Genetic Analysis of *Brassica carinata*

A thesis submitted to the College of
Graduate Studies and Research
in partial fulfillment of the requirements
for the degree of Master of Science
in the Department of Plant Science
University of Saskatchewan
Saskatoon

By

Gabriel Sze Yuen Wong

PERMISSION TO USE

In presenting this thesis/dissertation in partial fulfillment of the requirements for a Postgraduate degree from the University of Saskatchewan, I agree that the Libraries of this University may make it freely available for inspection. I further agree that permission for copying of this thesis/dissertation in any manner, in whole or in part, for scholarly purposes may be granted by the professor or professors who supervised my thesis/dissertation work or, in their absence, by the Head of the Department or the Dean of the College in which my thesis work was done. It is understood that any copying or publication or use of this thesis/dissertation or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of Saskatchewan in any scholarly use which may be made of any material in my thesis/dissertation.

DISCLAIMER

Reference in this thesis/dissertation to any specific commercial products, process, or service by trade name, trademark, manufacturer, or otherwise, does not constitute or imply its endorsement, recommendation, or favoring by the University of Saskatchewan. The views and opinions of the author expressed herein do not state or reflect those of the University of Saskatchewan, and shall not be used for advertising or product endorsement purposes.

Requests for permission to copy or to make other uses of materials in this thesis/dissertation in whole or part should be addressed to:

> Head of the Department of Plant Science
> 51 Campus Drive
> University of Saskatchewan
> Saskatoon, Saskatchewan S7N 5A8
> Canada
>
> OR
>
> Dean
> College of Graduate Studies and Research
> University of Saskatchewan
> 107 Administration Place
> Saskatoon, Saskatchewan S7N 5A2
> Canada

ABSTRACT

Brassica carinata is being actively pursued as a new industrial oil crop platform for the Canadian Prairies. A genetic assessment of *B. carinata* was performed to elucidate its evolutionary origins and create a genetic map to assist in locating genes and traits of interest that would help in marker-assisted breeding. First, genetic analysis using simple sequence repeat (SSR) markers, previously tested on *B. juncea* and *B. napus*, was performed, to examine the genetic diversity of 37 *B. carinata* lines. SSR analysis revealed world accessions were more diverse than lines conditioned to grow in the prairies. Diversity analysis revealed that the parental lines of a double haploid (DH) population, 179 and 345, obtained from the John Innes Centre (JIC), were among the more genetically diverse lines, supporting the use of this population for linkage mapping. Genetic markers created from 3' targeted SNP discovery between 179 and 345, were tested on the DH population resulting in the generation of a *B. carinata* genetic linkage map essentially with no prior sequence data knowledge. This genetic map contained 341 SNP and 86 SSR loci identifying eight linkage groups belonging to the B genome, nine belonging to the C genome and two unidentified groups spanning 2041 cM. Comparative mapping of polymorphic markers identified in the amphidiploid *B. carinata* indicated the orientation of B and C genomes coincide with that of other *Brassica* species, and the two genomes have remained essentially unaltered, with no major chromosomal rearrangements since the formation of *B. carinata*. A lesser number of polymorphic markers were detected in the C genome, which suggested the B genome is more genetically diverse in *B. carinata*. Limited field trials of the 179 x 345 DH population were performed during the 2011 and 2012 growing seasons. Preliminary quantitative trait loci (QTLs) for agronomic traits including flowering time (FT), plant height (PH), and seed quality were identified.

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. Introduction

Brassica carinata (Ethiopian mustard) is being actively pursued as a potential industrial oil crop platform for the Canadian Prairies. A lesser known species of the cruciferous and mustard Brassiceae tribe, *B. carinata* exhibits high environmental and disease tolerance.

Brassica carinata belongs to the Brassicaceae (Cruciferae or mustard) family. Native to Ethiopia, *B. carinata* is used as a leafy vegetable and the seeds are harvested for edible oil. Whereas many other *Brassica* species are used for commercial production for food and industrial use, there is little to no commercial production of *B. carinata* outside Ethiopia. Unlike conventional *Brassica* "<u>Can</u>adian <u>o</u>il, <u>l</u>ow <u>ac</u>id" (or canola) varieties, *B. carinata* is highly heat and drought tolerant (Getinet et al. 1996a; Kumar et al. 1984; Malik 1990; Schreiner et al. 2009). It has relatively good resistance to fungal pathogens; black leg disease (Gugel et al. 1990), *Sclerotinia sclerotiorum* (Sedun et al. 1989; Yang et al. 2010), white rust (Petrie 1988; Yitbarek 1992), Alternaria blackspot (Gómez-Campo et al. 1999; Yitbarek 1992), and powdery mildew (Singh et al. 1997; Tonguç and Griffiths 2004), to bacterial disease; black rot (Guo et al. 1991; Taylor et al. 2002; Vicente et al. 2002), and to insects; flea beetles (Bayeh and Gebre Medhin 1992) and aphids (Anand et al. 1985). It also has a relatively large seed size and is shatter resistant (Getinet et al. 1996a). Although, the seed oil content of many accessions of *B. carinata* evaluated to date are generally lower than canola grade *B. napus*, *B. carinata* seed quality lines have been identified with oil content above 40% (Ripley et al. 2006). Also, because of its large seed size, the protein content is generally higher in *B. carinata* than in canola and the crude fibre content is lower (Getinet et al. 1996a). There has been significant investment in agronomic improvement of *B. carinata;* however, there has been little research focusing on the genetics and genome organisation of this plant species.

A species found in the Triangle of U, *B. carinata* (BBCC) is one of three *Brassica* crop allotetraploids and was formed from the fusion of progenitors of modern day *B. nigra* (*Brassica* B genome) and *B. oleracea* (C genome) (U 1935). Molecular analysis of the related allotetraploids *B. napus* (AACC) and *B. juncea* (AABB) has determined that there have been no major chromosomal rearrangements since the fusion of their respective progenitor genomes (Panjabi et al. 2008; Parkin et al. 1995). The generation of a genetic map for *B. carinata* would allow the structure of the genome to be assessed giving insights into *Brassica* genome evolution as well as creating a valuable tool for marker assisted breeding for crop improvement.

Though *B. carinata* contains many useful traits, little is known regarding where these genes are located. Agronomic traits such as flowering time and plant height are important aspects in determining whether *B. carinata* is suitable for production in the Canadian Prairies. As well, seed components of *B. carinata* has the potential for many applications in food, energy and industrial uses creating interest in understanding the genetic architecture controlling these traits.

The main objective of this study is to analyse the genetics of *B. carinata*. This would involve assessing the genetic diversity within *B. carinata*, the generation of a genetic linkage map for *B. carinata*, and a phenotypic assessment involving a range of traits of interest.

2. Literature Review

2.1. Overview of *Brassica carinata*

Brassica carinata, commonly referred to as Ethiopian mustard, Abyssinian mustard and Gomenzer, is native to the highlands of Ethiopia. In Ethiopia, the plant is used as a leafy vegetable and the seeds are harvested for its edible oil. Outside Ethiopia, there is little to no commercial production of *B. carinata* (Taylor et al. 2010). The existence of genetic variation and desirable characteristics such as drought tolerance and oil profiles high in erucic acid make *B. carinata* a suitable industrial oil crop platform for the Prairies.

Brassica carinata belongs to the Brassicaceae (Cruciferae or mustard) family. The family has approximately 338 genera and 3,709 species of which *Brassica*, contains 39 species (Warwick et al. 2006a). Members of the Brassicaceae are described as having alternating leaves, rarely opposite, uniform flowers with bilateral symmetry consisting of four free sepals, and four distinct petals, with the androecium tetradynamous, consisting of two shorter outer stamen and 4 longer inner stamens and a bicarpellate ovary (Lysak and Koch 2011). Many of these species are grown in commercial production for food and industrial uses. Some of the more commonly known *Brassicas* are turnips and Chinese cabbage (*B. rapa)*; Indian mustard (*B. juncea*); kale, broccoli, and Brussels sprouts (*B. oleracea*); black mustard (*B. nigra*); rapeseed, rutabaga, and canola (*B. napus*).

Brassica carinata is unique in that it is an amphidiploid species derived from the fusion of the progenitors of modern day *B. nigra* (Brassica B genome) and *B. oleracea* (Brassica C genome) (U 1935). This characteristic gives *B. carinata* a distinct advantage in that it has the ability to use traits from both genomes for adaptability and survival.

3

2.1.1. Triangle of U

The Triangle of U is an evolutionary theory describing the relationship between six members of the plant genus *Brassica*. The theory linked three diploid species, containing three different genomes (Brassica A, B, C) with different numbers of chromosomes, *B. rapa* (n = 10; AA; ~550 Mbp)*, B. nigra* (n = 8; BB; ~630 Mbp), and *B. oleracea* (n = 9; CC; ~700 Mbp), with the convergence of three allotetraploid species, *B. juncea* (n = 18; AABB; ~1,100 Mbp)*, B. napus* (n = 19; AACC; ~1,130 Mbp), and *B. carinata* (n = 17; BBCC; ~1,300 Mbp) (Arumuganathan and Earle 1991; Johnston et al. 2005; U 1935) (Figure 2.1). Though the *Brassica* species evolved from a common ancestor, *B. rapa* and *B. oleracea* are believed to have diverged approximately 4 million years ago (Inaba and Nishio 2002). The polyploidization of *B. napus* has been suggested to have occurred under 10,000 years ago (Rana et al. 2004). While the age of the polyploidy event is not yet determined for *B. carinata*, no major chromosomal rearrangements have occurred for the allotetraploids *B. napus* (AACC) and *B. juncea* (AABB) since the fusion of the progenitor genomes (Panjabi et al. 2008; Parkin et al. 1995).

2.2. Uses

Brassica carinata has the potential for use in multiple applications. The leaves are used for human consumption and the seeds are used for edible oil production in the highland regions of Ethiopia. The oil in the seeds also has the potential to be used for the generation of bioindustrial oils such as biofuels, biolubricants and bioplastics (Bouaid et al. 2009; Cardone et al. 2003; Gasol et al. 2007; McVetty and Scarth 2002; Taylor et al. 2010). Defatted proteins from *B. carinata* are being studied for their potential benefits for human nutrition and animal feed (Chaudhary et al. 1998; Getinet et al. 1997; Pedroche et al. 2007). Other applications
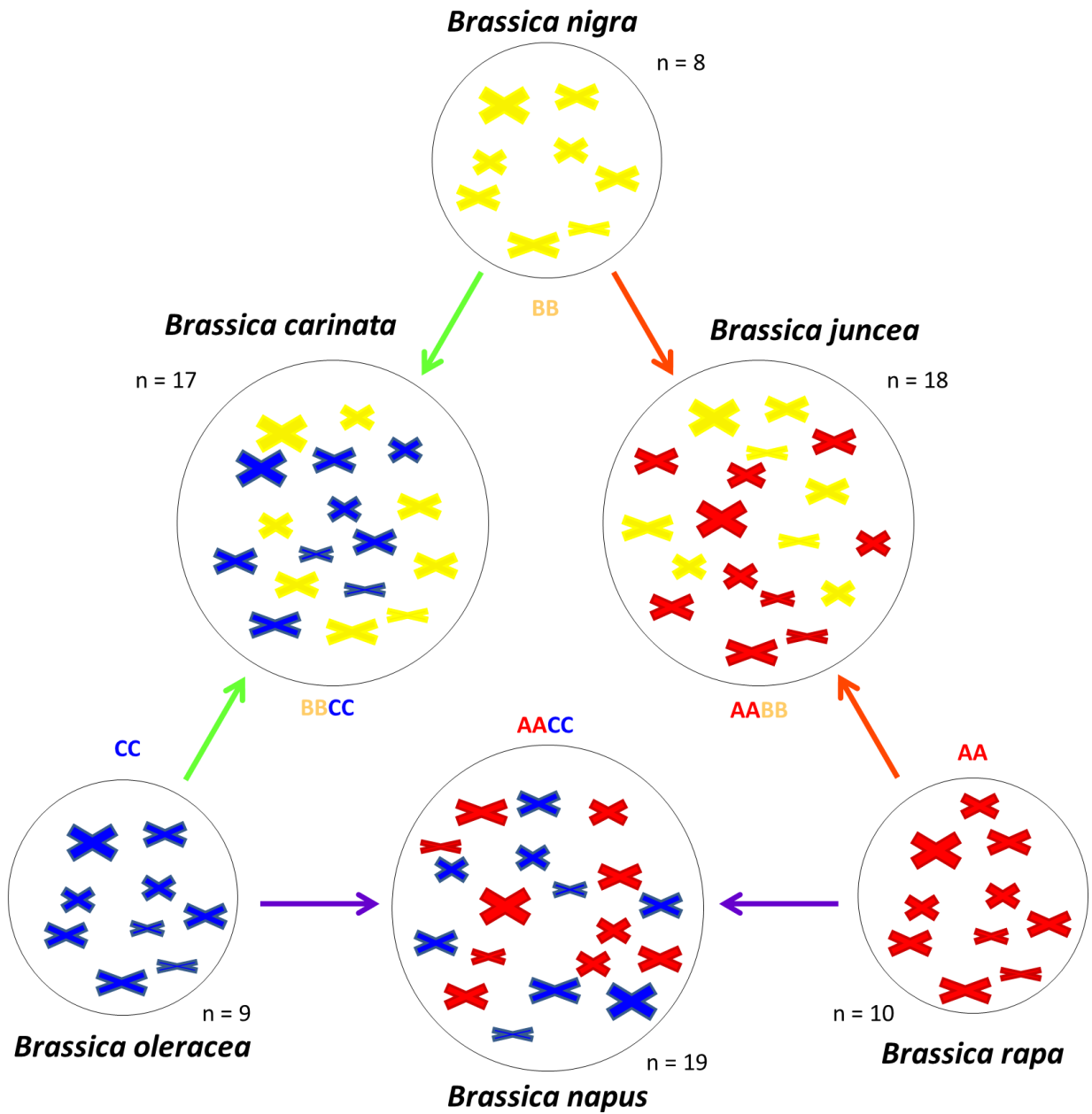
Figure 2.1. The Triangle of U depicting the evolution of the *Brassica* species. Each circle represents the haploid number (n) of chromosomes for each *Brassica* species. Diploid species *B. rapa*, *B. nigra* and *B. oleracea* are located at the apexes while amphidiploid species *B. carinata*, *B. juncea* and *B. napus* are on the sides of the triangle (U 1935).

include biopesticides, biopolymers, antioxidants, and heavy metal phytoremediation (Irtelli and Navari-Izzo 2008; Irtelli et al. 2009; Kirkegaard and Sarwar 1998; Pedroche et al. 2007; Ripley et al. 2006).

The many uses of *B. carinata* bring attention to its role in the food versus fuel debate. In industrialized countries, the potential for non-edible industrial products conflict with the current use as an oilseed crop in north-eastern Africa. The seeds generated from the manipulation of *B. carinata* for bio-industrial use could be unsuitable for human consumption. Should a bioindustrial *B. carinata* line come in contact with edible lines of *B. carinata*, the seeds produced from the cross pollination would also be unsuitable for human consumption. However, the limited use of *B. carinata* globally and its ability to adapt to heat and drought-like conditions of marginal land (Getinet et al. 1996a; Kumar et al. 1984; Malik 1990; Schreiner et al. 2009) make it a suitable candidate for bioindustrial uses, allowing for other well established edible crops to be grown in better, more fertile soil.

2.2.1. Oil

In Ethiopia, the leaves of *Brassica carinata* are used as a food source, where they are picked, boiled and consumed for nutrition. As well, the oil found in the seeds is edible. Outside of Ethiopia and the surrounding area, there is little or no commercial production (Taylor et al. 2010). To produce "<u>Can</u>adian <u>o</u>il, <u>l</u>ow <u>ac</u>id" (or canola) grade seed quality for human consumption, high oil content, early maturity, the removal of erucic acid and minimal amounts of glucosinolates are required. Accessions of *B. carinata* containing high seed oil content above 40% which is comparable to *B. napus* canola have been identified (Ripley et al. 2006). On selected lines, maturity was on average 5-7 days later than its *B. napus* canola counterparts

compared to unselected populations which on average matured 10-14 days later than *B. napus* (Taylor et al. 2010). Erucic acid, which make up ~40% of the total fatty acids in *B. carinata* seeds, is considered to be a potential health concern (Thomasson and Boldingh 1955). Animal studies revealed even low amounts of erucic acid over a long period of time will create lesions in the heart as well as amass fatty deposits around the heart (Yamashiro and Clandinin 1980). This had led to the development of "Zero Erucic Acid Ethiopian Mustard" (Getinet et al. 1994, 1997). Another health concern, when developing new *Brassica* lines for human consumption relates to the high levels of glucosinolates found in many natural varieties of *B. carinata*. Aliphatic glucosinolates are known to bind to iodine disabling the intake of iodine in the thyroidal gland (Zukalová and Vasak 2002). Low glucosinolate *B. carinata* lines have been developed using interspecific crosses [(*B. carinata* x *B. juncea*) x *B. carinata*] removing the production of 2-propenyl glucosinolate, which accounts for more than 95% of total alkenyl glucosinolate content alleviating the bitter taste from the oil and allowing the protein meal to be used as animal feed (Getinet et al. 1996b; Getinet et al. 1997). For commercialization of canola grade seed quality, the aim would be to create early maturing *B. carinata* lines with high oil, high protein and low erucic acid. Although there is potential of *B. carinata* being the next edible oilseed crop, support is strongest for its use as a bioindustrial crop platform (Ripley et al. 2006).

Globally, there is an increasing demand for biofuels. *Brassica carinata* has the potential to be a bioenergy crop making it a renewable energy source. *Brassica carinata* oil can be converted into biodiesel and is considered superior due to its stability in long term storage while avoiding oxidation (Bouaid et al. 2009). In an effort to lower greenhouse gas emissions environmental analysis has been performed on the growth of *B. carinata* as an alternative bioenergy crop for biodiesel and biofuel production. Biodiesel produced from *B. carinata* seed

oil displayed physical and chemical properties comparable to traditional diesel fuel with lower levels of particulate matter and less soot during engine performance (Cardone et al. 2003). In Canada, collaboration between the National Research Council, Agrisoma, Honeywell UOP Inc. and Saskatoon's Genome Prairie-led Prairie Gold project has led to testing on two commercial based planes, the Falcon20 and Lockheed T-33 Shooting Star chase plane, with an experimental bio-jet fuel mixture comprised of traditional petroleum-based fuel and a synthetic fuel produced from the oilseeds of *B. carinata* in the hopes of producing a more eco-friendly fuel source (Larson and Pilieci 2012). More recently, a Falcon20 completed the world's first flight powered using entirely bio jet fuel made from 100 percent *B. carinata* oilseed (Fougeres 2012). In-flight data collected revealed a 50% reduction in aerosol emissions when using the biofuel compared to conventional fuel. As well, static engine tests showed a 25% reduction in particles and 49% reduction in black carbon emissions in the biofuel compared to conventional fuel. There was also a 1.5% improvement in fuel consumption during steady state operations between the biofuel and conventional fuel (National Research Council Canada 2013).

2.2.2. Protein

Besides animal feed, the defatted seed meal of *B. carinata* is being researched as a potential functional protein source for bioactive peptides in human nutrition. *Brassica carinata* protein hydrolysates were obtained after sequential hydrolysis with digestive proteases, trypsin, chymotrypsin, and carboxy-peptidase A. The resulting products behaved as bioactive peptides having health promoting properties including antioxidative and hypocholesterolemic effects, as well as the inhibition of angiotensin converting enzyme activities, controlling high blood pressure (Pedroche et al. 2007).

Brassica carinata may serve as a viable alternative for the production of recombinant proteins. Hirudin, a blood anticoagulant protein from leeches, and *β*-glucuronidase, an enzyme that is able to break down complex carbohydrates, were produced from the seeds of transgenic *B. carinata* using oleosin, a fusion protein, as a carrier (Chaudhary et al. 1998).

2.2.3. Very Long Chain Fatty Acids

If human consumption is not in the best interest of the commercial production of *B. carinata*, increasing the levels of seed very long-chain fatty acids (VLCFA) such as erucic acid ($22:1\omega 9$) and nervonic acid ($24:1\omega 9$) are a viable option.

Erucamide, an erucic acid derivative, is mainly used as a slip promoting or anti-blocking agent for plastic films and surface-active additive in coatings. Erucic acid and its derivatives could be potentially used as lubricants, surfactants, fabric and textile softeners, cosmetics and heat transfer fluids (Katavic et al. 2000; Leonard 1992; McVetty and Scarth 2002).

The abundance of nervonic acid in white brain matter has suggested that it may play a role in enhanced brain function as well as the prevention of demyelination in nerve cells (Martínez and Mougan 1998). This has led to nervonic acid being pursued as a treatment for diseases which affect demyelination, such as Multiple Sclerosis (MS) and Adrenoleukodystrophy (ALD); for repairing and restoring brain and nerve tissue to reduce tremors caused by Parkinson's Disease; and for nervonic acid deficiencies such as those that cause Schizophrenia (Sargent et al. 1994). Nervonic acid is also being considered as a dietary supplement for preventative nerve degeneration as well as in periods of time, such as pregnancy, infancy and in youth, where myelinated nerve fibers and brain matter grow most rapidly (Martínez and Mougan 1998).

Fatty acid profiles with high concentrations of erucic acid (30.9 - 45.7%) were observed in 66 accessions of *B. carinata* seed (Warwick et al. 2006b). Only low levels of nervonic acid have been found. Substrate specific enzyme 3-ketoacyl-CoA synthase (KCS), which influences the chain length produced, was determined as the rate limiting step for VLCFA production in Brassicaceae*;* however, by introducing the KCS gene from plants with different substrate preferences, increases in certain VLCFA production could be favoured. Though, the regulation and properties are not well known for the KCS enzyme, transgenic *B. carinata* with the KCS gene from the plant species, *Lunaria annua* has exhibited affinity for multiple VLCFA in transgenic *B. carinata* by displaying high levels of both erucic acid and nervonic acid. However, the high levels of erucic acid made this system undesirable for human consumption. The continuing study using the KCS gene from *Caradamine graeca*, produced levels of nervonic acid of ~45%, while limiting erucic acid to less than 7%. Conversely, using the same method to produce higher levels of erucic acid, transgenic *B. carinata* with the KCS gene from *Crambe abyssinica* produced seed containing ~53% erucic acid (Taylor et al. 2010).

2.2.4. Glucosinolates

Glucosinolates, prevalent in *Brassicas*, are sulphur containing compounds derived from glucose and amino acid molecules. Conversely, rather than converting into canola quality oilseed, *B. carinata* lines could be selected for increased levels of 2-propenyl glucosinolate for use as a biopesticide. The glucosinolates, through hydrolysis of the myrosinase enzyme, can be converted to allyl iso-thio-cyanates for use as a biofumigant (Kirkegaard and Sarwar 1998; Ripley et al. 2006).

2.3. Traits of Interest

Brassica carinata boasts many valuable qualities compared to its *Brassica* counterparts, such as insect and microbial resistance as well as tolerance to semi-arid and drought-like conditions.

2.3.1. Pest and Disease Resistance

Leptosphaeria maculans, more commonly known as blackleg disease, is a fungal pathogen affecting *Brassica* crops all over the world. *Brassica carinata* is highly resistant to *L. maculans.* This trait was once thought to reside only in the B genome (Gugel et al. 1990; Roy 1978, 1984; Sjodin and Glimelius 1988) however, moderate resistance has been found in accessions of *Brassica* species containing the C genome at different plant life stages, so it is possible that the resistance resides on either the C or B genome of *B. carinata*. While resistance genes against *L. maculans* from the B genome were bred into *B. napus* through interspecific crosses with *B. juncea*, *B. carinata* has yet to be used as a source of resistance (Dolores Sacristan and Gerdemann 1986; Rimmer and van den Berg 1992; Roy 1984).

Ethiopian mustard has been reported to be moderately tolerant to *Sclerotinia sclerotiorum*, a fungal pathogen forming lesions causing stem rot disease. In one study, *B. carinata* and *B. napus* exhibited the slowest expansion of *S. sclerotiorum* lesions compared with other *Brassica* species (Sedun et al. 1989). A more recent study revealed that after stem inoculation treatments, lesions in *B. carinata* did not show any significant changes while in *B. napus* the lesions increased over time. Genes from signalling pathways, mediated by phytohormones jasmonic acid, ethylene, and salicylic acid, which are known to be involved in the defence of ubiquitous plant pathogens, exhibited increases in expression at different time

periods after infection of *S. sclerotiorum* in *B. carinata,* while they remained unchanged in *B. napus* (Yang et al. 2010).

Albugo candida or white rust, a fungal pathogen, has been found on many Crucifereae species in the Canadian western prairies. Compared to *B. napus*, accessions of *B. carinata* have been found to be resistant to many races of *A. candida* (Petrie 1988; Yitbarek 1992).

Alternaria blackspot or Alternaria blight is the common name for the fungal disease primarily caused by *Alternaria brassicae*. Black spots appear on stems and pods after infection of *B. rapa* and *B. napus*. The epicuticular wax produced from the C genome of *B. carinata* was suggested to be the first and best defence mechanism against Alternaria blight (Gómez-Campo et al. 1999; Yitbarek 1992). Even after detaching the leaves and removal of the waxy layer upon infection, *B. carinata* expressed the smallest lesion size amongst other triangle of U species (Bansal et al. 1990).

Another fungal pathogen, *Erysiphe polygoni* D.C., causing powdery mildew on select *Brassica* species, seems unable to infect *B. carinata*. The origin of resistance gene(s) to powdery mildew are not completely known since *B. juncea* (AABB) and many *B. oleracea* (CC) species are susceptible, but resistance breeding has been accomplished using *B. carinata*. (Singh et al. 1997; Tonguç and Griffiths 2004).

Xanthomonas campestris pv. *campestris* also known as Black rot is a bacterial disease that affects many crucifers worldwide. *Brassica oleracea* is the most susceptible as many of the race non-specific resistance genes reside in the A and B genomes. *Brassica carinata* accessions have been found to have both race specific and race non-specific resistance (Guo et al. 1991; Taylor et al. 2002; Vicente et al. 2002).

Brassica carinata confers insect resistance to *Phyllotreta cruciferae*, more commonly known as the flea beetle (Bayeh and Gebre Medhin 1992). The C genome identified as the source of leaf epicuticular wax production deters flea beetles from feeding. *Brassica* species not possessing the C genome, and with non-waxy leaves were prone to a higher rate of feeding (Bodnaryk 1992).

Ethiopian mustard has demonstrated aphid (*Lipaphis erysimi*, Kaltenbach) resistance (Bayeh and Gebre Medhin 1992). In naturally infested areas, *B. carinata* yield excelled that of *B. napus*, *B. rapa* and *B. juncea* (Anand et al. 1985). When aphids were allowed to reproduce on the leaves of *B. rapa*, *B. juncea* and *B. carinata*, *B. carinata* displayed the best antibiosis, reducing the survival and reproduction of aphids, compared with *B. rapa* and *B. juncea* (Amjad and Peters 1992).

2.3.2. Environmental and Climate Tolerance

Brassica carinata is able to thrive in semi-arid climates and drought-like conditions (Getinet et al. 1996a; Kumar et al. 1984; Malik 1990; Schreiner et al. 2009). In comparison to *B. napus*, reflectance caused by the abundance of wax on the leaf surface as well as the higher leaf conductance due to osmoregulation provided *B. carinata* with cooler leaves (Kumar et al. 1984). The production of *B. carinata* in semi-arid, drought tolerant conditions could be of importance due to its potential of expanding into areas that are deemed to be less fertile.

Ethiopian Mustard is able to withstand high aluminum concentrations in soil, a main contributing factor for acidic soil toxicity. Gene(s) controlling aluminum tolerance are believed to reside on the C genome and compared to other *Brassica* counterparts*, B. carinata* has the highest tolerance (Huang et al. 2002).

Brassica carinata and *B. napus* were both shown to be tolerant to sodium chloride; however, in field trials, *B. carinata* outperformed *B. napus, B. juncea,* and *B. rapa* under high salinity conditions (Ashraf and McNeilly 1990; Malik 1990).

The environmental tolerance of *B. carinata* also extends to heavy-metal conditions. *Brassica carinata* displayed the ability to uptake an excess of copper and arsenic suggesting a potential for heavy-metal phytoremediation qualities allowing for a cost-effective alternative to soil removal and washing (Irtelli and Navari-Izzo 2008; Irtelli et al. 2009).

2.3.3. Seed Quality Characteristics

Brassica carinata has a relatively large seed size compared to other *Brassica* species (Getinet et al. 1997). Ethiopian mustard also has good shatter resistance (Prakash and Chopra 1988). Yellow and brown seed variation has been found. Yellow seed of *B. carinata*, when compared to its brown seed counterpart had higher oil and protein content and less fibre content which correlates with studies performed on *B. juncea* (Getinet et al. 1996a; Woods 1980). *Brassica carinata* generally contain high levels of erucic acid and glucosinolates (Getinet et al. 1996a); however, zero erucic acid and low glucosinolate types have been developed (Getinet et al. 1996b; Getinet et al. 1994, 1997). Development of interspecific crosses have been accomplished to introgress traits from the C genome of *B. carinata* and the A genome in *B. juncea* to improve these qualities in *B. napus* (Jönsson 1978; Rashid et al. 1994).

2.4. Molecular Genetics in *Brassica*

Although breeding in plants to obtain desired characteristics has taken place for thousands of years, the work of Mendel in the mid-19th century was the first real evidence that

selected traits could be passed on from parents to progeny. It was not until 1953, with the

discovery of the double helical shape of DNA (Watson and Crick 1953) that an understanding of

the molecular control of Mendel's theories began. The inheritance of traits is now studied, not

only at a genetic level but more specifically, at a chromosomal level where the exact genome

sequences can be localized and associated with trait(s) of interest. Genetic markers designed to

exploit differences in DNA sequences, which include rearrangements, insertions, deletions or

base pair (bp) changes at known chromosome locations, are able to identify genes, individuals or

species.

In *Brassicas*, genetic markers have been used to create genetic linkage maps, help locate

and manipulate important agronomic traits, and have become valuable tools for not only marker

assisted breeding for crop improvement, but also to help unravel the evolution of the *Brassica*

species. Much of the *Brassica* research has focused on genetic traits beneficial to crops already

in commercial production. Advantages in yield, environmental adaptation and tolerance, and

pest resistance are valuable qualities that are being sought after and would benefit from the

application of marker technologies.

2.4.1. Technologies used to Develop Genetic Markers

The two major technologies used in developing genetic markers that have revolutionized

molecular genetics, were the application of the polymerase chain reaction (PCR) and the

development of DNA sequencing instruments. Polymerase chain reaction, a method used to

amplify DNA fragments, lowered the concentration of DNA needed for marker identification

and sequencing. This allowed for a shorter period of time to sequence DNA fragments. Primers

that bind to specific areas of DNA are able to distinguish differences between fragment length

and sequence bp changes. The development of Sanger sequencing became one of the most widely used methods to sequence DNA. In its infancy, ~50-100 bps were visually inspected on large sheets of x-ray film or gel images. Automation of Sanger sequencing drastically increased the number of bps that could be read and reduced the time required to read the sequence. This was followed more recently by second generation sequencing technologies that moved away from conventional Sanger sequencing methods. Second generation sequencing is able to produce millions of reads at an incredible rate allowing for the detection of vast amounts of polymorphisms right down to single bp changes.

