Kirti PB, Prakash S, Chopra VI (1998) Generation and characterization of monosomic chromosome addition lines of *Brassica campestris-B. oxyrrhina*. Theor Appl Genet 97:976-981

Stefansson BR, Downey RK (1995) Rapeseed. In: Slinkard AE and DR Knott (eds) Harvest of gold. University Extension Press, University of Saskatoon, SK, Canada, pp 140-152

Sun Z, Wang Z, Tu J, Zhang J, Yu F, McVetty PBE, Li G (2007) An ultradense genetic recombination map for *Brassica napus*, consisting of 13551 SRAP markers. Theor Appl Genet 114:1305-1317

Suwabe K, Iketani H, Nunome T, Kage T, Hirai M (2002) Isolation and characterization of microsatellites in *Brassica rapa* L. Theor Appl Genet 104:1092-1098

Suwabe K, Tsukazaki H, Iketani H, Hatakeyama K, Fujimura M, Nunome T, Fukuoka H, Matsumoto S, Hirai M (2003) Identification of two loci for resistance to clubroot (*Plasmodiophora Brassicae* Woronin) in *Brassica rapa* L. Theor Appl Genet 107:997-1002

Suwabe K, Tsukazaki H, Iketani H, Hatakeyama K, Kondo M, Fujimura M, Nunome T, Fukuoka H, Hirai M, Matsumoto S (2006) Simple sequence repeat-based comparative genomics between *Brassica rapa* and *Arabidopsis thaliana*: the genetic origin of clubroot resistance. Genetics 173:309-319

Szadkowski E, Eber F, Huteau V, Lodé M, Huneau C, Belcram H, Coriton O, Manzamares-Dauleux MJ, Delourme R, King GJ, Chalhoub B, Jenczewski E, Chévre AM (2010) The first meiosis of resynthesized *Brassica napus*, a genome blender. New Physiol 186:102-112

Takuno S, Kawahara T, Ohnishi O (2007) Phylogenetic relationships among cultivated types of *Brassica rapa* L. em. Metzg. as revealed by AFLP analysis. Gen Resour and Crop Evol 54:279-285

Talalay P (2000) Chemoprotection against cancer by induction of phase 2 enzymes. Bio Factors 12:5-11

Tanhuanpää PK, Vilkki JP, Vikki HJ (1996) Mapping a QTL for oleic acid concentration in spring turnip rapa (*Brassica rapa* spp. *oleifera*). Theor Appl Genet 92:952-956

Tanhuanpää PK, Vilkki JP, Vilkki HJ (1995) Association of a RAPD marker with linolenic acid concentration in the seed oil of rapeseed (*Brassica napus* L.). Genome 38:414-416

Teutonico RA, Osborn TC (1994) Mapping of RFLP and quantitative trait in *Brassica rapa* and comparison to the linkage maps of *B. napus*, *B. oleracea* and *Arabidopsis thaliana*. Theor Appl Genet 89:885-894

Textor S, de Kraker JW, Hause B, Gershenzon J, Tokuhisa JG (2007) *MAM3* catalyzes the formation of all aliphatic glucosinolate chain lengths in *Arabidopsis*. Plant Physiol 144(1):60-71

Tierens KF, Thomma BP, Brouwer M, Schmidt J, Kistner K, Porzel A, Mauch-Mani B, Cammue BP, Broekaert WF (2001) Study of the role of antimicrobial glucosinolate-derived isothiocyanates in resistance of *Arabidopsis* to microbial pathogens. Plant Physiol 125:1688-1699

Tommasini L, Batley J, Arnold GM, Cooke RJ, Donini P, Lee D, Law JR, Lowe C, Moule C, Trick M, Edwards KJ (2003) The development of multiplex simple sequence repeat (SSR) markers to complement distinctness, uniformity and stability testing of rapa (*Brassica napus* L.) varieties. Theor Appl Genet 106:1091-1101

Toroser D, Thormann CE, Osborn TC, Mithen R (1995) RFLP mapping of quantitative trait loci controlling seed aliphatic glucosinolate content in oilseed rape (*Brassica napus* L.) Theor Appl Genet 91:802-808

Traw MB (2002) Is induction response negatively correlated with constitutive resistance in black mustrard? Evolution 56:2196-2205

Tu Y, Sun J, Ge X, Li Z (2009) Chromosome elimination, addition and introgression in intertribal partial hybrids between *Brassica rapa* and *Isatis indigotica*. Annals of Botany 103:1039-1048

