Heterotic Pool Development in *Brassica napus* L.

By

Corey James Lees

A Thesis Submitted to the Faculty of Graduate Studies

In Partial Fulfillment of Requirements for the Degree of

Masters of Science

(M.Sc)

Department of Plant Science

University of Manitoba

Winnipeg, Manitoba

ACKNOWLEDGEMENTS

I would like to take this time to acknowledge all the people in my life that made this dream of higher education a reality:

First and foremost I would like to thank my loving wife Joanna Lynn Lees for all the years of support through this endeavour. Without you I never would have been able to materialize this dream, and for that, I will be forever grateful.

To my supervisor, Dr. Robert W. Duncan, thank you for all the time, coaching, guidance and support you have given throughout these years. You have been an incredible asset in my development and words really can't express my gratitude for all the work that has gone into my project.

To all the faculty members, support staff and technicians in the Department of Plant Science, so many people directly contributed to my success I would like to thank the entire department for all their roles big and small.

Thank-you to Bunge Canada, DL Seeds, NSERC Canada and the University of Manitoba for the financial support of this project, without continued research support my training would not have been possible.

To my committee members, Dr. Dilantha Fernando, Dr. Anne Worley and Dr. Bert Innes, thank-you for all your time and feedback on my project throughout these two years.

Last but not least, thank-you to all my friends and family for the continued encouragement, support and advice that has been so critical to my success over these two years.

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVEATIONS

ATP, adenosine triphosphate

CoA, coenzyme A

CMS, cytoplasmic male sterility

ddH_2O, double distilled water

dNTP, deoxynucloetide triphosphate

EU, European Union

FAE, fatty acid elongase

GBS, genotyping-by-sequencing

GMS, genic male sterility

INRA, Institut national de la recherche agronomique

KCS, beta-ketoacyl-coenzyme A (CoA) synthase

LPAAT, lysophosphatidic acid acyltransferase

Mt, million tons

ORF, open reading frame

PCA, principal component analysis

PCR, polymerase chain reaction

Rfo, Restorer gene

SN-2, Stereospecific number 2

SNP, Single nucleotide polymorphism

SRAP, Sequence related amplified polymorphisms

TAG, Triacylglyceride

ABSTRACT

Lees, Corey James. M.Sc. University of Manitoba, June 2015. Heterotic Pool Development in *Brassica napus* L. Major Professor: Dr. Robert W. Duncan, Department of Plant Science.

Hybrid development has become an essential component in many major crop species due to the exploitation of heterosis. However, parental combinations that exhibit high heterotic gains can be difficult to obtain if no prior genetic architecture is known. The current research focuses on heterotic pool development using genetic distance and multivariate cluster analysis. Seventy-nine inbred accessions were grouped using three different methods including; 1) phenotypic classification based on 20 qualitative and quantitative traits; 2) Sequence related amplified polymorphisms (SRAP) using 29 forward and reverse primer combinations; 3) genotyping-by-sequencing (GBS) using 80,005 single nucleotide polymorphisms. Both genotypic methods (SRAP and GBS) were compared to each other, and hierarchical clustering produced similar results with 68% homology between the two methods. Heterotic cluster accuracy was investigated between the two genotypic methods through hybrid performance. Forty-four hybrid combinations along with parental cultivars were grown at one location in Winnipeg, MB with three replicates in a randomized complete block design. Hybrid yield and heterosis were regressed against the genetic distance from each genotypic method. GBS was a better predictor of hybrids with high yield ($R = 0.47$, $p < .01$); however, SRAP was a better predictor for mid-parent and high-parent heterosis at $R = 0.53$ and $R = 0.61$, respectively. This research provides experimental evidence that SRAP and GBS heterotic pool definitions have utility in the prediction of high heterotic parental combinations.

TABLE OF CONTENTS

1.0 GENERAL INTRODUCTION

Canola/rapeseed (*Brassica napus* L.) is one of the most important oilseed crops around the world (Carré and Pouzet, 2014; Wittkop *et al.*, 2009; Oil World 2014; Lin *et al.*, 2013). During the last twenty years, *B. napus* production has increased by a factor of 2.4X worldwide (Carré and Pouzet, 2014). As of 2014 in Canada, this represents a 19.3 billion dollar crop planted on approximately 8 million hectares annually (Canola Council of Canada, 2015). This intensification has created a competitive market for the development of new genotypes and hybrid cultivars with specific oil profiles and end uses (Rahman, 2013). Additionally, the current breeding standard in *B. napus* includes the development of hybrid cultivars that can show significant heterosis compared to the inbred parental lines (Rahman, 2013; Xing *et al.*, 2014). However, significant time and resources are needed for the development of commercially viable hybrid cultivars that express high heterotic gains (Rahman, 2013; Xing *et al.*, 2014).

To alleviate some of this pressure, heterotic pool development has been suggested as a method to separate breeding material into distinct pools with the end goal of reducing the amount time, space and money required to find high-heterotic parental combinations (Rahman, 2013; Girke *et al*., 2012). Heterotic pool definition can be based on genetic distance and multivariate cluster analysis (Girke *et al*., 2012; Yu *et al*., 2005; Jesske *et al*., 2013). Once these definitions are complete, inbred parental lines from different pools or clusters can be crossed and evaluated for heterotic gain. Previous reports suggest that inter-cluster hybrids (crosses from different clusters or pools) exhibit higher heterosis than crosses made from intra-cluster hybrids (crosses from the same cluster or pool) (Riaz *et al*., 2001; Ahmad and Quiros, 2011). However, a unified theory that explains the driving mechanism behind the observed effects of heterosis has not been attained (Kaepple, 2012; Rahman, 2013; Li *et al*., 2015).

Following the hypothesis of Falconer and McKay (1996), genetic distance has been implicated as a contributor to heterotic gains. Genetic distance can be calculated using a range of morphological datasets (Ali *et al*., 1995) or using polymorphisms based on DNA markers and sequence data (Diers *et al*., 1996; Riaz *et al*., 2001; Yu *et al*., 2005). Genetic distance itself has a variety of mathematical formulas that infer different hypotheses about the mutation rates in a population over a given timespan (Nei, 1972; Cavalli-Sforza and Edwards, 1967; Reynolds *et al*., 1983; Tamura and Nei, 1993). In conjunction with different genetic distance methods, multivariate cluster analysis also contains a wide variety of clustering theorems that address different assumptions on how specific data should be clustered (UPGMA, Sokal and Michener, 1958), (neighbour joining, Saitou and Nei, 1987), (Ward's method, Ward, 1963). Combining all of these hierarchical levels allows researchers to devise a plethora of experimental designs that utilize different distance methods with different clustering techniques in an attempt to aid in the prediction of high heterotic hybrids.

Hybrid heterosis was first observed in maize (*Zea maize* L.) over a century ago (Shull, 1908; East, 1908; Melchinger, 2010). This phenomenon, which increases the fitness and vigour of progeny, has been the target of intense scientific investigation (Thiemann *et al*., 2009; Kaeppler, 2012). Therefore, current research is still heavily invested in the dissection and utilization of hybrid heterosis in agricultural crop breeding (Kaepple, 2012; Rahman, 2013). Specifically in *B. napus*, hybrid heterosis was first investigated by Grant and Beversdorf (1985) and was followed by a large research cascade during the development of *B. napus* hybrids (Diers *et al.*, 1996; Riaz *et al.*, 2001; Radoev *et al.*, 2008; Cuthbert *et al.*, 2009; Ahman and Quiros, 2011). *Brassica napus* hybrids have been shown to outperform their mid-parent seed yield values by up to 169%

(Riaz *et al*., 2001). Therefore, there is tremendous interest in harnessing hybrid heterosis for the development of commercially viable *B. napus* hybrid cultivars (Rahman, 2013).

The objectives of this research were:

1) Phenotype seventy-nine germplasm accessions based on twenty qualitative and quantitative traits, calculate Euclidian genetic distance and apply Ward's clustering technique to develop heterotic pools for the development of hybrid cultivars.
2) Genotype seventy-nine germplasm accessions using sequence related amplified polymorphisms and genotyping-by-sequencing, calculate genetic distance using Nei's standard genetic distance and the Tamura-Nei distance model respectively, cluster the results using the neighbour joining method and finally compare the heterotic pool assignments.
3) Calculate hybrid heterosis from forty-four hybrid combinations based on the two genotypic genetic distance methods to conclude if genetic distance influences heterosis and determined the efficacy of each method as a predictive model for hybrid heterosis. Lastly, combine those models to investigate if a consensus method can improve heterotic parental prediction.

2.0 LITERATURE REVIEW

2.1 *Brassicaceae* Family

The family of *Brassicaceae* (Cruciferae) is one of the most widely cultivated plant families, supplying a wide range of nutrients and phytochemicals for human nutritional requirements (Ghani *et al.*, 2013). The *Brassicaceae*, or mustard family, is comprised of about 338 genera and some 3709 species distributed across a wide variety of growing regions (Al-Shehbaz *et al.*, 2006). Some important domesticated crops within the *Brassicaceae* are cabbage, broccoli, cauliflower, turnip, rapeseed, mustard, radish, horseradish, cress, wasabi, and watercress (Rask *et al*., 2000). However, the *Brassicaceae* family is most known for *Aradopsis thaliana* L., the model organism for flowering plants (Al-Shehbaz *et al.*, 2006).

2.11 *Brassica* Genus

Within the family *Brassicaceae* is the genus *Brassica*, which contains several globally important cultivated crops such as *Brassica carinata* A., *Brassica juncea* L., *Brassica oleracea* L., *Brassica napus* L., *Brassica nigra* L. and *Brassica rapa* L. (Østergaard and King, 2008). Several combinations of these species can be cross pollinated to produce fertile hybrids suggesting common ancestry (Østergaard and King, 2008). This ancestral relationship was first theorised by U (1935) using the triangle of U which depicts how the three ancestral diploid species of *B. rapa, B. olerecea* and *B.nigra* gave rise to the three modern allotetraploid species of *B. napus, B. carinata* and *B. juncea* (Figure 2.1). Globally, *B. napus* is cultivated on the most hectares, and its global economic impact cannot be understated.

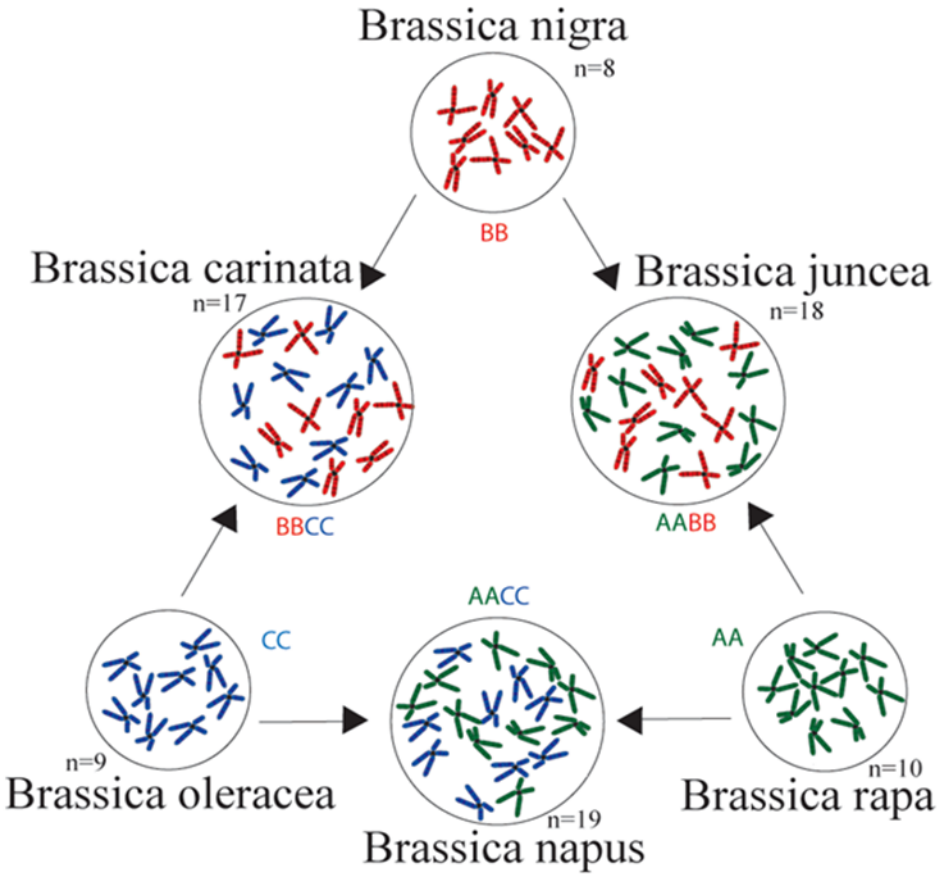

Figure 2.1: The triangle of U depicting the genetic relationship between Brassica species (U, 1935). Between the six species, three genomes are shared (A, B and C) and arrows indicate the direction of genome transfer. Image obtained from: http://en.wikipedia.org/wiki/Triangle_of_U

2.12 Evolutionary Theory of *Brassica napus*

Brassica napus is a close relative of *Aradopsis thaliana* and is theorized to have evolved through a spontaneous hybridization event between its two progenitor species, *B. rapa* (AA N=10) and *B. oleracea* (CC N=9) some 10,000 years ago within the Mediterranean region (Diers and Osborn, 1994; Cheung *et al.*, 2009; Chalhoub *et al.*, 2014). *Brassica rapa* and *B. oleracea* are also very close in relation to *A. thaliana* and have extensively triplicated genomes, a remnant of a common hexaploid ancestor (Harper *et al.*, 2012; Town *et al.*, 2006). Therefore, every gene within *A.*

thaliana is expected to have multiple copies (orthologs) within *B. napus* (Harper *et al.*, 2012; Chalhoub *et al.*, 2014). Additionally, the A and C genomes of *B. napus* are considered to be homeologous which constitutes genes in the A genome having a similar constituent in the C genome (Harper *et al.*, 2012; Chalhoub *et al.*, 2014). Unfortunately, no defining time or location can be specified for the evolution of *B. napus* due to the lack of wild *B. napus* landraces (Allender and King, 2010). However, genomic and cytological analysis has confirmed that *B. napus* is the product of a recent *B. rapa x B. oleracea* hybridization event (Snowdon *et al.*, 2002; Cheung *et al.*, 2009; Chalhoub *et al.*, 2014). This makes *B. napus* an allotetraploid (AACC N=19) that may have evolved from multiple hybridization events (Allender and King, 2010; Jesske *et al.*, 2013). Interestingly, *B. rapa* and *B. oleracea* can be artificially crossed to create fertile resynthesized lines that add genetic diversity to the *B. napus* gene pool (Allender and King, 2010; Jesske *et al.*, 2013).

2.2 *Brassica napus* Breeding History in Canada

Originally, oilseed rape was grown for its erucic acid and used as a slip agent for steam ships (McVetty and Scarth, 2012). During the Second World War, demand for plant oils rich in erucic acid drastically increased for use in warships and foodstuffs (McVetty *et al.*, 2009). To alleviate this demand, Argentine rape was imported to Canada from the United States and production grew steadily with the Canadian government supporting price guarantees for its supply (McVetty *et al.*, 2009). In the years following the war (1945-60), prices dropped and health concerns about erucic acid as an edible oil resulted in a drastic decline in oilseed rape production (McVetty *et al.*, 2009; Stefansson and Hougen, 1964). However, breeding programs for rapeseed were established at the Agriculture Canada Research Station, Saskatoon in 1944 and the University of Manitoba in 1952 and concentrated on improving agronomic traits including oil concentration

and lodging resistance (Steffansson, 1983). From these breeding programs, discoveries were made into the natural variation of fatty acid profiles that existed in *B. napus*. Through classical breeding techniques and gas chromatography, a rare genotype with no erucic acid in its oil profile was discovered (Stefansson *et al.*, 1961). It was determined that 2 genes controlled the accumulation of erucic acid and they acted in an additive manner (Stefansson *et al.*, 1961). This resulted in the first breeding material low in erucic acid, which was derived from the German cultivar Liho (Snowdon *et al.*, 2007; Stefansson *et al.*, 1961, Stefansson and Hougen, 1964). This renewed rapeseed production in the Canadian prairies although the crops usefulness was still hindered by the presence of high glucosinolates, an anti-nutritional compound that resides in the protein meal after oil extraction. In addition to Liho, another monumental breakthrough was discovered in 1969 in the polish rape cultivar Bronowski that was low in glucosinolate content (Krzymanski and Downey, 1969; Kondra and Stefansson, 1970). This trait was combined with low erucic acid material from Liho to give rise to the first double low (00) canola quality cultivar Tower in 1974 (Stefansson and Kondra, 1975; Snowdon *et al.*, 2007).

2.21 Current *Brassica napus* production

In recent years, oilseed rape (*B. napus*) has been heralded as one of the most important oilseed crops around the world, second only to soybean (*glycine max* L.) in oilseed production and 3rd for total global oil production behind palm oil and soybean oil (Carré and Pouzet, 2014; Wittkop *et al.*, 2009, Oil World, 2014, Lin *et al.*, 2013). As of 2014, global rapeseed production is estimated to be 63.99 million metric tons of winter and spring oilseed rape with Canada supplying a projected 16.5 million metric tons to the global oilseed market (Oil World, 2014). Rapeseed has steadily grown in popularity worldwide due to its temperate growing requirement,

winter and spring habit, high oil content (~ 40-45%), high crude protein content (~ 35%), and continued increases in yield (Bell, 1993; Carré and Pouzet, 2014).

2.3 *Brassica napus* Oil Profiles

2.31 Canola

The term Canola (a contraction between Canada and ola for oil), which was adopted in 1979, is now used to describe the double low erucic acid/glucosinolate trait (Canola Council of Canada, 2015). Canola quality plants must include trace amounts of erucic acid (less than 2%) in the extracted oil and less than 30 µmol/g of the four known aliphatic glucosinolates (gluconapin, progoitrin, glucobrassicanapin and napoleiferin) in the residual protein meal to retain their classification (Shahidi, 1990). Erucic acid in high doses is now known to cause myocardial lesions in rats; however, its effects on human health are unresolved and due to this stigma consumption of high amounts of erucic acid is generally avoided (Bellenand *et al.*, 1980; Sherazi *et al.*, 2013). Conversely, aliphatic glucosinolates are known to breakdown into toxic by-products in non-ruminant animals and can induce iodine deficiency, kidney and thyroid malformation and hypertrophy of the liver (Tripathi and Mishra, 2007; Rask *et al*., 2000). Therefore, with a reduction of glucosinolate content, rapeseed extracted meal (REM) can safely be used in moderation for feed in monogastric livestock animals (Auldist *et al.*, 2014; Khajali and Slominski, 2012; Quiniou *et al.*, 2012). The residual protein meal after oil extraction has an excellent balance of amino acids including a high content of sulphuric amino acids methionine and cysteine (Liu *et al.*, 2012; Downey and Bell, 1990). Although, due to the presence of aliphatic glucosinolates and other antinutritional compounds, canola meal is still ranked second behind soybean meal for total protein usage (Wittkop *et al.*, 2009).

Canola quality oil provides healthy edible oil for human consumption. Canola oil contains 0% trans fats, 7% saturated fat, 32% polyunsaturated and 61% monounsaturated fat giving it a variety of uses from cooking and frying oil to salad dressings (Lin *et al.*, 2013; Scarth and McVetty, 1999). However, *B. napus* is not limited to edible oil, commercial varieties are available that contribute specialty and industrial oil profiles to niche oleochemical markets.

2.32 High Oleic Acid Canola

High oleic acid (18:1) canola oil is specialty oil specifically designed for the commercial food industry (Scarth and McVetty, 1999). Akin to canola oil, high oleic canola oil contains the same level of trans fats and saturated fats at 0% and 7%, respectively; however, high oleic oil has a reduced ratio of polyunsaturated fats to monounsaturated fats (DeBonte *et al.*, 2012). Polyunsaturated fats are highly susceptible to oxidation and reduce the shelf-life of consumable foodstuffs, making these fats undesirable for the commercial food industry (Scarth and McVetty, 1999). As of 2012, Cargill, Bayer CropScience and Dow AgroSciences all offer high oleic canola varieties with differing amounts of oleic acid which covered 2 million acres and represented ~10% of the Canadian canola harvest (DeBonte *et al.*, 2012). Cargill's Clear Valley® 75 shows a distinct oil profile of 75% oleic, 10% linoleic and 5% linolenic acid (Scarth and McVetty, 1999) which differs slightly from Dow's Natreon™ at 73% Oleic, 15.7% linoleic and 2.1% linolenic acid (Matthäus, 2006). Currently, both companies market under new designations of Victory and Nexera respectively (Brewin and Malla, 2012). Due to its usefulness in food preparation, its oxidative stability and its lack of trans fats, demand for high oleic oil will only increase in the coming years (DeBonte *et al.*, 2012). Although high oleic canola oil dominates the oleochemical market (>90%), new varieties of high oleic sunflower (*Helianthus*

annuus L.) and soybean (*G. max*) are in production and will certainly impact future market growth (DeBonte *et al.*, 2012).