2.4.1.1. Polymerase Chain Reaction

Polymerase chain reaction (PCR), a technique to amplify small amounts of DNA into thousands and even millions of copies of a particular DNA sequence, has become common and essential to genetic research. A PCR cocktail made up of, DNA polymerase, deoxynucleoside triphosphates and forward and reverse primers targeting a sequence of DNA are needed for amplification. The heating and cooling of the PCR sample is defined in a series of temperature steps. PCR is initialized and the temperature is raised allowing for the target DNA to be denatured resulting in single strands. The temperature is lowered to allow for the primer to anneal onto the single strands of DNA. The temperature increases allowing for the DNA polymerase to bind and extend the DNA from the primer sequence. The cycling process is repeated until sufficient DNA has been amplified. Previously, DNA polymerase from *Escherichia coli* was added to each round of replication because the denaturing step inactivated the DNA polymerase. Taq polymerase, a heat stable DNA polymerase from the bacterium *Thermus aquaticus* replaced DNA polymerase in the PCR reaction which allowed for continuous

replication of DNA, as the enzyme is able to withstand temperatures over 90°C during the denaturing step (Brock and Freeze 1969; Chien et al. 1976).

2.4.1.2. Sanger Sequencing

In 1977, the Sanger sequencing method or DNA chain termination method established a technique which included the addition of deoxynucleotide analogues, dideoxynucleoside triphosphates (ddNTPs), in the amplification process for sequencing DNA. The missing 3′OH group in ddNTPs act as a chain terminator preventing the further addition of nucleotides. By incorporating radioactive phosphorus into the four distinct ddNTP reactions, radiolabelling was performed. The products were separated using acrylamide gel electrophoresis where the resulting bands identified by X-ray are read in reverse giving the sequence of the DNA (Sanger et al. 1977).

Although Sanger developed this method to sequence DNA before the invention of PCR, it was not until the process became automated that it became possible to sequence genes and eventually genomes. Sanger's principles were incorporated into dye-terminator sequencing where the specific fluorescent dye in each dideoxynucleotide chain terminator emits light at different wavelengths allowing for sequencing in a single reaction. The first generation DNA sequencers separated the DNA products by size using capillary electrophoresis. At its peak, first generation DNA sequencers were able to generate read lengths of up to 800 bps of DNA sequence per reaction. Genes could be sequenced in a fraction of the time and genome sequencing was possible, although very expensive (Kircher and Kelso 2010; Shendure and Ji 2008).

2.4.1.3.　　Current Sequencing Technologies

Previous first generation DNA sequencers, based on the Sanger method using capillary electrophoresis, produced reads in the thousands and produced high raw data accuracy (99.0% to >99.999%) (Kircher and Kelso 2010). Second generation sequencing technologies, although not as high in raw accuracies, generate larger numbers of reads and increased read depth (which increased accuracy via consensus). The 454 by Roche and the Illumina platforms will be discussed.

The first high throughput next generation sequencing platform was released in October, 2005. The 454 sequencing platform (Roche Applied Science) uses a pyrosequencing approach (also known as sequencing by synthesis and sequencing during extension) where a single labelled nucleotide species is added one at a time and washed sequentially over the copies of sequences to be determined in each well (Figure 2.2). Double stranded DNA is broken up into fragments and adaptors are attached. The fragments are then separated into single strands and each DNA fragment is placed onto individual beads and a process called emulsion PCR ensures there are multiple copies of the individual sequence to be determined. Each 28-μm bead is loaded into one of the two million wells on a picotiter plate. During sequencing by synthesis DNA polymerase incorporates the nucleotide onto the DNA if it is complementary to the template strand. For every nucleotide incorporated, a pyrophosphate is released and converted to ATP by ATP sulfurylase. The ATP propels the enzyme luciferase into a light reaction from which the light signal is measured. Multiple additions of the same nucleotide give higher light intensity. An apyrase wash follows, removing the unincorporated nucleotides, allowing for the next nucleotide to be added. This allows for ~400 nucleotides per read (bead) or ~750 Mb/day with an accuracy rate of 99.9 - 99.99% (Kircher and Kelso 2010).

A. B. C.

D. E. F.

G.

ATCTGATGCCCTAA

DNA Capture Bead

Primer

G

Sulfurylase

APS

PP$_i$

Luciferase

ATP

Luciferin

Light + oxy Luciferin

H.

Light Signal → Apyrase Wash → Light Signal
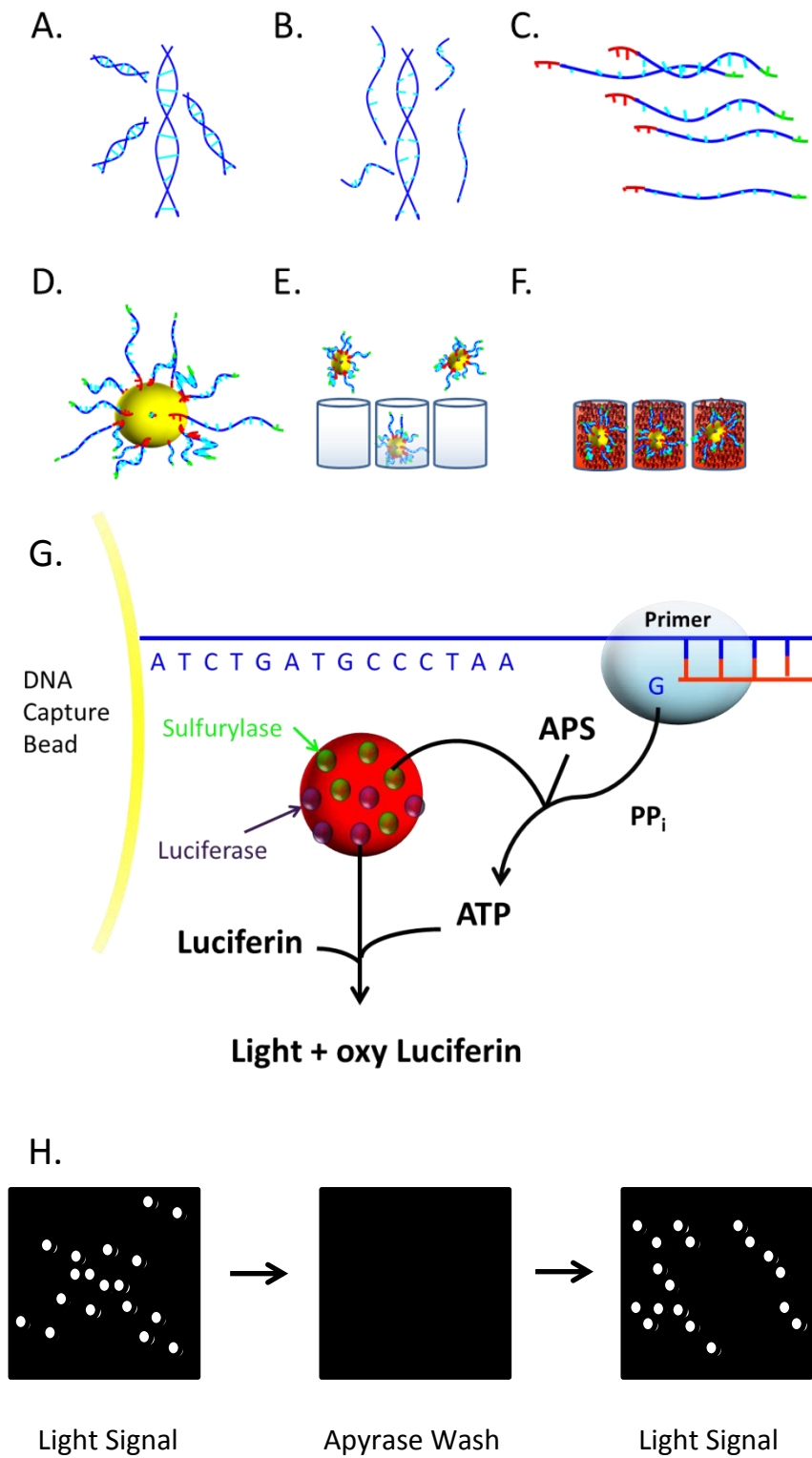
Figure 2.2. Overview of the 454 sequencing platform. (A.) Double stranded DNA are fragmented and (B.) separated into single strands. (C.) Adaptors are attached to the single strands. (D.) Each individual fragment binds onto one DNA capture bead where emulsion PCR is able to make multiple copies. (E.) The beads are placed in a picotiter plate and (F.) enzyme beads are added. (G.) Sequencing by synthesis occurs where DNA polymerase incorporates a nucleotide onto the DNA if it is complementary to the template strand. A pyrophosphate is released and converted to ATP by ATP Sulfurylase (APS). The combination of ATP and Luciferase cause a light reaction from which the light signal (H.) is released and measured. Intensity increases with multiple additions. Unincorporated nucleotides are removed with an apyrase wash allowing for the next nucleotide to be added (Shendure and Ji 2008).

Introduced in 2007, the Illumina platform uses reversible terminator technology. The sequencing by synthesis adopts the Sanger concept by stopping the reaction after each base is incorporated and the specific fluorescent dye read before continuing with the incorporation of the next base (Figure 2.3). Different adaptors for both ends are added to the double-stranded DNA fragment library. Sodium hydroxide is added to separate the DNA into single strands. Immobilized complementary oligonucleotides on the flow cell bind to the adaptor ends of the single stranded DNA. Unlabelled nucleotides and enzymes are incorporated and by reverse strand synthesis, a new strand is created. If the new strands are able to bend over and attach to another oligonucleotide commentary to the second adaptor sequence on the free end, the reverse strand can also be synthesized. This is referred to as bridge amplification. After the process is repeated numerous times, several thousand copies are generated within close proximity to each other. One of the forward or reverse strands are removed to prevent complementary base pairing or from hindering the extension reaction sterically. Sequencing primers complementary to the adaptor sequences bound to the single stranded DNA are added. Polymerases incorporate fluorescently labelled and terminated nucleotides, which stop the process after the addition of each complementary nucleotide. The polymerases and nucleotides are washed away and 4 different images are taken with different filters using 2 different lasers illuminating the fluorophores (or bases). The fluorophores and the terminators are removed and sequencing continues with the addition of another fluorescent/terminator nucleotide. This technology produces sequence read lengths of about 100 nucleotides with a throughput of 5,000 Mb/day and an accuracy rate of 99 – 99.9% (Kircher and Kelso 2010).

A.

B.

C.

D.

E.

F.

G.

H.

I.

J.

K.

L.

	T
	G
	A
	C
	C
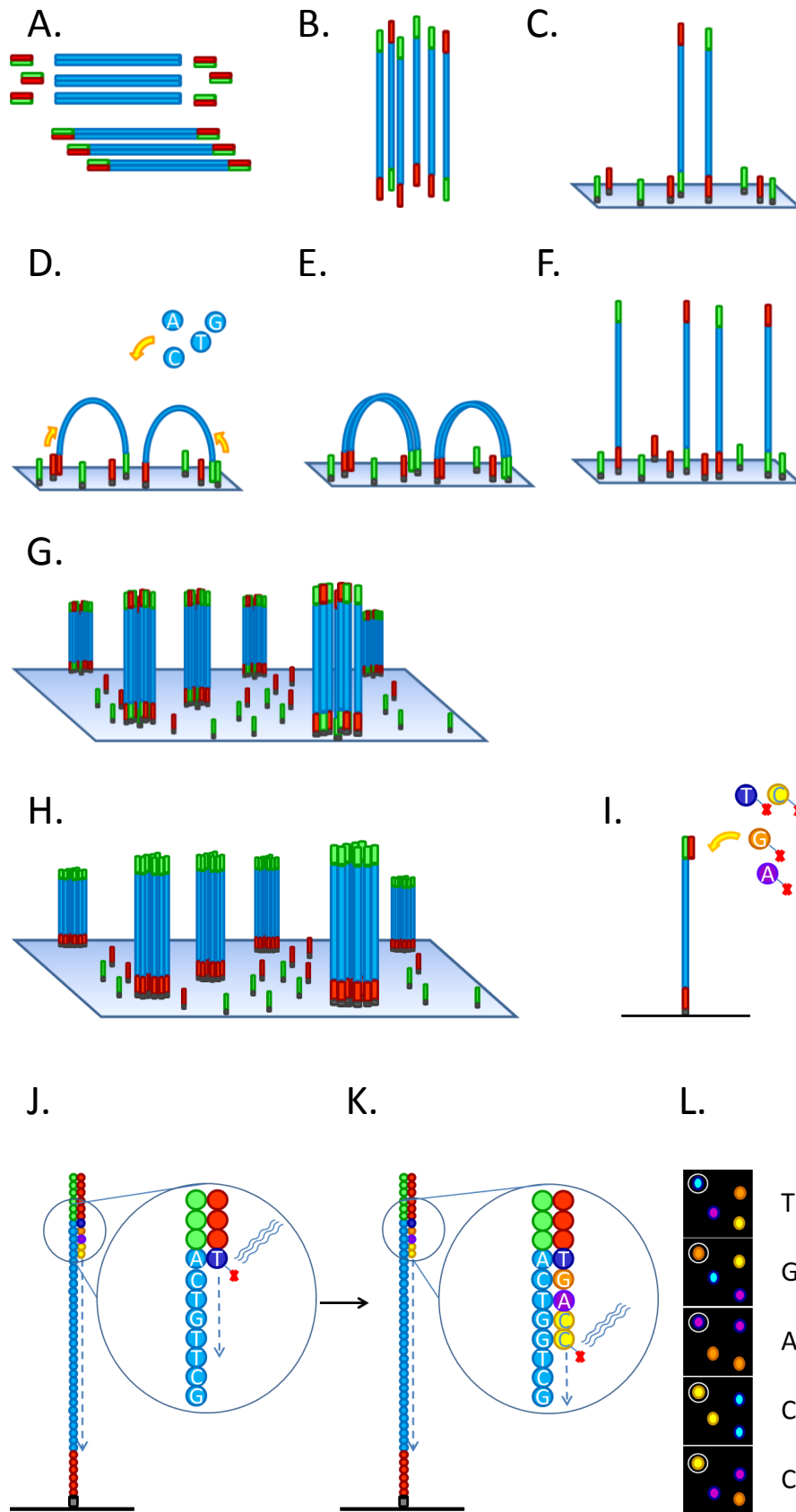
Figure 2.3. Overview of the Illumina sequencing platform – Bridge Amplification. (A.) Adaptors are ligated onto DNA fragments and (B.) separated into single strands. (C.) The adaptor end of the ssDNA binds to immobilized complementary oligonucleotides attached to the flow cell. (D.) The free end of the ssDNA binds onto another set of complementary oligonucleotides to form a bridge. Enzymes and unlabelled nucleotides are added to create a (E.) dsDNA bridge. (F.) The dsDNA is denatured leaving ssDNA anchored onto the flow cell. (G.) The bridging process is repeated several times, several thousand copies are generated within close proximity to each other forming clusters. (H.) The forward or reverse strands are removed. (I.) Sequencing primers complementary to the free end of the ssDNA are bound and (J.) polymerases incorporate a single fluorescently labelled, terminated nucleotide. A fluorescence image is captured, the fluorescence tag is cleaved and the terminator is unblocked ready for the next base to be added. (K.) The sequencing is repeated several times and the (L.) the DNA sequence is determined by the images taken (Kircher and Kelso 2010; Shendure and Ji 2008).

2.4.2. Genetic Marker types

Genetic markers are DNA sequences at a specific location on a chromosome. These markers are able to identify between individual lines and species through allelic variations in DNA structure, commonly referred to as polymorphisms. There are many types of genetic markers; restriction fragment length polymorphism (RFLP), amplified fragment length polymorphism (AFLP), variable number of tandem repeats (VNTRs), including mini and microsatellites and SNPs, are just a few examples.

Continuous advances in the development of new types of molecular markers have led to the generation of more useful and informative markers. Below are some examples of genetic markers which have contributed to molecular genetics.

2.4.2.1. Variable Number of Tandem Repeats

Variable Number of Tandem Repeats (VNTRs) are sequences of repeated DNA in tandem. VNTRs can be classified into 2 types; Minisatellites and or microsatellites. The allelic variation is determined through the differences in the length of repeats. VNTRs are highly polymorphic due to the high mutation rate affecting the number of repeats allowing for the identification of many alleles in a single locus.

Minisatellites are stretches of nucleotides varying between 10-100 nucleotide units long repeated in tandem spanning from 500 bp to several kilobases. The majority of minisatellites are GC rich with a strong strand asymmetry (Vergnaud and Denoeud 2000). Variation in minisatellite length can occur due to replication slippage, intramolecular recombination, unequal sister chromatid change, and unequal interallelic recombination or gene conversion (Armour and Jeffreys 1992). Identification of minisatellites involves restriction analyses and the use of DNA

probes and would be considered out-dated and difficult for large, high throughput genomic studies. PCR amplification has been incorporated but the disadvantage for minisatellites, compared to their microsatellite counterparts is that the alleles are too long for amplification to be consistently efficient.

Microsatellites or simple sequence repeats (SSRs) are stretches of 2-6 nucleotide units repeated in tandem and are spread across the genome. Allelic variation in the number of repeats can easily occur as an error during DNA replication. The ability for microsatellites to be easily amplified through PCR has made it quite popular as a genetic marker. The high variability of SSRs in transcribed and other single copy regions make them a powerful tool for genetic analysis.

In plant species, with the exception of tri-nucleotides, the insertion of microsatellites are under negative selective pressure in the coding region since they might interfere with protein synthesis (Morgante et al. 2002). The frequency of microsatellites in genomic DNA also differs significantly; *Arabidopsis thaliana* averages about 418.6 SSRs/Mb (~1 SSR / 2389bp), *Oryza sativa* (rice) about 363.3 SSRs/Mb (~1 SSR / 2753bp) and *Sorghum bicolor* about 175.4 SSRs/Mb (~1 SSR / 5701 bp) (Sonah et al. 2011). Aspects such as abundance, codominance, specificity, hypervariability and reproducibility once made SSRs the preferential molecular marker of choice.

2.4.2.2. Single Nucleotide Polymorphisms

A single nucleotide polymorphism (SNP) is a sequence variation occurring at a single nucleotide (or bp) position giving rise to different sequence alternatives or alleles between different individuals of the same species. SNPs are the most common type of allelic variation in

the genome. These allelic variances occur in high frequency and the abundance of SNPs makes

them particularly useful as genetic markers. There are two types of substitutions in SNPs;

transitions and transversions. About two-thirds of all SNPs are transition substitutions which

occur between purines, A⇔G, or between pyrimidines, C⇔T. Transversion substitutions occur

between a purine and a pyrimidine. These four substitutions are C⇔A, C⇔G, and T⇔A and

T⇔G (Brookes 1999; Vignal et al. 2002).

SNPs in the coding region can be nonsynonymous (where the amino acid is altered), or

synonymous (not causing a change in the amino acid), or they can simply occur in non-coding

regions. Although, a single SNP in a gene may or may not alter its structure or function,

multiple SNPs indicating multiple alleles at multiple loci, or a haplotype, enhances the

probability that the gene may not function correctly. This can influence promoter activity (gene

expression), messenger RNA (mRNA) confirmation (stability) and translational efficiency. Thus

SNPs may even play a direct role, with or without other factors, in the phenotypic expression of

diseases or traits.

SNP frequencies vary greatly in plants; in maize, 36 inbred lines revealed 1 SNP/124 bp

in the coding region and 1 SNP/31 bp in the non-coding region (Ching et al. 2002); 13 lines of

sugar beet (*Beta vulgaris* L.) showed 1 SNP/72 bp in expressed genes and 1 SNP/ 58 bp in the

non-coding region (Schneider et al. 2007); and 1 SNP/1.2 kb to 1 SNP/ 2.1 kb were found in 2

different lines of *B. napus* (Tapidor and Ningyou 7) (Trick et al. 2009). The ability to detect and

identify SNPs has increased dramatically with innovations in next generation sequencing.

SNPs were expensive to isolate because of the initial high cost of DNA sequencing.

Recent advances in sequencing technology, such as the Roche 454 and Illumina sequencing

platform, have reduced the cost of DNA sequencing significantly making SNP markers more

affordable (Kircher and Kelso 2010).

2.4.3. Applications

The use of genetic markers for *Brassica* has the potential to identify traits of interests.

Valuable traits which can be linked to one or many markers (or genes) can be isolated and bred

into other lines and species. Genetic linkage maps can be generated using molecular markers

and used to locate genes within the genome controlling agronomic traits. As well, genetic

linkage maps have helped elucidate the evolution of the *Brassica* species through comparative

mapping within its genus as well as with the model organism and family member *Arabidopsis*

thaliana.

2.4.3.1.   Genetic Linkage Maps

The generation of a sufficient quantity of molecular markers displaying polymorphisms

allows the production of a genetic linkage map. Genetic linkage maps are based on calculating

the frequency of recombination between molecular markers that have been assayed in

segregating populations. When determining the association between two genetic markers, the

greater the recombination frequency, the greater the genetic distance between the two markers.

Conversely, the closer the genetic markers are, the less likely a crossover event occurred between

them (Figure 2.4). In addition, some regions of the genome, particularly regions of low gene

density flanking centromeres and telomeres can have reduced recombination rates. With a

sufficient number of genetic markers displaying linkage and relatively evenly distributed across

the genome, linkage groups can be generated. These linkage groups, precursors to chromosome
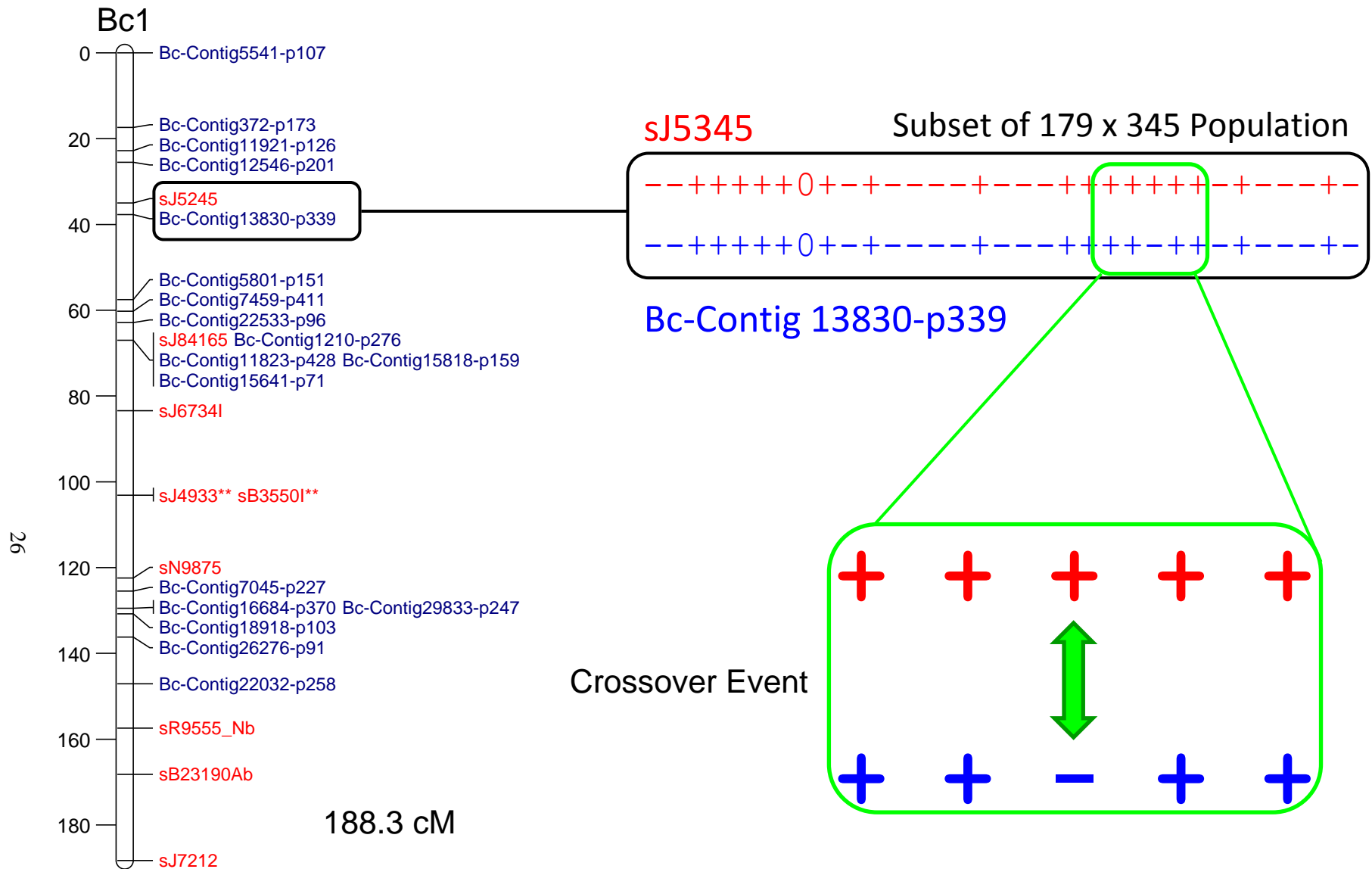
Figure 2.4. Genetic Mapping. (Left) Linkage group Bc1 generated from the 179 x 345 DH population. Map distance in centiMorgans (cM) are indicated on the left, marker names are indicated on the right. SSR and SNP markers coloured in red and green, respectively. (Top Right) Segregation patterns of SSR marker sJ5245 and SNP marker BC-Contig13830-p339 on a subset of the 179 x 345 population using a plus/minus matrix. The number "0" was used to indicate markers that did not amplify. The similarity in pattern is a result of the markers being in the same linkage group. (Bottom right) The difference in segregation pattern results from a recombination (or crossover) event. The frequency of such events determines the genetic distance between each marker.

designation, can help locate and manipulate important agronomic traits such as yield, disease and pest resistance and environmental and climate tolerance. Genetic linkage maps have been generated from all the *Brassica* species using various marker types (Axelsson et al. 2000; Chyi et al. 1992; Guo et al. 2012; Lagercrantz and Lydiate 1996; Landry et al. 1992; Lukens et al. 2003; Panjabi et al. 2008; Parkin et al. 1995; Slocum et al. 1990; Song et al. 1991; Teutonico and Osborn 1994) from the U's triangle.

2.4.3.1.1. Relationship to the model *Arabidopsis thaliana*

The model plant organism *Arabidopsis thaliana*, which also belongs to the Brassicaceae family, has been studied extensively and compared with *Brassica* species. It is believed that the genera *Arabidopsis* and *Brassica* split 20-24 million years ago (Koch et al. 2001; Koch et al. 2000; Lysak et al. 2005). Modern day *A. thaliana* contains five haploid chromosomes and is approximately 130 Mbps in size (Meyerowitz and Somerville 1994; Mozo et al. 1999), while diploid and amphidiploid *Brassica* species in U's triangle have 8-19 pairs of chromosomes and range between ~550-1,300 Mbp (Arumuganathan and Earle 1991; Johnston et al. 2005; U 1935). This divergence within the family occurred because of different evolutionary paths, involving multiple rounds of genome duplication and subsequent chromosomal rearrangements. In the year 2000, *A. thaliana* was the first plant genome to be sequenced (Theologis et al. 2000). The complete sequencing of the *Arabidopsis* genome has provided an understanding of the underlying genome complexity at the molecular level where individual genes and gene families can be studied in-depth in further detail. This breakthrough allowed for the development of specific molecular markers with the ability to identify homologous loci for comparative mapping across the genera and in this way define collinear chromosomal segments. In addition, genes of

27

interest in both *Arabidopsis* and *Brassica* species can be identified at a more rapid rate

generating a valuable tool to study genes and genome evolution.

2.4.3.2.   Comparative Mapping

A genetic linkage map derived from markers that are conserved within a family or

lineage is able to contribute to elucidating the evolution of a species. Genetic linkage maps of

Brassica species have been derived from markers with homology to sequences from the model

organism *A. thaliana*. The use of comparative mapping revealed *Brassica's* divergence from

Arabidopsis approximately 20-24 million years ago (Koch et al. 2001; Koch et al. 2000; Lysak et

al. 2005) involving the formation of a progenitor *Brassica* species through a whole genome

triplication event. It is from this ancestral hexaploid species that present-day diploid *Brassica*

species arose (Lagercrantz 1998; Lagercrantz and Lydiate 1996). Comparative mapping between

A. thaliana and *B. nigra* identified about 90 rearrangements since the divergence of the two

species.

In *Brassica*, many molecular markers from the diploid species can be found in the

amphidiploid species and vice-versa. Comparative mapping between the *Brassica* species using

genetic markers has allowed for the identification of homoeologous loci and collinear

chromosomal segments in diploid species *B. rapa* (A genome), *B. nigra* (B genome) and *B.

oleracea* (C genome) as well as allotetraploid species *B. napus* and *B. juncea* (Axelsson et al.

2000; Lagercrantz and Lydiate 1996; Parkin et al. 1995). Molecular analysis of the related

allotetraploids *B. napus* (AACC) and *B. juncea* (AABB) has determined that there have been no

major chromosomal rearrangements since the fusion of their respective progenitor genomes

(Panjabi et al. 2008; Parkin et al. 1995). Little is known about when polyploidization transpired

in *Brassica* species, but it has been suggested hybridization occurred during human cultivation less than 10,000 years ago (Rana et al. 2004). As well as elucidating the evolution of the *Brassica* species, comparative mapping has also identified genes of interest between species giving insight on the physiological modifications throughout history.

2.4.3.3. Traits of Interest

Distinguishing genotypes relevant to traits of interest has been a key goal in molecular genetics. Genetic markers based on DNA polymorphisms help isolate alleles at specific loci linked to genes of interest. Once these alleles are identified the gene can be bred or genetically engineered into lines and species that lack the trait of interest.

In *Brassicas*, traits of interest include increasing and maintaining yields of the material with commercial value, protecting against pest and diseases as well as unsuitable environmental and climate conditions. However, many of these traits have been studied in detail and the majority rely on multiple genes with small effect, collectively known as Quantitative Trait Loci (QTL).

In the Prairies, identifying molecular markers affecting factors such as early flowering and early maturity would be essential for production of *B. carinata* because of the short growing season.

2.4.3.4. Quantitative Trait Loci

In classic quantitative genetics, qualitative traits or "polygenes" were used to describe many genes having small effects influencing multiple genetic and environmental factors. (Mather

1941; Salvi and Tuberosa 2005). Quantitative Trait Locus (QTL) mapping uses the combination of molecular markers and trait data to link QTL and markers (genes) together.