U N (1935) Genome analysis in *Brassica* with special reference to the experimental formation of *B. napus* and peculiar mode of fertilization. Japanese J Bot 7:389-452

Udall J, Quijada P, Osborn TC (2004) Detection of chromosomal rearrangements derived from homologous recombination in four mapping populations of *Brassica napus* L. Genetics 169:967-979

Ugajin T, Takita K, Takahashi H, Muraoka S, Tada T, Mitsui T, Hayakawa T, Ohyama T, Hori H (2003) Increase in indole 3-acetic acid (IAA) level and nitrilase activity in

turnips induced by *Plasmodiophora Brassicae* infection. Plant Biotech 20:215-220

Uzunova M, Ecke W, Weissleder K, Robbelen G (1995) Mapping the genome of rapeseed (*Brassica napus* L.). I. Construction of an RFLP linkage map and localization of QTLs for seed glucosinolate content. Theor Appl Genet 90:194-204

Van Ooijen JW, Voorrips RE (2001) JoinMap® 3.0, Software for the calculation of genetic linkage maps. Plant Research International, Wageningen, the Netherlands

Velasco L, Becker HC (1998) Analysis of total glucosinolate content and individual glucosinolates in *Brassica* spp. by near-infrared reflectance spectroscopy. Plant Breed 117:97-102

Wang S, Basten CJ, Zeng ZB (2010) Windows QTL cartographer version 2.5, Department of Statistical Genetics, North Carolina State University, Raleigh, NC, USA

Wang Y, Sonntag K, Rudloff E, Wehling P, Snowdon RJ (2006) GISH analysis of Disomic *Brassica napus-Crambe abyssinica* chromosome addition lines produced by microspore culture from monosomic addition lines. Plant Cell Rep 25:35-40

Wang YP, Zhao XX, Sonntag K, Wehling P, Snowdon RJ (2005) Behaviour of *Sinapis alba* chromosomes in a *Brassica napus* background revealed by genomic *in-situ* hybridization. Chromo Research 13:819-826

Wentzell AM, Rowe HC, Hansen BG, Ticconi C, Halkier BA, Kliebenstein DJ (2007) Linking metabolic QTLs with network and cis-eQTLs controlling biosynthetic pathways. PLos Genet 3(9):1687-1701

Wink M (2003) Evolution of secondary metabolites from an ecological and molecular phylogenetic perspective. Phytochem 64:3-19

Wittstock U, Halkier BA (2000) Cytochrome P450 *CYP79A2* from *Arabidopsis thaliana* L. catalyzes the conversion of L-phenyl alanine to phenylacetaldoxime in the biosynthesis of the benzylglucosinolate. J Biol Chem 275:14659-14666

Wroblewski T, Coulibaly S, Sadowski J, Quiros C (2000) Variation and phylogenetic utility of the *Arabidopsis thaliana Rps2* homolog in various species of the tribe *Brassiceae*. Mol Phylogene and Evol 16(3):440-448

Yang TJ, Kwon SJ, Choi BS, Kim J, Jin M, Lim KB, Parl J, Kim JA, Lim MH, Kim HI, Lee HJ, Lim Y, Paterson A, Park BS (2007) Characterization of terminal-repeat

retrotransposon in miniature (TRIM) in *Brassica* relatives. Theor Appl Genet 114:627-636

Yang YW, Lai KN, Tai PY, Li WH (1999) Rates of nucleotide substitution in angiosperm mitochondrial DNA sequences and dates of divergence between *Brassica* and other angiosperm lineages. J Mol Evol 48:597-604

Yu SC, Wang YJ, Zheng XY (2003) Mapping and analysis QTL controlling heat tolerance in *Brassica campestris* L. spp. *pekinensis*. Acta Horti Sci 30:417-420

Zang YX, Kim HU, Kim JA, Lim MH, Jin M, Lee SC, Kwon SJ, Lee SI, Hong JK, Park TH, Mun JH, Seol YJ, Hong SB, Park BS (2009) Genome-wide identification of glucosinolate synthesis genes in *Brassica rapa*. FEBS J 276(13):3559-3574

Zang YX, Lim MH, Park BS, Hong SB, Kim DH (2008) Metabolic engineering of indole glucosinolates in Chinese cabbage plants by expression of *Arabidopsis CYP79B2, CYP79B3*, and *CYP83B1*. Mol Cells 25(2):231-241