2.33 High Erucic Acid Rapeseed

High erucic acid (22:1) rapeseed (HEAR) is typically grown for industrial oil end uses including erucamide, paints and emulsifiers, cosmetics, biodegradable lubricants and slip agents (Zanetti *et al.*, 2009; Sasongko and Möllers, 2005; Wittkop *et al.*, 2009; Fourmann *et al*., 1998; Sherazi *et al.*, 2013). As the world population is expected to grow over the coming years and petroleum stocks are in decline, demand for this renewable oil will only increase (Sherazi *et al.*, 2013). Therefore it is critical for *Brassica* breeding programs to continue to improve cultivars for erucic acid content, agronomic performance and other quality traits. However, simultaneous improvements for the multiple traits listed above can be very challenging. Currently, erucic acid production in HEAR is over 50% of the total oil content (Nath *et al.*, 2009; McVetty *et al.*, 2014). Therefore a significant increase in erucic acid content would greatly increase erucic acid yield and improve processing efficiencies. However, conventional HEAR is limited to a maximum erucic acid content of 66-67% in its oil profile due to the inability to esterify erucic acid in the stereospecific number 2 (*sn*-2) position of the triacylglyceride (TAG) head (Fourmann *et al.*, 1998; Sasongko and Möllers, 2005). The reason for this inability lies in the specificity of the *B. napus sn*-2 acyltransferase (LPAAT—lysophosphatidic acid acyltransferase), which does not accept erucoyl-CoA as a substrate. However, Nath *et al*. (2009) was able to increase erucic acid content in seeds up to 72% using transgenic techniques. Nath *et al.* (2009) created transgenic lines that incorporated the lysophosphatidic acid acyltransferase gene from *Limnanthes douglasii* (*Ld-LPAAT*), which enables the insertion of erucic acid into the *sn*-2 triglyceride position. The progress presented by Nath *et al*. (2009) represented a major

breakthrough in rapeseed breeding and the genetic engineering of specific oil profiles. Unfortunately, there is controversy with the use of transgenic approaches in crop production and the regulatory processes are simply cost prohibitive for traits of limited use. Therefore, classical breeding for many traits still has an advantage over transgenic approaches.

2.34 Erucic Acid Production

Classical genetic studies in *B. napus* have shown that the accumulation of erucic acid is controlled by 2 genes, *Bn.FAE1.1* and *Bn.FAE1.2* which act in an additive effect and were mapped to the E1 and E2 loci (Harvey and Downey, 1964; Fourmann, *et al.,* 1998; Nath *et al.*, 2009). The Fatty Acid Elongase1 (FAE1) gene creates a protein product called β-keto-acyl-CoA synthase (KCS), an enzyme responsible for a key step in erucic acid production (Fourmann *et al.*, 1998; Rahman *et al.*, 2008). Erucic acid is ultimately pulled from a reserve pool of mono-unsaturated oleic fatty acids (18:1) within the cytosol. These acyl precursors are esterified to Coenzyme A (CoA) through the action of acyl-CoA synthetase and are recruited by the fatty acid elongase (FAE) complex. This theorized four-step process involves KCS in the first rate-determining step binding with and condensing oleoyl-CoA and malonyl-CoA in the cytosol and catalyzing the reduction to β-hydroxy-CoA and the dehydration to a trans 2-3-enoyl-CoA (Fourmann *et al.,* 1998; Rahman *et al.*, 2008). Lastly, the final reduction step leads to an acyl-CoA elongated by 2 carbons, this sequence happens again at the eicosenoyl-CoA (20:1) length to erucoyl–CoA (22:1 Δ13-*cis*), which becomes available for triacylglycerol insertion (Fourmann *et al.*, 1998, Sasongko and Möllers, 2005; Rahman *et al.*, 2008).

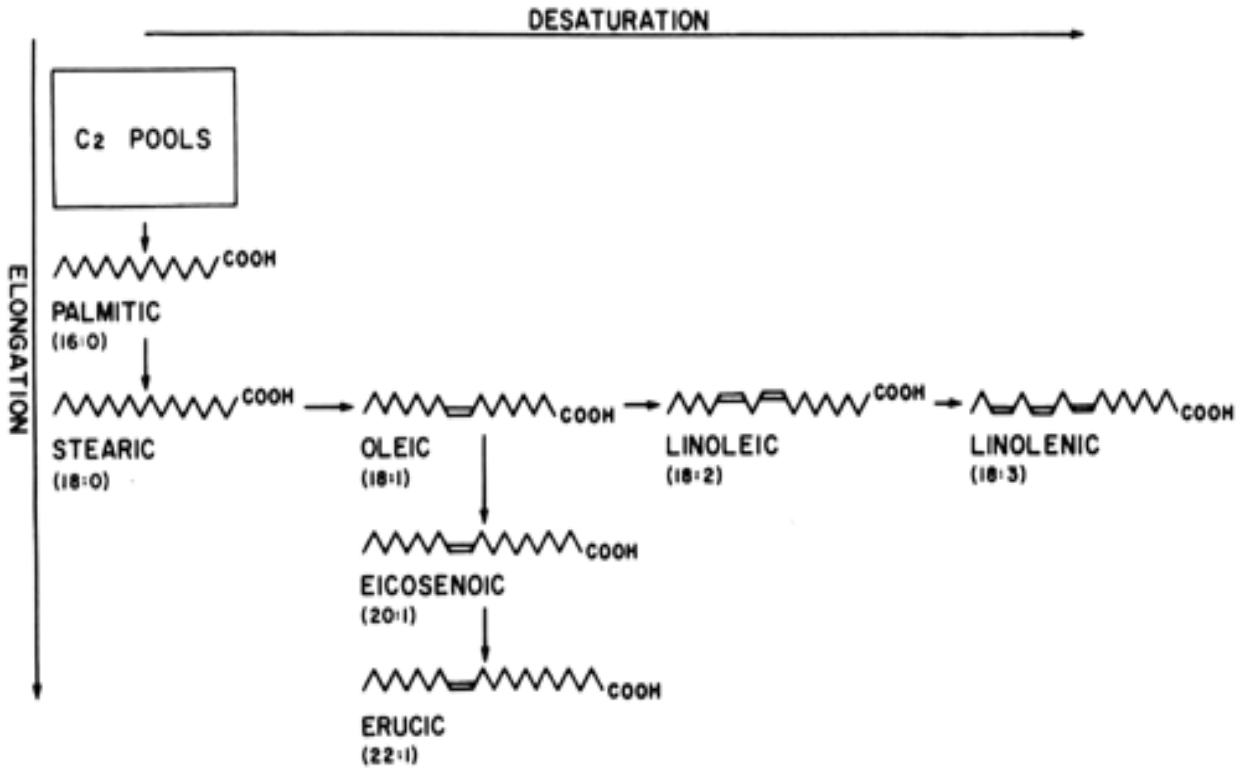

Figure 2.2: Sequential fatty acid elongation steps. Elongation (the addition of two carbon atoms) proceeds downwards whereas desaturation (the removal of a hydrogen atom and the creation of a double carbon-carbon bond) increases to the right. Image obtained from: https://www.hort.purdue.edu/newcrop/proceedings1990/v1-211.html.

Understanding how long chains fatty acids are synthesized and the enzymes involved is an important step to maximising the potential of *B. napus* in the renewable oleochemical world. Since 18:1 oleic acid is the acyl precursor for erucic acid production, one would expect that increasing the 18:1 acyl pool would result in higher erucic acid cultivars. Sasongko and Möllers (2005) investigated this assumption by crossing high erucic acid cultivars with high oleic cultivars in the hope of increasing the 18:1 acyl pool for higher erucic acid production. Unfortunately, Sasongko and Möllers (2005) found no significant difference in erucic acid production with elevated 18:1 acyl pools into the F_3 generation. Therefore it is evident that erucic

acid production and accumulation is governed by more than just having higher levels of precursors available and the exact regulatory mechanisms controlling total erucic acid content are still unclear (Hua *et al.*, 2012).

2.35 Improving Oil Traits

Without using transgenic approaches, breeders are forced to utilize the natural variation within the *B. napus* population for improved oil traits. One avenue of research which is gaining momentum for oil production improvement is determining how the maternal genotype affects oil production and also how the silique itself influences long chain fatty acid development and storage within the embryo (Hua *et al.*, 2012). Naturally, seeds are dependent on the mother plant for their nutrient supply during development and several studies have shown that the maternal genotype plays a critical role in determining seed oil content (Wu *et al.*, 2006; Wang *et al.*, 2010; Hua *et al.*, 2012). Hua *et al.* (2012) specifically investigated this maternal influence by reciprocal crosses with high and low oil producing *B. napus* cultivars. When a high oil cultivar was the maternal genotype and crossed with a low oil paternal genotype, the resulting F_1 was only 2% lower for total oil content compared with the same self-pollinated maternal F_1 seeds. The same result was seen in the low oil maternal cultivar crossed with a high oil paternal genotype, its F_1 seed was only 2% higher when compared with the same self-pollinated low oil maternal genotype (Hua *et al.*, 2012). Continuing this experiment, Hua *et al.* (2012) analysed silique tissue from each high and low oil content maternal cultivar. Providing genetic and physiological evidence, Hua *et al.* (2012) determined that local photosynthesis within each silique wall contributes critical energy needed for oil accumulation during seed growth and development. These results are consistent with Li *et al.*, (2006) who showed that seed grown in high light environments tend to produce higher oil content.

2.4 Disease and Resistance in *Brassica napus*

Like many other crops throughout the world, disease resistance to multiple pathogens is critical for high levels of *B. napus* production. Blackleg, one of the most important diseases of canola/oilseed rape is caused by the fungus *Leptosphaeria maculans* (Desm.) (Raman *et al*., 2012; Van de Wouw *et al*., 2014). This fungal pathogen resides in infected canola/rapeseed stubble and releases wind-borne ascospores as inoculum (Van de Wouw *et al*., 2014). Blackleg progressively damages the crop by girdling stems and restricting moisture and nutrient uptake, leading to yield loss. One major disease management practice has been to deploy disease resistant cultivars as *L. maculans* has a gene-for-gene interaction with *B. napus* (Raman *et al*., 2012; Van de Wouw *et al*., 2014). Each avirulence allele in the fungus renders the pathogen unable to infect *B. napus* cultivars with the corresponding resistance gene (Van de Wouw *et al*., 2014; Marcroft *et al*., 2012). However, single gene resistance is problematic as *L. maculans* has a high propensity to overcome single gene resistance as seen in France between 1996 and 1999 and more dramatically in Australia during 2000-03 which led to >90% yield losses (Raman *et al*., 2012; Van de Wouw *et al*., 2014; Marcroft *et al*., 2012). Currently, at least 17 resistance genes (*Rlm1-Rlm11, LepR1-LepR4, BLMR1* and *BLMR2*) have been identified and multi-resistant lines are now available which "stack" different compliments of resistance genes to deter infection (Van de Wouw *et al*., 2014; Marcroft *et al*., 2012).

Another economically important disease in *B. napus* is Sclerotinia stem rot caused by the fungal pathogen *Sclerotinia sclerotiorum* (Lib.) de Bary (Li *et al*., 2015). At best, the genetic characterization of host/pathogen interaction is poorly understood (Garg *et al*., 2013). This is most likely due to the multi-factorial defence responses that can occur in response to infection by this pathogen (Garg *et al*., 2013). Sclerotinia is a soil-borne pathogen and can remain viable in

the soil for several years (Taylor *et al*., 2015). When conditions are conducive for its growth, *S. sclerotiorum* can produce apothecia that release airborne ascospores (Taylor *et al*., 2015). Sclerotinia infects flowers and senescent leaves as it gradually moves into the stem eventually causing girdling of the stem and resulting in a reduction of yield and or plant death. To this day, fully resistant cultivars are unavailable (Li *et al*., 2015). However, with the recent release of the *B. napus* genome (Chalhoub *et al*., 2014), previous research which has characterized quantitative trait loci (QTL) for sclerotinia resistance can now be anchored to a physical map that will allow for putative gene identification (Li *et al*., 2015). This may hold promise for the development of resistance genes and their integration into new, resistant commercial cultivars.

Recently in Canada, the discovery of clubroot in Alberta in 2003 and in Saskatchewan in 2010 has caused alarm as the economic impact of this disease can be severe (Rahman *et al*., 2014; Hwang *et al*., 2014). Clubroot is caused by the soil borne obligate parasite *Plasmodiophora brassicae* Woronin, which systematically infects root tissue causing galls (Hwang *et al.*, 2012; Hwang *et al*., 2014). These galls disrupt nutrient and water uptake leading to stunting, wilting and premature senescence (Hwang *et al.*, 2012). *Plasmodiophora brassicea* can be encountered in over 60 countries around the world and all 330 genera and 3700 plant species in the family *Brassicaceae* are potential hosts for this pathogen (Hwang *et al.*, 2012). However, disease resistance to pathotype 3 of *P. brassicae* has been discovered in the winter canola cultivar 'Mendel' and is attributed to a single dominant gene (Rahman *et al*., 2014). Akin to blackleg resistance, single gene resistance can be eroded over time and therefore a long-term strategy is needed for the careful management of this soil-borne pathogen (Rahman *et al*., 2014; Hwang *et al*., 2014). Due to these diseases and their potential impact on *B. napus* production, breeding

efforts have substantially increased to predict and protect future canola cultivars from infection and yield loss (Vincourt, 2014; Barbetti *et al*., 2015).

2.5 Current and Future Breeding Objectives

Currently, the modification of the *B. napus* oil profiles and the concurrent reduction of antinutritional compounds within rapeseed meal will continue to be a top priority for breeders around the world. Future conventional breeding should concentrate on pod shatter resistance, as Peng *et al*. (2015) describes this trait to be a major objective within rapeseed breeding programs around the world. Drought tolerance is also a highly desired trait yet current research lacks appropriate selection criteria to develop fully tolerant lines (Moradi *et al*., 2015). Therefore, impending research is continually concentrating on selection criteria and its influence in drought trait development (Moradi *et al*., 2015). Disease resistance will also remain a forefront problem as selective pressure forces pathogens to adapt to the corresponding resistance genes. The development of novel resistances genes and the careful management of "stacked" resistance genes will continue to be a future breeding mandate (Vincourt, 2014).

2.6 *Brassica napus* Hybrid Breeding

Recently, many studies have been published detailing resynthesized *B. napus* lines and their incorporation into rapeseed breeding programs to combat low genetic diversity (Girke *et al.*, 2012; Jesske *et al.*, 2013). Due to its recent evolutionary history, the genetic diversity of *B. napus* is generally considered to be limited (Hasan *et al.*, 2006; Seyis *et al.*, 2003; Girke *et al.*, 2012). This low and limited diversity of *B. napus* has also been eroded by intense artificial selection by breeders and geneticists for uniformity in several important agronomic traits (Hasan *et al.*, 2006; Bus *et al.*, 2011). Bus *et al.* (2011) specifically found that the breeding of double

low (00) varieties of spring and winter rape has lowered the genetic diversity of modern cultivars compared to those released only a few decades ago. Therefore, genetic resource management should be a top priority in any breeding program to prevent future loss of diversity. Low genetic diversity may be problematic in future breeding programs that wish to exploit heterosis in *B. napus* hybrids (Becker *et al.,* 1995).

2.61 Heterosis

Heterosis or hybrid vigour is a phenomenon observed in nature, yet poorly understood, whereby progeny outperform their parents in several agronomical traits but specifically seed yield (Radoev *et al.*, 2008). Heterosis was first observed in maize (*Zea mays* L.) in 1908 by George Shull (Shull, 1908). In his paper, Shull (1908) commented on inbreeding and the loss of fitness with highly inbred lines. However, Shull (1908) discovered that hybrids from inbreds immediately regained fitness and vigour, surpassing both parental types. More than a century later, the underlying mechanism that controls heterotic gains is still not fully understood (Radoev *et al.*, 2008; Li *et al*., 2015). This has led to the development of two competing theories that underpin heterosis, dominance and overdominance (Crow, 1948). Dominance theory, first suggested by Bruce (1910) and Keeble and Pellew (1910) state there is a positive correlation with dominant alleles and a negative association with recessive alleles in regards to hybrid performance. Heterosis is achieved through the stacking of dominant positive alleles in the F_1 generation. Conversely, overdominance, first stated by Shull (1908) and East (1908), suggest that heterozygosity has a greater positive influence on heterosis than either of the two possible homozygous states. More recently, hybrid vigour has been suggested to be linked to the genetic distance between parental genotypes (Falconer and MacKay, 1996). In conjunction, several studies have supported this positive correlation between genetic distance and hybrid heterosis in

B. napus with yield increases of up to 169% (Ali *et al*., 1995; Diers *et al.*, 1996; Riaz *et al.*, 2001; Radoev *et al.*, 2008; Cuthbert *et al.*, 2009; Ahmad and Quiros, 2011). Therefore, *B. napus* hybrids have demonstrated an increased yield potential along with superior uniformity within the F_1 generation over their open-pollenated (OP) parents. These positive gains are extremely attractive to breeders and breeding companies, which is why *B. napus* hybrids have completely replaced OP cultivars in commercial production in Canada (Brewin and Malla, 2012). This fundamental change in breeding strategy could not have been accomplished without the development of novel pollination control methods, which allow for commercial scale production of hybrid seed. Most notable in *B. napus* are nuclear genic male sterility (GMS) and cytoplasmic male sterility (CMS) systems.

2.62 Genic Male Sterility

Male sterility is a common phenotype observed throughout a wide variety of higher plants (Song *et al.*, 2006). Male sterility consists of non-functional pollen and/or pollen abortion in hermaphroditic and dioecious flowers yet leaves the ovule fully fertile (Budar and Pellier, 2001). Genic male sterility (GMS) is controlled by genes in the nucleus (Huang *et al.*, 2007). Currently there are three types of systems in use that can be separated based on the inheritance of controlling factors (Huang *et al.*, 2007). These include double recessive GMS, the interaction of recessive alleles causing GMS and finally dominant gene GMS (Huang *et al.*, 2007). Of the genes characterized to cause genic male sterility, the *Ms* (male sterility) gene and the *Rf* (suppressor) gene are the most common (Vinod, 2005; Lu *et al.*, 2004). Expression of the *Ms* gene itself will cause male sterility, however, in the presences of *Rf*, suppression of *Ms* will occur and the resulting progeny will be fully fertile (Lu *et al.*, 2004). The advantage of a GMS system is that it can involve only two breeding lines and the genes responsible for male sterility

at the genetic level are relatively easy to transfer to any desired genetic background (Wei *et al.*, 2013; Huang *et al.*, 2007). This system provides an alternative to the more common cytoplasmic male sterility system (CMS) used in *B. napus* hybrid breeding (Yang and Zhang, 2013; Engelke *et al.*, 2010).

2.63 Cytoplasmic Male Sterility Systems

Cytoplasmic male sterility (CMS) is a well-documented trait throughout the plant world, seen in over 150 different plant species (Schnable and Wise, 1998). CMS is a maternally inherited condition that prevents the development of functional pollen yet female fertility remains viable (Chase, 2007). The use and development of CMS for pollen control has revolutionized classical breeding techniques in *B. napus* to facilitate large scale-production of hybrid seed. Due to the highly self-fertile nature of *B. napus*, outcrossing can be difficult and requires hand emasculation of fertile flowers which is laborious, time consuming and costly.With the use of CMS, highly self-fertile crops can now be specifically outcrossed to utilize and capitalize upon heterosis and heterotic gains. Fortunately, several CMS systems are available for *B. napus* hybrid seed production and include the *pol*, *nap* and *ogu* pollination control systems (Fu, 1981; Sernyk, 1983; Bannerot *et al.*, 1974). All three CMS systems are correlated to novel chimeric mitochondrial open reading frames (ORF) that interfere with the function and development of pollen (Schnable and Wise, 1998; Jean *et al.*, 1997). These novel ORF's are caused by mitochondrial rearrangements that are usually linked to positions very close to the ATP synthases (Bonhomme *et al.*, 1992; Hanson and Bentolila, 2004). In turn, this creates chimeric proteins that are non-functional leading to pollen sterility and abortion through the lack of ATP production (Pelletier and Budar, 2007). All three CMS systems have been associated to specific

ORF's within *B. napus* mitochondria and for the *pol*, *nap* and *ogu* systems, they have been designated ORF 224, 222 and 138, respectively (Engelke *et al.*, 2010).

Since CMS is inherited maternally, progeny in these systems are always sterile unless a dominant nuclear restorer gene is present which supresses the sterile phenotype and restores fertility to the F_1 progeny (Chase, 2007; Pelletier and Budar, 2007). For each CMS system, a restorer of fertility (*Rf*) gene has been found and characterised. For *nap*, *pol* and *ogu* they are referred to as *Rfp, Rfn* and *Rfo*, respectively (Yang and Zhang, 2013; Yamagishi and Bhat, 2014). Interactions between cytoplasmic sterility and restorer genes are poorly understood; however, in most cases, the protein product of the CMS-determining locus (ORF) fails to accumulate in the presence of a restorer gene (*Rf*) (Chase, 2007). This indicates a nuclear mechanism which controls or mediates mitochondrial expression or modifies the post-translational modification of novel mitochondrial ORF's mRNA (Chase, 2007; Pelletier and Budar, 2007).

2.64 Cytoplasmic Male Sterility Development

Since the early 1980's intense research has been devoted to the deployment of CMS systems for canola/rapeseed hybrid development. *Pol* was the first discovered natural sterile cytoplasm within the *B. napus* winter cultivar Polima (Fu, 1981). Since then, several studies have investigated the nuclear and mitochondrial interactions of the *pol* CMS system (Singh and Brown, 1991; Fang and McVetty, 1989). The *pol* cytoplasm causes CMS through a direct novel ORF associated with the ATPase subunit 6 gene and is co-transcribed with it (Singh and Brown, 1991). *Pol* showed promise, but upon implementation in commercial production, male sterility would breakdown under certain environmental conditions (Engelke *et al.*, 2010; Fan and Stefansson, 1986). Further investigations showed that temperature was a main factor in the

reversion of male-sterile plants to functional hermaphrodites (Fu, 1981; Fan and Stefansson, 1986; Burns *et al.*, 1991). The *nap* CMS system, which is very similar to the *pol* CMS system, also suffers from the same fertility reversion by environmental effects (Engelke *et al.*, 2010). Due to the unstable nature of the *pol* and *nap* systems, the *ogu*-INRA CMS system is currently the most prominent system used for *B. napus* hybrid seed production (Abbadi and Leckband, 2011).