Molecular markers reveal allelic differences in a specific locus that can be scored and mapped in a segregating population. While many of these segregating markers are in the non-coding regions which will not affect any trait directly, some will be linked affecting the trait of interest. The premise of QTL analysis is to determine where these linkages occur. Quantitative traits show a continuous range of variation in a population which is more or less normally distributed with no obvious discontinuities in this distribution such that one would find with a qualitative phenotype, controlled by a single gene, in say a F2 population, which would be expect to produce a 1:2:1 distribution. QTLs affecting the same trait can be on the same chromosome or spread across the genome (Kearsey 1998).

QTL mapping has been used in a number of *Brassica* species. QTLs for full and partial resistance to race 1 and 4 of the fungus *X. campestris pv. campestris* were located on linkage groups A06, A02 and A09 in *B. rapa* (Soengas et al. 2007). In *B. juncea*, 65 QTLs in 13 linkage groups influencing yield were found predominately in the A genome suggesting it played a more prominent role in domestication (Ramchiary et al. 2007). A study in *B. napus* revealed 13 QTL affecting fatty acid composition in seeds distributed across 10 linkage groups of which seven QTL also affected total seed oil content (Burns et al. 2003). To date, the only trait identified through QTL mapping in *B. carinata* is seed colour (Guo et al. 2012).

The use of comparative mapping combined with QTL mapping has also identified two major QTLs for genes controlling FT in *B. rapa* corresponding to two major QTLs identified in *B. napus*. By aligning these map positions, one of these pairs of QTLs show collinearity with the top of chromosome 5 of *A. thaliana*, which contains a candidate FT gene, while the other

displayed fractured collinearity with several regions of the *Arabidopsis* genome (Osborn et al. 1997).

2.5.   Summary

While many of the genetic studies have concentrated on the other *Brassica* species in the Triangle of U, little research has been focused on *B. carinata*. With the increasing world's population and the increasing need for more food, fuel and bio-industrial products, *B. carinata* should be explored in depth. To date, very little assessment of *B. carinata* on a molecular level has been carried out. A molecular genetic assessment is needed to understand this species to take advantage of the available quality traits and potential adaptation of *B. carinata* to relatively marginal areas. Research involving a genetic assessment to determine the genetic diversity of *B. carinata* lines and further the development of a genetic map would assist in locating genes and traits of interest to help in marker assisted breeding for variety development. In addition, such research could further elucidate the evolution of the *Brassica* species through comparative mapping of *B. carinata* with other *Brassica* species, as well as the model organism *A. thaliana*. A phenotypic assessment involving qualitative and quantitative traits would be beneficial in associating markers with traits of interest. This genetic assessment would be the first step in unlocking the potential of *B. carinata* as a crop platform for the Canadian Prairies and the world.

3. Assessing Genetic Diversity of *Brassica carinata*

3.1. Introduction

Genetic diversity is essential to continued improvement of a species. While the environment plays a significant role in influencing specific traits, it is the underlying genotypic differences that allow plant species to adapt and survive in changing environments. Many of these genetic changes may not affect gene expression; however, the combined effects of the resultant genotypic variations make each species and individual line unique. Genetic markers can be used to reveal the level of available genetic diversity within a species, which in turn can be associated with its ability to endure unfavourable climates and environments as well as qualities desired for agronomic improvement.

Brassica carinata until very recently was only grown by small-scale farmers in north-eastern Africa and has been paid little or no attention with regards to crop improvement (Hemingway 1976; Taylor et al. 2010). Although *B. carinata* is native to Ethiopia, many of the traits exhibited by Ethiopian mustard allow it to flourish in multiple environments. Geographically, Ethiopian mustard is well adapted to the highlands of northern Africa. However, *B. carinata* is not grown or sold outside of Ethiopia for commercial use (Taylor et al. 2010). The level of genetic diversity in *B. carinata* is unknown since there is limited knowledge and virtually no studies conducted at the genetic level.

Simple sequence repeat (SSR) markers are co-dominant, tend to be highly variable in nature and allow for the identification of many alleles at a single locus. An abundance of SSR sequences have been found in the *Brassica* genomes of *B. napus* and *B. juncea* (Andrew Sharpe, NRC Saskatoon, unpublished data) and more importantly they have been converted into robust

marker assays. The attributes of SSRs and their availability in related *Brassica* species made them excellent tools for studying genetic diversity in *B. carinata*.

A total of 39 *B. carinata* genotypes from world-wide germplasm collections including locally adapted varieties (Figure 3.1) were selected for diversity analysis. These genotypes were screened using *Brassica* SSR markers derived from the A and C genome of *B. napus* and the B genome of *B. juncea* to determine the levels of genetic diversity among the genotypes. SSRs from other *Brassica* species were used due to their availability and their ability to amplify loci in the *Brassica* B and C genome of *B. carinata*. SSR markers developed for the A genome were also tested due to its close relationship with the C genome (Harrison and Heslop-Harrison 1995; Howell et al. 2008; Lim et al. 2007).

Initially, the most diverse lines determined from this part of the study were planned to be used as parents to produce a DH population to generate a genetic linkage map for *B. carinata*. However, a DH population generated from the lines 179 and 345 was obtained from the JIC and these lines were included in the diversity study. If there was sufficient genetic diversity identified between 179 and 345, these lines along with the DH progeny would be used to generate a genetic linkage map.

The experiment tested the hypothesis whether previously generated SSRs were able to detect alleles from *B. napus* and *B. juncea*, and whether the same SSRs would be able to detect allelic variability from different *B. carinata* lines.
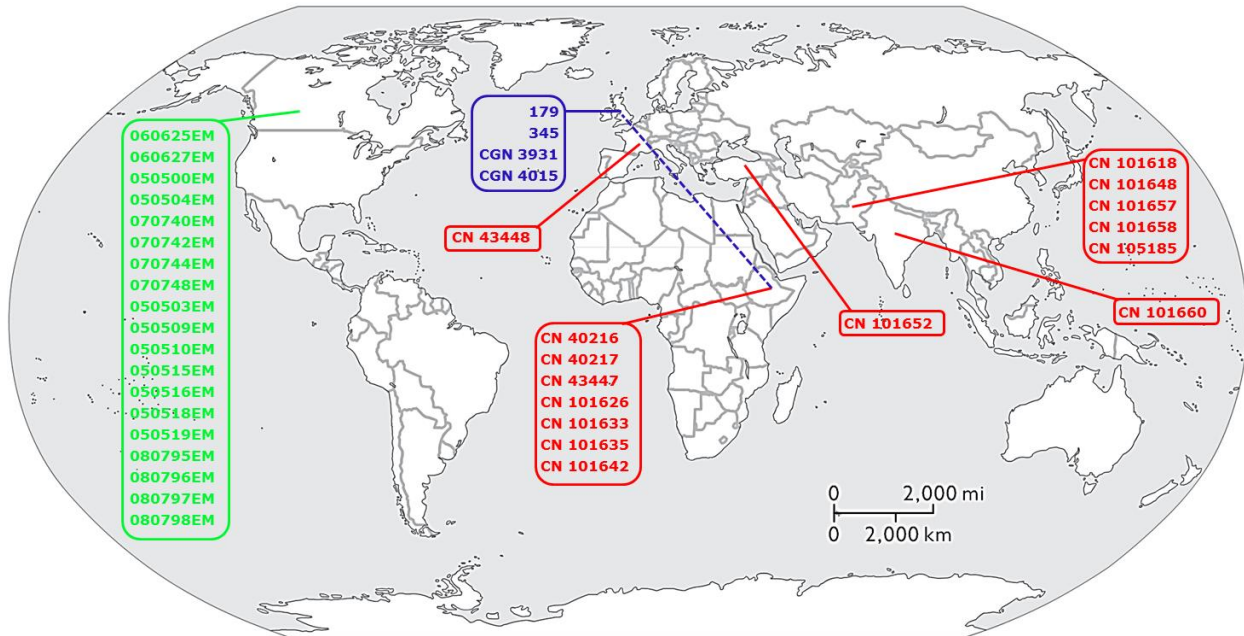
Figure 3.1. Source of *B. carinata* lines used to assess genetic diversity. ■ – represents AAFC breeding lines obtained from Kevin Falk in Saskatoon. ■ – represents World accessions obtained from Plant Genome Resources Canada. ■ – represents accessions obtained from the JIC, Norwich, UK (originated from Ethiopia). The location of accession CN 101665 was unknown.

3.2. Objectives

- To assess the genetic diversity among a range of *B. carinata* germplasm.

  - Genetic characterization of *B. carinata* lines for allelic variation using SSR markers.
  - To determine whether the genetic diversity between lines 179 and 345 was sufficient to generate a genetic linkage map.

3.3. Materials and Methods

3.3.1. Germplasm

A total of 39 *B. carinata* lines were chosen for this study (Table 3.1). Of the 39 lines used in the study 19 lines were provided by Dr. Kevin Falk of the Saskatoon Research Centre, Agriculture and Agri-Food Canada (AAFC), which were expected to be the most diverse among Falk's breeding lines, based on available phenotypic data. Sixteen accessions from around the world were obtained from Plant Gene Resources of Canada (PGRC) selected based on their geographic distribution. Of these 16 lines, seven were from Ethiopia, five from Pakistan, one each from Spain, India and Turkey. There was one accession, CN 101665, which was of unknown origin. The final four lines (179, 345, CGN 3931 and CGN 4015) originally from Ethiopia, were obtained from the JIC in Norwich, UK. DH line 179 was originally from a segregating population derived from CGN 4015, and was observed to be an average of 1.33 m in height (Section 5.4.1. Agronomic Traits and Seed Content), early flowering, and contained many branches with small pods. In contrast, DH line 345 was from a segregating population derived from CGN 3931 and is on average 2.45 m tall (Section 5.4.1. Agronomic Traits and Seed Content), late flowering, and contained few branches with large pods. A DH population consisting of 82 lines generated from a cross between 345 x 179 was also obtained from the JIC.

3.3.2. Isolation of Genomic DNA

Seeds of *B. carinata* were sown in the greenhouse at 18-20°C with an 18 hr day length with both natural and artificial light. About 50 mg of leaf tissue from each line was collected in a microtube and placed on ice prior to freeze-drying for 72 hr in a FreeZone 6 Liter Benchtop

Table 3.1. *Brassica carinata* accessions, origin and common name.

PGRC World Accession Lines	Alternative Names (Common)		Source
CN 40216	WIR 13-3/13		Ethiopia
CN 40217	WIR 66		Ethiopia
CN 43447			Ethiopia
CN 43448			Spain
CN 101618	Saryan	PAK 85876	Pakistan
CN 101626		R 796	Ethiopia
CN 101633		R 807	Ethiopia
CN 101635	Tamu	R 1826	Ethiopia
CN 101642		R 2815	Ethiopia
CN 101648		R 2822	Pakistan
CN 101652		R 2935	Turkey
CN 101657		R 2941	Pakistan
CN 101658		R 2942	Pakistan
CN 101660		R 2945	India
CN 101665		R 3624	Unknown
CN 105185	PI 181033	UC 77-916	Pakistan

AAFC Accession Lines	Alternative Name Used in Study	Source
060625EM	CR01	Canada
060627EM	CR02	Canada
050500EM	CR03	Canada
050504EM	CR04	Canada
070740EM	CR05	Canada
070742EM	CR06	Canada
070744EM	CR07	Canada
070748EM	CR08	Canada
050503EM	CR09	Canada
050509EM	CR10	Canada
050510EM	CR11	Canada
050515EM	CR12	Canada
050516EM	CR13	Canada
050518EM	CR14	Canada
050519EM	CR15	Canada
080795EM	CR101	Canada
080796EM	CR110	Canada
080797EM	CR123	Canada
080798EM	CR164	Canada

John Innes Centre Accession Lines	Source
179	England*
345	England*
CGN3931	England*
CGN4015	England*

* Originated from Ethiopia

Freeze Dry System (Labconco, Kansas City, MO, USA). A glass bead was added to the tissue

and the tissue was ground for 5 min at 20 rev/sec using the Retsch Mixer Mill. Two times CTAB

buffer containing 2% CTAB powder, 100 mM Tris-HCl pH 8, 20 mM EDTA, 1.4 mM NaCl, and

0.2% β-Mercaptoethanol (added prior to use) was added to the plant tissue and incubated at 65°C

for 1 hr. Then, 450 μL of chloroform was added and mixed for 5 min before being centrifuged at

5600g for 15 min. About 400 μL of the clear upper phase was collected and transferred carefully

to new tubes so as not to disturb the middle protein layer. Isopropanol, at 250 μL, was added and

the mixture inverted several times to precipitate the DNA. The samples were incubated at room

temperature for 10 min prior to centrifugation at 5600g for 15 min. The resultant pellet was

washed in 500 μL of 70% ethanol and centrifuged at 5600g for 10 min. The supernatant was

discarded and the pellets were dried at 65°C for 10 min. To resuspend the pellet, 50 μL of 1 x TE

buffer containing 10 mM Tris, HCl to pH 8.0 and 1 mM EDTA was added to 1 μL of RNase A

and incubated at 37°C for 30 min. DNA concentration was tested on the Nanodrop ND 100

Spectrophotometer and the DNA stored at -20°C.

3.3.3. Fluorescently-Labelled Tail PCR for Microsatellite Analysis

Fluorescently-labelled tail Polymerase Chain Reaction (PCR) was used to amplify SSRs

(Schuelke 2000; Wenz et al. 1998). SSRs from the A and C genome of *B. napus* and the B

genome of *B. juncea* generated by Dr. Andrew Sharpe (NRC Saskatoon) were used in this study.

The PCR mix contained 5 μM each of the forward, reverse and M13-21 fluorescently labelled

primer (VIC, FAM, or NED), 25-30 ng of *B. carinata* DNA and 2 x FideliTaq[TM] PCR Master

Mix (USB, Cleveland, OH, USA) in a final 8 μL reaction volume. Conditions for the PCR

amplifications were as follows: 95°C (5 min), 30 cycles of 94°C (15 sec)/ 50°C (15 sec)/ 68°C

(30 sec), followed by 8 cycles of 94°C (15 sec)/ 53°C (15 sec)/ 68°C (30 sec), and a final

extension of 68°C for 5 min. Three PCR reactions, each labelled with a different fluorescent

primer were pooled as follows: 1 μL VIC : 2 μL FAM : 3 μL NED and 6 μL ddH$_2$0.

Subsequently, 1 μL pooled PCR product was added to 0.075 μL GeneScan^{TM} ROX-500 size

standard (ABI, Warrington, U.K.) and 3.925 μL Hi-Di^{TM} Formamide (ABI, Warrington, U.K.)

and the samples were run by capillary electrophoresis on the ABI 3730XL DNA Analyzer. The

data were analyzed using the program Genographer v2.1.4 (Benham et al. 1999).

3.3.4. Genetic Distinctiveness Analysis

Average dissimilarity scores as described by Fu (Fu 2006) were used to calculate genetic

distinctiveness. For the total number of accessions n, the number of pairings each accession can

have is $n - 1$. SSR marker similarity S_{ij} between accession i and j is calculated as $(a + d)/(a + b +$

$c + d)$, where a is the total number of bands present in both i and j, b is the number present in i

and absent in j, c is the number present in j and absent in i, and d is the number absent in both i

and j. The SSR dissimilarity for each pair is therefore $1 - S_{ij}$. The mean (average) dissimilarity for

accession i is obtained by averaging all $n - 1$ SSR dissimilarities. Therefore, the higher the

average dissimilarity score (ADS), the greater the genetic distinctiveness (Fu 2006).

3.3.5. DARwin

Phylogenetic analysis was generated using DARwin (Perrier and Jacquemoud-Collet

2006). The dissimilarity between units i and j for allelic data was calculated using the simple

matching formula: $d_{ij} = 1 - \frac{1}{L}\sum_{l=1}^{L}\frac{m_l}{\pi}$ where L is the number of loci present, π is the ploidy and

m_l is the number of matching alleles for locus l (Perrier et al. 2003). A phylogenetic tree was

constructed based on the dissimilarity and the relationship between the lines by distance. The phylogenetic tree construction used a weighted neighbour-joining parameter after removing missing alleles.

3.4. Results

One hundred and sixty-five SSR markers were assessed on 39 *B. carinata* lines. Genetic variation as determined by differences in amplified fragment size was identified using capillary electrophoresis. Bands resulting from allelic variation in the length of the repeat element at SSR loci were assessed to identify the level of genetic diversity among the lines (Figure 3.2). If a marker identified an identically sized fragment in all lines and was thus determined to be monomorphic, then that SSR marker did not contribute to the assessment of genetic diversity within the lines. From the 165 tested markers, 81 (49%) were polymorphic, 53 (32%) were monomorphic while 31 (19%) could not be amplified. Each of the polymorphic markers were scored with the variable size of the fragments (bp) recorded for each line. Of the 39 lines tested, two lines, CN101665 and CR06, consistently showed poor amplification products, which was probably the result of degradation of the DNA, and were removed from the subsequent analyses. The data from all SSR markers were collected and comparisons were made between each pair of lines.

3.4.1. Genetic Diversity between the accession lines

Two different methods were used to measure the genetic variation within the collection of lines. The first calculates the average dissimilarity scores (ADS) between the lines (Fu 2006).
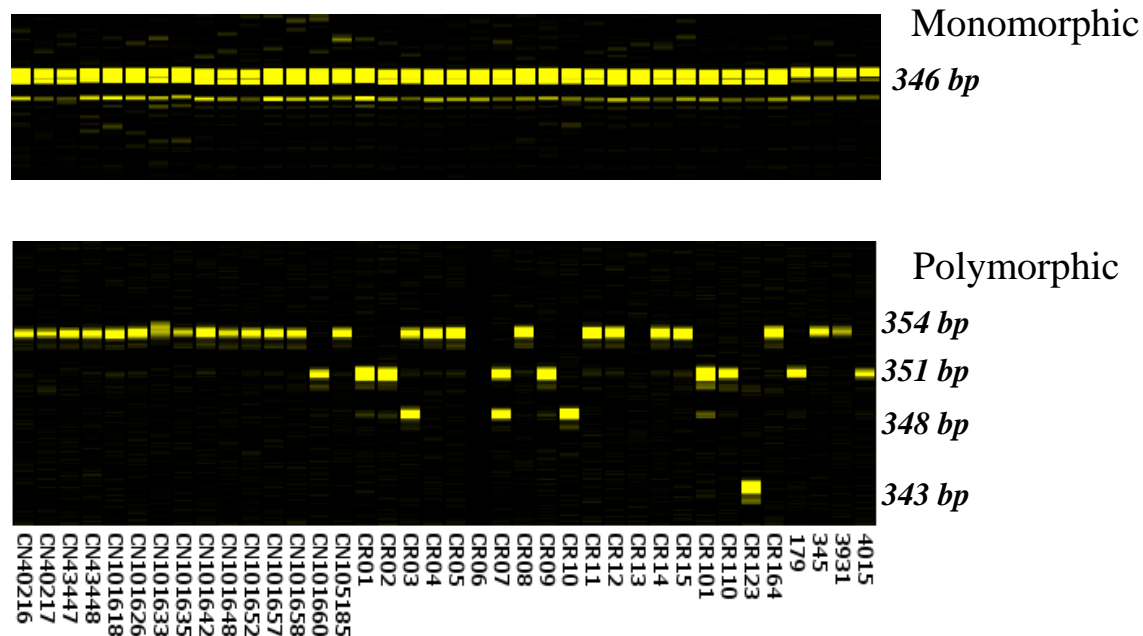
Figure 3.2. Amplification of SSR alleles in *B. carinata* lines. Banding pattern of selected *B. carinata* lines using SSR markers sJ6842 (Top) and sJ7104 (Bottom) showing fragments representing fluorescently labelled PCR products. Each band represents an SSR allele with a variable bp length. The top screen showing the same size fragments amplified in all the lines is considered monomorphic and uninformative. However, the bottom screen illustrates 4 different alleles at the same SSR locus. It is these polymorphic variations that allow us to assess genetic diversity. By studying many markers showing multiple polymorphic loci, it is possible to determine the genetic relatedness of the lines.

The second method using the software program DARwin allowed the development of a phylogenetic tree to analyze the relationship among the multiple *B. carinata* lines. Accessions containing missing data from SSR markers during pairwise comparison were discarded when at least one of the two values was missing.

3.4.1.1.   Average Dissimilarity Scores

The calculated mean of the dissimilarity scores (ADS) was determined from pairwise comparisons across all lines. The fragment sizes were converted to binary format and scoring was determined by the presence of a band as "1" and the absence of a band as "0". The abbreviation N/A was given if there were no bands present in the lane and was not included in the calculation. SSRs considered to be monomorphic across all lines were not included as they were uninformative. The ADS was then calculated to measure the overall genetic difference between the lines of interest. The lower mean dissimilarity, the more genetically redundant the accession is within the collection. A higher mean dissimilarity between unlinked SSR markers denotes that the specific line has a distinct genetic background compared to other tested lines (Fu 2006).

The ADS of the 37 accession lines ranged from $0.146 - 0.213$ with a mean of 0.173 and a standard deviation of 0.017 (Table 3.2) (Figure 3.3). There were seven *B. carinata* accessions with an ADS score of less than $0.156 (= 0.173 - 0.017)$ and four accessions with an ADS score greater than $0.190 (= 0.173 + 0.017)$. Lines with values below 0.156 or above 0.190 could then be considered genetically redundant or genetically distinct, respectively.

Of the seven genetic redundant lines (CR110, CR10, CN101658, CR01, CR101, CR13, and CN101660), five were from the AAFC breeding program. Of the six genetically distinct

Table 3.2. Average genetic diversity (ADS) of 37 *B. carinata* lines calculated using Fu's (2006) method. The ADS was calculated to measure the overall genetic difference present between the lines of interest. Calculated mean of the dissimilarity scores determined from pairwise comparisons across all lines. A higher ADS obtained between unlinked SSR markers indicates that the specific line is more genetically distinct compared to other lines.

AAFC		PGRC (World Accessions)		JIC	
Line	ADS	Line	ADS	Line	ADS
CR01	0.151	CN40216	0.176	179	0.205
CR02	0.166	CN40217	0.162	345	0.190
CR03	0.182	CN43447	0.168	CGN3931	0.174
CR04	0.181	CN43448	0.195	CGN4015	0.213
CR05	0.179	CN101618	0.200		
CR07	0.177	CN101626	0.181		
CR08	0.183	CN101633	0.173		
CR09	0.161	CN101635	0.200		
CR10	0.148	CN101642	0.171		
CR11	0.165	CN101648	0.169		
CR12	0.171	CN101652	0.167		
CR13	0.152	CN101657	0.168		
CR14	0.162	CN101658	0.149		
CR15	0.171	CN101660	0.153		
CR101	0.151	CN105185	0.173		
CR110	0.146				
CR123	0.194				
CR164	0.168				
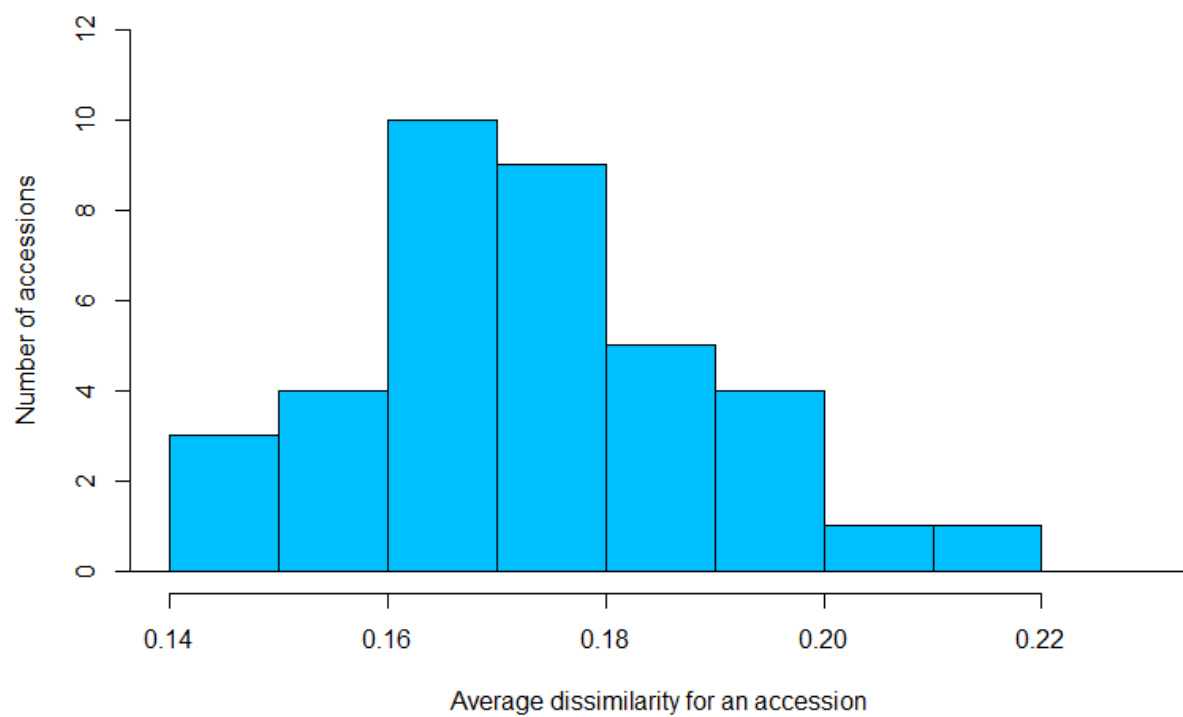
Figure 3.3. Frequency distribution of mean SSR dissimilarities for 37 *B. carinata* lines.

lines, three were from the world accession group (CN43448 – Spain, CN101618 – Pakistan and

CN101635 – Ethiopia), one was from AAFC (CR123) and two were from the JIC. The mean

ADS of the AAFC lines were 0.167 with a standard deviation of 0.014 and ranged from 0.146 –

0.194. The mean ADS of the World Accessions were 0.174 with a standard deviation of 0.015

and ranged from 0.149 – 0.200. The lines from the JIC had a mean ADS was 0.196 with a

standard deviation of 0.002 and a range of 0.174 – 0.213. Since a DH population was generated

from a cross between 179 and 345, these two lines warrant further study to determine the levels

of genetic dissimilarity. The ADS for 179 and 345 were 0.205 and 0.190 respectively, both are at

or higher than the mean and standard deviation for the other lines in diversity set. Thus, these

lines would be considered as genetically distinct allowing them and their derived DH population

to be used for further molecular genetic studies.

3.4.1.2.   DARwin

Another common method that can be used to assess genetic diversity data was

implemented in the DARwin software. A phylogenetic tree was generated using the software

program DARwin (Perrier and Jacquemoud-Collet 2006). As indicated before, lines CN101665

and CR06 were not included in the data analysis. The resulting phylogenetic tree from DARwin

showed five distinct clusters (Figure 3.4), which largely corroborated the ADS analysis. The first

branch contained 11 lines, 10 from AAFC and one from the PGRC world accession. The second

cluster contained four lines all from the AAFC accessions. The third cluster contained 11 lines of

which three were from AAFC, six were from PGRC and two lines, CGN3931 and 345, were

from JIC. The fourth cluster contained a total of six lines, one from AAFC, three from PGRC
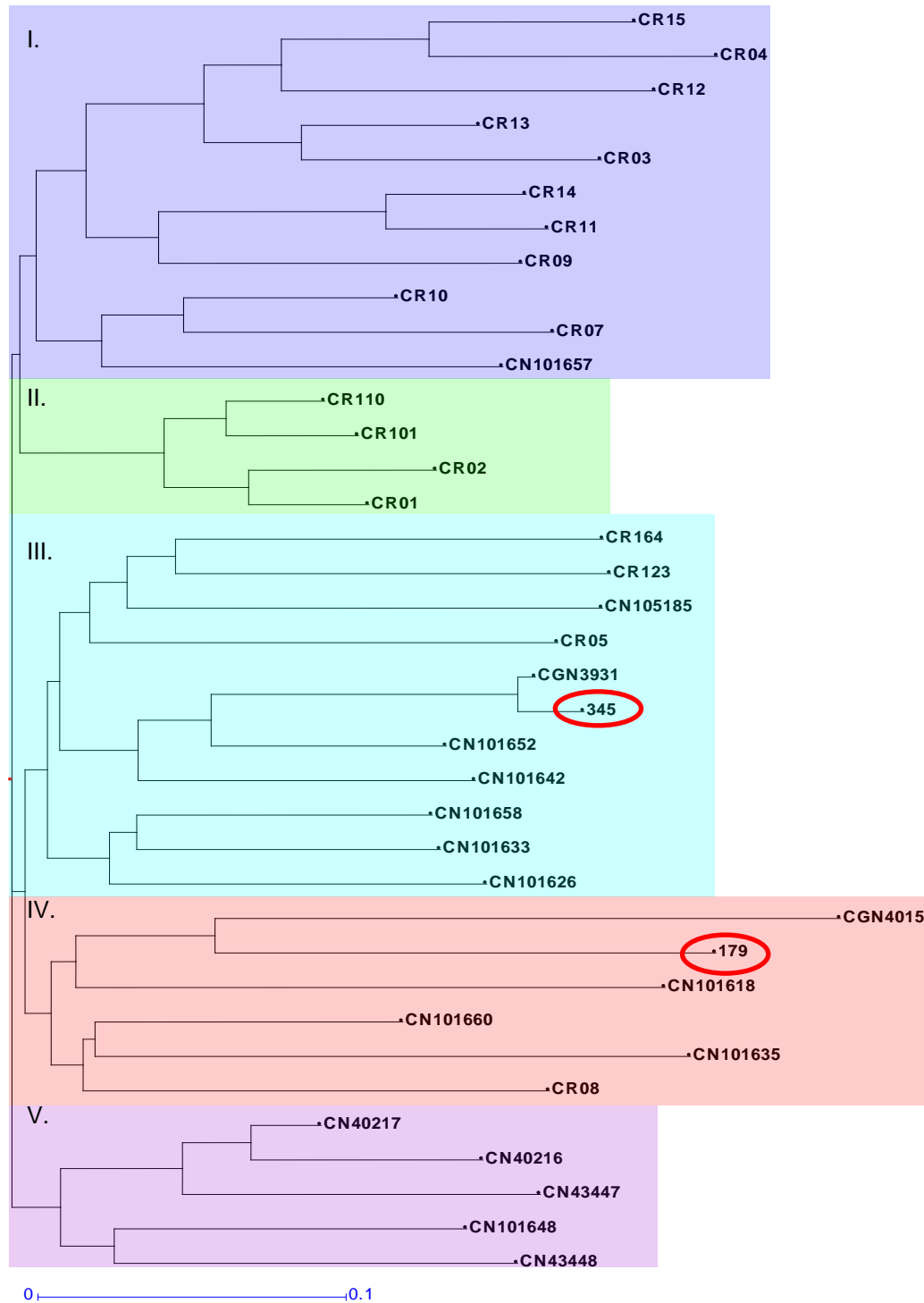
Figure 3.4. Phylogenetic Tree displaying the relationship of *B. carinata* from the analysis of SSR data. Scale in blue represents proportion of SSR dissimilarity between the lines. Parental lines 179 and 345 (circled red) were crossed to produce a DH population obtained from the JIC. *B. carinata* lines CN101665 and CR06 were omitted. Clusters I-V are labelled and separated by colour.

and two lines, CGN4015 and 179, from JIC. The fifth cluster contained 5 accessions, all from PGRC.