Zhang J, Lu Y, Yuan Y, Zhang X, Geng J, Chen Y, Cloutier S, McVetty PBE, Li G (2009) Map-based cloning and characterization of a gene controlling hairiness and seed coat color traits in *Brassica rapa*. Plant Mol Biol 69:553-563

Zhang Y, Talalay P, Cho C-G, Posner GH (1992) A major inducer of anticarcinogenic protective enzymes from broccoli: isolation and elucidation of structure. Proc Natl Acad Sci USA 89:2399-2403

Zheng L, Hammerlindl J, Keller W, McVetty PBE, Daayf F, Quiros CF, Li G (2010) *MAM* gene silencing leads to the induction of C3 and reduction of C4 and C5 side-chain aliphatic glucosinolates in *Brassica napus*. Mol Breed 27(4):467-478

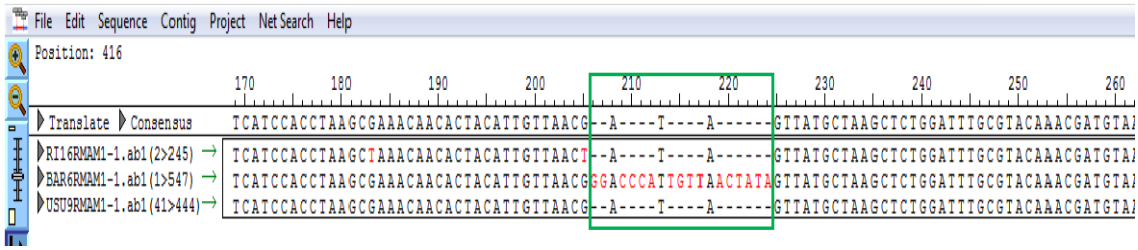
Zimmermann NS, Gerendas J, Krumbein A (2007) Identification of desulphoglucosinolates in *Brassicaceae* by LC/MS/MS: Comparison of ESI and atmospheric pressure chemical ionisation-MS. Mol Nutr Food Res 511:537-1546

9. APPENDICES

Appendix i

SCAR marker BrMAM1-1 based on a 16 bp deletion in BrMAM1-1 locus in Chinese cabbage RI16 and DH line USU9. (a) BrMAM1-1 gene sequence from RI16, BAR6 and USU9 showed deletion in RI16 and USU9. (b) Sequence alignment of BrMAM1-1 SCAR marker in RI16, BAR6 and USU9.

(a)



(b)