2.65 *Ogu*-INRA Cytoplasmic Male Sterility System

CMS was first described by Rhoades (1933) in maize and later by Ogura (1968) in *Raphanus sativus* L. whereby the flowers of the radish plant would cease to produce functional pollen yet the maternal fertility of the flower remained unaffected. This discovery gave rise to the notion of using CMS for hybrid seed production (Ogura, 1968). Continuing the work presented by Ogura, Bannerot *et al.*, (1974) introduced CMS to *B. napus* through interspecific crosses yielding a *B. napus* genome within a *R. sativus* cytoplasm. This technique conferred the CMS trait although the progeny suffered leaf chlorosis and were generally of poor agronomic quality. Fortunately, Pelletier *et al.*, (1983) was successful in utilizing somatic hybridization and protoplast fusion with *B. napus* CMS lines to greatly improve their agronomic performance. These newly developed cybrids contained a *B. napus* nucleus with *R. sativus* cytoplasm and *B. napus* cytoplasm (referred to as *ogu*). In conjunction, these mixed cytoplasms conferred CMS to *B. napus* and relieved the chlorosis attribute (Pelletier *et al.*, 1983). Continuing with the *ogu* research at the National Institute of Agronomic Research (INRA, France) Delourme *et al*., (1994) did extensive classical breeding to develop a double low restorer line for canola quality hybrid production. These developments led directly to the creation of the *ogu*-INRA hybrid system (Delourme and Eber, 1992; Gourret *et al.*, 1992; Delourme *et al.*, 1994). In *B. napus*,

ogu-INRA CMS is conferred by the mitochondrial ORF138 (Bonhomme *et al.*, 1992). Unlike most CMS ORF loci, the ORF138 protein does not form a chimeric polypeptide with other mitochondrial proteins, but forms larger complexes in the inner mitochondrial membrane (Dong *et al.*, 2013). Further studies have shown that ORF138 acts as a pore in the inner mitochondrial membrane of sterile plants and exerts a mild uncoupling effect on oxidative phosphorylation (Dong *et al.*, 2013). However, the exact mechanism that interferes with pollen production and fertility in the *ogu*-INRA system still remains unclear (Dong *et al.*, 2013).

2.66 Restorer Gene

In conjunction with CMS, fertility must be restored in order to create fertile hybrids. Restorer of fertility (*Rf*) genes are dominant nuclear genes which supress the sterile phenotype and fully restore male fertility. The *Rfo* gene was introgressed into *B. napus* from *R. sativus* through intergeneric hybridization (Heyn, 1976). Unfortunately, this introgression introduced a large portion of radish genomic DNA of up to 50 cM into *B. napus* conveying poor agronomic traits (Delourme *et al.*, 1998; Brown *et al.*, 2003; Hu *et al.*, 2008). Several years of breeding were required to facilitate a double low restorer line for use in canola quality hybrid production (Primard-Brisset *et al.*, 2005). Current molecular understanding of the *Rfo* gene introgressed from radish contains a pentatricopeptide repeat (PPR) in tandem arbitrarily named PPR-A, B and C (Uyttewaal *et al.*, 2008). Genetically, PPR-B has been isolated and shown to be the mechanism that causes restoration of fertility in the *ogu* system (Uyttewaal *et al.*, 2008; Brown *et al.*, 2003; Koizuka *et al.*, 2003). PPR-A is an active gene yet its functional mechanism is poorly understood and PPR-C has been shown to be a pseudo gene with a premature stop codon preventing any function (Brown *et al.*, 2003; Uyttewaal *et al.*, 2008). PPR-B encodes a polypeptide with a length of 687 amino acids, which contains 16 segments of a 35 amino acid repeat; the PPR motif is

predicted to target the mitochondria (Brown *et al.*, 2003). The *Rfo* introgression has been mapped to linkage group 19 or C9 in *B. napus* (Primard-Brisset *et al.*, 2005; Hu *et al.*, 2008). Currently, several marker systems allow for marker-assisted selection (MAS) for the *Rfo* gene, which has greatly enhanced breeding efficiency in the *ogu* system (Delourme *et al.*, 1998; Brown *et al.*, 2003; Hu *et al.*, 2008; Havlíčková *et al.*, 2014).

2.67 Implementation of the *ogu*-INRA CMS System in Hybrid Breeding

Utilizing the *ogu*-INRA CMS system in hybrid breeding is a relatively straightforward process, providing pollen control is well executed. The *ogu*-INRA system relies on the implementation of a three line inbred breeding system consisting of A-lines, B-lines and R-lines (Figure 2.3) (Friedt and Snowdon, 2010). A-lines are the male-sterile parent (female) which contain the *ogu* sterile cytoplasm. B-lines are called maintainer lines and are essentially isogenic to A-lines and donate pollen to the A-line. Since CMS is inherited maternally, crossing A-lines and B-lines with no *Rfo* gene will result in subsequent A-lines. In the final step, R-lines (restorers of fertility) donate pollen to the male-sterile A-lines (Friedt and Snowdon, 2010). This fertilization event will fuse genetic information from both parents into a hybrid. This hybrid will now contain the sterile cytoplasm from the mother plant but will receive the *Rfo* gene from the R-line father, thus, the resulting progeny will be fertile (Figure 2.3).

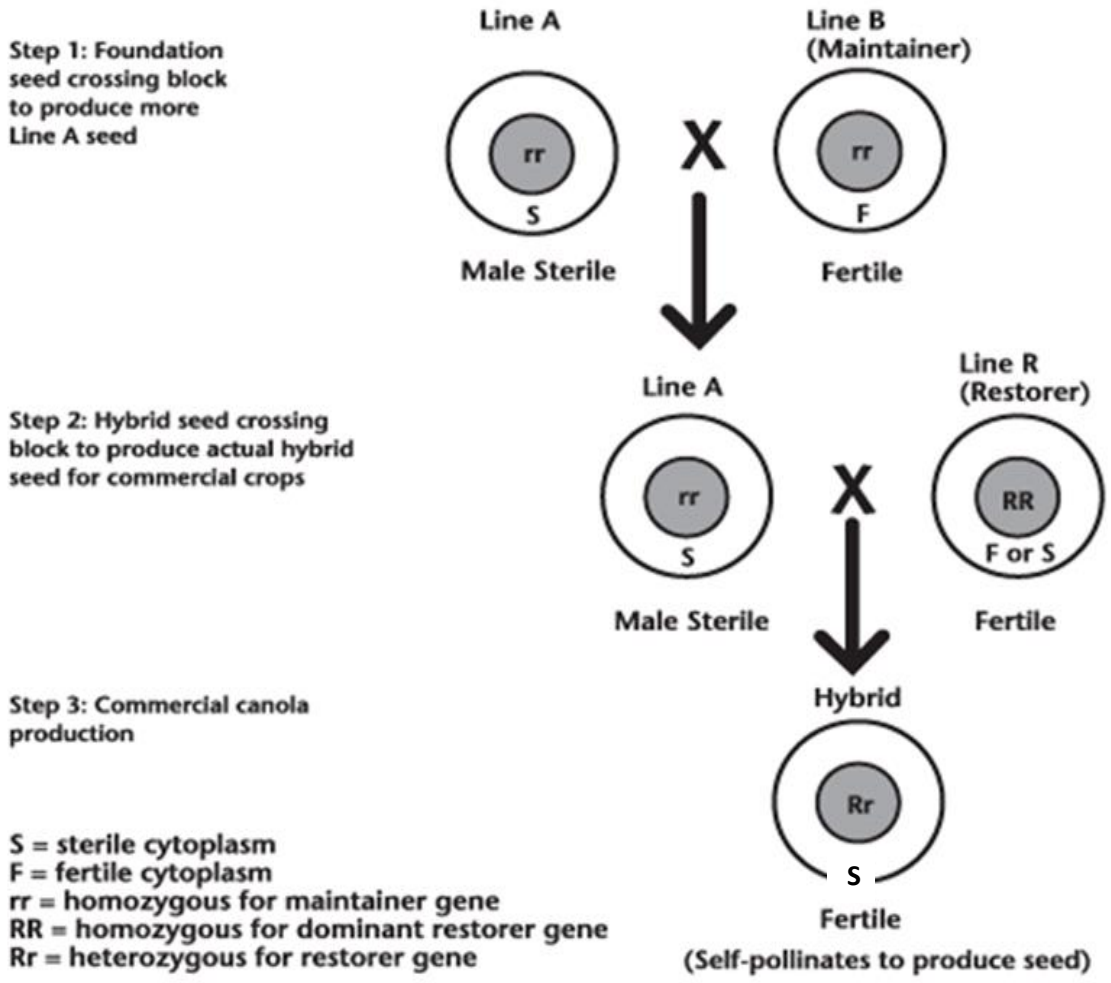

Figure 2.3: Schematic of hybrid breeding using the *ogu*-INRA CMS system. Step 1: A-lines are crossed with B-lines to create more A-lines. Step 2: A-lines are crossed to R-lines to create fertile hybrids for step 3. Figure obtained from: http://www.canolacouncil.org/crop-production/canola-grower's-manual-contents/chapter-2-canola-varieties/canola-varieties.

Typically, for research purposes, hybrid production tents are set up containing rows of A-lines and R-lines (Friedt and Snowdon, 2010). These tents prevent any outside pollen from contaminating specific A-line x R-line crosses. Once flowering has begun, captive insects can be released within the tent to facilitate cross pollination (Friedt and Snowdon, 2010). Only the A-line seeds are harvested and used as the F_1 hybrid (McVetty and Duncan, 2015).

2.68 *Brassica napus* Hybrids

Hybrid vigour in *B. napus* F_1 hybrids has been shown to be 40-60% higher than mid-parent values for seed yield (Diers *et al.*, 1996; Riaz *et al.*, 2001; Radoev *et al.*, 2008; Cuthbert *et al.*, 2009; Riaz *et al.*, 2011). Although these results vary, there is a strong tendency for crosses from different heterotic clusters or pools to produce hybrids that excel over hybrid crosses made within the same heterotic cluster or pool (Grant and Beversdorf, 1985; Riaz *et al.*, 2001). Therefore, accurate placement of breeding lines in defined heterotic clusters is essential to any hybrid program that wishes to maximise heterotic gains between parents (Girke *et al.*, 2011).

2.7 Genetic Distance

Genetic distance is a measure of genetic divergence between species or between populations of a single species (Nei, 1987). Therefore, different populations that contain very similar genes will have a short genetic distance indicating their relative similarity to one another. Over the years, several methods for measuring genetic distance have been developed including Nei's standard distance (Nei, 1972), Cavalli-Sforza and Edwards measure (Cavalli-Sforza and Edwards, 1967) and Reynolds, Weir and Cockerham's genetic distance (Reynolds *et al.*, 1983). In conjunction with genetic distance measures, new molecular methods have been utilized which yield genetic differences based on genetic polymorphisms between different genotypes. These include restriction fragment length polymorphism (RFLP, Botstein *et al.*, 1980), short sequence repeat (SSR, Webber and May, 1989), amplified fragment length polymorphism (AFLP, Vos *et al.*, 1995), sequence related amplified polymorphism (SRAP) (Li and Quiros, 2001) and more recently single nucleotide polymorphism (SNP) calling in genotyping-by-sequencing (GBS)

(Elshire *et al.*, 2011). Commonly, all of these methods can be used to distinguish molecular differences within genotypes and these differences can be used to calculate genetic distance.

2.71 Genetic Distance and Heterosis

Following the hypothesis of Falconer and MacKay (1996), several studies have investigated the effect of hybrid heterosis and genetic distance (Grant and Beversdorf, 1985; Diers *et al.*, 1996; Riaz *et al.*, 2001; Radoev *et al.*, 2008; Cuthbert *et al.*, 2009; Ahman and Quiros, 2011). The results of these studies support the conclusion that genetic distance positively influences heterosis in *B. napus* hybrids, although results have varied. Some of the best results have been published by Riaz *et al.*, (2001) where *B. napus* hybrids showed 169% seed yield over their mid-parent values and the correlation coefficient (r = 0.64) indicated a moderately strong relationship between genetic distance and hybrid yield. The coefficient of determination (R^2) was 40.43% for yield and genetic distance, which means that 40.43% of the variation seen in yield could be attributed to the genetic distance between parents (Riaz *et al.*, 2001). This is contradictory to several other studies that showed little to no correlation between genetic distance and hybrid heterosis in *B. napus* (Jesske *et al.*, 2013; Yu *et al.*, 2005). Cheres *et al.*, (2000) also investigated genetic distance and heterotic pool definition in sunflower (*Helianthus annuus* L.) and found no correlation between genetic distance and seed yield in F_1 hybrids. These differences may be attributed to environmental variation and effect, experimental design, molecular method, genetic distance method or hierarchical clustering method used.

2.72 Heterotic Pool Definition

The development of heterotic pools or clusters is essential for *B. napus* hybrid breeding (Girke *et al.*, 2011). To date, hierarchical clustering remains one of the most popular methods for

separating diverse genotypes into heterotic clusters that are displayed in the format of a dendrogram. Due to its popularity, several methods for clustering have been developed including unweighted pair group method with arithmetic mean (UPGMA, Sokal and Michener, 1958), neighbour joining (Saitou and Nei, 1987) and Ward's method (Ward, 1963). All three methods are very popular, simple agglomerative bottom up clustering techniques (Odong *et al.*, 2011; Tamura *et al.*, 2004). Using genotypic data from current molecular techniques, genetic distance can be calculated and hierarchical clustering algorithms can be implemented which separate or cluster genotypes based on genetic distance. Once these clusters or heterotic pools have been established, hybrid crosses can be explored which maximise the genetic distance between each parental genotype. Heterosis has been shown to highly correlate to inter-cluster crosses (crosses that take one parent from divergent or different clusters) opposed to intra-cluster crosses (crosses from the same cluster) (Grant and Beversdorf, 1985; Riaz *et al.*, 2001). Proper heterotic pools would greatly reduce the cost and time for experimental hybrid development through proper pairing of parental lines based on genetic distance and between genotypes that exhibit maximum heterosis (Havlíčková *et al.*, 2014).

2.73 General and Specific Combining Ability

Although genetic distance has a moderately strong correlation to hybrid heterosis, the combinational compatibility of each parent plays a role in the vigour and heterosis seen in the F_1 (Riaz *et al.*, 2001; Grant and Beversdorf, 1985; Sabaghnia *et al.*, 2010). The ability of one parent to combine well with many other partners is called general combining ability; whereby one parent produces high heterotic hybrids with many partners (Sleper and Poehlman, 2006). Specific combining ability on the other hand is the ability to achieve high heterosis through a single specific female x male or male x female cross (Sleper and Poehlman, 2006). Ideally,

heterotic crosses are designed to maximise heterotic gains and ultimately encompass the total amount of heterosis available.

2.8 Methods for determining Genetic Distance

2.81 Phenotyping

Since the dawn of agriculture, phenotyping has been a powerful tool to which humans have used to shape plant architecture and productivity (Hufford *et al.*, 2012). Phenotyping is simply observing the physical characterization of traits that represent the organism as a whole (Mahner and kary, 1997; Dawkins, 1999). The characterization of these traits can then be used to separate species from species or individuals from others within the same species (Burstin and Charcosset, 1997). By identifying the natural variation that exists within a population, phenotyping can be used to separate specific traits in a population and a population structure can be formed (Burstin and Charcosset, 1997). Once a set of traits has been established, a rating scale or presence/absence scores can be applied to a population and genetic distance can be measured based on any of the three formulas previously mentioned (Burstin and Charcosset, 1997). Phenotyping can be very effective at discovering or separating specific species, or genotypes within a population (Burstin and Charcosset, 1997). However, in today's post-modern genomic era many methods are available for the characterization and classification of organisms based on DNA sequence (Havlíčková *et al.*, 2014).

2.82 Genotyping

Today's high throughput next-generation sequencing technologies allow for the quantification of massive amounts of data for a relatively low cost (Hayward *et al.*, 2012). Sequence Related Amplified Polymorphism (SRAP) (Li and Quiros, 2001), the Brassica 60K SNP chip (Hayward

et al., 2012) and Genotyping-by-Sequence (GBS) (Elshire *et al.*, 2011) are three specific genomic methods for differentiating genetic diversity within a population of highly related individuals.

2.83 Sequence Related Amplified Polymorphism (SRAP)

SRAP is a simple polymerase chain reaction (PCR) method, which is designed to amplify ORF using a variable forward and reverse primer system (Li and Quiros, 2001). Each primer is 17-18 base pairs long with the forward primer containing a core sequence of CCGG. This forward core sequence targets ORF due to the known distribution that exons are GC rich (Li and Quiros, 2001). The reverse primer has a core sequence of AATT near the 3′ region to target introns and promoter regions, which are typically AT rich (Li and Quiros, 2001). Together, these primer combinations create polymorphic DNA bands that are separated through electrophoresis in polyacrylamide gels and visualized through autoradiography (Li and Quiros, 2001). Presence/absence scoring is then applied to the visualized polymorphic bands and genotypes can be separated based on scoring. SRAP can also be taken a step further with the sequencing of each polymorphic band to develop highly reproducible linkage maps (Li and Quiros, 2001).

2.84 Brassica 60K Illumina Infinium SNP Chip

The Illumina® GoldenGate and Infinium assays provide excellent utility in SNP detection and discovery of genetic variation for both breeding and research purposes (Hayward *et al.*, 2012). SNP's are single base pair changes within an organisms DNA and have quickly become the most popular method for fine mapping and heritable traits (Hayward *et al.*, 2012). There are three types of SNP's: transversions (C/G, A/T, C/A and T/G), transitions (C/T or G/A) and insertions/ deletions (indels), which insert a new base or delete a current base (Hayward *et al.*, 2012). Since

SNP's arise over time, mutation SNP's are excellent markers for species divergence (Hayward *et al.*, 2012; Berger *et al.*, 2001). The *Brassica* 60K Illumina® Infinium SNP chip from Illumina Inc. (San Diego, CA) consists of randomly packed beads, each containing thousands of copies of specific 50-mer oligonucleotide sequences (Clarke *et al.*, 2013). Each BeadArray is designed to "tag" the majority of DNA sequences in a given organism (Clarke *et al.*, 2013). Multiple BeadArrays collectively form a BeadChip, with a configuration of six (WG-6), eight (Ref-8), or 12 (HT-12) samples per chip for DNA products (Clarke *et al.*, 2013). During 2011-2012, an international consortium featuring partners from Australia, Europe, China, North America and South America was established to fund and design the development of a public high-density (50K-60K) Illumina® Infinium SNP array for *B. napus* (http://www.illumina.com/applications/agriculture/consortia.html; Hayward *et al.*, 2012). This development has facilitated very fast and low cost association mapping studies (Li *et al.*, 2014), molecular karyotyping (Mason *et al.*, 2014) and genetic diversity studies in several *Brassica* species (Mason *et al.*, 2014).

2.85 Genotyping-by-Sequencing (GBS)

Genotyping-by-Sequencing (GBS) is one of the most modern methods for measuring genetic diversity within species based on mutation SNP's (Elshire *et al.*, 2011). GBS is a highly multiplexed PCR method yet varies from the previous SRAP method. GBS uses a reduced representation of genome complexity through the use of restriction enzymes (RE) that are methylation sensitive (Elshire *et al.*, 2011). Heavy methylation is indicative of inactive and repetitive DNA and thus does not actively participate in gene expression (Elshire *et al.*, 2011). Reducing repetitive and inactive DNA greatly simplifies sequencing and alignment procedures allowing for deep coverage (Chen *et al.*, 2013; Elshire *et al.*, 2011). Following RE digestion,

adapter barcodes are ligated to the RE cut site allowing many samples to be pooled into one flow cell greatly reducing cost (Elshire *et al.*, 2011; Chen *et al.*, 2013). Generally, millions of sequence tags (64 bp reads) are generated and 10,000's to 100,000's of SNP's can be called with a very high degree of accuracy through a novel GBS computational pipeline, Tassel (Elshire *et al.*, 2011; Glaubitz *et al.*, 2014). Since SNP's are the most abundant polymorphism in a genome they are applicable to a wide range of processes including genome wide associated studies (GWAS), phylogenetic analysis and quantitative trait loci (QTL) detection (Sonah *et al.*, 2013). Since SNP's are so abundant, calculating genetic distance using computation algorithms is simple and allows for high definition of specific breeding pools essential for a hybrid breeding program. GBS can provide many advantages over other genotypic methods; it offers a simplified library preparation procedure with reduced genome representation; it can be performed with minute concentrations of initial DNA (100–200 ng) and it is versatile enough to handle a high level of multiplexing. Also, GBS provides highly accurate SNP calling in gene-rich regions of the genome in a highly cost-effective manner (Sonah *et al.*, 2013). Several studies have been published showing the robustness of the GBS platform in many agriculturally important species including maize, barley and soybean (Elshire *et al.*, 2011; Sonah *et al.*, 2013). Future goals with this novel technology will allow for the quick and cheap sequencing of entire populations that can be used for a wide range of genetic diversity and heterosis studies.