Many of the AAFC accessions were grouped together as shown in the top two clusters, while the PGRC world accessions did not show any specific pattern linking the accessions geographically. The phylogenetic tree separated the parents of the available DH mapping population, 179 and 345, into separate clusters. The number of shared alleles between 179 and CGN4015 as well as 345 and CGN3931 was also illustrated by their proximity to each other within the phylogenetic tree as expected, since 345 was derived from CGN3931 and 179 from CGN4015.

3.5. Discussion

Since there are close similarities between *Brassica* species, the use of *Brassica* SSRs generated from within the family would presumably allow the amplification of loci from the *B. carinata* genome (Lagercrantz et al. 1993). By using SSR markers previously tested on other *Brassica* species, genetic differences between *B. carinata* lines were displayed through the amplification of polymorphic bands. The *Brassica* SSRs tested identified 81 polymorphic markers, which was 49% of the markers tested, within the 37 *B. carinata* diversity lines. Average Dissimilarity Scores (ADS) between the lines were calculated to measure genetic distinctiveness. The mean ADS of the AAFC breeding lines (0.167) was lower than those observed for the accessions provided by the PGRC (0.174). Although the AAFC *B. carinata* lines were phenotypically diverse, the lower ADS values probably arises since these lines have been bred to be adapted to the Canadian Prairie environment and have been selected over a number of generations for good agronomic potential (Personal communication, Kevin Falk).

47

Breeding strategies often reduce genetic variability within species due to constant selection for a small number of valuable traits (Snowdon and Friedt 2004).

In contrast to other species it appears the relative level of genetic diversity within *B. carinata* is slightly lower. A study looking at 45 Canadian soybean cultivars tested against 37 SSR primers revealed ADS scores between 0.161 to 0.216 with a mean ADS of 0.185 while the ADS of 37 exotic germplasm were not surprisingly higher and ranged from 0.205 to 0.250 with a mean of 0.221 (Fu et al. 2007). A study between 219 barley accessions using 30 SSR markers revealed a dissimilarity range from 0.175 to 0.222 with a mean of 0.199 (Fu 2012). It is possible that although efforts were made to select as diverse a subset of material as possible based on available information, the current study is limited by numbers of accessions. However, the low level of diversity found within *B. carinata* lines could reflect its evolutionary origins, that is *B. carinata* may have been formed through a relatively small number of fusion events with limited exchanges having occurred with its diploid progenitors. Based on the previously tested SSR markers in other *Brassica* species, it would be presumed that 52 of these markers would amplify loci on the B genome and 29 on the C genome (Parkin, unpublished data). The mean ADS comparison between the B and C genome markers revealed that the mean ADS was lower in the C genome (0.164) than in the B genome (0.177). These results suggest that the C genome is less polymorphic than that of the B genome in *B. carinata* which would also contribute to a lower level of genetic diversity.

The ADS calculated for 179 and 345 indicated that these two lines are genetically distinct. The results suggest that had not there been a DH population constructed between 179 and 345, these genetically diverse lines of *B. carinata* would have been a suitable pair of

candidates that could be chosen for population development. This higher dissimilarity score supported the continued use of the 179 x 345 DH population.

Phylogenetic tree development using the software program DARwin was used to visualise the relationships among *B. carinata* lines. The close proximity of 345 with CGN3931 and 179 with CGN4015 confirmed that these lines were derived from the CGN3931 and CGN4015, respectively and validated the analyses. There was no differentiation based on geographical origin; however, it is often difficult to confirm the actual origin of lines within core germplasm collections. The limited breeding efforts thus far in *B. carinata* were reflected in the tree with the widest observed variation being that found within the original Ethiopian landraces, with lines from Ethiopia being observed in all branches of the tree apart from the two containing only the adapted Canadian lines.

Notably the two parents of a DH population, 179 and 345, fell into two separate branches within the phylogenetic tree which supports the use of the DH population to generate a genetic linkage map.

3.6.   Conclusion and Future Research

Previously generated SSR markers from *B. napus* and *B. juncea* were able to detect alleles in 37 *B. carinata* lines. Overall, these lines were less genetically diverse than previously studied species. As well, the amphidiploid nature of *B. carinata* revealed more polymorphisms in the B genome when compared to its C genome counterpart. While future studies on genetic diversity of *B. carinata* should involve more polymorphic markers and more accessions, a closer look at the diversity within the genomes warrant further study.

Accessing genetic diversity through the genetic screening of lines 179 and 345 revealed these lines were genetically distinct. This also allowed for the 179 x 345 DH population to generate a genetic linkage map by using the SSR markers polymorphic between 179 and 345. Polymorphic markers between 179 and 345 would allow for the generation of a low density genetic backbone. The putative map positions of the polymorphic SSR loci could be identified by studying previously generated linkage maps using the same SSR markers. It would be anticipated that in the allotetraploid genome of *B. carinata* SSR markers displaying multiple loci would likely have one locus on each of the *B. carinata* B and C genomes. These SSR markers could potentially generate a genetic template for *B. carinata* which will further be expanded with the inclusion of SNP markers in the next chapter.

4. Genetic Mapping of *Brassica carinata*

4.1. Introduction

Brassica carinata has many interesting traits that give it potential for use in multiple applications. These traits include pest and disease resistance, climate and environmental tolerances and other agronomic traits, which could turn *B. carinata* into a valuable crop platform. Creating a detailed genetic linkage map for *B. carinata* would reveal the genetic architecture as well as provide a tool to locate genes and QTLs controlling traits of interests. In addition, a molecular linkage map for *B. carinata* would help elucidate the evolution of the *Brassica* species through comparative mapping within its genus, as well as with the model organism and family member *Arabidopsis thaliana*.

Detailed genetic linkage maps have been generated for five out of the six *Brassica* species within the Triangle of U (Axelsson et al. 2000; Chyi et al. 1992; Lagercrantz and Lydiate 1996; Landry et al. 1992; Lukens et al. 2003; Panjabi et al. 2008; Parkin et al. 1995; Slocum et al. 1990; Song et al. 1991; Teutonico and Osborn 1994). Recently, a *B. carinata* genetic linkage map consisting of SSRs, intron based polymorphisms (IBPs), amplified fragment-length polymorphisms (AFLPs) and sequence related amplified polymorphic markers was generated (Guo et al. 2012). *Brassica napus* and *B. rapa* are the only two *Brassica* species to have genetic linkage maps with comprehensive SNP marker integration (Bancroft et al. 2011; Li et al. 2009).

Both SSR and SNP markers are highly variable and can be automated. However, the abundance of SNPs is an advantage for accessing both genetic variation and genome structure. The use of short-read sequencing technologies has allowed for reliable, cost-effective coverage of the transcriptome without sequencing the complete genome. Generating such sequence data from the parental lines of mapping populations can identify useful polymorphisms (SNPs) at a

much faster rate and at a high frequency compared to screening for other types of genetic markers. In addition, they can act as robust markers for genes in their vicinity and potentially represent alterations in protein structure if located within the coding regions. SNP variation tends to be unevenly distributed between the protein coding regions and the untranslated regions (UTR) because of the potential for functional mutation. (Rafalski 2002). Targeting the search for sequence variation to the 5' and 3' UTRs of genes will increase the frequency of identified polymorphisms up to 3 fold (Rafalski 2002). SNP discovery protocols have been established to generate SNP markers tagging genes of interest using a 3' region targeted approach (Eveland et al. 2008). In comparison to shotgun sequencing approaches, where randomly sheared cDNA fragments can be assembled into contiguous sequences (contigs) generating full coverage of the cDNA, targeting the 3'UTR sequence allows individual mRNAs to be distinguished without having to assemble the complete cDNA and importantly increases the depth of sequencing in the target region (Figure 4.1). Increasing the sequence depth allows the accuracy and certainty of the SNP identification to increase. Using 3'UTR transcript profiling also allows for increased resolution of near identical transcripts and has been shown to effectively distinguish evolutionary related duplicate genes or paralogs in paleopolyploid species (Eveland et al. 2008).

The availability of previously generated microsatellite markers allowed for the initial genetic screening of a range of *B. carinata* lines. This in-turn allowed the generation of a genetic framework utilising a DH population derived from a cross between parental *B. carinata* lines 179 and 345. Second generation sequencing of 3' targeted cDNA libraries has permitted the rapid generation of sequences from the parental lines to saturate the map with markers that have been mined to discover SNPs between the parental lines 179 and 345. These SNPs were converted to PCR based assays and the resultant markers were used to fill in gaps in the genetic framework

A. Shotgun
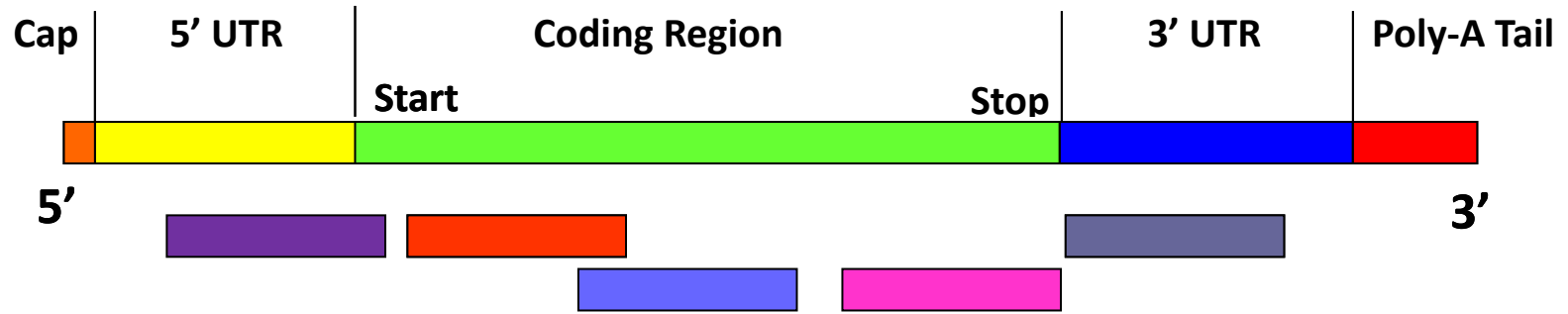
B. 3' Anchored
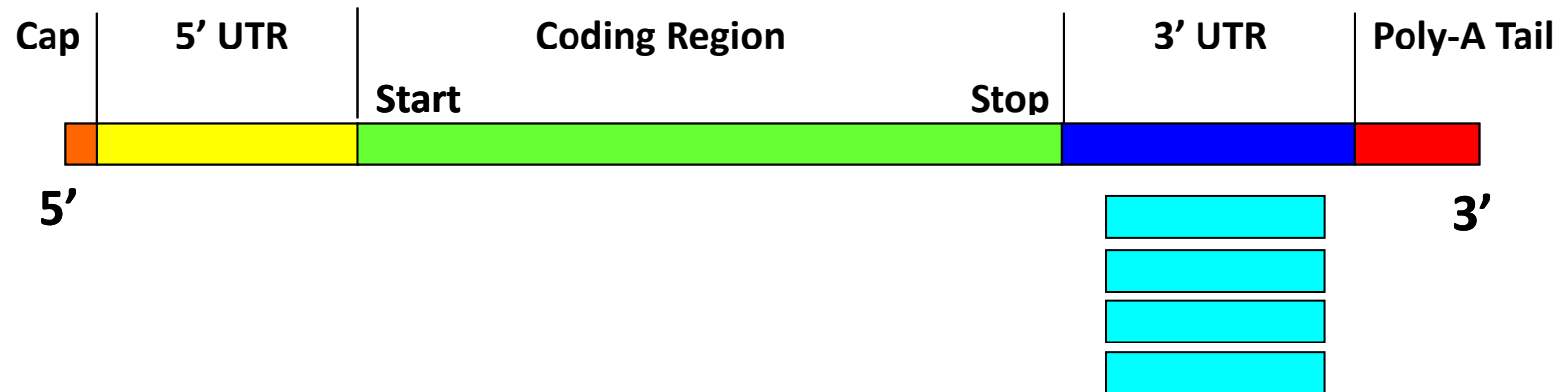
Figure 4.1. A comparison of transcript profiling of 454 sequences using the (A) shotgun and (B) 3'anchored approach. A, Contig assembly of random cDNA fragments providing full length coverage of cDNAs giving a lower average read depth. B, 3' UTR, rich in SNPs, are targeted providing increased depth for SNP discovery without the need to assemble full length cDNAs (Eveland et al. 2008).

created from allelic variations of SSRs for *B. carinata* lines 179 and 345 to develop a dense genetic map for *B. carinata*. Generating a genetic linkage map for *B. carinata* will provide a useful tool for marker-assisted breeding in this relatively unimproved crop species.

The hypothesis being tested was that since the formation of the allopolyploid *B. carinata* the genetic linkage groups of its progenitor genomes will have been maintained relatively unchanged, such that they closely resemble the genomes of current day *B. nigra* (BB) and *B. oleracea* (CC). In addition, it is also hypothesised that based on previous phenotypic assessment there would be sufficient polymorphism observed between the parental lines 179 and 345 to generate a dense genetic linkage map for a *B. carinata*.

4.2. Objectives

- To develop a dense genetic linkage map for *B. carinata*

  - Polymorphic SSR markers identified in the previous chapter between lines 179 and 345 will be used as the genetic framework to create the linkage map for *B. carinata.*

  - Genome wide SNP discovery was carried out using second generation sequencing technology. *Brassica carinata* lines 179 and 345 were targeted for SNP discovery. Validation of these SNP was performed on the segregating DH population (179 x 345) to generate a dense linkage map.

  - Comparative analysis of the linkage groups in *B. carinata* against known B and C genome *Brassica* species was used to determine if chromosomal rearrangements had occurred since the generation of the amphidiploid species *B. carinata*.

4.3.   Methods and Materials

4.3.1. SSR Analysis

SSR analysis on the 179 x 345 population was performed as mentioned in Section 3.3.2. Isolation of Genomic DNA and Section 3.3.3. Fluorescently-Labelled Tail PCR for Microsatellite Analysis. Amplification of alleles were converted and scored as a plus/minus (+/-) matrix.

4.3.2. cDNA

The *B. carinata* parental lines 179 and 345 were sown in the greenhouse at 18-20°C with an 18 hour day length of both natural and artificial light. Leaf tissue was collected and total RNA was isolated using the RNeasy Plant Mini Kit (QIAGEN). The concentration of the RNA sample was measured on the NanoDrop ND 1000 – Spectrophotometer (Thermo Fisher Scientific Inc.) A RiboMinus plant kit for RNA Sequencing (Invitrogen) was used to isolate the transcribed products by removing ribosomal RNA (rRNA) from the total RNA, which was then used for cDNA synthesis.

To synthesize the first strand of cDNA using reverse transcription (RT) a total of 5 µg of ribominus RNA and 100 pmol Oligo(dT) primer (Btin-B-R4 primer) in a 16 µL total reaction volume was incubated at 70°C for 10 min. Four µL of reverse transcription master mix consisting of 5 X ArrayScript RT buffer (Ambion), 2.5 mM dNTPs and 200 units of ArrayScript RT was added to each RNA sample and incubated at 42°C for 2 hr followed by 70°C for 15 min. Two µL of RNase Cocktail Enzyme Mix (Ambion) was added and the sample incubated at 37°C for 30 min. A Second Strand Master Mix consisting of 1.28X DNA Polymerase I Reaction

Buffer, 128 µM dNTPs, 30 units of *E. coli* DNA Polymerase I (Fermentas) and 1 unit of *E. coli*

RNase H (New England BioLabs Inc. (NEB)), total volume 78 µL, was added to each first strand

cDNA sample in a volume of 100 µL and placed at 15°C for 2 hr. Ten units of T4 DNA

Polymerase (Fermentas) was added to each sample and left at 15°C for 5 min followed by the

addition of 7 µL of 0.5 M EDTA, pH 8 to stop the reaction. A 1:1 total cDNA sample and

Phenol:Choroform:Iso-Amyl Alcohol (25:24:1, pH 8.0) was added to a phase-lock gel tube

(Inter Medico) and centrifuged at 10,600 x g for 5 min. The aqueous solution was transferred to a

clean tube along with 500 µL of Buffer PB (MinElute Kit, Qiagen). The sample was loaded onto

a Qiagen MinElute column and centrifuged at 10,600 x g for 1 min. The flow through was

discarded and 750 µL of Buffer PE (MinElute Kit, Qiagen) added and centrifuged twice at

10,600 x g for 1 min, each time discarding the flow through. A volume of 13 µL 1:2 ratio of

Buffer EB (MinElute Kit, Qiagen) and 5 mM Tris Buffer preheated to 55°C was added to the

column and centrifuged at 10,600 x g for 1 min. This elution step was repeated for a final volume

of 25 µL.

The next step was to prepare the 3' cDNA 454 Library for sequencing. Restriction

enzyme digestion occurred by adding 5 µL of 10X NEB buffer 3 and 5 units of *Aci*I (NEB) to the

25 µL cDNA sample, made up to 50 µL with dH$_2$O and incubated at 37°C for 1 hr. Next, 35 µL

of AMPure beads (Agencourt®) were added which bind to the cDNA. This allows for the DNA

bound magnetic beads to be separated from small fragments using a magnetic particle

concentrator (MPC). The supernatant was removed and the beads washed twice with 500 µL

70% ethanol. The pellet was placed at 37°C for 10 min to completely dry the beads after which

they were resuspended in 25 µL Elution Buffer (Agencourt®) to remove the cDNA from the

beads. The beads were pelleted using the MPC and the supernatant containing the cDNA transferred to a new tube.

Purification of the 3' cDNA fragments 50 µL of prepared DynaBeads M-270 streptavidin beads (Invitrogen) were added to the cDNA and placed in a tube rotator for 20 min. The supernatant was removed and the bead mixture washed twice with 50 µL of 1 X binding and washing (B&W) Buffer and then washed twice with 50 µL of 1 X Ligase Buffer (NEB). A mixture of 5 pmol A- adaptor, 20 µL 2X Quick Ligase Reaction Buffer and 2 µL Quick Ligase (NEB) in 40 µL reaction was added to the cDNA library carrying beads and incubated on a tube rotator at 25°C for 20 min. The beads were then washed twice with 50 µL of 1X B&W Buffer and twice with 50 µL of 1X ThermoPol Reaction Buffer (NEB). A mixture containing 1X ThermoPol Reaction Buffer, 1 X Bovine Serum Albumin (BSA), 0.2 mM dNTP mix and 24 units of *Bst* DNA Polymerase, and dH_2O to a final volume of 50 µL was added to the library-carrying beads. This mixture was incubated at 42°C for 20 min and washed twice with 50 µL of 1 X B&W Buffer and twice with 50 µL of dH_2O before all the liquid was removed. The final process was to isolate single stranded template DNA (sstDNA). A 50 µL Melt Solution containing 125 N sodium hydroxide (NaOH) and dH_2O was used to wash the library carrying beads. The solution was placed on the MPC and the supernatant containing the sstDNA removed and added to the Neutralization solution containing 60 mM Sodium Acetate pH 5.2 and 500 µL of PB Buffer (Qiagen). This step was repeated once. The neutralized sstDNA was cleaned using a MinElute column (Qiagen) as described above and the sstDNA eluted in 15 µL of EB Buffer. The resultant sstDNA library was sequenced at the National Research Council DNA Sequencing Laboratory (NRC-Saskatoon, Canada) on the 454 Genome Sequencer (Roche, Indianapolis, IN., USA).

4.3.3. SNP Discovery

Once the data was obtained from the 454 Genome Sequencer, the sequences were trimmed and analysed. *De novo* assembly of the reads from 179 was performed by DNAStar's SeqMan NGen 2.1.0 software (DNAStar, Madison, WI., USA) and used as reference. The stringency of the assembly was set at 95% sequence identity so that any homologues that were less than 95% identical were separated into distinct contigs. The reads were trimmed for the 454 specific key sequence and end quality. For each read, 10 bps were trimmed at the 5' end to remove the key sequence. A Phred score, which determines a quality value from an error probability of a particular base call, was set at minimum of 20 over a sliding 5 bp window to ensure read quality.

The 179 contigs were compared to the B genome of *B. nigra* (DH No100) and the A genome of *B. rapa* cv. Chiifu using Basic Local Alignment Search Tool (BLASTN) version 2.2.16 (Altschul et al. 1990), specific for searching nucleotide databases using a nucleotide query. At the time, a C genome reference was not available and the A genome was used instead due to its similarity to the C genome (Howell et al. 2008).

The reads from 345 were reference mapped to 179 contigs, using a 98% similarity threshold to account for sequence variation, using CLC Bio Genomics Workbench (CLC Bio, Denmark). Potential SNPs were identified by first filtering for those which showed complete agreement in the reads aligned at the SNP position. Ambiguous bases in the reference were also eliminated as potential SNP candidates. Robust SNPs with a read depth of at least four and 120 bp of flanking DNA (60 up/down stream of SNP) were compared to *B. rapa* and *B. nigra* using BLAST-Like Alignment Tool v. 0.34 (BLAT) (Kent 2002) to eliminate sequences crossing exon/intron splice sites and were selected for single-plex SNP analyses using the KASPar system

(KBioscience Ltd, Hoddesdon, U.K.). Each SNP assay contained two allele specific primers, one for each SNP allele and one common reverse primer, which were designed using the Primer Picker Lite for KASPar version 0.27 software (KBiosciences Ltd.). The SNP markers were named according to the bp position in the contig where they were identified. These SNPs were tested on the 179 x 345 DH population.

4.3.4. SNP Testing

The purified DNA from the parental lines 179 and 345 as well as the DH population derived from these lines (as described in Section 3.3.2. Isolation of Genomic DNA) was used to map the identified SNPs. The population contained 76 lines (Line 1-75, 77); however line 36 consistently did not amplify and was used as a negative control.

In the KASPar genotyping system once the DNA is denatured, the allele-specific primer binds to the complementary region upstream from the SNP with the 3'end of the primer positioned at the SNP nucleotide. The common reverse primer also anneals to its respective region and amplification begins. The primers are extended and the product once again denatured. The allele specific primers incorporate a tail sequence that is complementary to fluor-labelled oligos. The target sequence and the Fluorescence Resonance Energy Transfer (FRET) cassette are denatured allowing the annealing and incorporation of a fluor-labelled oligo to the specific complementary tail sequence. The 5'labelled oligos, FAM or HEX, are specific to each allele-specific primer. After multiple extensions of the fluor-labelled primer occur, the end product is read on a FRET-capable reader.

Each reaction volume consisted of 10 ng of DNA, 0.163 μM of each allele specific SNP primer, 0.407 μM of the common reverse SNP primer and 1X KASPar Reaction Mix

(Kbioscience Ltd.) for a final 4.055 µL reaction volume. Conditions for the PCR amplification were as follows: 94°C (15 min), 10 cycles of 94°C (20 sec)/ 65°C-57°C (60 sec) dropping 0.8 ^{o}C per cycle, followed by 26 cycles of 94°C (20 sec)/ 57°C (60 sec). The end point analyses were carried out on a CFX96 Touch^{TM} Real-Time PCR Detection using the Bio-Rad CFX Manager 2.1 software System (Bio-Rad, Hercules, CA., USA). HEX dye was represented as VIC during the SNP analyses since the excitation (nm) and emission (nm) values of HEX and VIC are very similar.

4.3.5. Linkage Map Construction

Once the allele specific sequence had been amplified and read on a FRET-capable reader, the results were converted and scored in a plus/minus (+/-) matrix, as with the SSR data. The SSR and SNP marker data was examined using the computer program Mapmaker 3.0 (Lander et al. 1987; Lincoln et al. 1992). This program carries out multipoint linkage analysis to assign markers to linkage groups and define preliminary orders. An initial LOD score of 5.0 was used to associate loci into linkage groups. Once generated, a considerable amount of manual checking of data is required to confirm group assignments and the final marker positions. The resulting order of linkage groups was analyzed through the Kosambi mapping function in Mapmaker 3.0 to determine the recombination distance in centiMorgan (cM). Deviations from the expected segregation ratios of the 179 x 345 DH population of 1:1 for each locus was calculated using χ^2 analysis.

SNPs were further assessed by comparing and aligning their flanking sequences with the B genome of *B. nigra* (DH No100) and a complete C genome sequence of *B. oleracea* (TO1000) using BLAT. BLAT compared the flanking sequences of the SNPs against each genome and

Bc-Contig 10148-p466

BLAT

B. nigra (B genome)

Alignment	Location	# of Matches	Length of SNP
1	scaff_14020_1	24	121
2	scaff_13934_1	121	121
3	scaff_13587_3	16	121
4	scaff_13629_2	16	121
5	scaff_13340_1	18	121

B. oleracea (C genome)

Alignment	Location	# of Matches	Length of SNP
1	C9	16	121
2	C8	17	121
3	C1	16	121
4	C1	16	121
5	C1	16	121
6	C5	17	121
7	C4	17	121
8	C4	16	121
9	C2	53	121

Best Match Scores

121/121 > 53/121

B genome

Figure 4.2. SNP marker comparison between the B and C genome of *B. carinata*. Alignments of SNP marker Bc-Contig 10148-p466 using BLAT to determine the greatest number of base matches (top hit) within each genome. The top hits are compared between genomes to determine the SNP sequence origin. In this instance, the top hit of the B genome is greater than the top hit of the C genome. Therefore the SNP Bc-Contig-10148-p466 and its flanking sequences is inferred to be from the B genome. The B genome alignment was determined using scaffolds from *B. nigra*. SNP flanking sequences were aligned to pseudomolecules corresponding to the C genome linkage groups of *B. oleracea.*

returned a series of alignments. For each sequence the top hit for each genome was compared and the genome with the higher resulting number of bp matching was inferred to be the genome of origin (Figure 4.2). These data were used to assist in anchoring the markers to a specific genome and linkage group. Map illustrations for each linkage group was drawn using MapChart v. 2.2 (Voorrips 2002).

4.4. Results

4.4.1. Generation of a Scoring Matrix using SSR Markers

The availability of the DH segregating population derived from the lines 179 and 345 allowed for the rapid development of a low density genetic linkage map for *B. carinata*. Plant tissue was collected from 76 lines of the 82 lines from the DH population, (Line 1-75, 77). Seeds from six lines did not germinate due to poor quality. The parental lines had previously been determined to be phenotypically diverse based on morphological traits and the initial SSR screening indicated there was sufficient genetic variation. However, only 33 polymorphic markers were identified between the parental lines from the original 165 tested. An additional 251 SSR markers were tested, which yielded in total 78 (27%) polymorphic markers, 206 monomorphic markers and 132 failed reactions. The markers, known to be polymorphic, were mapped in the segregating population. Since the population is DH, the expected SSR pattern was the presence of a single parental allele at each locus in each line. A scoring matrix was generated by assigning a score of "+" for the 179 parental allele and a "-" for the 345 parental allele (Figure 4.3). Confirmation of a truly doubled haploid population was provided by the absence of heterozygous genotypes.
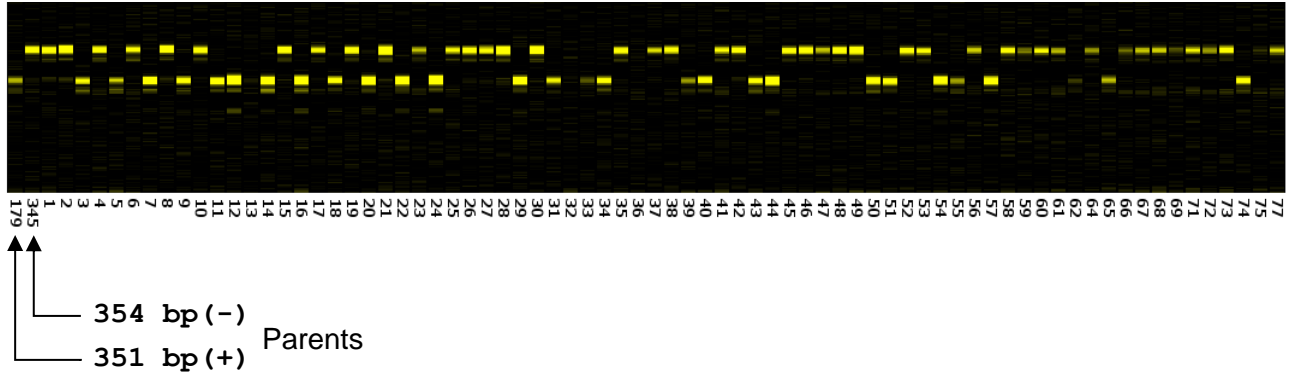
Figure 4.3. SSR marker sJ7104 on *B. carinata* 179 x 345 DH population. Alleles are assigned from each parent which then can be scored into a plus/minus matrix. The top band is scored as a plus while the bottom band will be scored with a minus. No heterozygous genotypes were found confirming this is a doubled haploid population.

By comparing the segregation patterns of the polymorphic markers, genetic linkage can be identified. Similar segregation patterns would indicate two loci map to the same chromosome. Minor differences indicate recombination events between markers. If there are significant differences the polymorphic markers are not linked. The 78 SSR markers polymorphic between 179 and 345 identified 86 loci, since some markers amplified multiple loci in the polyploid genome of *B. carinata*. SSRs were assigned to specific linkage groups by studying the marker positions of the mapped SSRs in previously generated genetic maps of the B and C genome of other *Brassica* species (Dr. Andrew Sharpe, unpublished data). Based on previous mapping in other *Brassica* species, 45 of these SSR markers would be presumed to amplify loci on the B genome and 33 on the C genome.

4.4.2. 3'-cDNA transcript profiling and SNP discovery

In order to create a dense genetic linkage map for *B. carinata*, parental lines 179 and 345 were selected for targeted 3'cDNA transcript profiling using 454 pyrosequencing technology. Isolated mRNA from seedling tissue of lines 179 and 345 was used to synthesise cDNA template libraries. The transcripts were targeted to the 3' UTR to generate specific sequence reads that allow for more efficient SNP discovery. The 454 pyrosequencing of cDNA libraries yielded 465,249 and 477,547 reads, with an average read length of 280 and 257 bp for lines 179 and 345, respectively. Since there is currently no available genome sequence for *B. carinata*, a reference *de novo* assembly from the 179 reads was produced using DNAStar's SeqMan NGen 2.1.0. software (DNAStar) at a stringency rate of 95% sequence identity. Homologues falling below the 95% threshold were separated and placed as independent contigs. The 454 specific key sequence and the low quality ends of the reads were trimmed to generate high quality reads. The

key sequence, found at the start of the read, was trimmed to eliminate artificial alignment at the beginning of the reads. The poor quality data found at the ends of the sequence reads were also clipped to enhance the overall quality of the reads. Of the 464,851 high quality reads from the raw data, 361,019 reads could be assembled using NGen to produce a reference assembly of 34,117 contigs (Table 4.1). The largest contig was 2,918 bp long while the highest number of contigs was in the range of 300-399 bp and the average length for the contigs was 387 bp (Figure 4.4). The 179 contigs were aligned and compared using BLASTN v. 2.2.16 against available genome assemblies for *B. nigra* (DH No100) and *B. rapa* cv. Chiifu to attempt to identify the genome of origin for each contig (Table 4.2). Since there was no *B. oleracea* or C genome reference available at the time, the A genome of *B. rapa* was used due to its close similarity to the C genome (Howell et al. 2008). The BLASTN alignment identified 18,167 contigs belonging to *B. nigra* (DH No100) and 15,059 contigs to *B. rapa* cv. Chiifu. These results would suggest similar transcription rates of both the B and C genome of *B. carinata*. The BLASTN alignment also identified 346 contigs with the same score between the two genomes. The results also produced 545 contigs that did not match to either genome, which would suggest that they may be specific to either the C genome or to *B. carinata*.