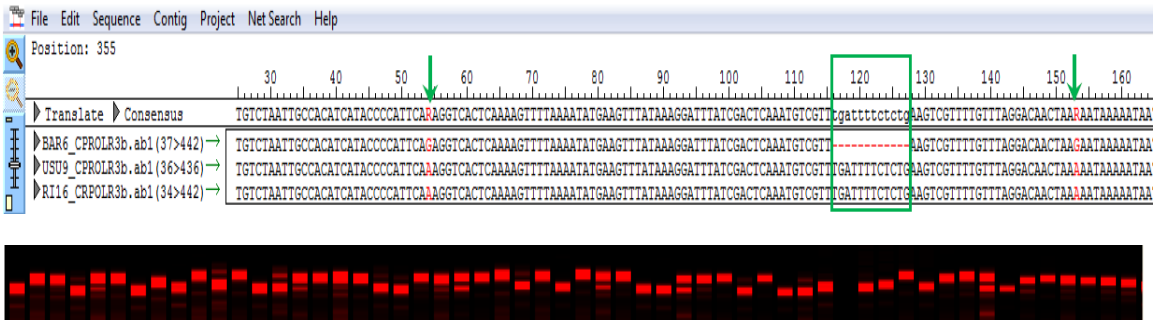
```
RI16            GTTCTTTGGTCTTTTTAACGTTAAATTTCATGCGAATGTCACTAGTAGATATGAATATGA 60
BAR6            -TTCTTAGGTCTTTTTAACGTTAAATTTCATGCGAATGTCACTAGTAGATATGAATATGA 59
USU9            -TTCTTAGGTCTTTTTAACGTTAAATTTCATGCGAATGTCACTAGTAGATATGAATATGA 59
                 ***** **************************************************

RI16            TTATTGTCTGCCTCTTGGACGTACTTCACTGCGTTCCCAAACCGCCTCGATGTCTCTATG 120
BAR6            TTATTGTCTGCCTCTTGG-CGTACTTCACTGCGTTCCCAAACCGCCTCGATGTCTCTATG 118
USU9            TTATTGTCTGCCTCTTGG-CGTACTTCACTGCGTTCCCAAACCGCCTCGATGTCTCTATG 118
                ****************** ****************************************

RI16            GTTGCATCGTGCGTAGGCCGCATATAACTGGAACATAACCTGTTTCTTCATCCACCTAAG 180
BAR6            GTTGCATCGTGCCTAGGCCGCATATAACTGGAACATAACCTGTTTCTTCATCCACCTAAG 178
USU9            GTTGCATCGTGCCTAGGCCGCATATAACTGGAACATAACCTGTTTCTTCATCCACCTAAG 178
                ************ ************************************************

RI16            CTAAACAACACTACATTGTTAACT--A----T----A------GTTATGCTAAGCTCTGG 224
BAR6            CGAAACAACACTACATTGTTAACGGGACCCATTGTTAACTATAGTTATGCTAAGCTCTGG 238
USU9            CGAAACAACACTACATTGTTAACG--A----T----A------GTTATGCTAAGCTCTGG 222
                * ********************   *    *    *      *****************

RI16            ATTTGCGTACAAACGATGTAAATATTAATCCCAATTTTTACCTAATGTCCCTGAAAAAAA 284
BAR6            ATTTGCGTACAAACGATGTAAATATTAATCCCAATTTTTACCTAATGTCCCTGAAAAAAA 298
USU9            ATTTGCGTACAAACGATGTAAATATTAATCCCAATTTTTACCTAATGTCCCTGAAAAAAA 282
                ************************************************************

RI16            AAGGTTAACTGTATAGTTTAAACATAAAATCTTTACGTTATATATTGAAGGAAGACATTA 344
BAR6            AAGGTTAACTGTATAGTTTAAACATAAAATCTTTACGTTATATATTGAAGGAAGACATTA 358
USU9            AAGGTTAACTGTATAGTTTAAACATAAAATCTTTACGTTATATATTGAAGGAAGACATTA 342
                ************************************************************

RI16            TTAACCTCATTCCCTACAGTCTTGGCAATGGTCTTGATGGTTTCGAACTCTTCCTCTGAT 404
BAR6            TTAACCTCATTCCCTACAGTCTTGGCAATGGTCTTGATGGTTTCGAACTCTTCCTCTGAT 418
USU9            TTAACCTCATTCCCTACAGTCTTGGCAATGGTCTTGATGGTTTCGAACTCTTCCTCTGAT 402
                ************************************************************
```

```
RI16            GACGCAGGGGAACCAACTTCCATGATGTCTACTCGGAGTTTGGCAAGCTGCCGGGCAATC 464
BAR6            GACGCAGGGGAACCAACTTCCATGATGTCTACTCGGAGTTTGGCAAGCTGCCGGGCAATC 478
USU9            GACGCAGGGAAACCAACTTCCATGATGTCTACTCGGAGTTTGGCAAGCTGCCGGGCAATC 462
                ********* **************************************************

RI16            TCTAGCTTCTGCGGTGGGGAAGTGCTGCACCAGGGGATTGTTCCCCGTCACGGAGCGTCG 524
BAR6            TCTAGCTTCTGCGGTGGGGAAGTGCTGCACCAGGGGATTGTTCCCCGTCACGGAGCGTCG 538
USU9            TCTAGCTTCTGAGGTGGGTAAGTGCTGCACCAGGGGATTGTTCCCCGTCACGGAGCGTCG 522
                *********** ****** ****************************************

RI16            TATC---------------------------------- 528
BAR6            TATC---------------------------------- 542
USU9            TATCAAAACACGCCATAGTTTTGTCGGGAGCTTGTCGG 560
                ****
```

Appendix ii

SCAR and SNP markers of *GSL-PRO* gene based on BAC sequence B59J16. (a) *GSL-PRO* gene SCAR (12 bp deletion) and SNP marker mapped on linkage group A9 on the SR-RILs linkage map. (b) Sequence of *GSL-PRO* gene display INDEL and SNP among RI16, BAR6 and USU9.