2.9 Heterotic Pairing Based on Cluster Analysis

As previously mentioned, cluster analysis in conjunction with genetic distance has been explored in *B. napus* hybrid development (Ali *et al*., 1995; Riaz *et al*., 2001; Ahman and Quiros, 2011). A general consensus from these studies indicates that inter-cluster hybrids out-perform intra-cluster hybrids when based on genetic distance (Ali *et al*., 1995; Riaz *et al*., 2001; Ahman and Quiros,

2011). Therefore, one of the first steps in the development of a *B. napus* hybrid breeding program should be to discover the level of genetic diversity present and cluster those genotypes into heterotic pools (Girke *et al*., 2012). Once these heterotic pools are defined, exploration into inter-cluster hybrids can occur. This may reduce the number of crosses required to find high heterotic parents as intra-cluster hybrids could be avoided saving time, space and money.

3.0 Phenotypic Heterotic Pool Characterization of *Brassica napus* L. Inbred Lines Using Multivariate Cluster Analysis

3.1 ABSTRACT

Information about the genetic structure of germplasm collections can be of great importance for the utilization of genetic resources in hybrid breeding projects. To date, genetic distance and multivariate cluster analysis remain one of the most popular methods for separating germplasm accessions into distinct breeding pools. Twenty phenotypic qualitative and quantitative traits from a field trial in Winnipeg, MB, during 2014 were used to determine the genetic structure of seventy-nine *Brassica napus* L. germplasm accessions using Ward's clustering technique. Cluster analysis grouped the seventy-nine genotypes into eight distinct clusters and two general phenotypic pools with no outliers. Principal component analysis was then undertaken to determine the interaction of each qualitative and quantitative trait. Principal component analysis found a strong correlation between pod length, oil content and seed yield and a negative correlation between those traits and total protein content and glucosinolate concentration. Clustering based on twenty qualitative and quantitative traits could prove useful for future breeding schemes that wish to exploit heterosis in *B. napus* hybrids through parental pairing of divergent genotypes separated by cluster analysis.

3.2 INTRODUCTION

Twenty first century plant breeding is a highly multidisciplinary science that requires expertise in genetics, breeding, pathology, entomology, agronomy and statistics with the goal of continued plant improvement (Baenziger *et al.*, 2008). Plant breeders are especially fundamental in the development of breeding objectives and resource management (Baenziger *et al.*, 2008; Gepts, 2006). Once a breeding objective is developed, it is necessary to find or create populations with significant variation, followed by selection for simultaneous goals. Depending upon the breeding objectives or goals, selection intensities may differ, but the selection strategy needs to be appropriate to limit excess crosses or evaluations that may waste time, space or lack commercial potential (Baenziger *et al.*, 2008; McKeown *et al*., 2013). The inclusion of genetic diversity is a valued resource for breeders, especially in hybrid cultivar development (Delucchi *et al*., 2012; Jesske *et al*., 2013). Therefore, characterizing germplasm into different heterotic pools is one of the first steps breeders should attempt when developing a hybrid breeding program (Girke *et al*., 2012). These heterotic pools may reduce redundant genetic diversity and simultaneously reduce the number of crosses required to find high heterotic parental combinations (Esposito *et al*., 2013).

Gene pools can be used to add genetic diversity and can be described as one of three groups: tertiary, secondary and primary (Figure 3.1) (Harlan and De Wet, 1971). The tertiary gene pool was defined by Harbered (1976), who grouped species and genera related to *Brassica* crops into 36 cytodemes which have the potential to exchange genetic material; however, extreme molecular methods are needed to stabilize chromosome arrangement (Branca, 2008). The secondary gene pool for *B. napus* includes *Brassica* species in direct contact with *B. napus* in the Triangle of U (U, 1935) (Branca, 2008). This secondary gene pool can be crossed with the

primary gene pool but with reduced fertility (Gepts, 2006). Within the *B. napus* primary gene pool, there are only the cultivated species due to the lack of naturally occurring wild populations and landraces. By stratifying this primary gene pool into distinct smaller heterotic pools using cluster analysis, suspected high heterotic crosses from divergent heterotic pools can be selected, saving money, space and the overall timeframe for discovering high heterotic parental combinations. Ideally, to maximise heterotic gains, each parent should belong to a different general heterotic pool (ie., one from A and the other from B). Riaz *et al.* (2001) found that crosses from different clusters within these general pools (A or B) may be sufficient to exploit high heterotic gains.

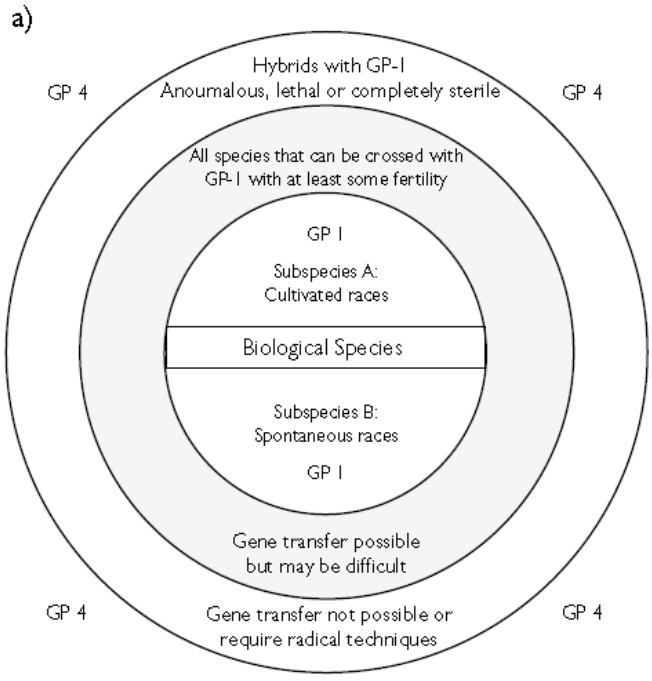

Figure 3.1: Theory of gene pools modified from Harlan and De Wet (1971). Gene pool 1 (GP1) consists of fertile hybrids. Gene pool 2 (dark circle) can create hybrids with GP1 with difficultly. Gene Pool 3 (white outer circle) can only make hybrids with GP1 with extreme molecular assistance. Gene pool 4 (GP4) transgenic techniques used to create hybrids.

Phenotyping is one method that allows plant breeders to separate genotypes into distinct breeding pools based on morphological characteristics. This characterization could aid in the selection strategy for cultivar improvement and hybrid development (Ali *et al.*, 1994). Phenotyping is simply the physical characterization of traits that represent the organism as a whole and can include complex plant traits such as growth, development, tolerance, resistance, architecture, physiology, ecology and yield (Ali *et al.*, 1994; Mahner and kary, 1997; Dawkins, 1999; Walter *et al.*, 2012). Plant phenotyping has been performed since the day humans started to domesticate plants and animals and increase yield or enhance other desirable traits (Araus and Cairns, 2014; Hufford *et al.*, 2012). In the past, phenotyping was broadly based on intuition along with experience and dealt with individual subjectivity for scoring or rating select plants (Walter *et al.*, 2012). Currently, phenotyping is still subject to an individual's perception for a trait or traits; however, the use of qualitative and quantitative data is widely used throughout the plant breeding world (Gixhari *et al.*, 2014; Bhattacharjee *et al.*, 2007; Upadhyaya *et al.*, 2001). In addition, high-throughput phenotyping is becoming common with the use of remote sensing and imaging and near-infrared reflectance spectroscopy (NIRS) analysis for yield components in many crop species (Araus and Cairns, 2014). These approaches may help plant breeders use qualitative and quantitative data for population diversity studies and population stratification studies (Araus and Cairns, 2014).

Quantitative measurements are simply measurements defined by specific numeric values (Lowhorn, 2007). Quantitative traits are therefore traits defined by specific measurements that span a continuous range (Walsh, 2001). This differs from qualitative traits, which are expressed qualitatively, meaning that the phenotype falls into distinct categories are not continuous. Using qualitative and quantitative traits plant breeders can apply statistical methods to compare traits

from one individual in a population to another or to compare different populations to each other (Walsh, 2001). Values from qualitative and quantitative traits can therefore be used to stratify a population and group phenotypes/genotypes that are similar. This grouping and stratification is the basis for hierarchical clustering (Anderberg, 1973).

Hierarchical clustering is a well cited statistical model used by plant and animal breeders to separate breeding lines or populations into distinct clusters (Gixhari *et al.*, 2014; Cuevas and Prom, 2013; Odong *et al.*, 2011; Morota *et al.*, 2014). Since clustering is a popular method, many different mathematical models have been developed to address certain assumptions about clustering parameters (Anderberg, 1973). Currently, three main clustering methods dominate modern plant breeding, those being Ward's method (Ward, 1963), Unweighted Pair Group Method with Arithmetic mean [UPGMA; (Sokal and Michener, 1958)] and neighbour joining [NJ; (Saitou and Nei, 1987)] which are all agglomerative (bottom up) clustering techniques. Each method has its own unique parameters for comparing data sets and clustering those data sets into hierarchical clusters (Mouchet *et al.*, 2008). However, despite the decades of use with these three methods, there is no scientific consensus for which model creates the most accurate set of clusters (Mouchet *et al.*, 2008; Mohammadi and Prasanna, 2003). UPGMA is by far the most cited method in plant breeding (Mohammadi and Prasanna, 2003); however, a recent study by Odong *et al.* (2011) demonstrated that Ward's method produced more complete and accurate hierarchical clusters when using molecular marker data. This is consistent with Lombard *et al.* (2000) who found Ward's method superior to UPGMA in *B. napus* using amplified fragment length polymorphisms. Using qualitative and quantitative data, several studies have shown that Ward's method is also effective at separating breeding lines/accessions into genetically distinct

heterotic clusters for breeding purposes and capturing genetic diversity in core collections (Gixhari *et al.*, 2014; Bhattacharjee *et al.*, 2007; Upadhyaya *et al.*, 2001).

Ward's minimum variance method is an agglomerative clustering technique that minimises within cluster variation through the use of error sum of squares based on Euclidian distance between two clusters (Ward, 1963). Ward's method begins with all clusters being singletons and mergers are based on the least amount of inter-cluster variation weighted against squared Euclidean distance between cluster centres (Ward, 1963). This successively merges clusters until the intra cluster variation exceeds the squared Euclidean distance between each cluster centre. Ward's method has been shown to excel in clustering phenotypes or genotypes into heterotic pools and therefore has been chosen as the most accurate method for separating *B. napus* germplasm in this study (Odong *et al.*, 2011; Gixhari *et al.*, 2014; Bhattacharjee *et al.*, 2006; Upadhyaya *et al.*, 2001). To maximize heterotic breeding potential and possibly limit the number of crosses required to find high heterotic parental combinations, a cluster analysis should be implemented in preparation for hybrid rapeseed development.

The objective of this study was to characterize seventy-nine *B. napus* spring habit genotypes through phenotypic methods and assign them to heterotic breeding pools using Ward's method of agglomerative clustering. This research is intended to improve high heterotic parental pairing through defined phenotypic heterotic pools.

3.3 MATERIALS AND METHODS

3.31 Plant Material

A diverse set of seventy-nine germplasm accessions were selected, comprised of mainly University of Manitoba (U of M) genotypes with the addition of several European and resynthesized lines (Table 3.1). Some genotypes are considered to be canola quality; however, a vast majority are considered to be high erucic acid rapeseed (HEAR). This germplasm set can also be divided into maintainer lines (-B) and restorer lines (-R) utilized within the *ogu*-INRA cytoplasmic male sterility pollination control system (Pelletier and Budar, 2015). All genotypes are considered spring habit.

Table 3.1: Seventy-nine *Brassica napus* genotypes with origin, quality and maintainer or restorer designation for the *ogu*-INRA cytoplasmic male sterility pollination control system.

Cultivar	Origin[1]	Quality[2]	Maintainer/Restorer
NEW-620-R	U of M	HEAR	Restorer
NEW-621-R	U of M	HEAR	Restorer
Castor-R	U of M	HEAR	Restorer
UMI-71-R	European	HEAR	Restorer
UMS-189-R	European	HEAR	Restorer
12DH378-R	U of M	HEAR	Restorer
12DH377 -R	U of M	HEAR	Restorer
RRHR503-R	U of M	HEAR	Restorer
RRHR204-R	U of M	HEAR	Restorer
RRHR404-R	U of M	HEAR	Restorer
RRHR5815-R	U of M	HEAR	Restorer
RRHR5819-R	U of M	HEAR	Restorer
08C702-R	U of M	HEAR	Restorer
08C712-R	U of M	HEAR	Restorer
08C847-R	U of M	HEAR	Restorer
Red River 1997-R	U of M	HEAR	Restorer
Red River 1826-R	U of M	HEAR	Restorer
Red River 1852-R	U of M	HEAR	Restorer
Red River 1997-R2	U of M	HEAR	Restorer
UMI-55-R	European	HEAR	Restorer
UMI-99-R	European	HEAR	Restorer
12DH384-R	U of M	HEAR	Restorer

Cultivar	Origin[1]	Quality[2]	Maintainer/Restorer
12DH430-R	U of M	HEAR	Restorer
12DH478-R	U of M	HEAR	Restorer
12DH915-R	U of M	HEAR	Restorer
12DH949-R	U of M	HEAR	Restorer
11DH91-R	U of M	HEAR	Restorer
11DH92-R	U of M	HEAR	Restorer
11DH97-R	U of M	HEAR	Restorer
11DH108-R	U of M	HEAR	Restorer
11DH109-R	U of M	HEAR	Restorer
11DH114-R	U of M	HEAR	Restorer
11DH122-R	U of M	HEAR	Restorer
11DH137-R	U of M	HEAR	Restorer
11DH144-R	U of M	HEAR	Restorer
11DH148-R	U of M	HEAR	Restorer
11DH149-R	U of M	HEAR	Restorer
11DH162-R	U of M	HEAR	Restorer
Red River 1826-B	U of M	HEAR	Maintainer
Red River 1852-B	U of M	HEAR	Maintainer
Red River 1997-B	U of M	HEAR	Maintainer
Red River 1861-B	U of M	HEAR	Maintainer
RRHR503-B	U of M	HEAR	Maintainer
RRHR204-B	U of M	HEAR	Maintainer
RRHR404-B	U of M	HEAR	Maintainer
RRHR5815-B	U of M	HEAR	Maintainer
RRHR5819-B	U of M	HEAR	Maintainer
08C702-B	U of M	HEAR	Maintainer
08C712-B	U of M	HEAR	Maintainer
08C847-B	U of M	HEAR	Maintainer
RRHR9707-B	U of M	HEAR	Maintainer
UMI-B	European	HEAR	Maintainer
UMSA-B	European	HEAR	Maintainer
UMSH-B	European	HEAR	Maintainer
Venus-B	U of M	HEAR	Maintainer
Neptune-B	U of M	HEAR	Maintainer
Castor-B	U of M	HEAR	Maintainer
Mill 03-B	U of M	HEAR	Maintainer
Mercury-B	U of M	HEAR	Maintainer
Hero-B	U of M	HEAR	Maintainer
08C344-B	U of M	HEAR	Maintainer
Reston-B	U of M	HEAR	Maintainer
79R713-B	U of M	HEAR	Maintainer
79R714-B	U of M	HEAR	Maintainer
79R728-B	U of M	HEAR	Maintainer
79R729-B	U of M	HEAR	Maintainer
ER2-B	Resynthesized	HEAR	Maintainer

Cultivar	Origin[1]	Quality[2]	Maintainer/Restorer
ER3-B	Resynthesized	HEAR	Maintainer
ER7-B	Resynthesized	HEAR	Maintainer
ER22-B	Resynthesized	HEAR	Maintainer
ZSDH2602-B	Resynthesized	HEAR	Maintainer
Topas-B	U of M	Canola	Maintainer
Polo-B	U of M	Canola	Maintainer
Sentry-B	U of M	Canola	Maintainer
Global-B	U of M	Canola	Maintainer
Westar-B	U of M	Canola	Maintainer
Apollo-B	U of M	Low linolenic	Maintainer
Stellar-B	U of M	Low linolenic	Maintainer
02R276-B	U of M	High oleic low linolenic	Maintainer

[1]U of M is the abbreviation for University of Manitoba.
[2]HEAR is the abbreviation for High Erucic Acid Rapeseed.

3.32 Field Evaluation

The seventy-nine *B. napus* accessions were grown during the 2014 growing season (May – September) Winnipeg, MB, Canada (Latitude: 49.80 | Longitude: -97.16). Each accession consisted of 0.5 grams of seed planted in a single 3-meter row with 30 cm between each row bordered by guard rows. The nursery rows were planted on May 28th, 2014. Soil type in southern Manitoba is generally considered to be slightly acid (pH 6.1-6.5) black chernozem fine clay (Li *et al.*, 2012) with the Winnipeg location being classified as Riverdale silty clay (Lewis and Gulden, 2014). Conventional management techniques were implemented which included fertilizer, herbicide, pesticide and fungicide use as needed. Fertilizer was supplied in late October, 2013 by Crop Production Services with a custom blend equivalent to 137 kg/ha of nitrogen broadcast (formula: 46-0-0 (42.7%), 11-52-0 (31%), 0-0-60 (7.5%) 20-0-0-24 (18.7%). Approximately at the same time in 2013, Edge™ granular herbicide (Dow AgroSciences Indianapolis, IN) was added to the site at 27.9 kg/ha and worked twice with a tandem disk. Roundup WeatherMAX® Herbicide (Monsanto, St. Louis, MO) was applied on May 20th and

May 29th, 2014 (2.47 L/ha), to kill emerging weeds. All mixing and handling procedures were followed as per the manufacturer's instructions. Decis® (group 3 insecticide) (Bayer CropScience, North Carolina, USA) was applied on June 10th to control flea beetle (*Phyllotreta cruciferae*) populations at 148 ml/ha. Poast-Muster-Lontrel herbicide mix (Muster 29.6 g/ha, Lontrel 0.84 L/ha, Poast Ultra 0.67 L/ha and Merge at 1% volume/volume (1 L merge in 100 L of water) was applied on June 14th, 2014. Lontrel 360/E (Group 4 herbicide) (Dow AgroSciences Indianapolis, IN) was applied June 20th with a backpack sprayer (15 litres @100mL/L) to spot control Canadian thistle (*Cirsium arvense* L.). All mixing procedures were followed as per the manufacturer's label. Lance® (BASF Canada Inc. Mississauga, Ont.) granular fungicide (Group 7 fungicide) was applied on July 14th, 2014 at 350 g/ha to control Sclerotinia stem rot (*Sclerotinia sclerotiorum* Lib.). The nursery rows were harvested with a Wintersteiger nursery harvester (Nursery master Classic; Wintersteiger, Salt Lake City, UT) on September 18 - 22, 2014.

3.33 Morphological Traits

Twenty morphological qualitative and quantitative traits were assessed to characterize and estimate genetic diversity among the seventy-nine *B. napus* genotypes. The 20 qualitative and quantitative traits included: leaf shape, leaf length, leaf width, leaf color, leaf margin shape, leaf attachment, time to flower, petal color, petal spacing, plant height (cm), lodging, pod length, pod angle, time to maturity, seed coat color, seed protein content, seed oil content, seed glucosinolate content, seed erucic acid (22:1) content and seed yield (g/row^{-1}). Of the twenty traits selected as measures of genetic diversity, seven traits followed the guidelines set by the Variety Registration Office of Canada, Objective Description Form, Canola/Rapeseed version date 2010: (http://www.inspection.gc.ca/plants/varietyregistration/eng/1299175847046/1299175906353).

These seven traits were leaf margin shape (2.16: 1: undulating, 2: rounded, 3: sharp) of three random plants per row, leaf attachment to stem (2.18: 1: complete clasping, 2: partial clasping, 3: non-clasping) of three random plants per row, leaf shape (2.5: ratio of leaf, width/length: >0.8 orbicular, 0.67-0.79 wide elliptical, < 0.66 narrow elliptical), leaf length (2.6: stem to leaf tip in cm) of the lowest intact leaf of three random plants per row, leaf width (2.7: across leaf in cm) of the lowest intact leaf of three random plants per row), pod (silique) angle (3.14: 1: erect, 3: semi-erect, 5: horizontal, 7: slightly drooping, 9: drooping) of three random plants per row, and petal spacing (3.8: 1: open, 3: not touching overlap, 5: touching, 7: slightly overlap, 9: strong overlap) of three random plants per row. Additionally, ten other morphological traits were measured: leaf color: (1: light green, 2: medium green, 3: dark green, 4: blue/green) of three random plants per row, days to flower: (timeline in days after planting for 50% of the plants in a row to exhibit first flowers), flower petal color: (1: white, 2: light yellow, 3: medium yellow, 4: dark yellow, 5: orange) of three random plants per row, row plant height: (mean row height in cm from soil to tip of plant, if lodging was encountered plants were physically held up to measure height) measured with a measuring stick (up to 150 cm), row lodging: (1: erect, 2: slight bend, 3: noticeable bending, 4: pronounced bending and strain, 5: flattened or broken stems), pod (silique) length: (based on cm) of three random pods per row, time to maturity: (timeline in days after planting for 50% of row pods to be mature), seed coat color: (1: black, 2: black/brown mix, 3: brown, 4: black brown yellow mix, 5: yellow) based on entire row yield and row yield: (total seed amount each row produced in g/row). Flower traits (time to flower, petal color and petal spacing) were all measured approximately during time to flower (40-45 days after seeding); however, some genotypes were considerable later (60-70 days post seeding). Leaf traits (leaf length, leaf width, leaf color, leaf attachment to stem and leaf margin shape) were all measured

after flowering approximately 60 days after seeding as leaf area trended toward maximum size. Leaf shape is a ratio of leaf width/leaf length. Harvest characteristics (plant height, lodging, pod length, pod angle, time to maturity and yield) were all measured at approximately at maturity (100-110 days post seeding).