Once the contigs were identified, reads from 345 were reference mapped onto the 179 contigs using CLC Bio Genomics Workbench (CLC Bio) resulting in 329,094 (80%) reads aligning to 22,699 contigs in the reference assembly (Table 4.3) and a total of 8,104 SNPs (Table 4.4) being called in 3,305 contigs (Table 4.5). BLASTN alignment identified 1,961 contigs containing SNPs preferentially matching to *B. nigra* and 1,320 contigs containing SNPs aligning to *B. rapa* while 16 contigs contained SNPs that were indistinguishable between the two species and eight that were not able to be identified.
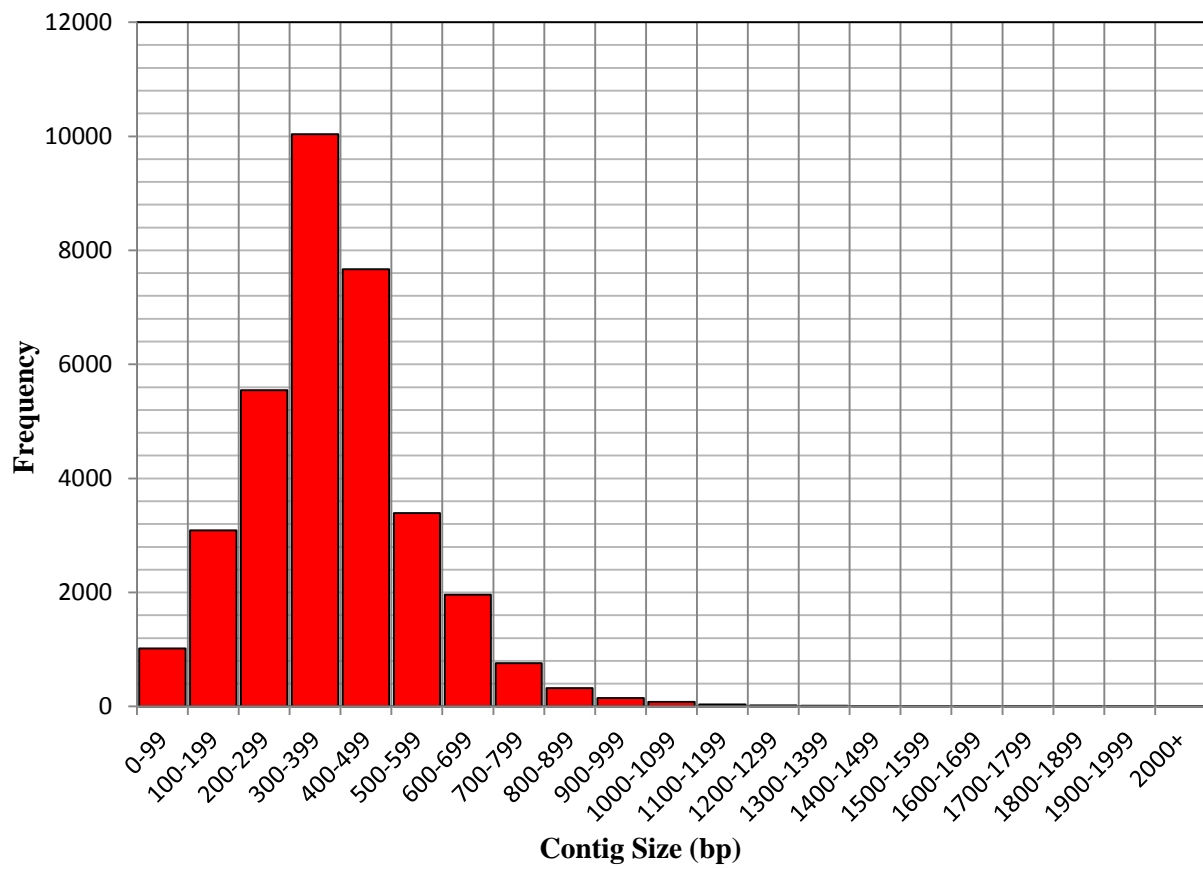
Figure 4.4. Frequency distribution of contig sizes generated from *B. carinata* line 179 using NGen.

Table 4.1. Assembly of the raw reads of *B. carinata* line 179 for contig production using NGen.

	# of Reads
Raw Reads (from 454 Genome Sequencer)	465,249
Quality Reads (NGen)	464,851
Assembled Reads	361,019
Unassembled Reads	103,832
Contigs	34,117

Table 4.2. BLASTN results of contigs generated from *B. carinata* line 179.

	# of Contigs
Contigs generated from NGen	34,117
Matched to the B genome	18,167
Matched to the C genome (Using the A genome as a surrogate)	15,059
Matched to the B/C genomes (same score in both genomes)	346
Did not match to either genome	545

Table 4.3. Assembly of raw reads of *B. carinata* line 345 reference mapped to 179.

	# of Reads
Raw Reads (from 454 Genome Sequencer)	477,547
Quality Reads (CLC Bio)	411,158
Assembled to 179 Reference (Mapped)	329,094
Non-matched Reads	82,064
Reference Assemblies Matched	22,699

Table 4.4. SNP remaining after filtering of 345 reads reference mapped to 179 contigs.

	SNPs
SNPs detected by CLC Bio	8,104
100% SNP calls in 179 (all bases at variant position are the same)	4,544
Corresponding position in reference sequence was a SNP (all bases at variant position are the same)	1,088
After filtering for full flanking sequences*	469
After filtering for unique flanking sequences (Duplicate removal)	442

* Flanking sequence of 120 bp was compared to *B. rapa* and *B. nigra* using BLAT to determine flanking sequences did not span exon/intron splice site.

Table 4.5. Contigs containing SNPs identified by genome based on inference using BLASTN.

	# of Contigs
Matched to *B. nigra*	1,961
Matched to *B. rapa*	1,320
Matched to both *B. nigra* and *B. rapa*	16
Unidentified	8
Total	3,305

Table 4.6. Genome inferred by BLAT using *B. nigra* and *B. oleracea* assemblies.

	SNPs
Mapped SNPs	341
Matched to B genome	230
Matched to C genome	108
Matched to B/C genomes	3

These SNPs were filtered to increase the likelihood for the calls to be true SNPs as well as to allow for SNP assays to be generated using the KASpar system (Kbioscience Ltd.) (Table 4.4). SNPs were filtered by eliminating bases that did not show complete agreement in reads aligned at the SNP position. The reference sequences displaying ambiguous bases at the SNP position were also removed as potential SNP candidates. SNPs found with a read depth of at least four and 120 bp of flanking DNA (60 up/down stream of SNP) were further compared to the *Brassica* genomes using BLAT to eliminate sequences crossing exon/intron splice sites. Of the remaining SNPs, duplicates were eliminated resulting in 442 unique SNPs to be tested on the 179 x 345 DH population.

4.4.3. SNP Detection and Validation

Genetic markers were designed with a high-level of stringency through extensive filtering of SNPs after alignment of the 345 reads, by the removal of ambiguous reference bases and the retention of SNPs containing bases that were uniform within each line. After removing the undesired SNP calls, a total of 442 potential SNP alleles remained. Alignment of the reference contigs through BLASTN to the *Brassica* genome sequences suggested that 59% of the SNPs were from the B genome, 28% from the C genome and the remaining 13% of SNP sequences were ambiguous in origin with equal similarity to both genomes.

SNP assays were then created and tested on the 179 x 345 DH population. The segregation of parental alleles 179 and 345 at these loci were scored and subjected to genetic linkage analysis. The SNP assays were performed using the KASpar genotyping system (KBioscience Ltd). Differential fluor-labelled oligos which associated specifically with the target SNP allowed the parental origin of the allele in each line to be assigned. To generate a scoring

matrix the SNP genotypes were assigned a "+" after inheriting the allele from line 179 and a "-" when from 345. As with the SSRs, the collective array of the alleles from the 76 lines of the 179 x 345 population generated a +/- matrix. Of the potential 442 SNP alleles, 341 markers could be scored reliably to generate a segregation pattern, while 101 assays were either monomorphic, failed to amplify, or gave an ambiguous clustering result (Figure 4.5).

4.4.4. Genetic Linkage Map

The genotypic data from the previously generated SSR markers and the newly generated SNP markers were assembled into a linkage map using Mapmaker 3.0 (Lander et al. 1987; Lincoln et al. 1992) with a LOD score of 5 and the Kosambi mapping function. The scoring matrixes from the resulting linkage groups were further assessed manually to reduce the presence of less likely double crossover events.

The SSRs which corresponded to previously mapped defined linkage groups in the B and C genome from other *Brassica* species were used to assign the names of the identified linkage groups. With fewer loci detected from the SSR markers of the C genome, it was considerably more difficult to distinguish these linkage groups from the genetic framework alone. After assigning SNP markers to the linkage groups, SNPs were further assessed by once more comparing and aligning their flanking sequences with the B genome of *B. nigra* (DH No100) and a complete C genome sequence of *B. oleracea* (TO1000), but this time using BLAT (Figure 4.2). BLAT returns a series of alignments for each query (or sequence containing the flanking sequences of SNP) of which the top hit is determined by the number of bases over the total length in the query. The top hit within each genome was then compared between the genomes.

A.
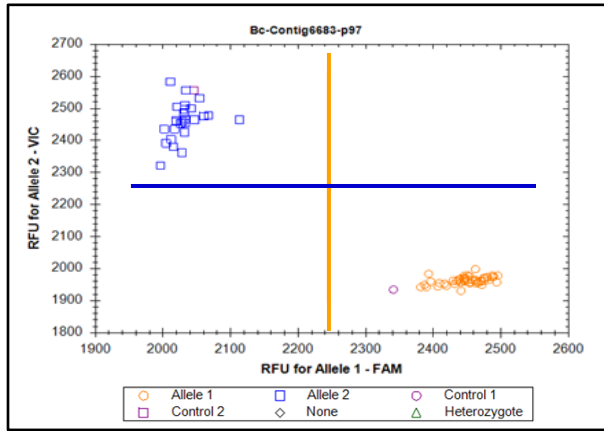
B.
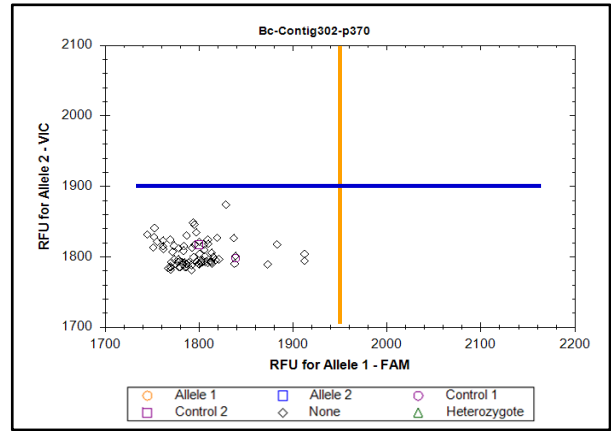
C.
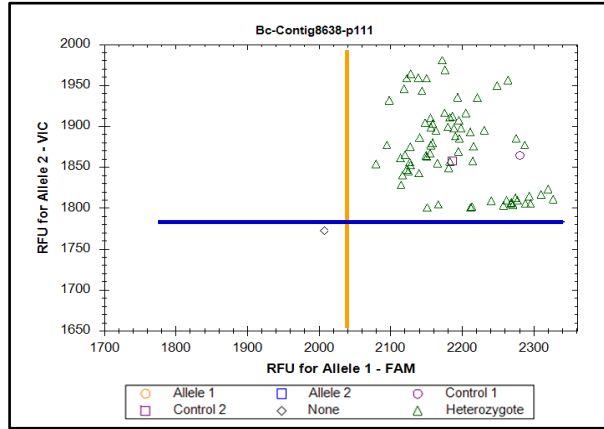
D.
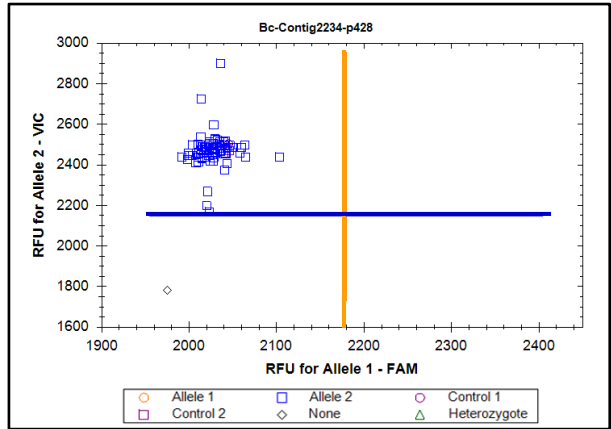
E.
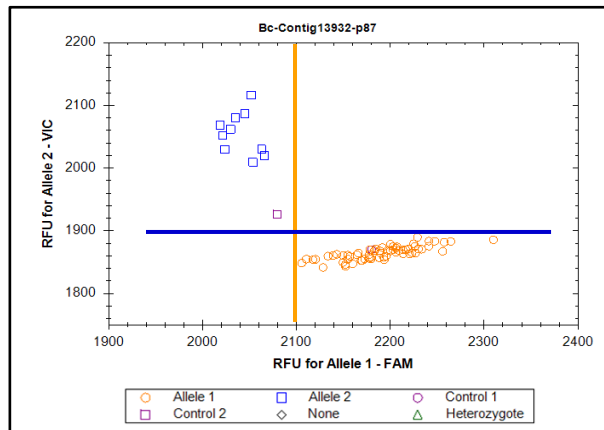
Figure 4.5. Various SNP Assays amplified by KASPar genotyping system on the 179 x 345 DH population using Bio-Rad CFX Manager 2.1 software. Differential fluor-oligos, targeting specific SNP alleles are measured in Relative Fluorescence Units (RFU). Each symbol (Allele 1 = ○, Allele 2 = □, Heterozygous = △) represents an individual line and the SNP allele it contains inherited from either 179 and 345 (○ = Control 1, □ = Control 2) parental lines. HEX is represented as VIC due to the excitation and emission values being very similar. SNP analysis shows A. amplification of 2 alleles, B. no amplification, C. heterozyogous, D. monomorphic alleles, E. ambiguous clustering.

The genome with the greater length of match was inferred to be the genome of origin. The SNP marker with the greatest number of identical bp was assigned to the corresponding genome to assist in anchoring the markers to a specific linkage group. From the 341 SNPs, 230 were best matched to the B genome and 108 were best matched to the C genome while three had an equal top hit score between the B and C genome (Table 4.6). With a completely anchored *B. oleracea* genome sequence (Parkin, unpublished data), it was possible to identify the most similar linkage group in *B. oleracea* and thus discover the corresponding *B. carinata* linkage group. The flanking sequences of the SNPs from the B genome could not be matched to the individual linkage groups of *B. nigra* since the sequencing of the B genome has not been completed. However, there were enough mapped SSR markers to identify the eight linkage groups of the B genome in *B. carinata*.

The segregation of parental alleles previously generated 86 loci from the microsatellite analysis which was combined with 341 SNP loci to generate 427 segregation patterns (Table 4.7). The resultant *B. carinata* genetic map included 368 polymorphic loci, 75 SSRs and 293 SNPs, with 17 substantial linkage groups (Bc1-Bc17) and 2 smaller groups of linked loci (Group 1 & 2) (Figure 4.6). The 17 linkage groups covered a total span of 2024 cM with an average of 5.67 cM between pair of markers. There were 59 unlinked markers, 11 SSRs and 48 SNPs.

The SSRs had previously been positioned on the B genome of *B. juncea* (Sharpe, unpublished data) and the C genome of *B. napus*, in addition the genome of *B. oleracea* was used to assist in annotating the 17 substantial linkage groups of *B. carinata*. The linkage groups equivalent to the B genome (B1-8) in *B. juncea*, J11-J18, were assigned Bc1-Bc-8 and the linkage groups of the C genome C1-9/N11-19, were assigned Bc9-Bc17.

Table 4.7. Analysis of linkage group composition derived from polymorphic SSR and SNP markers tested on the 179 x 345 DH population.

Linkage Groups (179 x 345)	Corresponding Linkage Group (*Brassica*)	Mapping Distance (cM)	# of SSRs	# of SNPs	Total # of Polymorphic Markers
Bc1	B1	188.3	9	18	27
Bc2	B2	301.7	10	38	48
Bc3	B3	130.3	6	26	32
Bc4	B4	135.0	7	9	16
Bc5	B5	103.9	1	20	21
Bc6	B6	131.5	6	27	33
Bc7	B7	102.3	3	17	20
Bc8	B8	121.0	6	39	45
Bc9	C1	43.6	1	10	11
Bc10	C2	115.9	1	8	9
Bc11	C3	70.7	1	10	11
Bc12	C4	137.4	5	11	16
Bc13	C5	147.9	6	11	17
Bc14	C6	90.7	2	16	18
Bc15	C7	63.1	2	9	11
Bc16	C8	97.3	2	12	14
Bc17	C9	43.2	4	4	8
Total		2023.8	72	285	357
Group 1		5.3	1	6	7
Group 2		11.6	2	2	4
Total		2040.7	75	293	368
Unlinked			11	48	59
Total			86	341	427

Bc1

cM	Marker
0	Bc-Contig5541-p107
20	Bc-Contig372-p173
	Bc-Contig11921-p126
	Bc-Contig12546-p201
	sJ5245
40	Bc-Contig13830-p339
	Bc-Contig5801-p151
	Bc-Contig7459-p411
60	Bc-Contig22533-p96
	sJ84165 Bc-Contig1210-p276
	Bc-Contig11823-p428 Bc-Contig15818-p159
	Bc-Contig15641-p71
80	sJ6734I
100	sJ4933** sB3550I**
120	sN9875
	Bc-Contig7045-p227
	Bc-Contig16684-p370 Bc-Contig29833-p247
	Bc-Contig18918-p103
140	Bc-Contig26276-p91
	Bc-Contig22032-p258
	sR9555_Nb
160	sB23190Ab
180	sJ7212

188.3

Figure 4.6. A genetic linkage map of *B. carinata* based on segregation in the 179 x 345 population of DH lines. Vertical lines represent 17 linkage groups (Bc1-Bc17) and 2 unidentified partial groups (Group 1-2) that have not been assigned. Numbers left of the horizontal lines describe the map distance in cM and to the right are the positions where the markers are located. SSR markers indicated in red and SNP markers in blue. The total map length of each linkage group is indicated. Unlinked markers are listed at the end of the figure. The degree of segregation distortion (** $P < 0.01$ and *** $P < 0.001$) is represented by an *asterisk* next to marker locus name.

Bc2

cM	Marker
0	Bc-Contig7926-p184
	sJ3302RI
20	Bc-Contig20174-p195
	Bc-Contig13485-p107
	Bc-Contig2095-p533 Bc-Contig9639-p150
40	sJ6846
	Bc-Contig11252-p99
	sN11670
	Bc-Contig17841-p375
60	sJ03104***
	Bc-Contig19667-p444 Bc-Contig32658-p68
	Bc-Contig7522-p327
80	sB0540R
100	Bc-Contig5006-p213**
	sJ3481
	Bc-Contig1779-p129 Bc-Contig3488-p382**
120	Bc-Contig6683-p97** Bc-Contig8208-p278**
	Bc-Contig4937-p89
	Bc-Contig26437-p142 Bc-Contig14212-p277
	Bc-Contig17577-p284
140	Bc-Contig3227-p418
	Bc-Contig3227-p216
	Bc-Contig23218-p175
	Bc-Contig2304-p64
	Bc-Contig11191-p72
160	sB0141 Bc-Contig25241-p197
	sJ4458I Bc-Contig2332-p435
180	
	Bc-Contig9252-p490
200	
	sJ7977b*** Bc-Contig7969-p363***
	Bc-Contig8792-p321*** Bc-Contig13606-p394***
220	sJ6730I***
	Bc-Contig3559-p247**
240	Bc-Contig9996-p70
260	Bc-Contig12889-p155***
280	Bc-Contig2191-p221***
	Bc-Contig9993-p297***
	Bc-Contig30620-p348***
300	Bc-Contig15512-p400*** Bc-Contig2643-p263***

301.7

Bc3

- 0 — Bc-Contig2425-p356**
- Bc-Contig2572-p101 sJ8232**
- Bc-Contig12726-p192 sB23190Aa
- Bc-Contig17274-p255
- Bc-Contig6954-p413
- sJ3627R**
- 20 — Bc-Contig6893-p373
- Bc-Contig5405-p104
- Bc-Contig13940-p369***
- Bc-Contig5248-p220***
- Bc-Contig14917-p338**
- 40 — Bc-Contig20674-p97
- Bc-Contig5143-p245**
- Bc-Contig1567-p118** Bc-Contig15346-p190***
- Bc-Contig9192-p283***
- 60 — Bc-Contig2324-p137**
- Bc-Contig23825-p72***
- Bc-Contig6788-p184
- Bc-Contig14372-p595**
- 80 — Bc-Contig21500-p155
- Bc-Contig6428-p130
- sJ8469
- 100 — Bc-Contig10016-p266
- Bc-Contig7846-p272
- Bc-Contig12761-p102
- 120 — Bc-Contig26953-p125
- sJ7588a
- sJ7588b
- Bc-Contig3396-p161

130.3 cM

Bc4

- 0 — sB2177R
- sN1939R
- Bc-Contig612-p539 Bc-Contig612-p837
- 20 — sJ5250
- Bc-Contig16116-p251
- 40 — Bc-Contig8018-p302
- Bc-Contig3037-p110
- 60 —
- sJ6111a**
- 80 — Bc-Contig8692-p150
- Bc-Contig31149-p304
- 100 — sB1935A
- sJ3381
- Bc-Contig18259-p377
- 120 — sJ7648
- Bc-Contig4358-p517

135.0 cM

Bc5

- 0 — Bc-Contig10825-p191
- Bc-Contig8035-p289 Bc-Contig22376-p76
- Bc-Contig33087-p85 Bc-Contig8035-p193
- Bc-Contig6164-p76
- 20 —
- Bc-Contig7992-p180***
- Bc-Contig10148-p466*** Bc-Contig26542-p153***
- 40 — Bc-Contig11067-p189***
- Bc-Contig7504-p165**
- Bc-Contig2442-p400***
- Bc-Contig24171-p140** Bc-Contig3236-p580**
- Bc-Contig25755-p193
- 60 — Bc-Contig17462-p149
- Bc-Contig5089-p248 Bc-Contig5089-p163
- Bc-Contig26536-p190
- 80 — sB51173
- 100 — Bc-Contig12803-p134

103.9 cM

Bc6

- Bc-Contig10812-p228***
- Bc-Contig13185-p194***
- Bc-Contig2551-p448
- Bc-Contig2551-p376
- Bc-Contig13165-p86
- Bc-Contig4722-p393
- Bc-Contig9934-p322 Bc-Contig22494-p80
- sJ7104 sB2519
- Bc-Contig33487-p110
- Bc-Contig11513-p103*** Bc-Contig22354-p94
- Bc-Contig8511-p63
- sJ0164b
- Bc-Contig19809-p122 sJ0164a
- sN1988_Nb
- Bc-Contig10957-p137***
- Bc-Contig23168-p214
- Bc-Contig4652-p289
- sJ3588A
- Bc-Contig24666-p258 Bc-Contig7199-p282
- Bc-Contig17492-p101
- Bc-Contig896-p453 Bc-Contig2571-p296
- Bc-Contig896-p159 Bc-Contig1172-p345
- Bc-Contig12167-p375
- Bc-Contig7171-p337
- Bc-Contig20432-p242
- Bc-Contig1339-p153

131.5 cM

Bc7

- Bc-Contig14285-p340
- Bc-Contig4565-p273 Bc-Contig10248-p193
- sJ6733
- Bc-Contig31763-p139
- Bc-Contig12879-p278 Bc-Contig1468-p520**
- Bc-Contig21191-p150
- sJ2971
- Bc-Contig29692-p288 Bc-Contig13076-p215
- Bc-Contig7241-p414
- sJ7977a
- Bc-Contig11861-p502 Bc-Contig11861-p412
- Bc-Contig30790-p141
- Bc-Contig25035-p284 Bc-Contig30805-p113
- Bc-Contig4175-p324
- Bc-Contig10944-p118

102.3 cM

Bc8

- Bc-Contig7610-p352 Bc-Contig23654-p191
- Bc-Contig7610-p457 Bc-Contig10965-p127
- Bc-Contig10040-p90
- Bc-Contig1679-p272 Bc-Contig14440-p205
- Bc-Contig1679-p389 Bc-Contig1679-p191
- Bc-Contig17655-p142
- Bc-Contig5835-p502
- Bc-Contig2195-p366 Bc-Contig23189-p155
- Bc-Contig2195-p222
- Bc-Contig20392-p129 Bc-Contig17773-p198
- Bc-Contig31104-p198 Bc-Contig6910-p288
- sN12574
- Bc-Contig9219-p202 Bc-Contig18265-p65
- Bc-Contig11177-p241
- Bc-Contig8107-p317
- sA1090
- Bc-Contig2769-p179
- Bc-Contig8164-p361
- Bc-Contig5406-p210
- Bc-Contig4320-p645 Bc-Contig4320-p403
- Bc-Contig5315-p184
- sJ3790
- sJ1073
- Bc-Contig6478-p454
- Bc-Contig15191-p489 Bc-Contig6579-p677
- Bc-Contig11646-p132 Bc-Contig17660-p196
- Bc-Contig21286-p165
- Bc-Contig20545-p75
- Bc-Contig3981-p431
- sJ8269 Bc-Contig14116-p328
- sB5162a Bc-Contig7562-p325
- Bc-Contig16283-p267

121.0 cM

Bc9

- Bc-Contig12725-p384
- Bc-Contig23017-p140
- Bc-Contig8433-p188 Bc-Contig22088-p257
- Bc-Contig29468-p387
- Bc-Contig6531-p344
- Bc-Contig24280-p159
- Bc-Contig7377-p300
- Bc-Contig2106-p206 Bc-Contig2106-p305
- sN3523R

43.6 cM

Bc10

Position	Marker
0	Bc-Contig3704-p171
	Bc-Contig2619-p330
	Bc-Contig5599-p308***
	sN3761***
	Bc-Contig12665-p330**
	Bc-Contig2574-p209***
	Bc-Contig20435-p223 Bc-Contig14724-p344**
	Bc-Contig14724-p76**

115.9 cM

Bc11

Position	Marker
0	Bc-Contig24936-p149
	Bc-Contig12916-p337***
	Bc-Contig12791-p112***
	Bc-Contig12791-p178***
	Bc-Contig12974-p177***
	Bc-Contig14101-p420***
	Bc-Contig7484-p339***
	sR6803***
	Bc-Contig9105-p344***
	Bc-Contig10256-p64***
	Bc-Contig23291-p171**

70.7 cM

Bc13

Position	Marker
0	Bc-Contig20063-p303
	Bc-Contig2383-p451
	Bc-Contig743-p700
	Bc-Contig5787-p296 Bc-Contig5787-p200
	Bc-Contig7717-p269
	sB2363
	Bc-Contig2421-p95
	sB4711
	Bc-Contig8527-p96
	Bc-Contig4233-p345
	sS2129
	Bc-Contig10495-p512
	sS2131
	sS1732
	Bc-Contig24281-p331 sR9477

137.4 cM

Bc12

Position	Marker
0	sS2277**
	Bc-Contig10883-p234***
	Bc-Contig10916-p269
	sORG31
	Bc-Contig18903-p138
	sN0464
	Bc-Contig13320-p164
	sORG90 Bc-Contig20849-p67
	Bc-Contig4370-p231
	Bc-Contig8914-p78
	Bc-Contig15341-p86
	Bc-Contig2507-p271
	Bc-Contig1041-p160**
	Bc-Contig1041-p784***
	sORE95

147.9 cM

Bc14

- Bc-Contig2835-p167
- Bc-Contig29305-p149
- Bc-Contig8340-p76
- Bc-Contig12580-p161
- Bc-Contig19077-p260
- Bc-Contig7563-p194
- Bc-Contig14802-p357
- Bc-Contig11003-p150
- Bc-Contig6722-p87
- Bc-Contig22519-p81 Bc-Contig1896-p506
- Bc-Contig18314-p362
- Bc-Contig5299-p220** Bc-Contig22345-p95**
- Bc-Contig8321-p63** Bc-Contig30200-p68**
- sN2195
- sS2486

90.7 cM

Bc15

- Bc-Contig24056-p199 Bc-Contig24056-p116
- Bc-Contig31527-p191 Bc-Contig21904-p144
- sN3825J
- Bc-Contig19209-p147
- Bc-Contig6162-p418
- Bc-Contig17206-p401 Bc-Contig5844-p320
- Bc-Contig13373-p218 sN1919

63.1 cM

Bc17

- sA1315
- Bc-Contig19725-p250
- Bc-Contig6271-p200
- sNRG42 sJ6111b
- sORB37
- Bc-Contig17061-p121 Bc-Contig27718-p142

43.2 cM

Bc16

- Bc-Contig15135-p321***
- Bc-Contig6507-p260***
- Bc-Contig2513-p139
- Bc-Contig9449-p75
- Bc-Contig1059-p548 Bc-Contig1373-p578
- Bc-Contig6268-p426
- sJ6117
- Bc-Contig13394-p148**
- Bc-Contig5577-p239
- sB4749I**
- Bc-Contig25835-p131***
- Bc-Contig2522-p314
- Bc-Contig10337-p243

97.3 cM

Group 2

- Bc-Contig5505-p294***
- sB2141AI***
- Bc-Contig29184-p319***
- sB0372***

11.6 cM

Group 1

- Bc-Contig6788-p394 Bc-Contig7876-p154
- Bc-Contig7876-p86
- Bc-Contig11574-p265 sJ5379
- Bc-Contig21207-p92 Bc-Contig11574-p467

5.3 cM

Unlinked

Bc-Contig109-p190***	Bc-Contig14838-p127	Bc-Contig2117-p486**	Bc-Contig5865-p321***	sB5162b
Bc-Contig11804-p71***	Bc-Contig15023-p102***	Bc-Contig21588-p368***	Bc-Contig6189-p258***	sJ0358***
Bc-Contig12103-p103***	Bc-Contig15023-p306***	Bc-Contig23281-p219***	Bc-Contig6365-p473***	sJ0644**
Bc-Contig12454-p227***	Bc-Contig15611-p101***	Bc-Contig26764-p62***	Bc-Contig6438-p659***	sJ1325
Bc-Contig13334-p303***	Bc-Contig15673-p281***	Bc-Contig27911-p91***	Bc-Contig6705-p92	sJ3396R
Bc-Contig13392-p128***	Bc-Contig15971-p365***	Bc-Contig285-p156***	Bc-Contig7198-p375	sJ4042**
Bc-Contig13582-p102***	Bc-Contig15971-p437***	Bc-Contig28948-p431***	Bc-Contig7881-p463***	sJ6272**
Bc-Contig13628-p109***	Bc-Contig1738-p474***	Bc-Contig3014-p277***	Bc-Contig8061-p297***	sJ7128***
Bc-Contig13628-p185***	Bc-Contig18060-p363***	Bc-Contig3460-p217***	Bc-Contig8640-p118***	sN1988_Na***
Bc-Contig13725-p123***	Bc-Contig18927-p176***	Bc-Contig3460-p351***	Bc-Contig8656-p292	sR12095**
Bc-Contig13932-p87***	Bc-Contig1991-p135***	Bc-Contig3520-p114***	Bc-Contig8935-p464***	sR9555_Na***
Bc-Contig14545-p207***	Bc-Contig19993-p230***	Bc-Contig5185-p380**	Bc-Contig9639-p299***	

The two smaller groups, Group 1 and Group 2 could not be definitively identified through computational analysis and manual checking. Group 1 SSR marker sJ5379 has been identified in the linkage group B3 of the B genome or J13 in *B. juncea* and Group 2 SSR markers sB2141AI and sB0372 in linkage group B4 or J14. However, even after reducing the mapping stringency it was not possible to conclude these smaller groups were linked to Bc3 and Bc4 respectively. The addition of further markers or the availability of a fully anchored B genome sequence would shed light on the likelihood these smaller groups belong to a certain linkage group.