(a)



(b)

```
RI16            GTAGTTAAATC-GATTTCGACTTTGTCTAATTGCCACATCATACCCCATTCAAAGGTCAC 59
USU9            GTAGTTAAATCAGATTTCGACTTTGTCTAATTGCCACATCATACCCCATTCAAAGGTCAC 60
BAR6            GTAGTTAAATCAGATTTTGCCTTTGTCTAATTGCCACATCATACCCCATTCAGAGGTCAC 60
                *********** ***** * ******************************** *******

RI16            TCAAAAGTTTTAAAATATGAAGTTTATAAAGGATTTATCGACTCAAATGTCGTTTGATTT 119
USU9            TCAAAAGTTTTAAAATATGAAGTTTATAAAGGATTTATCGACTCAAATGTCGTTTGATTT 120
BAR6            TCAAAAGTTTTAAAATATGAAGTTTATAAAGGATTTATCGACTCAAATGTCGTT------ 114
                ************************************************

RI16            TCTCTGAAGTCGTTTTGTTTAGGACAACTAAAAATAAAAATAATCATAAAAATAAAAGCC 179
USU9            TCTCTGAAGTCGTTTTGTTTAGGACAACTAAAAATAAAAATAATCATAAAAATAAAAGCC 180
BAR6            ------AAGTCGTTTTGTTTAGGACAACTAAGAATAAAAATAATCATAAAAATAAAAGCC 168
                      ************************ ****************************



RI16            GGTGTATAAAATCAGAGGATCTGATATGATTTCTAATTTTGGTGTGAACTTTTTTTGAGC 239
```

```
USU9            GGTGTATAAAATCAGAGGATCTGATATGATTTCTAATTTTGGTGTGAACTTTTTTTGAGC 240
BAR6            GGTGTATAAAATCAGAGGATCTGATATGATTTCTAATTTTGGTGTGAACTTTTTTTGAGC 228
                ************************************************************

RI16            AAACATGTCAGATCCCAAATTATGTTCCTTCAGTTATATATCCATTATTAATCAAATATG 299
USU9            AAACATGTCAGATCCCAAATTATGTTCCTTCAGTTATATATCCATTATTAATCAAATATG 300
BAR6            AAACATGTCAGATCCCAAACTATGTTTCTTCAGTTATATATCCATTATTAATCAAATATG 288
                ******************* ****** ********************************

RI16            TCTTACTAGTTATTCCTTTGGCTAAGGTCAGGAGCATGATTAATCTAACATAGATTCGTA 359
USU9            TCTTACTAGTTATTCCTTTGGCTAAGGTCAGGAGCATGATTAATCTAACATAGATTCGTA 360
BAR6            TCTTGCTAGTTATTCCTTTTGCTAAGGTCAGGAGCATGATTAATCTAACATAGATTAGTA 348
                **** ************** ************************************ ***

RI16            GGAGACC-CTCTTTGACAGAATCTTGGAAT---------------------------- 388
USU9            GGAGAC--CTCTTTGACAGGATCTTGGAAT---------------------------- 388
BAR6            GGAGACAACTCTTTGACAGGATCTTTGAAATAAACGACA-CATAGATTA-TAGGAGACAA 406
                ******  *********** ***** ***

RI16            ----------------------------------------
USU9            ----------------------------------------
BAR6            CTCTTTGACAGGATCTTTGAAATAAACGACATGCA-CTGGA 446
```

Appendix iii

Fragment of daughter ions detected for individual desulphoglucosinolate by LC-MS in the parent BAR6 and USU9.

Parents	GSL	$[M+RCNOH]^+$	$[M+C_6H_{10}O_5]^+$	$[M+H]^+$	$[M+Na]^+$	$[M+K]^+$
BAR6	Gluconapin	-	132.1	294.1	316.2	-
	4-OH	189.1	223.1	385.3	407.2	-
	2-Methylpropyl		134.1	296.2	318.3	334.2
	Glucoerucin	-	180.3	-	364.2	380.3
	Glucobrassicin	-	207.1	369.2	391.2	-
USU9	Progoitrin	-	-	310.3	332.3	348.3
	Glucoraphanin	-	196.1	-	380.3	396.2
	Gluconapin	-	132.1	294.1	316.1	-
	4-OH	-	223.1	-	-	385.2
	Glucoerucin	-	180.1	342.1	364.2	380.2
	Glucobrassicin	173.2	207.1	369.3	391.3	-