3.34 Seed Quality Analysis

In addition to morphological traits, seed quality traits were also evaluated. Total oil content (the proportion of each seed containing lipids as a percent), crude protein content (the proportion of each seed containing protein as a percent), fatty acid content (the proportion of total oil that is comprised of each fatty acid as a percent) and glucosinolate content (the proportion of aliphatic glucosinolates after oil extraction in protein meal as μmol/g) were all assessed post-harvest (> 120 days post seeding) from yield samples obtained from each genotype.

Fatty acid quantification was conducted with gas chromatography (GC) and was carried out on a Varian model 3900 fitted with a CP-Wax 52 CB capillary column and a flame ionization detector (Hougen and Bobo, 1973; Liu, 1994; Kim *et al.*, 2007). The column was 15 m x .32 mm i.d. fused silica coated with a .025 micron polyethylene glycol phase (Varian, Walnut Creek, USA). The carrier gas was Ultra High Purity (UHP) Helium at a flow rate of 2.0 ml/min. Column oven temperature was programmed from 190 C to 240 C, and the detector temperature was 280 C. Peak areas were measured using the Varian Star Workstation software system. Reference standard, GLC # 421, purchased from Nu-Check Prep (Elysian, Minnesota) was used to ensure proper GC operation. For the purpose of the current study, only fatty acid data for erucic acid content will be presented. In short, 300mg of seed was crushed with a test cylinder from Carver Press (Carver Inc., IN USA). The crushed seed was placed into a 13x100 mm test tube and 3 ml

heptane was added. This solution was left to stand overnight to extract oil. The supernatant was decanted into a clean 13x100 mm test tube. Next, 500 µl of 0.5N sodium methoxide reagent (13.5 g of sodium methylate powder in 500 ml anhydrous methanol) was added and shaken for 30 min. This step is the transesterification of triglycerides to fatty acid methyl esters. 100 µl of acidified water (0.3% acetic acid) was added and mixed gently. The solution was left to stand for 1-2 hours at 4 °C to clear. Approximately 500 µl of reaction mixture was pipetted into a 2 ml autosampler vial.

Total oil, protein and glucosinolate content were measured using Near Infrared Spectroscopy (NIR) with the procedure following Daun *et al.* (1994). The NIR spectroscopic analysis was performed using a near infrared scanning monochromator (NIRSystem model 6500, Foss NIRSystems Inc., MD, USA). Both GC and NIR analyses were conducted in the University of Manitoba Seed Quality Lab that is annually certified by the Canadian Grain Commission.

3.35 Distance Matrix and Dendrogram

Raw data based on the twenty morphological categories (tables 3.2-3.5) was entered into JMP Genomics software (JMP®, Version 6.3. SAS Institute Inc., Cary, NC, 1989-2007) after which a multivariate clustering algorithm was implemented with a standardized scale where data in each column was first standardized by subtracting the column mean and dividing by the column standard deviation. Ward's distance matrix and hierarchical clustering was done simultaneously as calculated with 20 qualitative and quantitative trait values from the 2014 growing season with all data points of equal weight in JMP Genomics. Ward's method is described below. Lowercase symbols generally pertain to observations and uppercase symbols to clusters (JMP®, Version *6.3*.). Where n is the number of observations, v is the number of variables, x_i is the ith

observation. At each step in Ward's method a new cluster C_{KL} with the smallest possible internal variance is created by merging the two clusters C_K and C_L that have the minimum variance between them. The variance between C_K and C_L is computed by Ward's method where D_{KL} is the *K*th cluster, subset of (1, 2,..., *n*), N_K is the number of observations in C_K, X is the sample mean vector, X_K is the mean vector for cluster C_K and X_L and N_L are defined similarly for C_L and $\|x\|$ is the square root of the sum of the squares of the elements of **x** (the Euclidean length of the vector **x**) (Zhang *et al*., 2010).

Distance for Ward's method is represented by:

$$D_{KL} = \frac{\left\|\overline{x_K} - \overline{x_L}\right\|^2}{\frac{1}{N_K} + \frac{1}{N_L}}$$ (Taken from the JMP 6.3 user manual, Zhang *et al*., 2010)

This equation compares the individual observations for each variable against the cluster means for that variable. The goal of Ward's method is to join clusters so that the merger has its variance within the new cluster minimized. To do this, each accession begins as its own cluster and clusters are then merged in such a way as to reduce the variability within a cluster. Specifically, two clusters are merged if this merger results in the minimum increase in the error sum of squares. Therefore, accessions with little variation between qualitative and quantitative traits should cluster together (Ward, 1963; Lee and Willcox, 2014).

3.36 Principal Component Analysis

A principal component analysis (PCA) was conducted in JMP genomics (V.6.3) using the twenty qualitative and quantitative traits obtained from the field data in 2014. The goal of PCA is to extract the important information from the data and to express this information as a set of new

orthogonal variables called principal components (Abdi and Williams, 2010). This analysis was used in the current study to investigate correlations between the morphological traits.

3.37 Statistical Analyses

Statistical analysis for mean (μ) and coeffiecent of variation (CV) for all data tables was completed with Excel 2010. Mean (μ) values were obtained with the Averge function in Excel and standard deviation (σ) was obtained with the STDEV.P function in Excel. CV was obtained as dictated by the formula $CV = \frac{\sigma}{\mu}$ x 100% in Excel.

3.4 RESULTS

3.41 Qualitative and Quantitative Trait Scores

For ease of viewing, all twenty phenotypic trait scores have been separated into four generalized tables (Tables 3.2-3.5) and highlight the twenty qualitative and quantitative values obtained for each genotype in the growing season of 2014. Table 3.2 contains flowering characteristics for all genotypes, including time to anthesis (flower), petal color and petal spacing.

Table 3.2: Flower characteristics for 79 *Brassica napus* genotypes grown during the 2014 field season in Winnipeg, MB.

Genotype[1]	Time to Flower[2] (days)	Petal Color[3] (1-5)	Petal Spacing[4] (1-9)
NEW-620-R	44	4	5
NEW-621-R	45	4	7
Castor-R	44	3	5
UMI-71-R	46	3	5
UMS-189-R	40	3	5
12DH378-R	45	4	7
12DH377 -R	44	3	9
RRHR503-R	47	3	7
RRHR204-R	43	4	7
RRHR404-R	48	4	9
RRHR5815-R	44	4	9
RRHR5819-R	39	4	9
08C702-R	40	3	1
08C712-R	40	4	7
08C847-R	45	3	1
Red River 1997-R	45	4	1
Red River 1826-R	43	2	7
Red River 1852-R	46	3	5
Red River 1997-R2	46	3	3
UMI-55-R	39	3	3
UMI-99-R	39	4	5
12DH384-R	46	4	9
12DH430-R	48	4	7
12DH478-R	43	4	5
12DH915-R	48	3	5
12DH949-R	41	3	3
11DH91-R	44	3	1

Genotype[1]	Time to Flower[2] (days)	Petal Color[3] (1-5)	Petal Spacing[4] (1-9)
11DH92-R	48	4	1
11DH97-R	47	3	3
11DH108-R	43	3	5
11DH109-R	41	3	5
11DH114-R	42	4	7
11DH122-R	47	4	1
11DH137-R	44	4	5
11DH144-R	44	4	1
11DH148-R	42	4	7
11DH149-R	47	4	1
11DH162-R	44	4	1
Red River 1826-B	42	3	3
Red River 1852-B	44	3	7
Red River 1997-B	42	3	5
Red River 1861-B	43	3	3
RRHR503-B	44	4	5
RRHR204-B	39	3	5
RRHR404-B	42	4	5
RRHR5815-B	39	4	7
RRHR5819-B	40	4	7
08C702-B	41	3	1
08C712-B	40	4	5
08C847-B	41	3	5
RRHR9707-B	42	2	3
UMI-B	47	4	5
UMSA-B	47	4	7
UMSH-B	58	4	7
Venus-B	40	3	7
Neptune-B	41	3	9
Castor-B	41	3	9
Mil 03-B	41	4	7
Mercury-B	42	3	9
Hero-B	41	4	7
08C344-B	39	3	1
Reston-B	42	3	7
79R713-B	40	3	9
79R714-B	40	3	9
79R728-B	39	3	7
79R729-B	44	3	5
ER2-B	74	2	9
ER3-B	54	4	3
ER7-B	67	3	5
ER22-B	51	2	5
ZSDH2602-B	54	2	9

Genotype[1]	Time to Flower[2] (days)	Petal Color[3] (1-5)	Petal Spacing[4] (1-9)
Topas-B	46	3	3
Polo-B	71	3	3
Sentry-B	44	3	7
Global-B	42	4	7
Westar-B	44	3	3
Apollo-B	41	3	3
Stellar-B	44	3	5
02R276-B	42	4	9
Mean (μ)	44.63	3.37	5.33
Standard Deviation (σ)	6.31	0.60	2.53
CV (%)	14.13	17.80	47.46

[1] R = restorer genotype, B = maintainer genotype for *ogu*-INRA hybrid system
[2] Time to flower is in days after planting for 50% of a row to exhibit first flower.
[3] Petal color is the mean flower color of three random plants per row.
[4] Petal spacing is the mean of three random plants per row.

Time to flower is an extremely valued phenotypic trait as the Canadian growing season can be relatively short. Mean time to flower was 44.6 days and several genotypes greatly exceeded 45 days with Polo and the ER genotypes extending flowering time into the 60 and 70-day period. These can be classified as un-adapted genotypes for the Canadian prairies as their flowering period will extend into the high temperatures in July and August. Petal color was investigated on a 1-5 scale and no genotypes exhibited white or orange petal colors. Petal color had a mean of 3.37, equivalent to yellow for all genotypes. Despite the rather uniform petal color, petal spacing was morphologically varied and all petal spacing morphologies were observed ranging from 1-9 with a mean petal spacing of 5.33 or touching, overlapping flowers.

Leaf characteristics (Table 3.3) were observed approximately 60 days after planting as leaf area trends towards maximum. This allows for a variety of measurements and phenotypic observations that add additional genotypic information for the separation of individual genotypes.

Table 3.3: Leaf characteristics for 79 *Brassica napus* genotypes grown during the 2014 growing season in Winnipeg, MB.

Genotype[1]	Leaf Shape[2] (W/L)	Leaf Length[3] (L)(cm)	Leaf Width[4] (W)(cm)	Leaf Color[5] (1-4)	Leaf Margin[6] (1-3)	Leaf Attachment[7] (1-3)
NEW-620-R	0.53	18.0	9.5	3	3	2
NEW-621-R	0.43	15.0	6.5	2	3	2
Castor-R	0.37	17.5	6.5	1	2	2
UMI-71-R	0.35	25.5	9.0	3	1	2
UMS-189-R	0.31	11.5	3.6	3	3	2
12DH378-R	0.41	13.5	5.5	3	3	2
12DH377 -R	0.41	14.5	6.0	3	3	2
RRHR503-R	0.51	18.5	9.5	2	3	2
RRHR204-R	0.32	11.0	3.5	2	3	2
RRHR404-R	0.46	13.0	6.0	3	3	2
RRHR5815-R	0.35	17.0	6.0	3	3	2
RRHR5819-R	0.32	14.0	4.5	3	1	2
08C702-R	0.42	19.0	8.0	2	3	2
08C712-R	0.27	11.0	3.0	3	3	2
08C847-R	0.46	18.5	8.5	2	2	2
Red River 1997-R	0.51	20.5	10.5	1	3	2
Red River 1826-R	0.28	14.5	4.0	2	1	2
Red River 1852-R	0.33	13.5	4.5	3	2	2
Red River 1997-R2	0.40	17.5	7.0	2	1	2
UMI-55-R	0.23	19.5	4.5	2	2	3
UMI-99-R	0.31	16.0	5.0	3	2	2
12DH384-R	0.35	11.5	4.0	3	1	2
12DH430-R	0.41	20.5	8.5	2	1	3
12DH478-R	0.31	21.0	6.5	3	3	2
12DH915-R	0.41	16.0	6.5	1	2	2
12DH949-R	0.33	18.0	6.0	1	2	2
11DH91-R	0.24	16.5	4.0	1	1	2
11DH92-R	0.23	13.0	3.0	1	2	2
11DH97-R	0.41	19.5	8.0	1	1	2
11DH108-R	0.47	17.0	8.0	1	1	2
11DH109-R	0.30	16.5	5.0	3	3	2
11DH114-R	0.25	16.0	4.0	3	3	2
11DH122-R	0.43	21.0	9.0	3	2	2
11DH137-R	0.38	18.5	7.0	3	3	2
11DH144-R	0.44	18.0	8.0	3	3	2
11DH148-R	0.36	19.5	7.0	3	3	2
11DH149-R	0.32	19.0	6.0	3	3	2
11DH162-R	0.42	18.0	7.5	3	3	2
Red River 1826-B	0.29	17.5	5.0	3	3	2
Red River 1852-B	0.23	13.0	3.0	3	2	2
Red River 1997-B	0.43	17.5	7.5	3	3	2

Genotype[1]	Leaf Shape[2] (W/L)	Leaf Length[3] (L)(cm)	Leaf Width[4] (W)(cm)	Leaf Color[5] (1-4)	Leaf Margin[6] (1-3)	Leaf Attachment[7] (1-3)
Red River 1861-B	0.32	18.5	6.0	3	3	2
RRHR503-B	0.22	18.0	4.0	3	3	2
RRHR204-B	0.27	13.0	3.5	3	2	2
RRHR404-B	0.43	11.5	5.0	3	3	2
RRHR5815-B	0.39	18.0	7.0	3	3	2
RRHR5819-B	0.24	17.0	4.0	2	3	2
08C702-B	0.30	20.0	6.0	3	3	2
08C712-B	0.23	15.0	3.5	1	3	2
08C847-B	0.38	16.0	6.0	3	2	2
RRHR9707-B	0.33	15.0	5.0	3	2	2
UMI-B	0.39	25.5	10	3	2	2
UMSA-B	0.15	23.5	3.5	3	1	2
UMSH-B	0.42	29.5	12.5	3	1	1
Venus-B	0.35	13.0	4.5	2	2	2
Neptune-B	0.48	14.5	7.0	2	2	2
Castor-B	0.44	17.0	7.5	2	1	2
Mil 03-B	0.18	19.0	3.5	3	3	2
Mercury-B	0.31	16.0	5.0	3	2	2
Hero-B	0.53	15.0	8.0	3	1	2
08C344-B	0.31	13.0	4.0	3	3	2
Reston-B	0.35	17.0	6.0	3	1	2
79R713-B	0.46	14.0	6.5	3	2	2
79R714-B	0.48	11.5	5.5	3	1	2
79R728-B	0.21	14.0	3.0	3	3	2
79R729-B	0.31	13.0	4.0	2	1	2
ER2-B	0.30	25.0	7.5	2	1	2
ER3-B	0.49	29.5	14.5	2	1	2
ER7-B	0.41	49.0	20.0	2	1	2
ER22-B	0.34	35.0	12.0	3	1	2
ZSDH2602-B	0.38	30.0	11.5	2	1	2
Topas-B	0.32	22.0	7.0	3	3	2
Polo-B	0.39	14.0	5.5	2	3	2
Sentry-B	0.29	17.5	5.0	2	1	2
Global-B	0.21	19.0	4.0	3	2	2
Westar-B	0.17	17.5	3.0	2	1	2
Apollo-B	0.30	10.0	3.0	3	1	2
Stellar-B	0.39	14.0	5.5	3	1	2
02R276-B	0.44	13.5	6.0	3	2	2
Mean (μ)	0.35	17.70	6.32	2.51	2.11	2.01
Standard Dev (σ)	0.09	5.80	2.85	0.69	0.86	0.19
CV (%)	25.71	32.70	45.10	27.50	40.75	9.45

[1] R = restorer genotype, B = maintainer genotype for *ogu*-INRA hybrid system.[2]Leaf shape is the ratio of leaf width divided by length.[3,4]Leaf length and width are mean values for 3 random

leaves per row measured in cm.[5]Leaf color is based on color in the entire row.[6,7]Leaf margin shape and leaf attachment are the mean of three observations per row.

Leaf morphology for the seventy-nine genotypes exhibited a wide range of diversity (Table 3.3). Diversity was encountered for leaf length ranging from 11 cm – 29.5 cm with a mean of 17.70 cm. Greater variation was seen in leaf width, with a range of 3 cm – 14.5 cm and a mean 6.32 cm. However, despite the range in leaf length and width, only one class of leaf shape was observed by dividing leaf width by leaf length. This represented a narrow elliptical leaf shape (>0.66) as defined by the Variety Registration Office of Canada, Rapeseed Objective Description Form. This classification showed that although wide differences were seen in leaf length and leaf width, all varieties conformed to a similar leaf shape pattern. Leaf color had a mean of 2.51, or a light green phenotype. No genotypes exhibited the dark green/blue phenotype for leaf color. Leaf margin shape exhibited the full range of variation from 1-smooth edges to 3-sharp serrated edges. To the contrary of other leaf characteristics, leaf attachment was generally uniform within the seventy-nine genotypes displaying the partially clasping phenotype and only three genotypes (UMI-55-R, 12DH430-R and UMSH-B) exhibited different clasping morphology. UMI-55-R and 12DH430-R exhibited non-clasping leaf morphology while UMSH-B was the only genotype to exhibit the complete clasping phenotype.

Harvest characteristics (Table 3.4) were measured at approximately time of maturity. This category contains several very important agronomic components, specifically lodging, time to maturity and yield. These three components have dramatic influence for the Canadian growing region as the growing season can be short and many producers prefer to direct harvest *B. napus* varieties saving time and money.

Table 3.4: Harvest characteristics and yield components for 79 *Brassica napus* genotypes grown during 2014 in Winnipeg, MB.

Genotype[1]	Plant Height[2] (cm)	Lodging[3] (1-5)	Pod Length[4] (cm)	Pod Angle[5] (1-9)	Time to Maturity[6] (days)	Yield[7] (g/row)
NEW-620-R	85	2	6.0	1	107	212
NEW-621-R	80	1	4.5	1	107	100
Castor-R	93	1	6.5	5	106	218
UMI-71-R	97	2	6.5	1	106	179
UMS-189-R	85	1	4.5	1	107	162
12DH378-R	90	1	6.5	3	108	230
12DH377 -R	94	1	6.0	3	109	254
RRHR503-R	96	1	5.0	1	111	195
RRHR204-R	85	1	5.0	1	109	156
RRHR404-R	112	1	4.0	1	107	164
RRHR5815-R	85	3	5.0	1	107	48
RRHR5819-R	80	3	5.0	1	106	124
08C702-R	92	3	5.5	1	106	305
08C712-R	87	2	4.0	1	108	78
08C847-R	98	2	4.0	1	108	279
Red River 1997-R	100	2	4.0	1	109	189
Red River 1826-R	90	1	4.0	1	109	92
Red River 1852-R	87	1	5.0	3	108	261
Red River 1997-R2	105	1	5.5	1	109	82
UMI-55-R	95	2	5.5	1	108	231
UMI-99-R	94	1	4.0	1	106	131
12DH384-R	84	3	4.5	1	106	261
12DH430-R	87	3	5.0	1	107	65
12DH478-R	100	1	4.0	3	108	313
12DH915-R	80	1	5.0	3	109	39
12DH949-R	80	3	5.5	1	110	124
11DH91-R	90	2	6.0	1	110	171
11DH92-R	90	1	6.0	1	111	139
11DH97-R	93	1	5.0	1	106	288
11DH108-R	81	3	5.5	1	107	211
11DH109-R	90	3	6.5	1	106	187
11DH114-R	85	3	6.0	1	107	233
11DH122-R	110	2	5.5	1	108	271
11DH137-R	110	1	5.0	1	108	237
11DH144-R	100	1	5.0	1	108	186
11DH148-R	100	1	5.0	1	109	220
11DH149-R	112	1	4.5	1	109	224
11DH162-R	90	1	5.0	3	106	153
Red River 1826-B	90	1	6.5	3	110	391
Red River 1852-B	95	1	6.0	1	110	311
Red River 1997-B	85	2	6.5	5	109	293

Genotype[1]	Plant Height[2] (cm)	Lodging[3] (1-5)	Pod Length[4] (cm)	Pod Angle[5] (1-9)	Time to Maturity[6] (days)	Yield[7] (g/row)
Red River 1861-B	90	3	6.0	3	106	228
RRHR503-B	90	2	5.5	1	108	373
RRHR204-B	83	4	5.5	1	107	199
RRHR404-B	95	2	6.0	1	106	302
RRHR5815-B	80	3	6.5	1	106	127
RRHR5819-B	85	2	7.0	1	106	258
08C702-B	100	1	8.0	5	107	335
08C712-B	85	2	7.5	5	106	246
08C847-B	86	2	7.0	3	107	196
RRHR9707-B	85	2	8.0	5	107	334
UMI-B	100	1	5.5	1	107	234
UMSA-B	90	2	5.0	1	107	183
UMSH-B	106	2	6.0	1	108	103
Venus-B	80	4	6.5	1	109	89
Neptune-B	82	3	7.0	3	109	280
Castor-B	82	3	7.0	1	108	313
Mil 03-B	80	3	7.5	3	107	213
Mercury-B	90	2	7.0	3	107	346
Hero-B	81	3	5.5	1	106	243
08C344-B	90	4	6.0	1	107	105
Reston-B	92	4	5.5	3	108	205
79R713-B	87	2	5.0	1	109	136
79R714-B	109	3	5.5	1	109	139
79R728-B	80	1	6.0	1	109	137
79R729-B	87	2	5.5	1	109	138
ER2-B	120	3	6.0	5	116	0
ER3-B	120	1	5.0	5	116	7
ER7-B	150	1	7.0	5	116	0
ER22-B	140	1	4.5	5	116	35
ZSDH2602-B	85	1	5.0	5	116	9
Topas-B	100	1	4.5	5	110	72
Polo-B	110	1	5.0	1	112	6
Sentry-B	81	2	4.0	1	110	50
Global-B	90	1	6.0	3	106	330
Westar-B	81	2	6.0	3	110	113
Apollo-B	90	2	5.0	1	107	241
Stellar-B	80	1	5.0	1	110	85
02R276-B	84	2	7.0	1	109	245
Mean (μ)	92.8	1.9	5.6	1.9	108.4	185.6
Standard Dev (σ)	12.9	0.9	1.0	1.5	2.4	95.7
CV(%)	13.9	48.2	17.5	74.7	2.3	51.6

[1] R = restorer genotype, B = maintainer genotype for *ogu*-INRA hybrid system.[2]Plant height is the row mean from soil to tip.[3]Lodging is the row mean using a 1-5 scale.[4]Pod length is the mean

of three random pods per row.[4]Pod angle is the mean of three random plants per row.[5]Time to maturity is measured in days post seeding for 50% of ripe pods per row.[6]Yield is grams of seed harvested per row.