Segregation distortion for each marker locus tested using the χ^2 Goodness of fit revealed 31% (131/427) loci with significant levels ($P < 0.01$) of allelic imbalance outside the 1:1 Mendelian ratio. Of these marker loci, 21% (79/368) were found on the linkage groups (Figure 4.6). The loci with segregation distortion were not randomly distributed throughout the genome but mostly concentrated on regions of 12 of the 17 mapped linkage groups and one of the two unidentified groups. The marker loci showing deviations from the Mendelian segregation ratio showed an even bias towards either parental line (179 (35/79) and 345 (44/79)).

A total of 59 polymorphic markers on the 179 x 345 DH population were unlinked. Of these unlinked markers, 88% (52/59) had loci with significant levels ($P < 0.01$) of allelic imbalance outside the 1:1 Mendelian ratio, while 78% (46/59) deviate towards the parental line 345. Though, this may not rule out the possibility that these markers belong to a particular linkage group, they may be associated with segregation distortion through gametic selection and chromosomal rearrangements. Segregation ratios may be skewed during the generation of DH lines through selection pressure during in vitro androgenesis or unintentional selection of seeds or flowers which exhibit better fertility (Foisset and Delourme 1996; Iniguez-Luy et al. 2009). It is also possible that some SNP assays may amplify loci in both genomes which could create erroneous segregation patterns and explain the unlinked loci.

SSR and SNP markers were compared to the B genome and C genome of the *Brassica* genus to identify linkage groups. The loci determined from the common set of SSR markers, derived from the *Brassica* genome, were compared to the previously mapped B genome of *B. juncea*, and the C genome of *B. napus*. This comparison was used as a genetic template for the naming of the genetic linkage groups in *B. carinata*. However, since the number of SSRs for the C genome were marginal in determining the linkage groups, the flanking sequence of each SNP was used to determine how similar they were to the sequence of either the B or C genome. The comparative differences in sequence between the B and C genome would be a result of evolutionary sequence divergence (Table 4.8). The comparison of the average percent sequence identity between the calls identified as originating from the B or C genome (99%) and their opposing non-calls (85-86%) from the flanking sequence of the SNP markers suggests how significantly the B genome and C genome have changed since their divergence from a common *Brassica* ancestor. By aligning the flanking sequences of the C genome *B. carinata* SNP

Table 4.8. Average Percent Sequence Identity of SNP flanking sequences for genome calls using BLAT.

Calls	B genome % match	C genome % match
B. nigra (B genome)	98.70	85.25
B. oleracea (C genome)	86.03	99.10

markers to the genome sequence of *B. oleracea*, alleles belonging to loci on a specific linkage group could be identified. A cluster of markers with similar segregation patterns aligned to a specific linkage group in *B. oleracea*, were identified as the corresponding linkage group in *B. carinata*.

4.5. Discussion

A *B. carinata* genetic linkage map, consisting of 368 loci from SSR and SNP alleles, positioned on 17 substantial linkage groups and 2 smaller groups, has been generated. These polymorphic SNP loci were recognized between lines 179 and 345 and mapped using a DH population generated from the same parental lines. The 17 substantial linkage groups of *B. carinata* ascertained from this study agree with the evolutionary theory set forth by U (U 1935), with eight linkage groups belonging to the B genome and nine to the C genome, confirming *B. carinata* is an amphidiploid/allotetraploid.

The use of next generation sequencing technologies has provided a platform for generating SNP markers allowing the creation of dense genetic linkage maps for plants with un-sequenced genomes. Contigs generated from *B. carinata* line 179 were used as a reference to align line 345 reads and determine SNP positions in order to create SNP markers for the development of a genetic map. Although the available genome sequence data was used to determine the origin of each linkage group, a set of SNP markers was essentially developed with no prior sequence data for *B. carinata*. Next generation sequencing technologies would be able to generate SNP markers between any two *B. carinata* varieties with sufficient genetic diversity, as well as within other *Brassica* species encompassing evolutionary divergence.

The *B. carinata* linkage map identified 242 loci in the B genome spanning 1214.0 cM and 115 loci in the C genome covering 809.8 cM. With noticeably fewer loci in the C genome, the observed map length is smaller than the denser B genome. However, while the C genome of *B. carinata* falls in line with the recombination distance of most *Brassica* C genome species in previously generated linkage maps, (Babula et al. 2003; Bohuon et al. 1996; Camargo et al. 1997; Gao et al. 2007; Guo et al. 2012; Iniguez-Luy et al. 2009; Kianian and Quiros 1992; Landry et al. 1992; Ramsay et al. 1996; Sebastian et al. 2000; Slocum et al. 1990), the B genome is comparatively larger than previously developed linkage maps (Guo et al. 2012; Lagercrantz and Lydiate 1995; Truco and Quiros 1994) (Table 4.9). The increase in number of polymorphic markers could contribute to a denser, more detailed linkage map and therefore support the difference in recombination distances. However, the noticeable discrepancy in the number of polymorphic markers in the C genome compared to the B genome could suggest that the C genome is less polymorphic than the B genome in *B. carinata.* In *B. napus* a similar phenomenon has been observed for the *Brassica* A genome, with higher levels of genetic diversity found in the *Brassica* A genome than the C genome in a wide range of *B. napus* genotypes (Bus et al, 2011). It was speculated that the more prevalent use of *B. rapa*, rather than *B. oleracea*, in breeding strategies for *B. napus* had led to the observed differences. However, presumably the recent origins leading to a narrow genetic base of *B. carinata* coupled with limited domestication and cultivation pertaining to breeding strategies would suggest the fusion of the progenitor C genome in *Brassica* amphidiploids were genetically limited during formation. In the paper published by Guo et al (Guo et al. 2012), the *B. carinata* linkage map contained 214 loci consisting of mostly SSR markers (151/212) of which only 16 SSR markers were originally mapped from species containing a B genome (*B. nigra* = 13, *B. juncea* = 3).

84

Table 4.9. Genetic linkage studies on the B genome of *B. nigra* and C genome of *B. oleracea*.

| *Brassica* Species | Marker Type | # of Loci | Distance (cM) | | Author |
			B	C	
B. carinata	SSR, IBP*, AFLP, SRAP**	212	756	947	(Guo et al. 2012)
B. nigra	RFLP	288	855		(Lagercrantz and Lydiate 1995)
B. nigra	RFLP, RAPD	124	677		(Truco and Quiros 1994)
B. oleracea	RFLP	258		820	(Slocum et al. 1990)
B. oleracea	RFLP	201		1112	(Landry et al. 1992)
B. oleracea	RFLP, Isozyme	108		747	(Kianian and Quiros 1992)
B. oleracea	RFLP, RAPD, Izozyme	138		747	(Ramsay et al. 1996)
B. oleracea	RFLP	303		875	(Bohuon et al. 1996)
B. oleracea	RFLP, RAPD	159		921	(Camargo et al. 1997)
B. oleracea	RFLP, AFLP, SSR	547		893	(Sebastian et al. 2000)
B. oleracea	RFLP	212		1806	(Babula et al. 2003)
B. oleracea	SRAP**, SSR, BAC, Known *Brassica* genes	1,257		709	(Gao et al. 2007)
B. oleracea	RFLP, SSR, Phenotypic	280		891	(Iniguez-Luy et al. 2009)

*IBP – Intron-Based Polymorphism
**SRAP – Sequence Related Amplified Polymorphism

While the SSR markers were derived from the other five *Brassica* species in the Triangle of U, similarities between the A and C genome (Howell et al. 2008), could result in additional polymorphic repeats favouring identification of loci on the C genome. Although the bias in levels of polymorphism between genomes does not seem to be apparent, Guo notes a lower level of polymorphisms in *B. carinata* compared to other *Brassica* species in the analysis.

The same method to comparatively map the C genome of *B. carinata* was used on the B genome with the exception of matching the SNPs to the known chromosome locations in its *B. nigra* counterpart. The *B. nigra*, B genome has not been sequenced yet and therefore, there is insufficient data to match the sequences to specific linkage groups. However, using the SSRs as genetic backbone and comparing these markers to known locations in the B genome, as well as identifying SNPs more closely related to the B genome, the linkage groups of the B genome could be clearly defined. Based on the assembled map and comparison to previous genetic maps and the C genome sequence it can be inferred that both the B and C genome in *B. carinata* have remained essentially unaltered with no major rearrangements since the formation of the amphidiploid species. This result is consistent with studies in the amphidiploid species *B. napus* and *B. juncea* (Axelsson et al. 2000; Parkin et al. 1995).

4.6.   Conclusions and Future Research

The use of next generation sequencing has allowed for the production of SNPs without prior sequence knowledge. Coupled with previously generated SSR markers a genetic linkage map of *B. carinata* has been generated based on segregation in the 179 x 345 DH population. Analysis of common SSRs and SNP marker sequences resulted in the production of 17 substantial linkage groups, with eight aligning to the B genome and nine aligning to the C

genome of previous generated *Brassica* genetic linkage maps and genome sequences. In doing so, this genetic linkage map elucidates the evolution of the *Brassica* species as well as reiterates what has been known of the Triangle of U. Comparative mapping has suggested that the B and C genome remained essentially unchanged since the formation of *B. carinata*. This *B. carinata* genetic map will be able to assist in future research for specific genes and QTLs of agronomic interest and in-depth comparative map studies with *Arabidopsis thaliana* and other *Brassica* species.

The smaller groups in the *B. carinata* genetic linkage map could benefit from the identification of additional polymorphic markers to potentially bridge the gap with one of the substantial linkage groups. The addition of polymorphic markers and as well as using a larger segregation population would increase the density and resolution of the map.

Sequencing of the genome would not only confirm the current study, but allow for an extremely in-depth analysis of the *B. carinata* genome. Currently, with second generation sequencing technologies available there is the potential to sequence the genome in a relatively short amount of time, which would be ideal for studying the genetic architecture of *B. carinata* and other *Brassica* species.

This genetic map is the first SNP based map for *B. carinata* and the information attained will advance research in the *Brassica* species, as well as potentially allowing for a more commercially viable crop platform to be developed. The next phase in this study is to conduct a field assessment of phenotypic traits between lines 179 and 345, and within the associated DH population, which combined with the map would allow QTL analysis of genetic characteristics beneficial to the *B. carinata* crop.

5. Phenotypic Analysis of *Brassica carinata*

5.1.   Introduction

Variance in phenotypes can be attributed to genotypic variation, the environment and the interaction of both. Early Mendelian genetics have demonstrated that traits, controlled by genetic factors or genes, could be inherited from generation to generation. At the simplest level, phenotypic qualities such as seed colour, seed shape, and many others were categorized as qualitative traits. Further research into important traits revealed that many characteristics, such as plant height and disease resistance, are not always qualitative but are quantitative in nature and rely on multiple genes. Quantitative traits cannot be divided into distinct categories or groups, but the values are distributed over a continuous range throughout a population. The difficulty therein lies in determining the number, location and effect of the genes controlling a specific quantitative trait.

Gene and QTL mapping studies have previously been performed on many plant species including many of the *Brassicas* (Axelsson et al. 2001; Burns et al. 2003; Osborn et al. 1997; Ramchiary et al. 2007; Soengas et al. 2007). In contrast, a limited number of *B. carinata* studies have concentrated on observing and mapping loci controlling phenotypic traits (Guo et al. 2012). The DH population obtained from the JIC, which was used to generate the molecular map described in Chapter 4, showed considerable variation for a number of phenotypic traits. The molecular linkage map provides an opportunity to potentially locate genes and QTLs controlling these traits. Exploring natural phenotypic variation in conjunction with a high-density genetic linkage map allows for genes and QTLs controlling a phenotype to be identified at the genetic level which provides insights into adaptation and evolution, and can facilitate crop improvement.

QTL mapping projects have utilized a number of population designs, including recombinant inbred lines (RIL), backcross populations, intercross populations and DH populations, which involve crosses between two lines that differ in a specific trait(s) of interest. In comparison to the other population designs, lines in a DH population are able to reach homozygosity in only two generations. DH lines can be propagated without needing to consider further segregation in subsequent generations, allowing for more precise measurements of quantitative traits by using replicated trials.

Agronomic traits such as plant height (PH) and flowering time (FT) are important aspects in determining whether crops would be suitable for production in the Prairies. In *B. napus*, the discovery of genes and QTLs for PH could mitigate lodging issues (Foisset et al. 1995; Mei et al. 2009). The study of FT not only would improve our understanding of the plant life cycle in *Brassica* species, but also its evolution. It has been proposed that many of the FT genes in the *Brassica* genome result from replication of single ancestral genes potentially contributing to the evolution of the *Brassica* species (Axelsson et al. 2001). In addition, the seed content of *B. carinata* has the potential for many applications (as described in Section 2.2 Literature Review - Uses) so there is interest in understanding the underlying genetic architecture controlling these traits.

The study tested the hypothesis that QTL for PH and FT will be detected in the 179 x 345 DH population and that there are significant variations for quality traits in the 179 x 345 DH population.

5.2. Objectives

- A phenotypic assessment of the 179 x 345 DH population for qualitative and quantitative traits was carried out for the 2011 and 2012 field seasons

    - o Identify potential QTL associated with FT and PH using a DH population derived from the 179 x 345 cross.
    - o A preliminary analysis of seed quality traits such as oil, protein, fibre, unsaturated fatty acids, and glucosinolates will be evaluated for potential QTLs

5.3.   Materials and Methods

5.3.1.   Plant Material

Double haploid (DH) lines of *B. carinata* from a cross between 179 and 345 were grown in the greenhouse until maturity (as described in Section 3.3.1. Germplasm). The seeds were collected and stored until field conditions permitted seeding. Of the 76 lines generated from the 179 x 345 genetic map, 64 lines produced enough seed for the field trial in 2011 and 73 lines for 2012.

5.3.2.   Field Trial

A phenotypic assessment of the 179 x 345 DH population was initiated in the summers of 2011 and 2012 near Saskatoon, Saskatchewan. The mean temperature and total precipitation is presented in Table 5.1.

The field trials used approximately 100 seeds for each line, which consisted of the two parental accessions along with 64 DH lines in 2011 and 73 DH lines in 2012. The lines were grown in single rows in a randomized block design with 3 replicates. *Brassica juncea* cv. 'AC Vulcan' was used as a check and incorporated in the randomized complete block design. Seeding occurred in the third week of May, 2011 and observations were made and recorded throughout the growing season up until the third week of October, 2011 when all the lines were harvested. Seeding in 2012 occurred in the third week in May and observations were made throughout the growing season; however, the lines could not be harvested before the first snowfall and therefore no data was collected for seed content.

Table 5.1. Mean temperature and total precipitation during the growing season in Saskatoon, Saskatchewan, Canada in 2011 and 2012.

Month	2011		2012	
	Mean temp (^{o}C)	Total Precipitation (mm)	Mean temp (^{o}C)	Total Precipitation (mm)
May	11.1	17.5	10.1	108.0
June	15.7	94.4	15.8	121.1
July	18.6	68.6	19.7	80.9
August	17.3	16.5	17.3	48.5
September	14.8	6.0	13.0	0.8
October	6.5	35.6	1.7	8.8
Total		238.6		368.1

Source: http://www.climate.weatheroffice.gc.ca/climateData/canada_e.html

5.3.3. Phenotypic Measurement

Flowering time (FT) was measured from the sowing day to the day the first flower had opened on at least 50% of the *B. carinata* plants in the row. Plant height (PH) was measured from the soil level to the tip of the central axis of the *B. carinata* plant at harvest maturity.

Seed composition was measured using NIRS 6500 Feed and Forage Analyser (Foss, Denmark). Constituents included oil content, protein, fibre content (NDF, ADF ADL), IV, and total glucosinolates (GLU). Iodine value (IV) measures the reaction between iodine and double bonds in fatty acids (C=C) or the amount of unsaturation (Ramos et al. 2009). Neutral detergent fibre (NDF) which contains hemicellulose, cellulose and lignin are separated from other cell components such as proteins starch, sugars, organic acids and pectin through a neutral detergent digest. Hemicellulose is removed with an acid detergent digestion leaving the acid detergent fibre (ADF) constituents' cellulose and lignin. A further digestion using sulfuric acid separates the indigestible acid detergent lignin (ADL) and cellulose.

5.3.4. Statistical Analysis and QTL Mapping

Data analysis on phenotypic variance between replications was performed using a single factor analysis of variance (ANOVA) with the General Linear Model procedure using the R v.2.14.1. software. The traits measured in subsequent years were evaluated using Levene's test to test the homogeneity of variance.

The 179 x 345 genetic linkage map (Section 4.4.4.) was utilized for the QTL analysis. QTL analysis was performed by composite interval mapping (CIM) using Windows QTL Cartographer v. 2.5 (Wang et al. 2012). The Kosambi mapping function was used for uniformity of mapping distances between the generated *B. carinata* map (Section 4.3.5. Linkage Map

Construction 4.4.4) and the determination of the possible location of QTLs. The standard model (Model 6) of Windows QTL Cartographer was used in the analysis with a genome scan interval of 1 cM. A backward regression method was used to select background markers. The threshold for determining the presence of a QTL was set at a likelihood ratio of 11.5 or a LOD score of 2.5.

5.4.   Results

In 2011, the mean temperature during the growing season (May-October), 14.0^{o}C with a standard deviation of 5.5, was higher than the 30-year average (1971-2000) of 13.3^{o}C with a standard deviation of 1.72, and the total precipitation of 238.6 mm was 7% lower than the 30-year average of 256.8 mm with a standard deviation of 34.1. In 2012, the mean temperature, 12.9^{o}C with a standard deviation of 6.7, was lower and there was 43% more precipitation (368.1 mm) than the 30-year average.

Field trials in the 2011 season contained 64 out of the 76 lines from the mapped 179 x 345 DH population. Lines, 4, 15, 16, 19, 24, 28, 32, 36, 37, 53, 62, and 71 were not used due to the lack of available seed. Line 25 was omitted since data could only be collected in one replicate. Lines 28, 36 and 62 were not included in the 2012 season as continued efforts to produce enough seed in the greenhouse were not successful. The trait data between subsequent years was analysed separately due to the difference in number of lines.

Seed for the 2012 season could not be collected due to inclement weather and therefore the findings related to seed quality traits of the 179 x 345 DH population are only preliminary.

5.4.1. Agronomic Traits and Seed Content

The summary of agronomic traits including FT and PH of the 179 x 345 *B. carinata* DH population for the 2011 and 2012 growing seasons are presented in Table 5.2. FT of the population for 2012 took an average of five days longer than the previous year and the parental lines 179 and 345 required four and nine days longer to flower, respectively. The PH was shorter by an average of 27 cm in 2012 with the parental lines 179 and 345 showing a 26% and 9% decrease, respectively, in 2012.

The two parental lines 179 and 345 contrasted in FT and PH (Figure 5.1). Using a two sample t-test, the parental means of each trait was statistically significant ($P < 0.05$) in 2011 and 2012 (Table 5.3). In 2011, a difference of 112 cm in PH and 30 days separated the FT of 179 and 345 (Table 5.2). The following year, a 35 day difference in FT and 120 cm difference in PH set apart the parental lines 179 and 345.

An ANOVA was used to measure the variation between the replicates for each trait during each year (Table 5.4). The means between each replicate for each trait determined that the replicates were not statistically different from one another, with the exception of oil content ($P < 0.05$). All traits were analysed for possible QTLs. Levene's test, which tests the assumption that variances are equal across groups, was used to test the homogeneity of variance for the traits FT and PH. Both traits exhibited statistical equivalence between the 2011 and 2012 growing seasons ($P < 0.05$) (Table 5.5). However, since the number of lines were different between subsequent years the data were analysed separately.

Linkage groups generated by Mapmaker were exported to Windows QTL Cartographer and the phenotypic data was analysed (Table 5.6; Figure 5.2). No significant QTLs for FT and PH could be identified in the 2011 season; however, in 2012, there were two QTL's identified

Table 5.2. Phenotypic value of the parents and the DH population from each trial in 2011 and 2012.

Character	2011					2012				
			DH Lines					DH Lines		
	179	345	Mean	Minimum	Maximum	179	345	Mean	Minimum	Maximum
Field										
Days to Flower (50%)	46	76	56	44	77	50	85	61	50	86
Plant Height (cm)	133	245	162	93	230	98	218	135	77	210
Seed Content										
Oil (%)	37.2	32.5	35.9	30.2	42.5					
Protein (%)	30.1	27.2	29.9	23.8	34.2					
Neutral Detergent Fibre (NDF) (g/100g)	19.1	17.7	17.3	14.1	21.9					
Acid Detergent Fibre (ADF) (g/100g)	13.7	11.3	11.7	8.8	13.3					
Acid Detergent Lignin (ADL) (g/100g)	5.3	3.0	3.7	1.7	5.1					
Iodine (g I_2/100 g)	113.7	101.5	112.0	99.3	137.3					
Total Glucosinolates (µmol/g)	99.8	70.8	97.4	44.8	130.9					

Table 5.3. Two sampled t-test comparing means of parental lines 179 and 345 for various traits during the 2011 and 2012 growing season.

Trait	Year	*df*	*T-Critical Value*	*P-value*
FT	2011	4	2.78	0.00006
FT	2012	4	2.78	0.00006
PH	2011	4	2.78	0.00010
PH	2012	4	2.78	0.00004
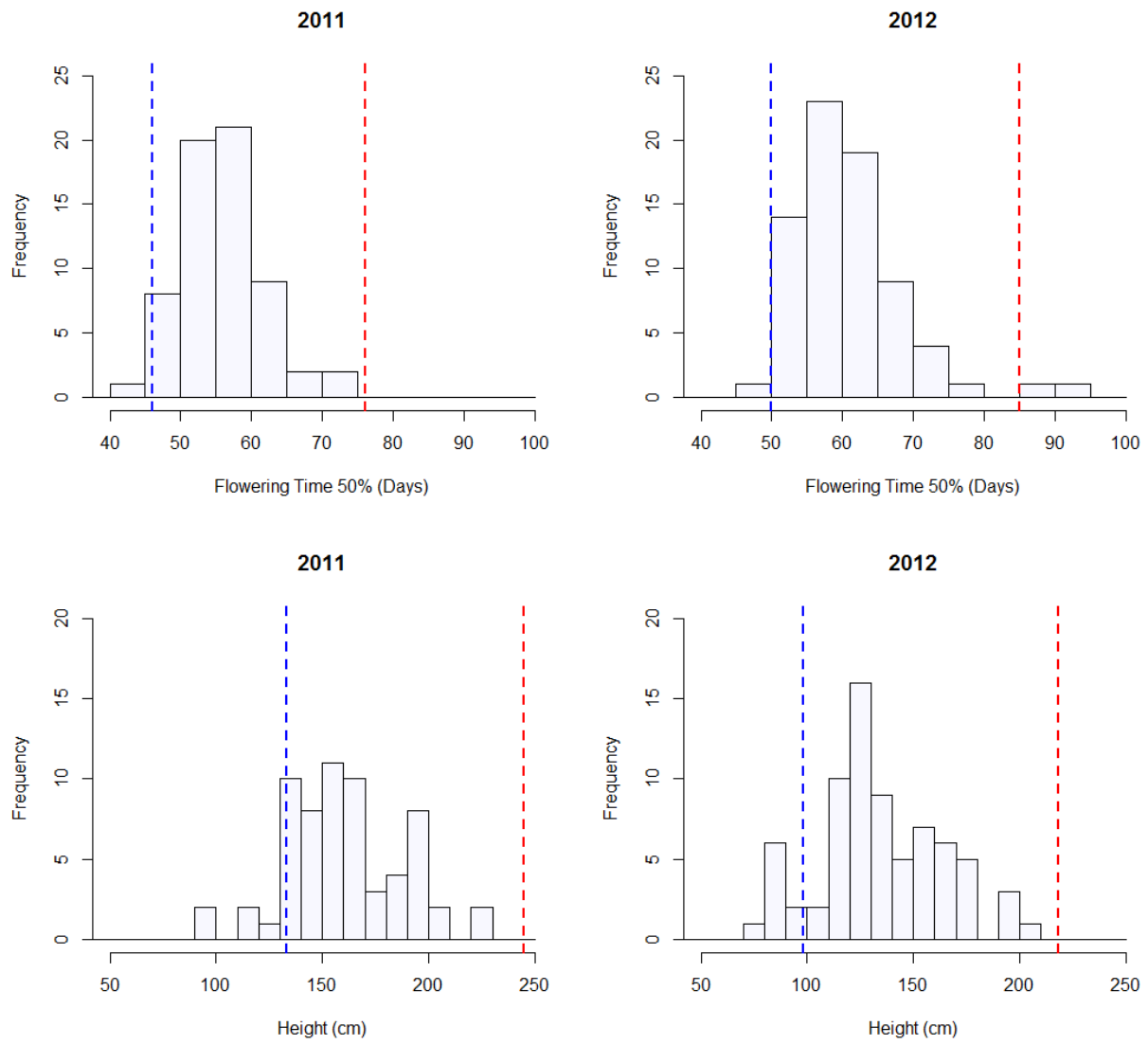
Figure 5.1. Frequency distribution of FT and PH for 2011 and 2012 season. Vertical lines represents values of parents; blue = 179, red = 345.

Table 5.4. ANOVA of the phenotypic traits within the replicates of the *B. carinata* 179 x 345 DH population during the 2011 and 2012 growing season.

Trait	Year	Source of Variation	Sum Squares	df	Mean Squares	F	P-Value
FT	2011	Replicates	3.0	2	1.47	0.040	0.960
		Lines	4087.3	62	65.29	2.533	5.27E-06
FT	2012	Replicates	22.0	2	10.90	0.248	0.781
		Lines	7866.4	71	110.80	14.740	3.11E-38
PH	2011	Replicates	1294.0	2	647.00	0.693	0.501
		Lines	145651.9	62	2349.22	10.120	1.18E-27
PH	2012	Replicates	33.0	2	16.30	0.017	0.983
		Lines	169566.0	71	2388.25	11.37	1.07E-32
Oil	2011	Replicates	77.4	2	38.68	5.004	0.008
		Lines	1056.3	62	17.04	4.677	1.15E-13
Protein	2011	Replicates	8.5	2	4.25	0.660	0.518
		Lines	892.7	62	14.40	5.801	3.9E-17
NDF	2011	Replicates	1.8	2	0.92	0.475	0.623
		Lines	225.7	62	3.64	3.398	3.04E-09
ADF	2011	Replicates	1.6	2	0.81	0.628	0.535
		Lines	168.6	62	2.72	4.739	7.24E-14
ADL	2011	Replicates	0.8	2	0.39	0.440	0.645
		Lines	142.1	62	2.29	12.257	1.04E-31
IV	2011	Replicates	11.0	2	5.50	0.166	0.847
		Lines	4534.7	62	73.14	5.655	1.05E-16
GSL	2011	Replicates	60.0	2	30.00	0.080	0.923
		Lines	37501.2	62	604.85	2.354	2.51E-05

Table 5.5. Levene's test for homogenity of variance across phenotypic traits between growing seasons.

Trait	Source of Variation	df	F	P-Value
FT	Between 2011 & 2012	1	1.224	0.271
	Total	134		
PH	Between 2011 & 2012	1	0.012	0.912
	Total	134		

Table 5.6. QTLs detected using composite interval mapping with QTL Cartographer for agronomic and seed content traits measured in 2011 and 2012.