Plant height for all 79 genotypes had a range of 80-150 cm and a mean height 92.8 cm. All ER lines were significantly higher than the mean and ER7 was the tallest genotype at 150 cm. Lodging was generally quite low for all 79 genotypes with a mean lodging value of 1.89 and a range of 1-4. No genotypes achieved a rating of 5 (flattened or broken stems). Pod length (selected from 3 random pods per row) had a mean value of 5.59 cm. RRHR9707-B had the longest pod mean at 8 cm of all genotypes. Pod angle (1-9) was generally uniform amongst the 79 genotypes with a mean value of 1.9 and a range of 1-5. No genotypes exhibited the 7-slightly droopy or 9-droopy classification. Time to maturity was delayed due to heavy rains in late August and early September of 2014. As a result, maturity dates are inflated due to the excess moisture. Time to maturity at the Winnipeg location had a mean of 108.4 and a range of 106-116 days. Genotypes ER-2 and ER-7 did not reach maturity by harvest date and failed to produce viable seed due to their unadapted background. From a producer standpoint, yield is one of the most important traits. Mean row yield for all 79 genotypes was 185.6 g. Red River 1826-B had the highest row yield with 391.0 g, more than double the mean. Eight other maintainer genotypes (Global-B, Mercury-B, Castor-B, RRHR9707-B, 08C702-B, RRHR404-B, RRHR503-B, Red River 1852-B) and two restorer genotypes (08C702-R, 12DH478-R) also produced over 300 g/row.

Seed quality components (Table 3.5) must be monitored and are necessary to ensure that specific genotypes retain certain aspects of quality. Currently in Canada, varietal registrations for canola quality genotypes follow strict guidelines for total oil and protein concentration in addition to oil

profile and glucosinolate restrictions. High erucic acid rapeseed genotypes are also restricted and cannot contain less than 50% erucic acid content.

Table 3.5: Seed quality components for 79 *Brassica napus* genotypes grown during 2014 in Winnipeg, MB.

Genotype[1]	Seed Coat Color[2] (1-5)	Protein[3] (%)	Oil[4] (%)	Glucosinolates[5] (µmol/g)	Erucic Acid[6] (%)
NEW-620-R	2	33.3	38.5	24.8	49.9
NEW-621-R	2	31.9	33.7	34.8	46.9
Castor-R	2	31.3	39.5	19.2	52.0
UMI-71-R	2	32.5	40.1	28.9	54.1
UMS-189-R	2	29.6	39.9	23.3	47.1
12DH378-R	1	29.4	43.0	18.0	42.2
12DH377-R	2	29.6	43.9	17.5	43.8
RRHR503-R	2	31.5	41.8	18.7	48.3
RRHR204-R	2	30.7	42.0	18.6	42.8
RRHR404-R	4	32.7	37.0	19.2	50.2
RRHR5815-R	4	36.9	33.5	27.3	39.9
RRHR5819-R	2	30.5	40.8	25.2	42.4
08C702-R	3	29.5	42.7	17.0	55.2
08C712-R	2	32.3	42.2	19.2	52.7
08C847-R	2	30.3	45.0	12.7	51.0
Red River 1997-R	2	33.1	40.3	19.2	54.0
Red River 1826-R	2	29.9	42.3	13.3	50.7
Red River 1852-R	2	28.3	42.1	11.7	51.0
Red River 1997-R2	2	31.1	42.3	16.5	54.5
UMI-55-R	4	29.3	43.6	14.3	52.1
UMI-99-R	4	30.8	40.9	29.5	50.1
12DH384-R	2	29.3	43.0	19.1	47.9
12DH430-R	2	33.6	40.3	16.6	49.7
12DH478-R	3	30.9	42.6	16.5	55.6
12DH915-R	2	31.3	40.2	16.8	51.0
12DH949-R	2	32.7	39.6	13.1	51.2
11DH91-R	2	28.6	41.2	19.4	48.8
11DH92-R	2	31.5	39.3	19.8	46.8
11DH97-R	1	27.2	42.3	17.8	44.6
11DH108-R	2	27.9	42.0	15.8	43.4
11DH109-R	2	29.5	42.2	17.9	42.5
11DH114-R	1	27.4	44.5	17.1	48.6
11DH122-R	2	29.5	42.8	19.4	47.4
11DH137-R	2	29.6	41.9	18.0	49.4

Genotype[1]	Seed Coat Color[2] (1-5)	Protein[3] (%)	Oil[4] (%)	Glucosinolates[5] (µmol/g)	Erucic Acid[6] (%)
11DH144-R	1	30.9	42.3	23.5	41.9
11DH148-R	2	28.6	42.9	21.4	47.6
11DH149-R	2	31.2	41.8	28.9	46.0
11DH162-R	2	30.7	42.3	18.0	49.3
Red River 1826-B	1	27.3	47.5	8.7	55.8
Red River 1852-B	1	27.8	47.0	14.9	50.8
Red River 1997-B	4	27.4	42.5	10.9	51.1
Red River 1861-B	2	29.3	44.5	11.7	52.8
RRHR503-B	2	29.7	44.3	21.1	50.6
RRHR204-B	4	30.7	42.6	19.5	47.2
RRHR404-B	4	27.5	47.2	10.8	55.6
RRHR5815-B	2	28.8	42.5	20.9	46.7
RRHR5819-B	4	30.5	42.8	14.6	54.4
08C702-B	2	28.4	46.7	13.6	54.9
08C712-B	2	29.4	45.8	16.6	54.1
08C847-B	2	26.8	47.1	8.7	54.3
RRHR9707-B	2	28.0	46.9	13.0	52.8
UMI-B	2	29.4	42.2	17.4	50.2
UMSA-B	4	28.8	44.9	18.6	50.1
UMSH-B	2	29.6	41.4	20.2	50.9
Venus-B	2	29.0	41.9	17.0	53.7
Neptune-B	2	27.5	43.7	14.2	55.7
Castor-B	2	30.0	41.7	11.9	53.6
Mil 03-B	2	30.4	42.6	11.1	54.8
Mercury-B	2	28.0	43.2	13.9	46.0
Hero-B	1	29.0	41.7	14.4	53.2
08C344-B	2	30.1	41.7	19.5	52.1
Reston-B	2	28.7	41.2	69.6	42.6
79R713-B	3	31.3	37.3	36.1	52.6
79R714-B	2	32.4	39.0	26.5	55.9
79R728-B	3	31.9	39.0	24.3	54.1
79R729-B	3	32.1	38.8	25.3	56.7
ER2-B	NA	NA	NA	NA	51.8
ER3-B	2	31.6	42.6	18.5	51.0
ER7-B	NA	NA	NA	NA	56.5
ER22-B	3	31.7	37.2	29.9	57.6
ZSDH2602-B	2	40.4	51.5	9.4	52.1
Topas-B	2	31.9	38.4	20.7	1.1
Polo-B	3	31.4	37.7	37.1	1.2
Sentry-B	2	31.2	37.9	49.0	1.3
Global-B	2	28.9	39.4	65.1	0.2
Westar-B	2	29.0	40.5	21.6	0.8
Apollo-B	1	28.1	42.9	14.6	0.9

Genotype[1]	Seed Coat Color[2] (1-5)	Protein[3] (%)	Oil[4] (%)	Glucosinolates[5] (µmol/g)	Erucic Acid[6] (%)
Stellar-B	2	29.2	40.8	25.7	0.9
02R276-B	4	28.1	45.9	12.4	5.5
Mean (µ)	2.4	30.2	41.9	20.5	45.6
Standard Dev (σ)	1.1	2.2	3.0	10.3	15.4
CV (%)	44.7	7.2	7.1	50.3	33.8

[1] R = restorer genotype, B = maintainer genotype for *ogu*-INRA hybrid system.[2]Seed coat color is the mean seed color for an entire row using a 1-5 scale.[3,4,5,6]Protein percent, oil percent, glucosinolates (µmol/g**)** and erucic acid percent are all mean values taken using a sample from the total collected seed.

Seed coat color (Table 3.5) had a variety of ranges with all phenotypes exept pure yellow. Protein content ranged from 26.8% in genotype 08C847-B to 40.4% in genotype ZSDH2602-B and a mean value of 30.2% for all genotypes. Oil content ranged from 33.5% in genotype RRHR5815-R to 51.5% in genotype ZSDH2602-B with a mean oil content of 41.9% for all genotypes. Glucosinolates were lowest in genotype Red River 1826-B at 8.7 µmol/g and highest in Reston at 69.6 µmol/g with a mean of 20.5 µmol/g for all genotypes. Erucic acid content distinguishes canola quality genotypes from high erucic acid genotypes with canola quality genotypes ranging from < 2% erucic acid and found in genotypes Topas, Polo, Sentry, Global, Westar, Apollo and Stellar. High erucic acid rapeseed genotypes ranged from 39.9% in RRHR5815-R to 57.6% for ER-22 for total erucic acid content. For ER-2 and ER-7, which did not reach maturity in time for the 2014 harvest, erucic acid content was substituted with erucic acid measurements from greenhouse produced seed. Unfortunately no other quality measurements could be collected due to lack of seed for ER-2 and ER-7. Therefore the eight missing data points in Table 3.5 for genotypes ER-2 and ER-7 were substituted during dendrogram creation using greenhouse seed.

3.42 Principal Component Analysis

Principal component analysis of twenty qualitative and quantitative morphological traits was completed on the 79 *B. napus* genotypes in JMP Genomics (V.6.3) to investigate trait-by-trait interaction. Figure 3.2 shows the interaction of all twenty morphological traits. Yield, pod length and oil content were all highly correlated and glucosinolate content and protein content were highly correlated as well. Protein content was negatively correlated to oil content and negatively correlated to yield. Lodging was negatively correlated to plant height as well as time to maturity. Time to flower, time to maturity, plant height and leaf length and width were all correlated (Figure 3.2).

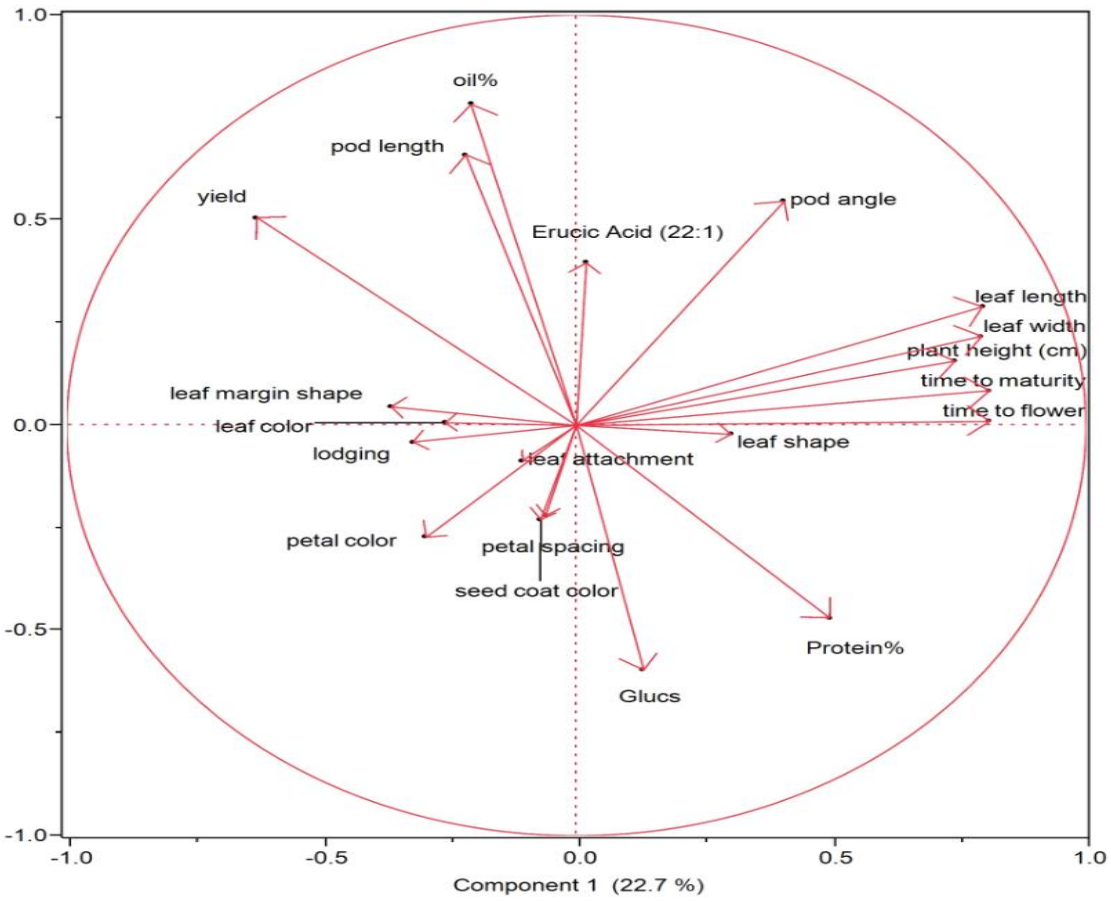

Figure 3.2: Principal component analysis of twenty qualitative and quantitative morphological traits obtained in 2014 on 79 *Brassica napus* genotypes completed in JMP Genomics (V.6.3).

3.43 Genetic distance

Genetic distance as calculated by Euclidean distance based on Ward's method was completed in JMP Genomics (V.6.3) and displayed a Euclidean distance range of 9.21 between cultivars RRHR5815-B and RRHR5819-R and 395.3 between cultivars ER7-B and Red River 1826-B. Table 3.6 presents a select set of distance values obtained.

Table 3.6: Distance matrix between select *Brassica napus* genotypes showing the Euclidean distance between genotypes based on twenty qualitative and quantitative traits as calculated by Ward's method completed in JMP genomics (V.6.3).

Genotype[1]										
NEW-620-R	0									
NEW-621-R	112.8	0								
Castor -R	12.9	120.1	0							
UMI-71-R	36.6	82.4	41.6	0						
UMS-189-R	51.2	64.0	57.7	28.1	0					
12DH378-R	23.0	131.9	17.4	55.7	68.9	0				
12DH377-R	44.9	156.0	37.8	77.8	93.0	24.5	0			
RRHR503-R	22.1	98.2	25.0	22.0	37.3	36.9	59.7	0		
RRHR204-R	57.7	59.5	63.9	34.4	10.1	74.3	98.6	42.2	0	
RRHR404-R	56.0	73.5	58.1	27.8	29.6	70.6	92.5	36.3	30.5	0

[1]Subsample of select genotypes from a larger distance matrix. 0 equals the Euclidian distance between each genotype and itself.

3.44 Multivariate Cluster Analysis

Ward's method clustered all 79 genotypes into eight distinct clusters and two general pools with no outliers (Figure 3.3). Each distinct cluster is colour coded for ease of viewing; however each color has no bearing on clusters constituents. Each genotype within each cluster is also tagged with a –B or –R designation signifying maintainer and restorer genotypes, respectively. These designations are used for parental combinations in the *ogu*-INRA pollination control system (Delourme and Eber, 1992; Gourret *et al.*, 1992; Delourme *et al.*, 1994).

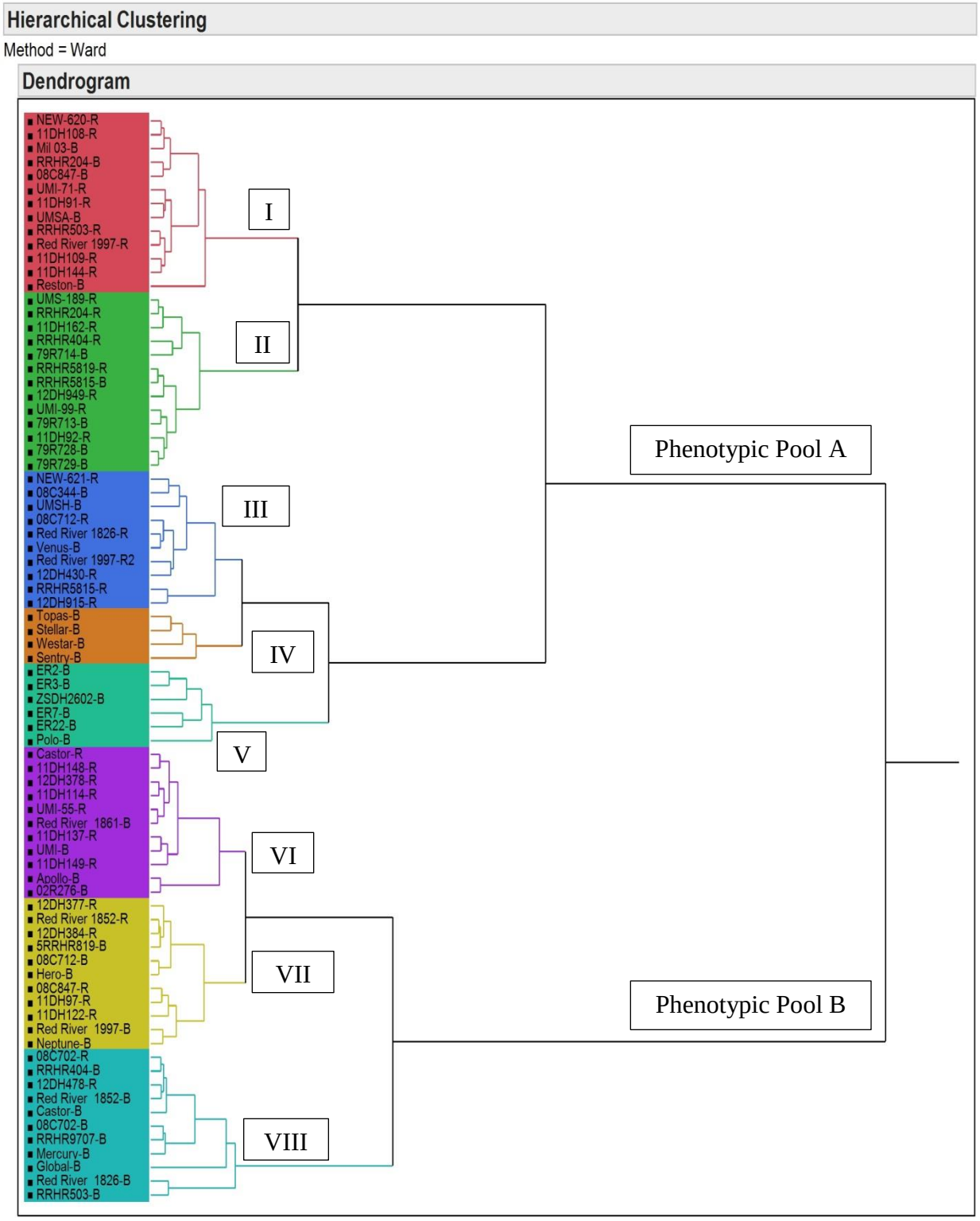

Figure 3.3: Phenotypic dendrogram for 79 *Brassica napus* genotypes based on Euclidean genetic distance and clustered with Ward's methods of hierarchical clustering completed in JMP genomics (V.6.3). R = restorer genotype, B = maintainer genotype for *ogu*-INRA hybrid system.

Phenotypic pool A (Figure 3.3) as clustered by Ward's method consists of five clusters containing forty-six of the seventy-nine genotypes. Cluster I contains 13 genotypes, 8 restorers and 5 maintainers. Cluster II contains 13 genotypes, 8 restorers and 5 maintainers. Of the 5 maintainers in cluster II, four out of five are 79R700. These genotypes have similar pedigree and the clustering of these four genotypes was expected and may lend credibility to the phenotypic evaluation and Ward's clustering method. Cluster III contains 10 genotypes, 7 restorers and 3 maintainers. Cluster IV contains 4 genotypes, all maintainer genotypes and all canola quality genotypes. This cluster is essentially an out-group based on erucic acid content, as canola quality and rapeseed quality are mutually exclusive. Cluster V contains 6 genotypes, all were maintainer genotypes and five out of six are resynthesized *B. napus* genotypes with the sixth represented by the canola cultivar Polo. Phenotypic pool B as clustered by Ward's method results in a smaller grouping of genotypes, consisting of thirty-three out of seventy-nine genotypes in three distinct clusters. Cluster VI contains 11 genotypes, 7 restorers and 4 maintainers. Cluster VII contains 11 genotypes, six restorers and 5 maintainers. This cluster has the most even distribution of restorers and maintainers with almost equal numbers. Cluster VIII contains 11 genotypes, 2 restorers and 9 maintainers. The high propensity of maintainer lines in this cluster suggests that these genotypes have similar performance despite being a genetically unrelated group. The maintainer genotypes in this cluster could represent pairings for restorer genotypes in clusters I and II based on the calculated Euclidian distance (genetic distance).

3.5 DISCUSSION

Phenotyping plant germplasm is, and will continue to be, an important method for discerning agronomically important information (Sultan, 2000). Smith and Smith (1989) suggest that morphological characterization is the first step in the description and classification of the germplasm. In addition, Rabbani *et al*. (1998) suggested that germplasm characterization is imperative for the efficient and effective utilization of germplasm resources.

From the twenty qualitative and quantitative traits several important breeding traits will be discussed. Flowering time is an extremely important phenotypic trait for the Canadian Prairies and adapted cultivars often flower between 40-45 days after planting (DAP) (McVetty *et al*., 2012; McVetty *et al*., 2014). Interestingly, UMI-55-R and UMI-99-R were among the earliest flowering genotypes at 39 DAP, yet are from European origins (Table 3.2). Some of the latest flowering time genotypes were designated as "ER", which are re-synthesized genotypes and flowered 50, 60 or even 70 DAP (Table 3.2). The 60 and 70-day flowering period may be too late to use as breeding material for the Canadian Prairies; however, genotypes within the 50-day mark may be useful. Spring habit *B. napus* cultivars mature approximately 95-100 DAP (McVetty *et al*., 2012; McVetty *et al*., 2014), which is imperative to avoid late season frost (McVetty and Duncan, 2015). Time to maturity was over inflated due to environmental effects for all 79 genotypes and does not accurately represent the true maturity date for these genotypes. Luckily, no late season frost occurred during 2014; however these values extend harvest into late September where frost risk is a serious concern.

From the 79 genotypes phenotyped in 2014, highlights include several important agronomic traits for ZSDH2602-B, a resynthesized genotype (Cluster V). ZSDH2602-B had the highest

protein content (40.4%), second lowest glucosinolate content (9.4 μmol/g), highest oil content (51.5%) and high erucic acid content (52.1%). Morphologically, ZSDH2602-B was a shorter plant (85 cm) compared too many other genotypes and showed excellent lodging resistance (1). ZSDH2602-B has a later flowering date (54 DAP), later harvest date (116 DAP) and low yield (9 g/row). These drawbacks may hinder the use of ZSDH2602-B in a main stream breeding capacity; however, this genotype has excellent potential in a pre-breeding capacity and should be considered for oil content, protein content, height reduction and lodging resistance. Jesske *et al.* (2013) investigated the use of resynthesized lines in hybrid development and found that hybrids of domesticated *B. napus* and resynthesized *B. napus* produced poor quality hybrids. However, Jesske *et al*. (2013) found that resynthesized lines possess a genetic diversity not seen in domesticated *B. napus* breeding material. The current study, grouped the resynthesized lines into a distinct cluster, supporting the conclusions of Jesske *et al.* (2013) that resynthesized are phenotypically different and diverse. Thus, this diversity could be used in a pre-breeding capacity to broaden the genetic diversity in elite breeding lines and introduce novel and/or other agronomically important traits.

A principal component analysis (PCA) was conducted on the twenty qualitative and quantitative traits to investigate any correlation between phenotypic characteristics. PCA found a high positive correlation between yield, total oil content and pod length. Yield was negatively correlated to total protein content and glucosinolate content. Several studies have shown that the interaction between oil and protein correlation is a negative correlation in *B. napus* (Zhao *et al*., 2006; Gül *et al*., 2003). From a commercial breeding standpoint, yield is the most important trait and this correlation may help drive yield gains with the concurrent reduction in protein content and glucosinolate content. Pod length was also highly correlated to yield, which one would

expect, as longer pods equate to more seed, which in turn increases yield (Yang *et al*., 2012). Lodging was negatively correlated to plant height, which would be expected if plants that were lodging were not held up when plant height was measured. However, for this experiment, if lodging was encountered during plant height measurement, plants were physically held up to the maximum height. This correlation is in opposition of the general consensus that plant height has a close association with lodging (Wang *et al*., 2015; Udall *et al*., 2006). One hypothesis for these results could be that as lodging occurred it prevented proper elongation of the plant stem and therefore higher lodging genotypes remained shorter than non-lodging genotypes. Flowering time was highly correlated to maturity time and again this is expected as later flowering can cause later maturity (Amiri-Oghan *et al*., 2009). This is also in agreement with the findings from Amiri-Oghan *et al*. (2009) who found a direct correlation between early and late flowering varieties with maturity. Despite the positive association between certain traits, the coefficient of variation (CV) was extremely high for several traits (ie lodging, pod angle and yield) and these correlations may not be statistically valid and should be repeated.

Euclidean distance was calculated using all twenty qualitative and quantities traits with equal weight and Table 3.6 showcases the genetic distance for select lines. The smallest equitable distance between two genotypes from all 79 genotypes was 9.2 between genotypes RRHR5815-B and RRHR5819-R, a maintainer and restorer, respectively. Although these genotypes differ in their designation (-B and -R), we can infer that they are highly related, not only based on pedigree, but using the phenotypic scores that were collected in this study. In the past, restorer lines were generally of poor agronomic quality (Pelletier *et al.*, 1983). However, a great deal of effort has gone into the development of high quality *ogu*-INRA restorers (Primard-Brisset *et al*., 2005) and this study supports these conclusions as these represent a maintainer and restorer with

similar agronomic performance. The largest Euclidean distance was 395.3 calculated between genotypes ER7-B and Red River 1826-B. Several studies support this conclusion as resynthesized *B. napus* lines (ER-7) possess distinct allele combinations when compared to elite *B. napus* breeding material (Jesske *et al.*, 2013; Seyis *et al*., 2003).

It is theorised that high heterotic crosses can be obtained by selectively crossing genotypes from different clusters using heterotic pools based on genetic distance (Grant and Beversdorf, 1985; Riaz *et al.*, 2001). Ali *et al.* (1995) had significant success predicting hybrid heterosis in seed yield with Euclidian genetic distance based on morphological yield components (r = 0.70). In addition, Xing *et al*. (2014) found a correlation between parents and hybrids for seed yield at r = 0.76 (P = 0.01) in *B. napus*. However, Yu *et al*. (2005) found the prediction of hybrid heterosis based on the genetic distance from three different methods (morphological, isozymes and RAPD) to be inconsistent. Using the current dendrogram based on qualitative and quantitative cluster analysis, cluster VIII displays a high proportion of high yielding maintainer lines which could be suitable for crosses with restorer lines in clusters I and II. These crosses would maximize the Euclidian distance (genetic distance) between parental lines and may offer high heterotic yield potential as all crosses would be from different clusters and different heterotic pools (Riaz *et al.*, 2001). Cluster V, which contains all resynthesized lines except for polo-B, may provide useful and/or novel agronomic traits. Cluster V could be used to increase the genetic diversity of an elite *B. napus* gene pool; however, their usefulness in direct hybrid breeding may be of limited potential due to the poor agronomic performance of hybrids directly from resynthesized genotypes and elite breeding lines due to unwanted linkage drag from unadapted genotypes (Jesske *et al.*, 2013).

Overall, it is difficult to compare these results with other published dendrograms within *B. napus.* A majority of cluster analyses are based on DNA fingerprinting (Mailer *et al.*, 1994; Hasan *et al.*, 2006; Becker *et al.*, 1995) and few use phenotypic measurements (Ali *et al.*, 1995) and include spring or winter canola quality genotypes as opposed to rapeseed quality dendrograms. Diers and Osborne (1994) do have a cluster analysis that includes some genotypes in this study. Compared to Diers and Osborne (1994), the genetic analysis based on restriction fragment length polymorphisms (RFLPs), the genotypes Stellar, Hero and Westar cluster together with Topas one cluster away. In the current cluster analysis, Topas, Stellar and Westar cluster together (cluster IV) and Hero is located in cluster VII. Ultimately the success of heterotic pool placement will need to be investigated by creating hybrids using parents from different clusters along with the yield evaluation of the inbred parental lines and hybrids. These experiments will need to be evaluated in multiple environments in order to truly evaluate the influence phenotypic pool placement has on hybrid heterosis and the predictive power for high heterotic parental combinations. Chapter 5 explores this hypothesis with genetic distance measurements between parental genotypes along with hybrid yield.

Following heterotic pool accuracy testing, model optimization should be untaken. For this study, twenty qualitative and quantitative traits were used with equal weight for cluster construction. However, Ali *et al.* (1995) had great success predicting heterosis for seed yield using only yield components in *B. napus*. Ali *et al*. (1995) found that other characteristics such as leaf length and width contributed little toward cluster differentiation. In addition, Rabbani *et al*. (1998) using 35 morphological traits and PCA, found 21 traits contributed little to the apparent variation in *Brassica juncea* L., and could be discarded. Based on these findings, and integrating the results from the coefficient of variation (CV) from each table in this study, 13 traits (petal color, petal

spacing, leaf shape, leaf length, leaf width, leaf color, leaf margin, seed color, lodging, pod length, pod angle, glucosinolates and yield) have a CV over 15% and should be discarded. Future investigations should concentrate on calculating Euclidean genetic distance based on the remaining seven traits with the aim of improving high heterotic hybrid prediction.

Phenotypic classification using qualitative and quantitative traits has some level of applicability in determining genetic diversity and for separating genotypes into heterotic pools. Comparison of parental combinations based on this cluster analysis will allow for a better understanding of cluster accuracy and may generate an improved predictive model, ultimately reducing the number of crosses required to discover high heterotic parental combinations.

The next chapter, Chapter 4: Genotypic Characterization of *Brassica napus* L. Genotypes Utilizing SRAP and GBS in Association with Cluster Analysis will conduct a similar evaluation with genetic distance based on two genotypic methods. Chapter 5 will discuss the evaluation of hybrid performance based on genetic distance and cluster analysis to gauge if genetic distance can be used as a high heterotic predictive model.

4.0 Genotypic Characterization of *Brassica napus* L. Genotypes Utilizing SRAP and GBS in Association with Cluster Analysis

4.1 ABSTRACT

Identifying high heterotic combinations is a constant target for commercial *Brassica napus* L. hybrid development programs. Finding high heterotic parental combinations may require hundreds of testcrosses and years of yield evaluation. Heterotic pool development could be used to divide breeding material into specific breeding pools and focus the number of parental combinations created. Here, we report the genotypic characterization of seventy-nine *B. napus* genotypes by calculating genetic distance based on sequence related amplified polymorphism (SRAP) and genotyping-by-sequencing (GBS) in association with a neighbour joining clustering algorithm. Despite the different genotypic methods performed, neighbour joining cluster analysis based on genetic distance of SRAP and GBS produced similar heterotic clusters. Homology between SRAP and GBS clusters was approximately 77% when manually comparing clusters and 68% when comparing clusters using the software Compare2trees. This research demonstrates that SRAP can be as effective as next generation sequencing technologies for heterotic pool classification. This information may provide an important breeding scaffold for the development of hybrid cultivars based on genetic distance and cluster analysis.

4.2 INTRODUCTION

Brassica napus L. (AACC n =19) has been heralded as one of the most important global oilseed crops, second only to soybean (*Glycine max* L.) for total annual production (Carré and Pouzet, 2014; Wittkop *et al.*, 2009, Oil World, 2014, Lin *et al.*, 2013). This recent boom in production is attributed to the improvement of canola quality (<2% erucic acid, < 30 μmol/g glucosinolates) cultivars developed in Canada that produce edible oil for human consumption and meal for animal feed (Rahman, 2013; Li *et al*., 2014). Additionally, high erucic acid rapeseed (HEAR) cultivars also contribute valuable industrial oil for lubricants and slip agents for a niche oleochemical market (Wittkop *et al*., 2009). Currently, hybrid *B. napus* cultivars have replaced their open-pollenated counterparts due to the exploitation of heterosis (Rahman, 2013). Heterosis occurs when progeny outperform both parents in a variety of agronomic traits but specifically seed yield (Shull, 1908; Radoev *et al*., 2008). Therefore, hybrid cultivars that possess increased yield and superior uniformity in the F_1 generation are highly sought after by producers to continually match global demand (Ahmad and Quiros, 2011). Finding parental combinations that exhibit high heterotic gains may require hundreds of test crosses and years of yield evaluation. To alleviate this problem, heterotic pool development has been suggested as a method to separate germplasm accessions and breeding lines into distinct heterotic pools for breeding purposes (Rahman, 2013; Odong *et al*., 2011; Girke *et al*., 2011).

Distinguishing genotypes and the development of heterotic pools can be based on genetic distance and cluster analysis (Ali *et al*., 1995; Ahmad and Quiros, 2011; Jesske *et al*., 2013). Genetic distance is a measure of genetic divergence between species or between individuals and populations within the same species (Nei, 1987). Genetic distance can be calculated using a variety of genotypic data sets produced using multiple molecular characterization methods (Jain

et al., 1994; Becker *et al*., 1995; Lombard *et al*., 2000) along with numerous mathematical formulas (Nei, 1972; Cavalli-Sforza and Edwards, 1967; Reynolds *et al*., 1983; Tamura and Nei, 1993). Several forms of cluster analysis exist including Ward's method (Ward, 1963), Unweighted Pair Group Method with Arithmetic mean [UPGMA; (Sokal and Michener, 1958)] and neighbour joining [NJ; (Saitou and Nei, 1987)] which are all agglomerative (bottom-up) clustering techniques. Applying clustering algorithms to genetic distance allows the grouping of genotypes with a short genetic distance together into a cluster or pool (Odong *et al*., 2011). Genetic distance and cluster analysis has been employed in *B. napus* hybrid development (Diers *et al*., 1996; Riaz *et al*., 2001; Yu *et al*., 2005). Previous research has shown that *B. napus* parental crosses from different pools or clusters exhibit higher heterosis than those using parents from the same cluster or pool (Ali *et al*., 1995; Ahmad and Quiros, 2011). Although clustering techniques have been used for decades in plant breeding, there is no scientific consensus on which clustering method produces the most accurate cluster set for breeding purposes (Mouchet *et al*., 2008; Odong *et al*., 2011). Therefore, the molecular characterization method, genetic distance method and clustering method create a large combination of experimental designs that can be used to develop and explore heterotic pools for hybrid development.

Sequence Related Amplified Polymorphism (SRAP, Li and Quiros, 2001) has been a popular molecular method for the identification of genetic diversity in many crop species including *Cucurbita pepo* L. (Ferriol *et al*., 2003), *Prunus persica* L. (Ahmad *et al*., 2004) and *Lycopersicon esculentum* L. (Ruiz *et al*., 2005). In addition, SRAP is less complex than amplified fragment length polymorphism (Li and Quiros, 2001) and has previously been successful in associating genetic distance calculated on presence/absence markers with hybrid heterosis based on cluster analysis in *B. napus* (Riaz *et al*., 2001; Ahmad and Quiros, 2011).

However, today's next-generation sequencing technologies have reduced the cost of DNA sequencing, allowing the ability to evaluate genetic diversity based on single nucleotide polymorphisms (SNPs) (Nielson *et al*., 2011; Elshire *et al.*, 2011). Genotyping-by-sequencing (GBS) is one such next generation method that can be used for SNP discovery which uses reduced genome representation with methylation sensitive restriction enzymes (Elshire *et al.*, 2011). Through methylation sensitivity, repetitive regions of the genome can be avoided which greatly simplifies the computational challenge of genome alignment with large, polyploidy genomes (Elshire *et al.*, 2011). To date, several important crop species including corn (*Zea maize* L.), barley (*Hordeum vulgare* L.) and rice (*Oryza sativa* L.) have used GBS for a multitude of downstream applications (Elshire *et al.*, 2011; Spindel *et al*., 2013; Liu *et al*., 2014). However, no published reports into *B. napus* diversity and heterotic pool development have been completed using GBS.

The objective of this study was to calculate genetic distance between 79 *B. napus* genotypes using Nei's standard genetic distance based on SRAP presence/absence genotyping and the Tamuri-Nei genetic distance formula based on SNP's discovered through GBS. A neighbour joining clustering method was then used on each separate genetic distance method and the results were compared. Ultimately, the goal of this research is to improve heterotic pools definitions through multiple techniques and to find a system that will reduce the overall time and space required for discovering high heterotic parental combinations.

4.3 MATERIALS AND METHODS

4.31 *Brassica napus* Genotypes

Seventy-nine *B. napus* genotypes were selected for this study (Table 4.1). Most genotypes are considered spring habit germplasm developed for western Canada, with the addition of several European and resynthesized *B. napus* genotypes. Several genotypes are of canola quality, while the vast majority are considered as high erucic acid rapeseed (HEAR). Of the 79 genotypes, 38 are *ogu*-INRA restorers and 41 are open pollinated genotypes (maintainers).

Table 4.1: Seventy-nine spring habit *Brassica napus* genotypes with origin, quality and maintainer or restorer designation for the *ogu*-INRA hybrid system.

Identification	Origin[1]	Quality[2]	Maintainer/Restorer
NEW-620-R	U of M	HEAR	Restorer
NEW-621-R	U of M	HEAR	Restorer
Castor-R	U of M	HEAR	Restorer
UMI-71-R	European	HEAR	Restorer
UMS-189-R	European	HEAR	Restorer
12DH378-R	U of M	HEAR	Restorer
12DH377 -R	U of M	HEAR	Restorer
RRHR503-R	U of M	HEAR	Restorer
RRHR204-R	U of M	HEAR	Restorer
RRHR404-R	U of M	HEAR	Restorer
RRHR5815-R	U of M	HEAR	Restorer
RRHR5819-R	U of M	HEAR	Restorer
08C702-R	U of M	HEAR	Restorer
08C712-R	U of M	HEAR	Restorer
08C847-R	U of M	HEAR	Restorer
Red River 1997-R	U of M	HEAR	Restorer
Red River 1826-R	U of M	HEAR	Restorer
Red River 1852-R	U of M	HEAR	Restorer
Red River 1997-R2	U of M	HEAR	Restorer
UMI-55-R	European	HEAR	Restorer
UMI-99-R	European	HEAR	Restorer
12DH384-R	U of M	HEAR	Restorer
12DH430-R	U of M	HEAR	Restorer
12DH478-R	U of M	HEAR	Restorer
12DH915-R	U of M	HEAR	Restorer
12DH949-R	U of M	HEAR	Restorer

Identification	Origin[1]	Quality[2]	Maintainer/Restorer
11DH91-R	U of M	HEAR	Restorer
11DH92-R	U of M	HEAR	Restorer
11DH97-R	U of M	HEAR	Restorer
11DH108-R	U of M	HEAR	Restorer
11DH109-R	U of M	HEAR	Restorer
11DH114-R	U of M	HEAR	Restorer
11DH122-R	U of M	HEAR	Restorer
11DH137-R	U of M	HEAR	Restorer
11DH144-R	U of M	HEAR	Restorer
11DH148-R	U of M	HEAR	Restorer
11DH149-R	U of M	HEAR	Restorer
11DH162-R	U of M	HEAR	Restorer
Red River 1826-B	U of M	HEAR	Maintainer
Red River 1852-B	U of M	HEAR	Maintainer
Red River 1997-B	U of M	HEAR	Maintainer
Red River 1861-B	U of M	HEAR	Maintainer
RRHR503-B	U of M	HEAR	Maintainer
RRHR204-B	U of M	HEAR	Maintainer
RRHR404-B	U of M	HEAR	Maintainer
RRHR5815-B	U of M	HEAR	Maintainer
RRHR5819-B	U of M	HEAR	Maintainer
08C702-B	U of M	HEAR	Maintainer
08C712-B	U of M	HEAR	Maintainer
08C847-B	U of M	HEAR	Maintainer
RRHR9707-B	U of M	HEAR	Maintainer
UMI-B	European	HEAR	Maintainer
UMSA-B	European	HEAR	Maintainer
UMSH-B	European	HEAR	Maintainer
Venus-B	U of M	HEAR	Maintainer
Neptune-B	U of M	HEAR	Maintainer
Castor-B	U of M	HEAR	Maintainer
Mill 03-B	U of M	HEAR	Maintainer
Mercury-B	U of M	HEAR	Maintainer
Hero-B	U of M	HEAR	Maintainer
08C344-B	U of M	HEAR	Maintainer
Reston-B	U of M	HEAR	Maintainer
79R713-B	U of M	HEAR	Maintainer
79R714-B	U of M	HEAR	Maintainer
79R728-B	U of M	HEAR	Maintainer
79R729-B	U of M	HEAR	Maintainer
ER2-B	Resynthesized	HEAR	Maintainer
ER3-B	Resynthesized	HEAR	Maintainer
ER7-B	Resynthesized	HEAR	Maintainer
ER22-B	Resynthesized	HEAR	Maintainer
ZSDH2602-B	Resynthesized	HEAR	Maintainer

Identification	Origin[1]	Quality[2]	Maintainer/Restorer
Topas-B	U of M	Canola	Maintainer
Polo-B	U of M	Canola	Maintainer
Sentry-B	U of M	Canola	Maintainer
Global-B	U of M	Canola	Maintainer
Westar-B	U of M	Canola	Maintainer
Apollo-B	U of M	Low linolenic	Maintainer
Stellar-B	U of M	Low linolenic	Maintainer
02R276-B	U of M	High oleic low linolenic	Maintainer

[1]U of M is the abbreviation for University of Manitoba.
[2]HEAR is the abbreviation for High Erucic Acid Rapeseed.