Trait	QTL	Year	Linkage Group	Position (cM)*	Confidence Interval (cM)	LOD Score	R^2 (%)**	Additive Effect***
FT	FT.7.1	2012	Bc7	54.6	45.1-58.0	3.7	29.1	20.1
	FT.13.1		Bc13	97.9	85.6-108.2	3.2	14.5	6.9
Total variation****							8.5	
PH	PH.7.1	2012	Bc7	75.3	58.9-88.5	2.6	26.0	-31.4
	PH.17.1		Bc17	17.0	14.9-43.2	2.5	10.5	-28.8
Total variation							12.8	
Oil	Oil.6.1	2011	Bc6	53.3	26.6-59.0	2.5	11.3	1.7
NDF	NDF.2.1	2011	Bc2	107.4	90.3-115.9	2.8	18.0	1.0
	NDF.13.1		Bc13	47.1	46.6-73.7	2.6	15.0	0.9
Total variation							6.6	
ADF	ADF.U2.1	2011	U2	4.4	0.0-10.2	3.1	15.3	0.9
ADL	ADL.2.1	2011	Bc2	103.5	90.4-115.8	3.1	25.5	0.9
IV	IV.2.1	2011	Bc2	214.8	211.7-217.4	5.4	18.8	-18.8
	IV.16.1	2011	Bc16	44.5	39.5-44.9	4.3	22.6	-15.6
	IV.16.2	2011	Bc16	57.7	53.0-63.6	5.9	30.4	9.6
	IV.U1.1	2011	U1	4.1	0.0-5.1	3.4	12.5	3.6
Total variation							13.7	

* Position – Position of peak LOD score
** R2 - phenotypic variation
*** Positive values attributed to 179, negative values, 345.
****Amount of phenotypic variation simultaneously explained by all makers found for each trait determined using stepwise regression.
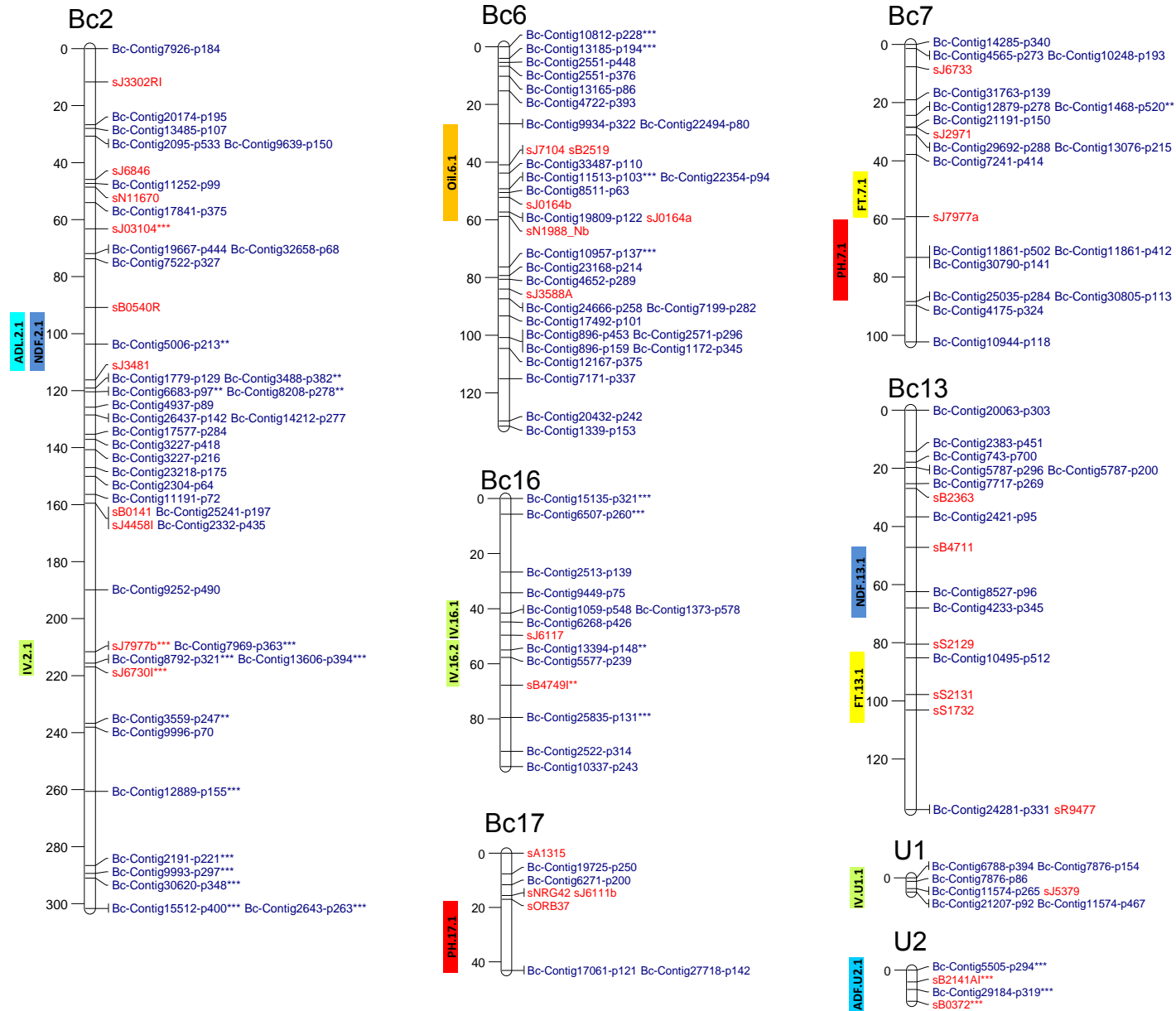
Figure 5.2. Mapping of QTLs for FT, PH, Oil, NDF, ADF, ADL, and IV on the SSR/SNP linkage map of *B. carinata* generated from Chapter 4. Distance between markers in Kosambi cM. The position of each QTL is indicated by a coloured box to the left of the linkage group. Length of the colour bar correlates to 1 LOD (tenfold) likelihood confidence interval.

for FT, located on Bc7 and Bc13 (Figure 5.3), which accounted for 9% of the total phenotypic variation (Table 5.6). Also in 2012, two QTL's for PH located on Bc7 and Bc17 (Figure 5.4) and accounted for about 13% of the total phenotypic variation.

An assessment of seed content was performed using NIRS for the seed harvested from the 2011 field trial. Seed data could only be measured for the 2011 season, so identified QTL relating to seed content are provisional. The parental lines were on contrasting sides of the mean with the exception of NDF (Figure 5.5). Preliminary analysis of QTL controlling seed content from the 2011 trial season resulted in nine possible QTLs on four *B. carinata* linkage groups and two of the unknown groups. No QTLs were located for protein and glucosinolate content. A QTL was detected for oil content on Bc6 (Figure 5.6) accounting for 11% of the phenotypic variation (Table 5.6). There were four QTLs discovered for fibre, two were from NDF on Bc2 and Bc13 accounting for 7% of the total phenotypic variation and one each for ADF on U2 and ADL on Bc2 accounting for 15% and 26% of the phenotypic variation respectively (Figure 5.7). There were four IV QTLs detected, with two located on Bc16 and one each on Bc2 and U1 (Figure 5.8) in 2011, which may indicate the variation of unsaturated fatty acid levels in *B. carinata*.

5.5.   Discussion

Replicates of the *B. carinata* 179 x 345 DH population were grown for two seasons in Saskatoon, Saskatchewan. Soil conditions and day length were not significant factors as each replicate was grown near the same area around the same time for each season. The average days to flower (FT), affected mainly by temperature, varied between 2011 and 2012 which could be attributed to the warmer drier growing season in 2011. The average PH was shorter in 2012 than
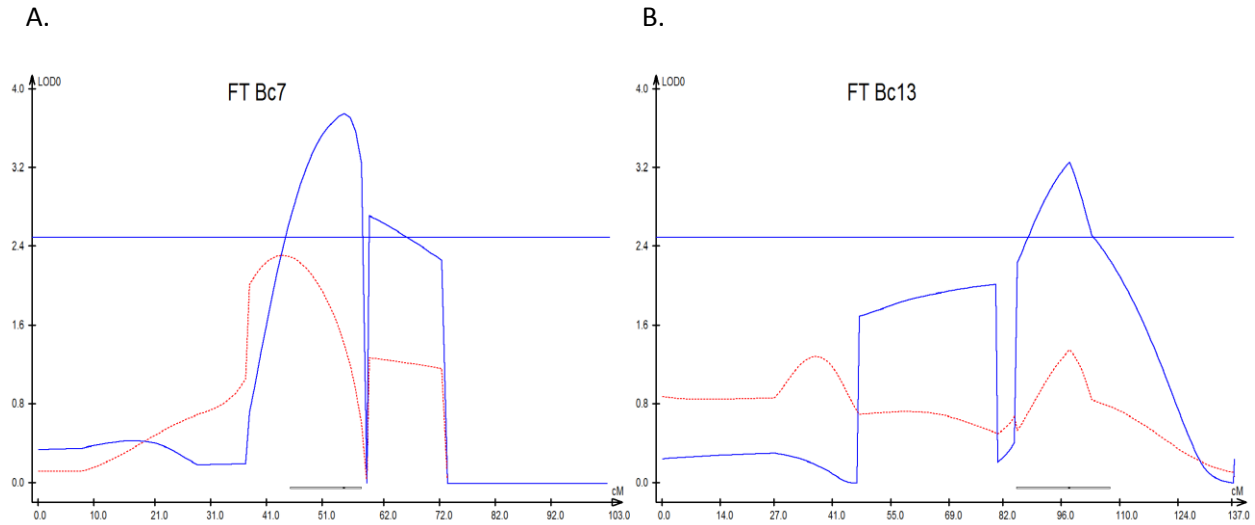
Figure 5.3. Quantitative trait loci (QTL) likelihood plot associated with FT on linkage groups Bc7 (left) and Bc13 (right). The vertical axis indicates the LOD score and the horizontal axis indicates distances in cM based on composite interval mapping. The horizontal bar above the x-axis indicates QTL region. The horizontal line indicates the LOD threshold (2.5) for the FT trait. Red = 2011, Blue = 2012.
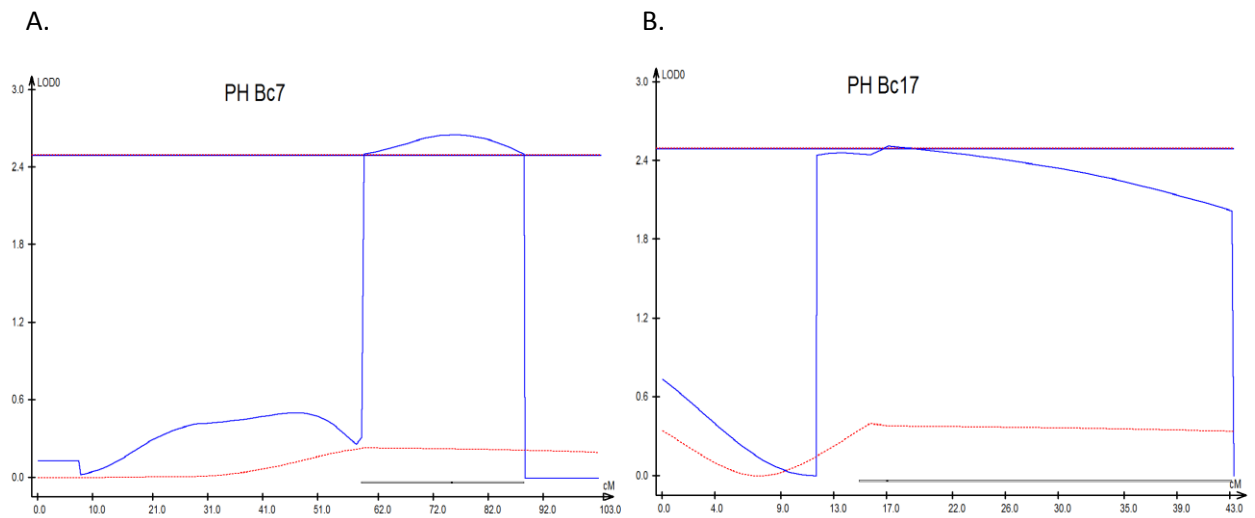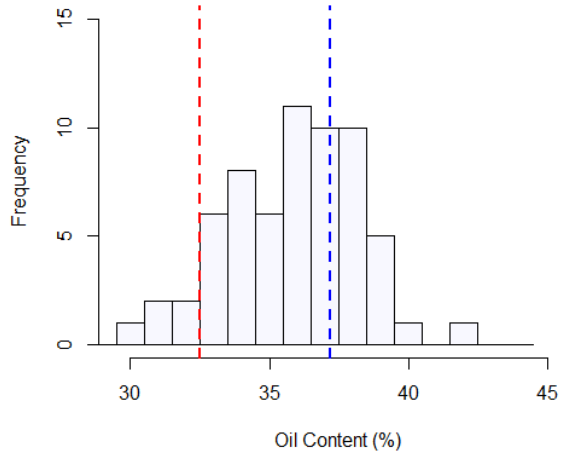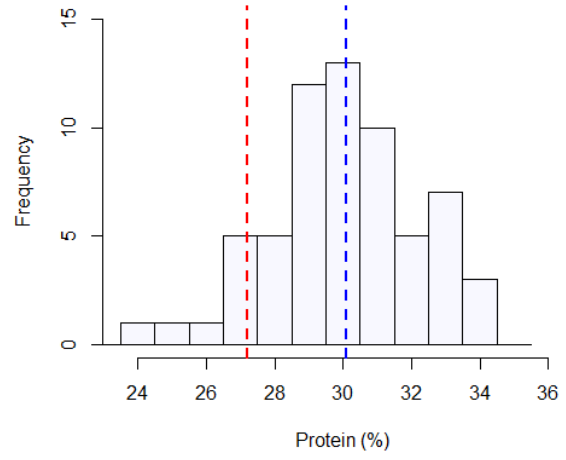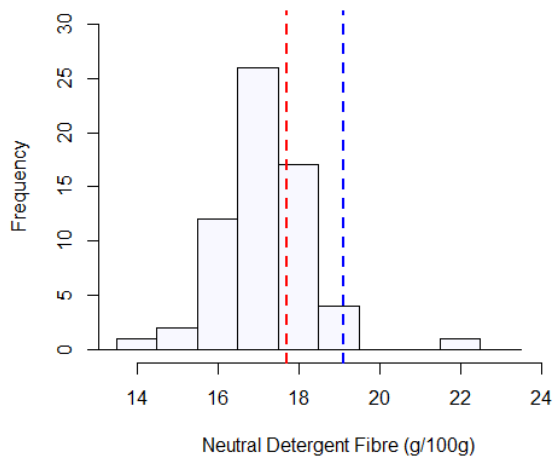

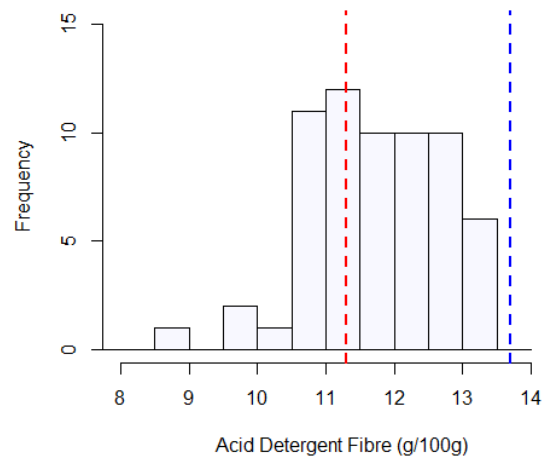Figure 5.4. Quantitative trait loci (QTL) likelihood plot associated with PH on linkage groups Bc7 (left) and Bc17 (right). The vertical axis indicates the LOD score and the horizontal axis indicates distances in cM based on composite interval mapping. The horizontal bar above the x-axis indicates QTL region. The horizontal line indicates the LOD threshold (2.5) for the PH trait. Red = 2011, Blue = 2012.
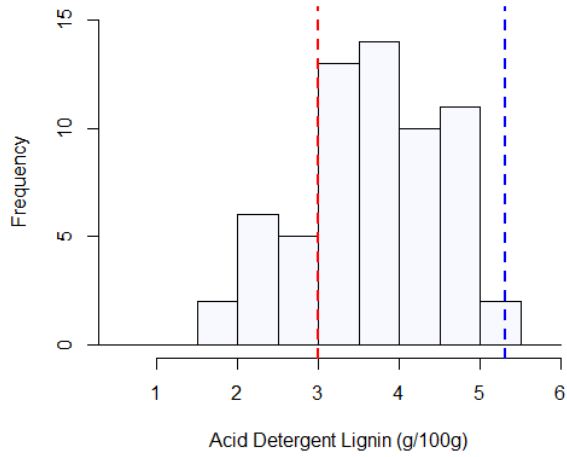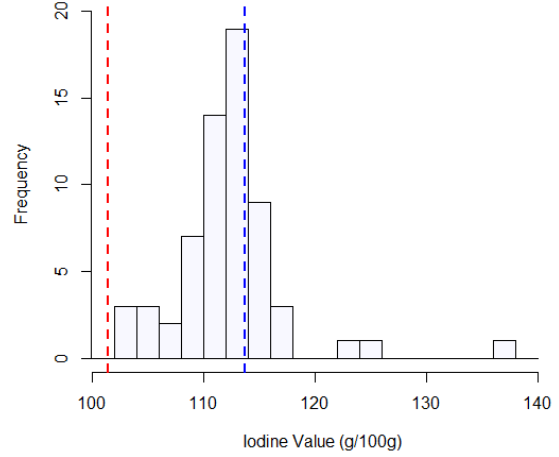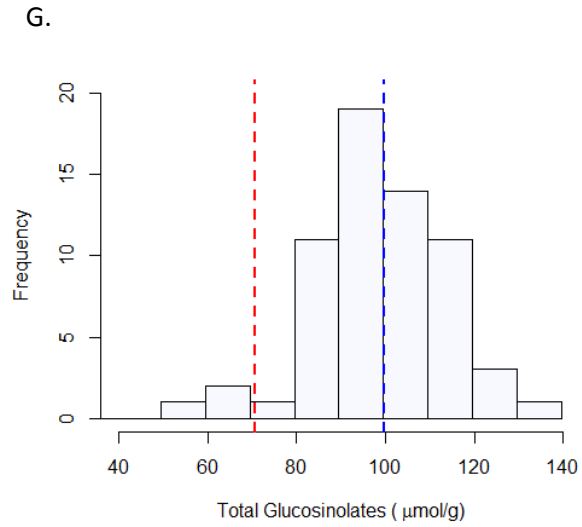
G.

Figure 5.5. Frequency distribution of 63 *B. carinata* lines from the 179 x 345 DH population and their seed content determined by NIRS for the 2011 season. A) Oil Content. B) Protein. C) NDF. D) ADF. E) ADL. F) Iodine. G) Glucosinolates. Vertical lines represent value of parental lines; blue = 179, red = 345.
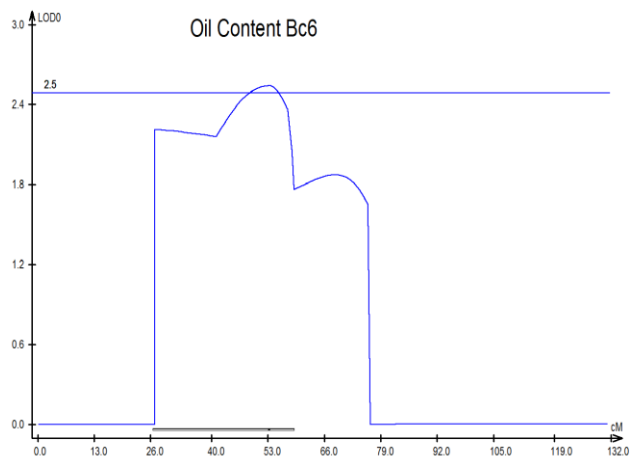
Figure 5.6. Quantitative trait loci (QTL) likelihood plot associated with oil content on linkage group Bc6 during the 2011 season. The vertical axis indicates the LOD score and the horizontal axis indicates distances in cM based on composite interval mapping. The horizontal bar above the x-axis indicates QTL region. The horizontal line indicates the LOD threshold (2.5) for the oil content trait.
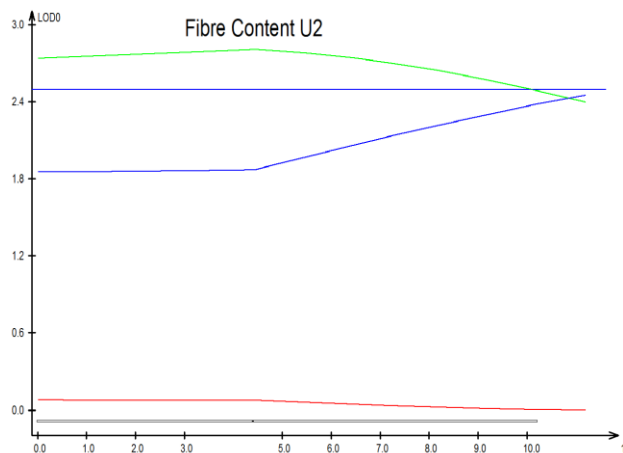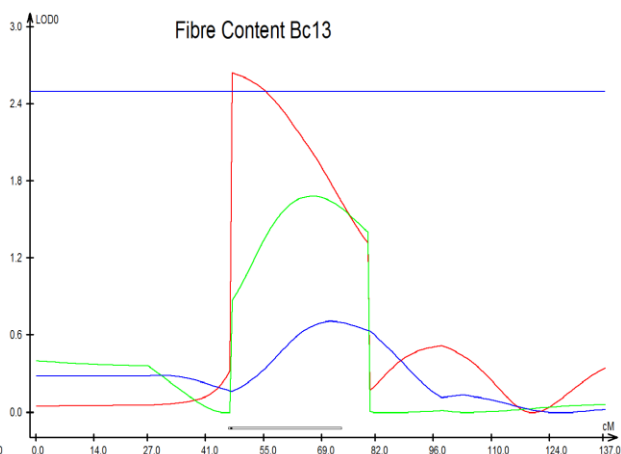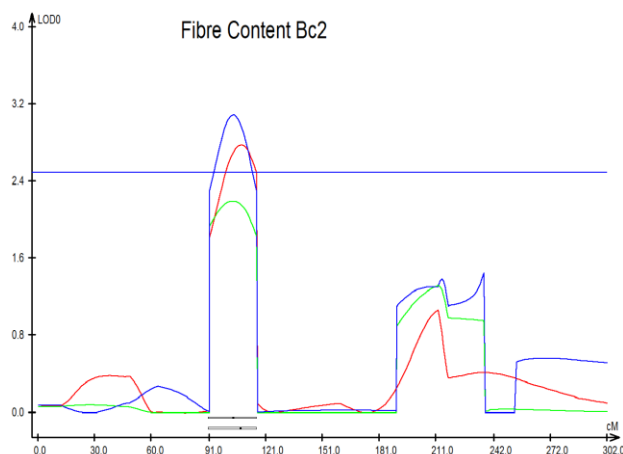

Figure 5.7. Quantitative trait loci (QTL) likelihood plot associated with fibre content on linkage groups Bc2, Bc13, and U2 during the 2011 season. The vertical axis indicates the LOD score and the horizontal axis indicates distances in cM based on composite inter composite interval mapping. The horizontal bar above the x-axis indicates QTL region. The horizontal line indicates the LOD threshold (2.5) for the fibre content trait. Red = NDF, Green = ADF, and Blue = ADL.

107

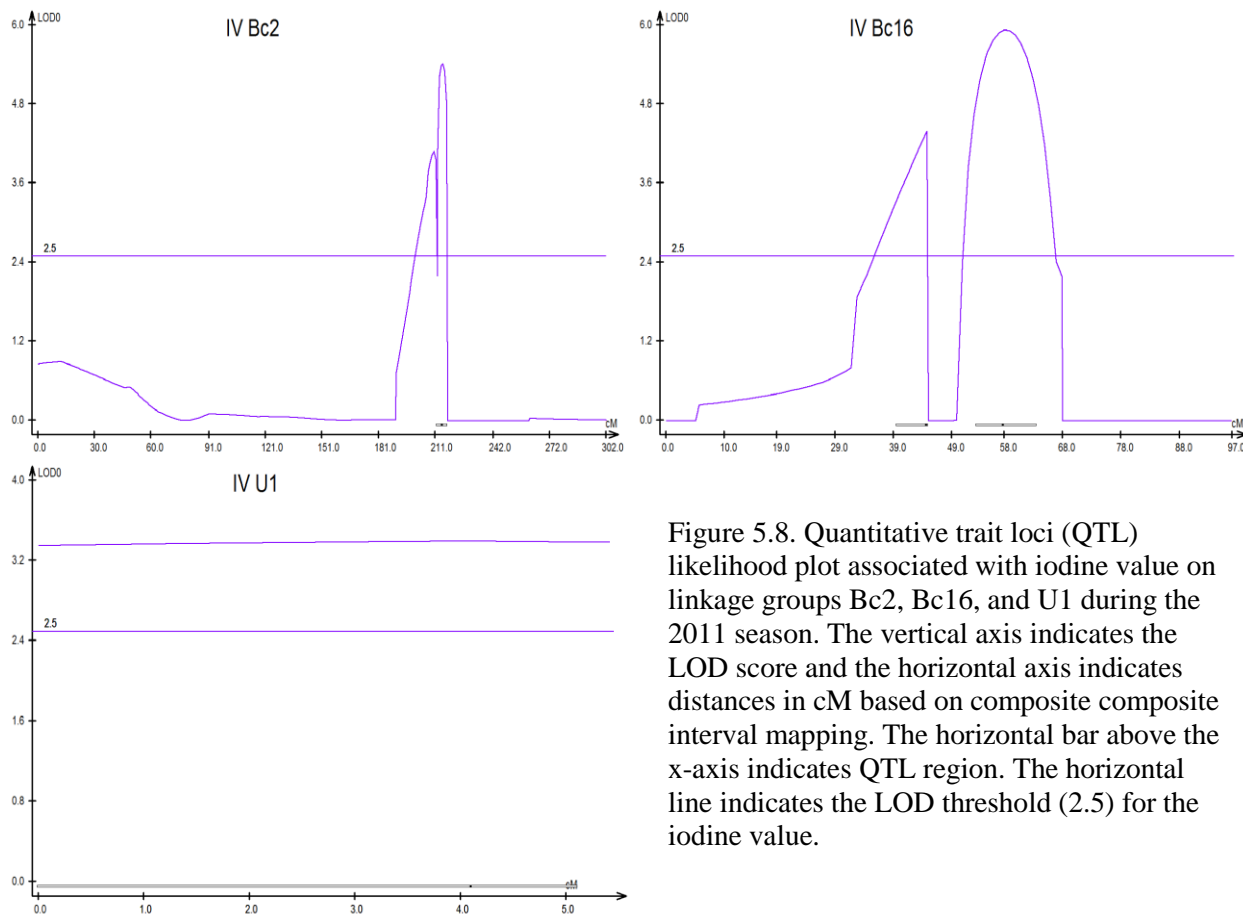Figure 5.8. Quantitative trait loci (QTL) likelihood plot associated with iodine value on linkage groups Bc2, Bc16, and U1 during the 2011 season. The vertical axis indicates the LOD score and the horizontal axis indicates distances in cM based on composite composite interval mapping. The horizontal bar above the x-axis indicates QTL region. The horizontal line indicates the LOD threshold (2.5) for the iodine value.

the previous year even though there was an increase in precipitation. This would suggest cooler temperatures also played a role in PH.

The loci controlling FT and PH could not be detected in 2011. QTLs were more significant for the same agronomic traits in the following year. Although analysis of FT data in the 2011 growing season did indicated a LOD peak of 2.4 in the same confidence interval on Bc7 for the FT QTL detected in 2012 (Figure 5.3). While the environmental conditions may have affected the level of expression, the results may be attributed to the difference in the number of lines between the seasons.

A potential QTL was located for oil content in Bc6, however the ANOVA analysis showed a significant difference between the replicates (P <0.05), additional data will need to be collected to confirm these results.

The *B. carinata* 179 x 345 DH population yielded a potential QTL for NDF and ADL in the same confidence interval of 90-116 cM on Bc2. Though not considered significant, a LOD peak of 2.2 was found for ADF in the same area on Bc2 (Figure 5.7). This may suggest a region controlling the amount of dietary fibre produced, with specificity towards lignin production.

The amount of unsaturation within the fatty acid seed content could potentially influence one of the many possible uses for *B. carinata*. Studies in Europe on *B. carinata* to assess its possible use as a biodiesel feedstock identified IV's between 122-128 g I_2/100g (Bouaid et al. 2009; Cardone et al. 2003), this is of concern since the European Union has introduced standards for biodiesel, which must have an IV below 120 g I_2/100g (Ramos et al. 2009). The heating of higher unsaturated fatty acid levels potentially results in polymerization of glycerides forming deposits or deterioration of the lubricating oil (Mittelbach 1996). The majority of the seed content within the lines of 179 x 345 population have exhibited IV below this level suggesting

this *B. carinata* germplasm would have utility for biodiesel production. The possible QTLs identified could influence the production of saturated and unsaturated fatty acids and may be a useful trait for further study.

The potential QTLs ascertained in *B. carinata* could not be easily compared to previously mapped *Brassica* QTL's in the B and C genome because of the differences in markers used as well as the size and orientation of the map. However, comparison of the *Brassica* linkage group nomenclature (Axelsson et al. 2000; Bohuon et al. 1996; Parkin et al. 1995) revealed that two QTLs for FT have been found in *B. oleracea* on O5 (Bohuon et al. 1998; Rae et al. 1999) which corresponds to the same linkage group as Bc13, as well, a QTL was mapped to N19 of *B. napus* for PH, which corresponds to Bc17 (Mei et al. 2009). Although this may suggest the possibility the QTLs are equivalent, further analysis will need to be carried out to compare the actual regions flanking the QTLs.

5.6.   Conclusions and Future Research

The results of the QTL study provide a better understanding of the genomic regions controlling the agronomic traits FT and PH, as well a preliminary framework for QTLs derived for seed content. An overview of the traits revealed the varying ranges of the 179 x 345 DH population as well as the contrasting phenotypes in PH and FT (Table 5.2). The locations of QTLs in FT and PH were more prominent in 2012, suggesting environmental factors influenced the differences between seasons, coupled with the increase in the number of lines in 2012 which contributed to additional informative data. Further trials are needed to assess seed content to determine the legitimacy of the identified QTLs. Comparisons between *B. carinata* and other

110

Brassica species using polymorphic markers to bridge the areas within linkage groups would allow for the potential identification of identical QTLs between the *Brassica* species.

A subset of traits has been evaluated in the *B. carinata* 179 x 345 DH population. Further research into locating genes and traits of interest would be beneficial towards developing *B. carinata* into a sustainable crop platform for the prairies. Additional studies on environmental tolerance, disease resistance, and traits of economic importance could now be evaluated for *B. carinata* as well as other *Brassica* species. Comparative mapping and identifying common genes between *Arabidopsis* and *Brassica* species would help elucidate candidate genes. Future work will focus on the efficacy of useful genes with emphasis towards promoting industry use and commercial viability of *B. carinata*.

6. General Discussion and Conclusions

A genetic overview of *B. carinata* has been accomplished in this study. Resistance to particular insects and pathogens, as well as *B. carinata's* ability to thrive in drought and water limited climates has suggested the potential of *B. carinata* as a new, renewable natural resource for food, energy and industrial uses. With the focus directed towards fuels and industrial purposes, emphasis on producing suitable yields in the Canadian Prairies to accommodate these uses can begin through studies at the genetic level. A genotypic and phenotypic assessment has illuminated the origins, genetic framework, and allelic variation for potentially useful traits for *B. carinata*.

The initial step in the genotypic assessment of *B. carinata* was evaluating the genetic diversity within a subset of lines. The use of microsatellite markers derived from *B. napus* and *B. juncea* were able to distinguish amongst 37 individual *B. carinata* lines. As expected, accessions from around the world were more genetically diverse than the accessions conditioned to grow specifically in the Canadian prairies. As well, the genetic diversity of *B. carinata* was lower compared to other plant species, which suggests only a limited number of interspecific hybridization events occurred during its recent formation from its progenitor species. The genetic screen also determined that the genetic diversity found in lines 179 and 345 was suitable to allow creation of a genetic linkage map.

SNP discovery for *B. carinata* was achieved by identifying SNP alleles between parental lines 179 and 345. The developed SNP markers were tested on the 179 x 345 DH population and the segregation patterns were aligned to generate genetic linkage groups, creating the first *B. carinata* SNP genetic linkage map with essentially no prior sequence knowledge.

112

The comparison of the genetic sequences between *B. carinata* SNP markers and the B and C genome of the *Brassica* genus, along with map positions of previously tested SSR markers, facilitated in identifying the 17 linkage groups of *B. carinata*. These data suggested the B genome of *B. juncea* and the C genome of *B. napus* are closely related to the B and C genome of *B. carinata,* further elucidating the origins of the *Brassica* species. This assembled genetic linkage map coincides with previously generated *Brassica* maps of amphidiploid species, with the two genomes remaining essentially unaltered with no major chromosomal rearrangements since their hybridisation.

The number of loci amplified between each genome using both SSR and SNP markers would suggest that the B genome of *B. carinata* is more polymorphic than the C genome of this amphidiploid species. It has been suggested that the use of *B. rapa* in breeding strategies has made the A genome more genetically diverse than the C genome in *B. napus* (Bus et al. 2011), alternatively it could be proposed that the progenitor B genome has undergone similar introgressions during the history of *B. carinata* cultivation. Elucidation of these observations would require further research.

The *B. carinata* lines 179 and 345 from the JIC contrasted in both their genotype and phenotype. The genetic diversity of these two lines determined in the genetic screen warranted their use in the generation of a genetic linkage map. Differences in agronomic traits, FT and PH, between parental lines were evident in the field trials. However, FT and PH QTL could not be detected from the 2011 data, which was primarily attributed to the fewer number of lines in the population. Potential QTLs for each trait, FT and PH, were detected in the 2012 growing season. The seed content analyses suggested potential variant alleles for oil content, fatty acid concentration and fibre content. These preliminary results will require further research.

This research has shed light on the genetic diversity of *B. carinata*. The generation of a genetic linkage map has located potential QTLs and has the ability to assist in locating other genes and traits of interest for marker assisted breeding. Comparative mapping of *B. carinata* with other *Brassica* genomes has further elucidated the evolution of the *Brassica* species within the Triangle of U revealing the chromosomes have remained relatively unchanged since their formation. This genetic assessment of *B. carinata* is the initial step towards future research into developing a viable industrial crop platform for the Canadian Prairies and the world.

7. References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. Journal of molecular biology 215:403-410

Amjad M, Peters DC (1992) Survival, development, and reproduction of turnip aphids (*Homoptera: Aphididae*) on oilseed *Brassica*. Journal of economic entomology 85:2003-2007

Anand IJ, Singh JP, Malik RS (1985) *B. carinata* - A potential oilseed crop for rainfed agriculture. Crucifer Newsletter 10:76-79

Armour JAL, Jeffreys AJ (1992) Biology and applications of human minisatellite loci. Current Opinion in Genetics and Development 2:850-856

Arumuganathan K, Earle E (1991) Nuclear DNA content of some important plant species. Plant Molecular Biology Reporter 9:208-218

Ashraf M, McNeilly T (1990) Responses of four *Brassica* species to sodium chloride. Environmental and Experimental Botany 30:475-487

Axelsson T, Bowman CM, Sharpe AG, Lydiate DJ, Lagercrantz U (2000) Amphidiploid *Brassica juncea* contains conserved progenitor genomes. Genome 43:679-688

Axelsson T, Shavorskaya O, Lagercrantz U (2001) Multiple flowering time QTLs within several *Brassica* species could be the result of duplicated copies of one ancestral gene. Genome 44:856-864

Babula D, Kaczmarek M, Barakat A, Delseny M, Quiros CF, Sadowski J (2003) Chromosomal mapping of *Brassica oleracea* based on ESTs from *Arabidopsis thaliana*: complexity of the comparative map. Mol Gen Genomics 268:656-665

Bancroft I, Morgan C, Fraser F, Higgins J, Wells R, Clissold L, Baker D, Long Y, Meng J, Wang X, Liu S, Trick M (2011) Dissecting the genome of the polyploid crop oilseed rape by transcriptome sequencing. Nat Biotech 29:762-766

Bansal VK, Seguin-Swartz F, Rakow GFW, Petrie GA (1990) Reaction of *Brassica* species to infection by *Alternaria brassicae*. Canadian Journal of Plant Science 70:1159-1162

Bayeh M, Gebre Medhin T (1992) Insect pests of noug, linseed and *Brassica* in Ethiopia. *Oilseeds Research and Development in Ethiopia,* Proceedings of the First National Oilseeds Workshop Abeba, Addis, Ethiopia, pp 174-178

Benham J, Jeung JU, Jasieniuk M, Kanazin V, Blake T (1999) Genographer: a graphical tool for automated fluorescent AFLP and microsatellite analysis. Journal of Agricultural Genomics 4:1-3

Bodnaryk RP (1992) Leaf epicuticular wax, an antixenotic factor in Brassicaceae that affects the rate and pattern of feeding of flea beetles, *Phyllotreta cruciferae* (Goeze). Canadian Journal of Plant Science 72:1295-1303

Bohuon EJR, Keith DJ, Parkin IAP, Sharpe AG, Lydiate DJ (1996) Alignment of the conserved C genomes of *Brassica oleracea* and *Brassica napus*. Theoretical and Applied Genetics 93:833-839

Bohuon EJR, Ramsay LD, Craft JA, Arthur AE, Marshall DF, Lydiate DJ, Kearsey MJ (1998) The association of flowering time quantitative trait loci with duplicated regions and candidate loci in *Brassica oleracea*. Genetics 150:393-401

Bouaid A, Martinez M, Aracil J (2009) Production of biodiesel from bioethanol and *Brassica carinata* oil: Oxidation stability study. Bioresource Technology 100:2234-2239

Brock TD, Freeze H (1969) Thermus aquaticus gen. n. and sp. n., a nonsporulating extreme thermophile. Journal of Bacteriology 98:289-297

Brookes AJ (1999) The essence of SNPs. Gene 234:177-186

Burns MJ, Barnes SR, Bowman JG, Clarke MHE, Werner CP, Kearsey MJ (2003) QTL analysis of an intervarietal set of substitution lines in *Brassica napus*: (i) Seed oil content and fatty acid composition. Heredity 90:39-48

Bus A, Körber N, Snowdon RJ, Stich B (2011) Patterns of molecular variation in a species-wide germplasm set of *Brassica napus*. Theoretical and Applied Genetics 123:1413-1423

Camargo LEA, Savides L, Jung G, Nienhuis J, Osborn TC (1997) Location of the self-incompatibility locus in an RFLP and RAPD map of *Brassica oleracea*. Journal of Heredity 88:57-60

Cardone M, Mazzoncini M, Menini S, Rocco V, Senatore A, Seggiani M, Vitolo S (2003) *Brassica carinata* as an alternative oil crop for the production of biodiesel in Italy: agronomic evaluation, fuel production by transesterification and characterization. Biomass and Bioenergy 25:623-636

Chaudhary S, Parmenter DL, Moloney MM (1998) Transgenic *Brassica carinata* as a vehicle for the production of recombinant proteins in seeds. Plant Cell Reports 17:195-200

Chien A, Edgar DB, Trela JM (1976) Deoxyribonucleic acid polymerase from the extreme thermophile Thermus aquaticus. Journal of Bacteriology 127:1550-1557

Ching A, Caldwell K, Jung M, Dolan M, Smith O, Tingey S, Morgante M, Rafalski A (2002) SNP frequency, haplotype structure and linkage disequilibrium in elite maize inbred lines. BMC Genetics 3:19

Chyi YS, Hoenecke ME, Sernyk JL (1992) A genetic linkage map of restriction fragment length polymorphism loci for *Brassica rapa* (syn. *campestris*). Genome 35:746-757

Dolores Sacristan M, Gerdemann M (1986) Different behavior of *Brassica juncea* and *B. carinata* as sources of *Phoma lingam* resistance in experiments of interspecific transfer to *B. napus*. Plant Breeding 97:304-314

Eveland AL, McCarty DR, Koch KE (2008) Transcript profiling by 3′-untranslated region sequencing resolves expression of gene families. Plant physiology 146:32-44

Foisset N, Delourme R (1996) Segregation distortion in androgenic plants. In: Jain SM, Sopory SK, Veilleux RE (eds) In vitro Haploid Production in Higher Plants: Volume 2: Applications. Springer, pp 189-201

Foisset N, Delourme R, Barret P, Renard M (1995) Molecular tagging of the dwarf BREIZH (Bzh) gene in *Brassica napus*. Theoretical and Applied Genetics 91:756-761

Fougeres S (2012) First solely-biofuel jet flight raises clean travel hopes. AFP

Fu Y-B (2012) Genetic structure in a core subset of cultivated barley germplasm. Crop Sci 52:1195-1208

Fu Y-B, Peterson GW, Morrison MJ (2007) Genetic diversity of Canadian soybean cultivars and exotic germplasm revealed by simple sequence repeat markers Crop Sci 47:1947-1954

Fu YB (2006) Redundancy and distinctness in flax germplasm as revealed by RAPD dissimilarity. Plant Genetic Resources 4:117-124

Gao M, Li G, Yang B, Qiu D, Farnham M, Quiros C (2007) High-density *Brassica oleracea* linkage map: identification of useful new linkages. TAG Theoretical and Applied Genetics 115:277-287

Gasol CM, Gabarrell X, Anton A, Rigola M, Carrasco J, Ciria P, Solano ML, Rieradevall J (2007) Life cycle assessment of a *Brassica carinata* bioenergy cropping system in southern Europe. Biomass and Bioenergy 31:543-555

Getinet A, Rakow G, Downey RK (1996a) Agronomic performance and seed quality of Ethiopian mustard in Saskatchewan. Canadian Journal of Plant Science 76:387-392

Getinet A, Rakow G, Raney JP (1996b) Glucosinolate Content Variation in *Brassica carinata* A. Braun Germplasm Grown at Holetta Ethiopia. Crucifer Newsletter 18:84-85

Getinet A, Rakow G, Raney JP, Downey RK (1994) Development of zero erucic acid Ethiopian mustard through an interspecific cross with zero erucic acid Oriental mustard. Canadian Journal Of Plant Science 74:793-795

Getinet A, Rakow G, Raney JP, Downey RK (1997) Glucosinolate content in interspecific crosses of *Brassica carinata* with *B. juncea* and *B. napus*. Plant Breeding 116:39-46

Gómez-Campo C, Tortosa ME, Tewari I, Tewari JP (1999) Epicuticular wax columns in cultivated *Brassica* species and in their close wild relatives. Annals of Botany 83:515-519

Gugel RK, Seguin-Swartz F, Petrie GA (1990) Pathogenicity of three isolates of *Leptosphaeria maculans* on *Brassica* species and other crucifers. Canadian Journal of Plant Pathology 12:75-82

Guo H, Dickson MH, Hunter JE (1991) *Brassica napus* sources of resistance to black rot in crucifers and inheritance of resistance. HortScience 26:1545-1547

Guo S, Zou J, Li R, Long Y, Chen S, Meng J (2012) A genetic linkage map of *Brassica carinata* constructed with a doubled haploid population. Theoretical and Applied Genetics 125:1113-1124

Harrison GE, Heslop-Harrison JS (1995) Centromeric repetitive DNA sequences in the genus *Brassica*. TAG Theoretical and Applied Genetics 90:157-165

Hemingway JS (1976) Mustards - *Brassica spp.* and *Sinapis alba* (Cruciferae). In: Simmonds NW (ed) Evolution of Crop Plants. Longman Inc., New York, pp 56-59

Howell EC, Kearsey MJ, Jones GH, King GJ, Armstrong SJ (2008) A and C genome distinction and chromosome identification in *Brassica napus* by sequential fluorescence *in situ* hybridization and genomic *in situ* hybridization. Genetics 180:1849-1857

Huang B, Liu Y, Xue X, Chang L (2002) Comparison of aluminium tolerance in the *Brassicas* and related species. Plant Breeding 121:360-362

Inaba R, Nishio T (2002) Phylogenetic analysis of *Brassiceae* based on the nucleotide sequences of the *S*-locus related gene, *SLR1*. TAG Theoretical and Applied Genetics 105:1159-1165

Iniguez-Luy F, Lukens L, Farnham M, Amasino R, Osborn T (2009) Development of public immortal mapping populations, molecular markers and linkage maps for rapid cycling *Brassica rapa* and *B. oleracea*. TAG Theoretical and Applied Genetics 120:31-43

Irtelli B, Navari-Izzo F (2008) Uptake kinetics of different arsenic species by *Brassica carinata*. Plant and Soil 303:105-113

Irtelli B, Petrucci WA, Navari-Izzo F (2009) Nicotianamine and histidine/proline are, respectively, the most important copper chelators in xylem sap of *Brassica carinata* under conditions of copper deficiency and excess. Journal of Experimental Botany 60:269-277

Johnston JS, Pepper AE, Hall AE, Chen ZJ, Hodnett G, Drabek J, Lopez R, Price HJ (2005) Evolution of genome size in Brassicaceae. Annals of Botany 95:229-235

Jönsson R (1978) Breeding for improved oil and meal quality in rape (*Brassica napus* L.) and

    turnip rape (*Brassica campestris* L.). Hereditas 87:205-218

Katavic V, Friesen W, Barton DL, Gossen KK, Giblin EM, Luciw T, An J, Zou J, MacKenzie

    SL, Keller WA, Males D, Taylor DC (2000) Utility of the Arabidopsis FAE1 and yeast

    SLC1-1 genes for improvements in erucic acid and oil content in rapeseed. Biochem Soc

    Trans 28:935-937

Kearsey MJ (1998) The principles of QTL analysis (a minimal mathematics approach). Journal

    of Experimental Botany 49:1619-1623

Kent WJ (2002) BLAT—the BLAST-like alignment tool. Genome research 12:656-664

Kianian SF, Quiros CF (1992) Generation of a *Brassica oleracea* composite RFLP map: linkage

    arrangements among various populations and evolutionary implications. Theoretical and

    Applied Genetics 84:544-554

Kircher M, Kelso J (2010) High-throughput DNA sequencing – concepts and limitations.

    BioEssays 32:524-536

Kirkegaard JA, Sarwar M (1998) Biofumigation potential of *Brassicas*. Plant and Soil 201:71-89

Koch M, Haubold B, Mitchell-Olds T (2001) Molecular systematics of the Brassicaceae:

    evidence from coding plastidic matK and nuclear Chs sequences. American Journal of

    Botany 88:534-544

Koch MA, Haubold B, Mitchell-Olds T (2000) Comparative evolutionary analysis of chalcone

    synthase and alcohol dehydrogenase loci in Arabidopsis, Arabis, and related genera

    (Brassicaceae). Molecular Biology and Evolution 17:1483-1498

Kumar A, Singh P, Singh DP, Singh H, Sharma HC (1984) Differences in osmoregulation in

    Brassica species. Annals of Botany 54:537-541

Lagercrantz U (1998) Comparative mapping between *Arabidopsis thaliana* and *Brassica nigra* indicates that *Brassica* genomes have evolved through extensive genome replication accompanied by chromosome fusions and frequent rearrangements. Genetics 150:1217-1228

Lagercrantz U, Ellegren H, Andersson L (1993) The abundance of various polymorphic microsatellite motifs differs between plants and vertebrates. Nucleic Acids Research 21:1111-1115

Lagercrantz U, Lydiate DJ (1995) RFLP mapping in *Brassica nigra* indicates differing recombination rates in male and female meioses. Genome 38:255-264

Lagercrantz U, Lydiate DJ (1996) Comparative genome mapping in *Brassica*. Genetics 144:1903-1910

Lander ES, Green P, Abrahamson J, Barlow A, Daly M, Lincoln SE, Newburg L (1987) Mapmaker: An interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. Genomics 1:174-181

Landry BS, Hubert N, Crete R, Chang MS, Lincoln SE, Etoh T (1992) A genetic map for *Brassica oleracea* based on RFLP markers detected with expressed DNA sequences and mapping of resistance genes to race 2 of *Plasmodiophora brassicae* (Woronin). Genome 35:409-420

Larson S, Pilieci V (2012) Grown in Saskatchewan, flown in Ottawa. Ottawa Citizen, Ottawa

Leonard EC (1992) High-erucic vegetable oils. Industrial Crops and Products 1:119-123

Li F, Kitashiba H, Inaba K, Nishio T (2009) A *Brassica rapa* linkage map of EST-based SNP markers for identification of candidate genes controlling flowering time and leaf morphological traits. DNA research 16:311-323

Lim K-B, Yang T-J, Hwang Y-J, Kim JS, Park J-Y, Kwon S-J, Kim J, Choi B-S, Lim M-H, Jin M, Kim H-I, de Jong H, Bancroft I, Lim Y, Park B-S (2007) Characterization of the centromere and peri-centromere retrotransposons in *Brassica rapa* and their distribution in related *Brassica* species. The Plant Journal 49:173-183

Lincoln SE, Daly M, Lander ES (1992) Constructing genetic maps with Mapmaker/Exp 3.0. 3.0 edn. Whitehead Institute Technical Report, Cambridge, MA.

Lukens L, Fei Z, Lydiate DJ, Parkin I, Osborn T (2003) Comparison of a *Brassica oleracea* genetic map with the genome of *Arabidopsis thaliana*. Genetics 164:359-372

Lysak MA, Koch MA (2011) Phylogeny, genome, and karyotype evolution of Crucifers (Brassicaceae). In: Schmidt R, Bancroft I (eds) Genetics and genomics of the Brassicaceae. Springer New York, pp 1-31

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

Malik RS (1990) Prospects for *Brassica Carinata* as an Oilseed crop in India. Experimental Agriculture 26:125-129

Martínez M, Mougan I (1998) Fatty acid composition of human brain phospholipids during normal development. Journal of Neurochemistry 71:2528-2533

Mather K (1941) Variation and selection of polygenic characters. Journal of Genetics 41:159-193

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

Mei DS, Wang HZ, Hu Q, Li YD, Xu YS, Li YC (2009) QTL analysis on plant height and flowering time in *Brassica napus*. Plant Breeding 128:458-465

Meyerowitz EM, Somerville CR (1994) Arabidopsis. Cold Spring Harbor Laboratory Press

Mittelbach M (1996) Diesel fuel derived from vegetable oils, VI: Specifications and quality
control of biodiesel. Bioresource Technology 56:7-11

Morgante M, Hanafey M, Powell W (2002) Microsatellites are preferentially associated with
nonrepetitive DNA in plant genomes. Nature genetics 30:194-200

Mozo T, Dewar K, Dunn P, Ecker JR, Fischer S, Kloska S, Lehrach H, Marra M, Martienssen R,
Meier-Ewert S, Altmann T (1999) A complete BAC-based physical map of the
Arabidopsis thaliana genome. Nature genetics 22:271-275

National Research Council Canada (2013) Analysis of 100 percent biofuel tested in flight by
NRC reveals reduction in emissions. Government of Canada, Ottawa

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

Panjabi 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 9:113-131

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

Pedroche J, Yust MM, Lqari H, Megias C, Girón-Calle J, Alaiz M, Vioque J, Millán F (2007)
Obtaining of *Brassica carinata* protein hydrolysates enriched in bioactive peptides using
immobilized digestive proteases. Food Research International 40:931-938

Perrier X, Flori A, Bonnot F (2003) Data analysis methods. In: Hamon P, Seguin M, Perrier X, Glaszmann JC (eds) Genetic diversity of cultivated tropical plants. Science Publishers, Inc., Montpellier, pp 43-76

Perrier X, Jacquemoud-Collet JP (2006) DARwin software. V.5.0.158 edn

Petrie GA (1988) Races of *Albugo Candida* (white rust and staghead) on cultivated Cruciferae in Saskatchewan. Canadian Journal of Plant Pathology 10:142-150

Prakash S, Chopra VL (1988) Introgression of resistance to shattering in *Brassica napus* from *Brassica juncea* through non-homologous recombination. Plant Breeding 101:167-168

Rae AM, Howell EC, Kearsey MJ (1999) More QTL for flowering time revealed by substitution lines in *Brassica oleracea*. Heredity 83:586-596

Rafalski A (2002) Applications of single nucleotide polymorphisms in crop genetics. Current Opinions in Plant Biology 5:94-100

Ramchiary N, Padmaja K, Sharma S, Gupta V, Sodhi Y, Mukhopadhyay A, Arumugam N, Pental D, Pradhan A (2007) Mapping of yield influencing QTL in *Brassica juncea* implications for breeding of a major oilseed crop of dryland areas. TAG Theoretical and Applied Genetics 115:807-817

Ramos MJ, Fernández CM, Casas A, Rodríguez L, Pérez Á (2009) Influence of fatty acid composition of raw materials on biodiesel properties. Bioresource Technology 100:261-268

Ramsay LD, Jennings DE, Kearsey MJ, Marshall DF, Bohuon EJR, Arthur AE, Lydiate DJ (1996) The construction of a substitution library of recombinant backcross lines in *Brassica oleracea* for the precision mapping of quantitative trait loci. Genome 39:558-567

Rana D, van den Boogaart T, O'Neill CM, Hynes L, Bent E, Macpherson L, Park JY, Lim YP, Bancroft I (2004) Conservation of the microstructure of genome segments in *Brassica napus* and its diploid relatives. The Plant Journal 40:725-733

Rashid A, Rakow G, Downey RK (1994) Development of yellow seeded *Brassica napus* through interspecific crosses. Plant Breeding 112:127-134

Rimmer SR, van den Berg CGJ (1992) Resistance of oilseed *Brassica* spp. to blackleg caused by *Leptosphaeria maculans*. Canadian Journal of Plant Pathology 14:56-66

Ripley VL, Falk KC, Roslinsky V (2006) Genetic diversity of *Brassica carinata* populations selected for improved seed quality. Acta Hort 706:135-140

Roy NN (1978) A study on disease variation in the populations of an interspecific cross of *Brassica juncea* L. x *B. napus* L. Euphytica 27:145-149

Roy NN (1984) Interspecific transfer of *Brassica juncea*-type high blackleg resistance to *Brassica napus*. Euphytica 33:295-303

Salvi S, Tuberosa R (2005) To clone or not to clone plant QTLs: present and future challenges. Trends in Plant Science 10:297-304

Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. Proceedings of the National Academy of Sciences 74:5463-5467

Sargent JR, Coupland K, Wilson R (1994) Nervonic acid and demyelinating disease. Medical Hypotheses 42:237-242

Schneider K, Kulosa D, Soerensen T, Möhring S, Heine M, Durstewitz G, Polley A, Weber E, Jamsari J, Lein J, Hohmann U, Tahiro E, Weisshaar B, Schulz B, Koch G, Jung C, Ganal M (2007) Analysis of DNA polymorphisms in sugar beet (*Beta vulgaris* L.) and

development of an SNP-based map of expressed genes. TAG Theoretical and Applied Genetics 115:601-615

Schreiner M, Beyene B, Krumbein A, Stützel H (2009) Ontogenetic changes of 2-propenyl and 3-indolylmethyl glucosinolates in *Brassica carinata* leaves as affected by water supply. Journal of Agricultural and Food Chemistry 57:7259-7263

Schuelke M (2000) An economic method for the fluorescent labeling of PCR fragments. Nat Biotech 18:233-234

Sebastian RL, Howell EC, King GJ, Marshall DF, Kearsey MJ (2000) An integrated AFLP and RFLP *Brassica oleracea* linkage map from two morphologically distinct doubled-haploid mapping populations. Theoretical and Applied Genetics 100:75-81

Sedun FS, Seguin-Swartz G, Rakow GFW (1989) Genetic variation in reaction to Sclerotinia stem rot in *Brassica* species. Canadian Journal of Plant Science 69:229-232

Shendure J, Ji H (2008) Next-generation DNA sequencing. Nat Biotech 26:1135-1145

Singh D, Chandra N, Gupta PP (1997) Inheritance of powdery mildew resistance in interspecific crosses of Indian and Ethiopian Mustard. Annals of Biology 13:73-77

Sjodin C, Glimelius K (1988) Screening for resistance to Blackleg *Phoma lingam* (Tode ex Fr.) Desm. within *Brassicaceae*. Journal of Phytopathology 123:322-332

Slocum MK, Figdore SS, Kennard WC, Suzuki JY, Osborn TC (1990) Linkage arrangement of restriction fragment length polymorphism loci in *Brassica oleracea*. TAG Theoretical and Applied Genetics 80:57-64

Snowdon RJ, Friedt W (2004) Molecular markers in *Brassica* oilseed breeding: current status and future possibilities. Plant Breeding 123:1-8

Soengas P, Hand P, Vicente J, Pole J, Pink D (2007) Identification of quantitative trait loci for resistance to *Xanthomonas campestris* pv. *campestris* in *Brassica rapa*. TAG Theoretical and Applied Genetics 114:637-645

Sonah H, Deshmukh RK, Sharma A, Singh VP, Gupta DK, Gacche RN, Rana JC, Singh NK, Sharma TR (2011) Genome-wide distribution and organization of microsatellites in plants: An insight into marker development in *Brachypodium*. PLoS ONE 6:e21298

Song K, Suzuki J, Slocum M, Williams P, Osborn T (1991) A linkage map of *Brassica rapa* (syn *campestris*) based on restriction fragment length polymorphism loci. TAG Theoretical and Applied Genetics 82:296-304

Taylor DC, Falk KC, Palmer CD, Hammerlindl J, Babic V, Mietkiewska E, Jadhav A, Marillia E-F, Francis T, Hoffman T, Giblin EM, Katavic V, Keller WA (2010) *Brassica carinata* − a new molecular farming platform for delivering bio-industrial oil feedstocks: case studies of genetic modifications to improve very long-chain fatty acid and oil content in seeds. Biofuels, Bioproducts and Biorefining 4:538-561

Taylor JD, Conway J, Roberts SJ, Astley D, Vicente JG (2002) Sources and origin of resistance to *Xanthomonas campestris pv. campestris* in *Brassica* Genomes. Phytopathology 92:105-111

Teutonico RA, Osborn T (1994) Mapping of RFLP and qualitative trait loci in *Brassica rapa* and comparison to the linkage maps of *B. napus*, *B. oleracea*, and *Arabidopsis thaliana*. Theoretical and Applied Genetics 89:885-894

Theologis A, Ecker JR, Palm CJ, Federspiel NA, Kaul S, White O, Alonso J, Altafi H, Araujo R, Bowman CL, Brooks SY, Buehler E, Chan A, Chao Q, Chen H, Cheuk RF, Chin CW, Chung MK, Conn L, Conway AB, Conway AR, Creasy TH, Dewar K, Dunn P, Etgu P,

Feldblyum TV, Feng J, Fong B, Fujii CY, Gill JE, Goldsmith AD, Haas B, Hansen NF, Hughes B, Huizar L, Hunter JL, Jenkins J, Johnson-Hopson C, Khan S, Khaykin E, Kim CJ, Koo HL, Kremenetskaia I, Kurtz DB, Kwan A, Lam B, Langin-Hooper S, Lee A, Lee JM, Lenz CA, Li JH, Li Y, Lin X, Liu SX, Liu ZA, Luros JS, Maiti R, Marziali A, Militscher J, Miranda M, Nguyen M, Nierman WC, Osborne BI, Pai G, Peterson J, Pham PK, Rizzo M, Rooney T, Rowley D, Sakano H, Salzberg SL, Schwartz JR, Shinn P, Southwick AM, Sun H, Tallon LJ, Tambunga G, Toriumi MJ, Town CD, Utterback T, Van Aken S, Vaysberg M, Vysotskaia VS, Walker M, Wu D, Yu G, Fraser CM, Venter JC, Davis RW (2000) Sequence and analysis of chromosome 1 of the plant *Arabidopsis thaliana*. Nature 408:816-820

Thomasson HJ, Boldingh J (1955) The biological value of oils and fats. II. The growth-retarding substance of rapeseed oil. J Nutr 56:469-475

Tonguç M, Griffiths PD (2004) Transfer of powdery mildew resistance from *Brassica carinata* to *Brassica oleracea* through embryo rescue. Plant Breeding 123:587-589

Trick M, Long Y, Meng J, Bancroft I (2009) Single nucleotide polymorphism (SNP) discovery in the polyploid *Brassica napus* using Solexa transcriptome sequencing. Plant Biotechnology Journal 7:334-346

Truco MJ, Quiros CF (1994) Structure and organization of the B genome based on a linkage map in *Brassica nigra*. TAG Theoretical and Applied Genetics 89:590-598

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

Vergnaud G, Denoeud F (2000) Minisatellites: Mutability and genome architecture. Genome research 10:899-907

Vicente JG, Taylor JD, Sharpe AG, Parkin IAP, Lydiate DJ, King GJ (2002) Inheritance of race-specific resistance to *Xanthomonas campestris* pv. campestris in *Brassica* Genomes. Phytopathology 92:1134-1141

Vignal A, Milan D, SanCristobal M, Eggen A (2002) A review on SNP and other types of molecular markers and their use in animal genetics. Genetics Selection Evolution 34:275 - 305

Voorrips RE (2002) MapChart: Software for the graphical presentation of linkage maps and QTLs. Journal of Heredity 93:77-78

Wang S, Basten CJ, Zeng ZB (2012) Windows QTL Cartographer 2.5. Department of Statistics, North Carolina State University, Raleigh, NC.

Warwick S, Francis A, Al-Shehbaz I (2006a) Brassicaceae: Species checklist and database on CD-Rom. Plant Systematics and Evolution 259:249-258

Warwick S, Gugel R, McDonald T, Falk K (2006b) Genetic variation of Ethiopian Mustard (*Brassica carinata* A. Braun) germplasm in Western Canada. Genetic Resources and Crop Evolution 53:297-312

Watson JD, Crick FHC (1953) A structure for deoxyribose nucleic acid. Nature 171

Wenz H, Robertson JM, Menchen S, Oaks F, Demorest DM, Scheibler D, Rosenblum BB, Wike C, Gilbert DA, Efcavitch JW (1998) High-precision genotyping by denaturing capillary electrophoresis. Genome research 8:69-80

Woods DL (1980) Association of yellow seed coat colour with other characteristics in mustard (*Brassica juncea*). Crucifer Newsletter 5:23-24

Yamashiro S, Clandinin MT (1980) Myocardial ultrastructure of rats fed high and low erucic acid rapeseed oils. Experimental and Molecular Pathology 33:55-64

Yang B, Rahman M, Liang Y, Shah S, Kav N (2010) Characterization of defense signaling

pathways of *Brassica napus* and *Brassica carinata* in response to *Sclerotinia*

sclerotiorum challenge. Plant Molecular Biology Reporter 28:253-263

Yitbarek S (1992) Pathological research on noug, linseed, gomenzer and rapeseed in Ethiopia.

Oilseeds Research and Development in Ethiopia, Proceedings of the First National

Oilseeds Workshop Abeba, Addis, Ethiopia, pp 151-161

Zukalová H, Vasak J (2002) The role and effects of glucosinolates of *Brassica* species - a

review. Rostlinna Vyroba 48:175-180