4.32 Greenhouse Production

Initially, three replicates for each genotype were planted at a depth of 1 cm into plastic 4 x 3 well cell packs (13 cm x 13 cm x 5 cm) containing Sunshine Metro Mix potting soil (Sungro® Horticulture, 770 Silver Street Agawam, MA, USA) during spring of 2014. Plants were grown in a growth chamber (temperature: day: 22ºC, night: 18ºC, light cycle: 16 hours light, 8 hours dark) and watered daily. At the two-leaf stage (14 days after planting [DAP]) each plant was transferred to a plastic growers pot (14.5 cm x 15 cm) using Sunshine Metro Mix potting soil. Pots were transferred to an Argus controlled greenhouse (Argus Control Systems Ltd., Surrey, BC Canada) with the following specifications (temperature: high 25 ºC, low 22 ºC, relative humidity 40-50%, light cycle: 16 hours light, 8 hours dark) and watered daily. Fertilizer application was applied twice, once at the time of transplanting (14 DAP) and once during flowering (50 DAP) using Plant-Prod® water soluble fertilizer (10-52-10) at a concentration of 15 g / 3.78 L. Insect populations were controlled with Intercept™60 WP (imidacloprid, Bayer Environmental Science, Research Triangle Park, NC, USA) added several days after transplant (20 DAP) with a concentration of 4.1 g / 1000 seedlings. All mixing and application procedures were followed as per the manufacturer's instructions on label.

4.33 DNA Extraction

For the SRAP genotyping method, DNA was extracted from fresh leaves by the cetyltrimethylammonium bromide (CTAB) method (Doyle and Doyle, 1990) with minor modifications. A 500–600 mg leaf sample was crushed using a mortar and pestle with liquid nitrogen. The ground leaf tissue was added to a 15 ml centrifuge tube. Six ml of preheated extraction buffer (500 ml of 2% CTAB, 100 mM Tris–HCl, pH 8.0, 1.4 M NaCl, 20 mM EDTA) was then added and the tube and was vortexed and incubated for 90 min in a 65ºC water bath. The tube was cooled to room temperature and 7 ml of chloroform–isoamyl alcohol (24:1) was added. The solution was mixed gently for 10 min and centrifuged at 4600 RPM for 16 min. The supernatant was transferred to a new 15 ml tube, 0.5 volumes of 2-propanol were added and mixed gently to precipitate the DNA. The mixture was centrifuged at 4600 RPM for 5 min, the supernatant was removed and the DNA pellet was washed with 8 ml 70% (v/v) ethanol. The pellet was then air-dried and resuspended in 3 ml of double distilled water.

Due to purity requirements for GBS, a Qiagen® DNeasy Plant Mini Kit (Qiagen, Valencia, CA, USA) was utilized to ensure ultra-pure DNA. All procedures were followed as per manufacturer's instructions for fresh tissue DNA extraction and purification with the addition of 10 min total elusion time (2 x 5 min).

4.34 DNA Concentration

DNA quantity for SRAP was determined using a Thermo Scientific NanoDrop 2000 spectrophotometer in conjunction with software NanoDrop 2000 (Thermo Fisher Scientific Inc., MA, USA) with all protocols followed as per manufacturer instructions for double stranded DNA quantification. All samples achieved a minimum DNA concentration of 30 ng/µL with a

260/280 ratio of ≥ 1.8. All DNA concentrations for GBS were quantified using a Life Technologies Qubit® Fluorometer (Life Technologies, Carlsbad, CA, USA) with all protocols followed as per manufacturer instructions for double stranded DNA quantification. All samples reached a minimum concentration of 50 ng/μL as per GBS requirements (http://www.biotech.cornell.edu/brc/genomic-diversity-facility).

4.35 Sequence Related Amplified Polymorphism

Sequence related amplified polymorphisms (SRAP) is a polymerase chain reaction (PCR) method, which is designed to amplify open reading frames (ORF's) using a variable forward and reverse primer system (Li and Quiros, 2001). Each primer is 17-18 base pairs long with the forward primer containing a core sequence of CCGG. This forward core sequence targets ORF's due to the known distribution that exons are GC rich (Li and Quiros, 2001). The reverse primer has a core sequence of AATT near the 3′ region to target introns and promoter regions that are typically AT rich (Li and Quiros, 2001). Together these primer combinations create polymorphic DNA bands that are separated through electrophoresis in polyacrylamide gels and visualized through autoradiography (Li and Quiros, 2001). Presence/absence scoring is then applied to the visualized polymorphic bands and genotypes can be separated based on scoring (Li and Quiros, 2001).

4.36 Sequence Related Amplified Polymorphism Primers

The utilization of primers followed the protocol of Li and Quiros (2001) for primer design. Table 4.2 displays the base pair sequence of all primers used for this study. Primer combinations consisted of 29 forward and reverse primer sets (Table 4.3).

Table 4.2: Forward and reverse sequence related amplified polymorphism primers for the polymerase chain reaction used in combination with each other to produce 293 polymorphic bands.

Forward Primers:	
SA7	5'-CGCAAGACCCACCACAA-3'
EM1	5'-GACTGCGTACGAATTCAAT-3'
bg23	5'-ATTCAAGGAGAGTGCGTGG-3'
ME2	5'-TGAGTCCAAACCGGAGC-3'
ODD3	5'-CCAAAACCTAAAACCAGGA-3'

Reverse Primers:			
BG01	5'-TTTCAGGAGCAGATGGTGG-3'	BG10	5'-AGTTGGACATTATTGGCAGC-3'
BG11	5'-AGTTGGACATTATTGGCAGC-3'	BG32	5'-AGTTGGACATTATTGGCAGC-3'
BG33	5'-AGTTGGACATTATTGGCAGC-3'	BG35	5'-AGTTGGACATTATTGGCAGC-3'
BG37	5'-AGTTGGACATTATTGGCAGC-3'	BG38	5'-AGTTGGACATTATTGGCAGC-3'
BG39	5'-AGTTGGACATTATTGGCAGC-3'	BG40	5'-AGTTGGACATTATTGGCAGC-3'
BG41	5'-AGTTGGACATTATTGGCAGC-3'	BG45	5'-AGTTGGACATTATTGGCAGC-3'
BG60	5'-AGTTGGACATTATTGGCAGC-3'	BG62	5'-AGTTGGACATTATTGGCAGC-3'
BG70	5'-AGTTGGACATTATTGGCAGC-3'	BG72	5'-AGTTGGACATTATTGGCAGC-3'
BG4	5'-AGTTGGACATTATTGGCAGC-3'	BG76	5'-AGTTGGACATTATTGGCAGC-3'
BG75	5'-AGTTGGACATTATTGGCAGC-3'	PM018	5'-AAGCGATCAAAGCGGGTG-3'
PM029	5'-GATGAGAAACTGAACGAGG-3'	PM032	5'-AAGTATGGGGTTGGGTTTC-3'
PM034	5'-GGGACGAACAGACAACGA-3'	PM117	5'-GTCAGCAAGTTTCACGGTT-3'

Table 4.3: Forward and reverse primer combinations used for sequence related amplified polymorphism to generate 293 polymorphic bands amplified by PCR with 79 *Brassica napus* genotypes.

Forward Primer				
EM1	bg23	Sa7	ME2	ODD3
Reverse Primer				
BG10	BG33	BG1	BG11	PM32
BG11	BG4	BG60	BG62	PM34
BG33	BG41	BG35	BG70	
BG45	BG62	BG39		
BG72	BG37	BG40		
BG75	BG38	BG41		
BG76	PM18	PM18		
BG32	PM117	PM29		

4.37 Polymerase Chain Reaction (PCR)

Ten µl aliquots of PCR master mix were allocated into a 384 well plate. The master mix was composed of 8.6 µl ddH_2O, 1 µl 10X PCR buffer (500 mM KCl, 100 mM Tris, 1% Triton, 1.5 mM $MgCl_2$, pH 9.3), 0.15 µl dNTP's, 0.15 µl forward primer (labelled, Table 4.2), 0.15 µl reverse primer (not labelled, Table 4.2) and 0.15 µl *Taq* polymerase. DNA was transferred via a stainless steel 96 spike stamping plate and sealed with a PCR plate cover for the PCR procedure. PCR was completed using the following parameters: temperature cycle initiated at 94ºC for 3 min with cycle 2 at 94ºC for 55 s, cycle 3 was 35ºC for 55 s and cycle 4 was 72ºC for 55 s. Cycles 2-4 were repeated five times. Cycle 5 was set to 94ºC for 55 s, cycle 6 was 50ºC for 55 s and cycle 7 was set to 72ºC for 55 s. Cycles 5-7 were repeated thirty times. After the final cycle was completed the PCR products were separated by denaturing acrylamide gels and detected by autoradiography with an ABI Prism 3130XL in association with GenScan® software (V.3.7) (Applied Biosystems, Life Technologies, Carlsbad, CA, USA) (Li and Quiros, 2001).

4.38 Genotyping-By-Sequencing

Genotyping-by-sequencing (GBS) is one of the most current methods for measuring single nucleotide polymorphisms (SNP's) and utilizes the Illumina® next generation sequencing technology (Elshire *et al.*, 2011). GBS is a highly multiplexed PCR method that uses a reduced representation of genome complexity through the use of restriction enzymes (RE) that are methylation sensitive (Elshire *et al.*, 2011). This greatly simplifies sequencing and alignment procedures allowing for deep coverage in gene rich regions (Chen *et al.*, 2013, Elshire *et al.*, 2011). Following RE digestion, adapter barcodes are ligated to the RE cut site allowing many samples to be pooled into one Illumina flow cell greatly reducing cost (Elshire *et al.*, 2011, Chen *et al.*, 2013). Generally, millions of sequence tags (64 bp reads) are generated and 10,000's to

100,000's of SNP's can be called with a very high degree of accuracy through a novel GBS computational pipeline, Tassel (Bradbury *et al*., 2007; Elshire *et al.,* 2011, Glaubitz *et al.,* 2014).

One 96-well plate (Eppendorf twintec PCR plate 96 well) (caps: thermo scientific PCR 8 strip flat caps) with 95 *B. napus* genotypes was submitted to Cornell University Institute of Biotechnology (http://www.biotech.cornell.edu/brc/genomic-diversity-facility) where GBS was completed as per the protocol defined by Elshire *et al*., (2011). *Ape*KI (GCWGC, where W is A or T) was the restriction enzyme chosen at a 95-plex level.

4.39 Genotyping-by-Sequencing Bioinformatics

All bioinformatics analysis (SNP calling) was completed by Cornell University Institute of Biotechnology using the Tassel computation pipeline V. 3.0.166 (Bradbury *et al*., 2007; Glaubitz *et al.,* 2014) and the *B. napus* reference genome (Chalhoub *et al.,* 2014). Genome alignment was generated with Burrows-Wheeler transform algorithm (BWA) version: 0.7.8-r455 (Li and Durbin, 2010; Li and Homer, 2010).

4.40 Cluster Analysis

For SRAP, each presence/absence marker was scored using a binary system of 1 for present and 0 for absence creating a matrix. Genetic distance using Nei's standard genetic distance (Nei, 1972) formula was calculated using the software Powermarker (V3.25) (Liu and Muse, 2005, http://statgen.ncsu.edu/powermarker/) based upon the binary matrix. This calculation created a new matrix of 79 x 79 genotypes with the genetic distance between each genotype displayed. Neighbour joining cluster analysis (Saitou and Nei, 1987) was then applied to the genetic distance matrix using the software Powermarker (V3.25). This created a reference tree with branch lengths between genotypes approximately equal to the genetic distance between

genotypes. Tree robustness was tested with the generation of 1000 bootstrapping replicates (Felsenstein, 1985). The reference tree file and the 1000 bootstrapping replicate trees were exported as Newick format into Mega 6 (V.6.06 [6140226] Tamura *et al*., 2013, http://www.megasoftware.net/) and exported as Newick file to be compatible with the software Geneious V.8.05 (Kearse *et al*., 2012, http://www.geneious.com/download). Consensus tree construction based on the 1000 bootstrapping replicates was completed in Geneious V. 8.05 with the following parameters: support threshold set to 0, topology threshold set to 0, burn in set to 0, and a greedy clustering model.

For GBS, 80,005 filtered biallelic SNP's were imported into the software program Geneious V.8.05. Only 79 of the 95 genotypes were analyzed to match the SRAP analysis. Genetic distance was calculated by Geneious based on the Tamura-Nei distance model (Tamura and Nei, 1993) and the subsequent distance matrix was clustered using neighbour joining method (Saitou and Nei, 1987). This created a reference tree with branch lengths between genotypes approximately equal to the genetic distance between genotypes. Tree robustness was tested with 1000 bootstrapping replicates with the following parameters: support threshold set to 0, topology threshold set to 0, burn in set to 0, and a greedy clustering model.

4.41 Cluster Similarity

Cluster similarity was explored on a cluster-by-cluster level, where each conserved cluster (*K*) was compared to each other for *n* number of matches and each match for each cluster was signified as a match percentage for each cluster. This consisted of a branch-to-branch comparison of conserved clusters within each tree. This allows a computation of the similarity of each conserved cluster and the overall similarity between dendrograms using a match percent

despite changes in topology. Secondly, cluster similarity was computed using the Java applet compare2trees (Nye *et al*., 2006; http://www.mrc-bsu.cam.ac.uk/personal/thomas/phylo comparison/comparison page.html). In short, two phylogenetic trees, T_1 and T_2 that share the same set of leaves (L) can be compared and can be either rooted or unrooted. The comparison algorithm has two stages, first every pair of edges (i, j) with $i \in T_1$ and $j \in T_2$ is assigned a score $s(i, j)$, that reflects the topological similarity of the branches i and j. Secondly, branches in the two trees are paired up to optimize the overall score creating a branch-to-branch comparison (Nye *et al*., 2006).

4.5 RESULTS

4.51 Sequence Related Amplified Polymorphism

Twenty-nine forward and reverse primer combinations amplified through PCR and visualised through autoradiography resulted in 293 polymorphic bands between the 79 *B. napus* genotypes (Figure 4.1). On average, each primer combination amplified 10 polymorphic bands per genotype.

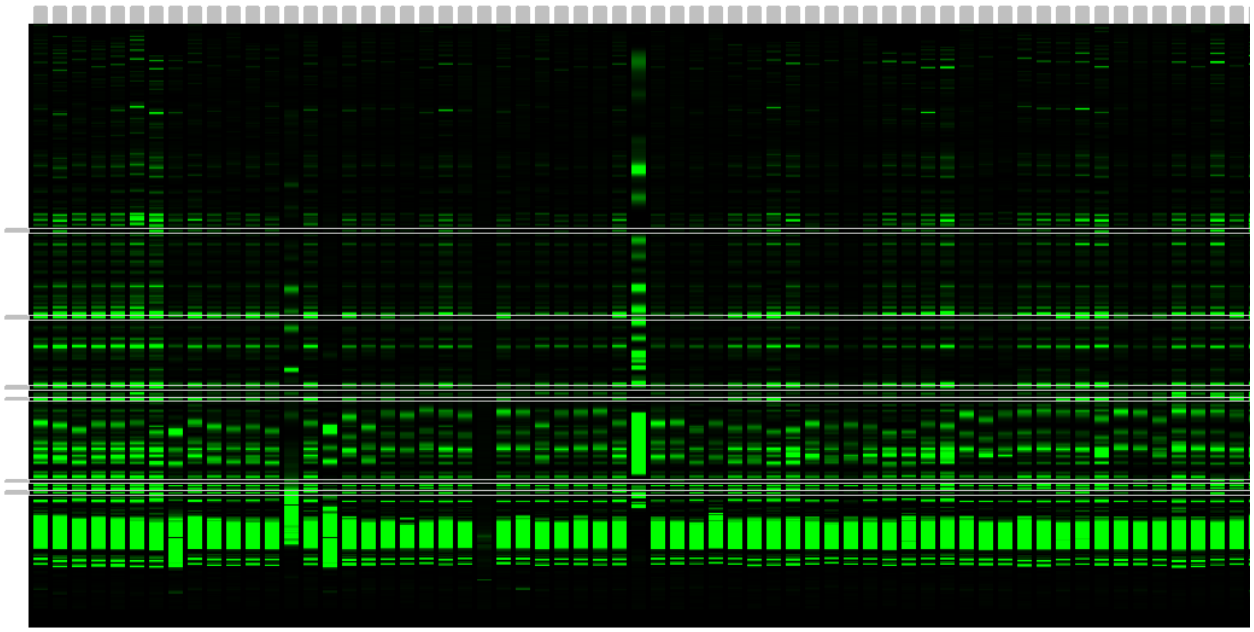

Figure 4.1: Acrylamide gel featuring polymorphic DNA bands amplified using sequence related amplified polymorphism through the polymerase chain reaction with primers EM1 and BG11 visualised through autoradiography with an ABI Prism 3130XL in association with GenScan® software (V.3.7). Grey rows (6 rows) represent polymorphic bands chosen to differentiate 79 *Brassica napus* genotypes.

4.52 Genotyping-by-sequencing (GBS)

The first GBS run produced ~126 million barcode reads, 84% of the minimum 150 million barcode reads that Cornell's Institute of Biotechnology has set as their standard. As a result, our

GBS material was sequenced a second time generating an additional ~116 million barcode reads and combined with the first sequencing run (Table 4.4). This combined pool generated ~8,110,000 and ~6,580,000 unique sequence tags, respectively, for a combined total of 1,631,637 filtered sequence tags.

Table 4.4: Total number of sequence reads, good barcoded reads and resulting tags for each sequence run of genotyping-by-sequencing on 95 *Brassica napus* genotypes.

	Barcodes found in lane	Total # of reads per lane[1]	Total number of good barcoded reads[2]	Resulting # of Tags[3]
Run 1	96	132278340	126177590	8110178
Run 2	96	121989805	116807543	6580155

[1]A read is a single sequence generated by the GBS assay.
[2]A good barcoded read was recorded if 1) the read perfectly matched the barcode sequence and the four base remnant *Ape*K1 cut site, 2) there were no adapter/adapter dimers and 3) contained no N's (missing) up to the trim length (Glaubitz *et al*., 2014).
[3]A tag is a unique sequence from the good barcoded reads.

Filtered sequence tags are tags at or above a defined threshold for all taxa (samples) in the experiment and were used for genome alignment (Glaubitz *et al*., 2014). A total of 1,631,637 filtered tags (64 bp) were sequenced from the 95 *B. napus* genotypes. Of those sequence tags, 925,657 (56.7%) aligned to unique positions, 420,244 (25.8%) aligned to multiple positions and 285,736 (17.5%) could not be aligned to the reference genome. From the alignment results, all unique aligned sequence tags (925,657 or 56.7%) were used for SNP calling based on the reference genome provided (Chalhoub *et al*., 2014).

The resulting SNPs called from the unique aligned positions were divided into 3 distinct categories [VCF, HapMap (unfiltered), HapMap (filtered)]. The VCF SNPs and HapMap SNPs were called independently and variation between the two formats was expected. VCF SNPs were called using the algorithm from Catchen *et al*. (2011) called Stacks, whereas HapMap SNPs were called in Tassel (Bradbury *et al*., 2007; Glaubitz *et al.*, 2014). Stacks SNP calling resulted in

382,560 VCF SNPs. VCFtools version (V.0.1.11) (Danecek *et al*., 2011) was used to calculate depth and missingness from the unfiltered VCF SNP file. Table 4.5 highlights the statistics from the VCF SNP file for depth coverage and missingness. Tassel SNP calling generated 179,974 unfiltered SNPs stored in HapMap format. HapMap SNPs were filtered on missingness and allele frequency which generated 80,005 high quality bi-allelic SNPs.

Table 4.5: Depth and mean missingness of unfiltered VCF SNPs obtained from Cornell's Institute of Biotechnology on 95 *Brassica napus* genotypes using VCFtools version (V.0.1.11).

	Mean	**Median**	**Standard Deviation**
Individual depth	10.231	9.536	5.042
Site depth	10.231	6.854	14.336
Individual missingness	0.238	0.213	0.109
Site missingness	0.238	0.104	0.272

4.53 Genetic distance

SRAP genetic distance based on 293 polymorphic bands had a range of 0.08 GD between genotypes 11DH-91-R and 11DH-114-R, to 0.74 GD between genotypes ER2-B and NEW-620-R (Figure 4.2). GBS genetic distance was calculated using 80,005 SNPs, which had a range of 0.0047 GD between genotypes 11DH-91-R and 11DH-114-R, to 0.629 GD between genotypes ER3-B and 08C847-R (Figure 4.4).

4.54 Cluster Analysis

The neighbour joining cluster analysis based on the SRAP genetic distance matrix produced a reference dendrogram with 11 distinct heterotic clusters (Figure 4.2). Dendrogram robustness was tested through 1000 bootstrapping replications (Figure 4.3). The bootstrapping dendrogram also produced 11 heterotic clusters. However, only 5 clusters (2, 4, 5, 7 and 9) remained identical over 1000 replications. These can be considered high confidence heterotic clusters, although only

minor genotype movement was observed throughout the other non-identical clusters. The neighbour joining cluster analysis based on the GBS distance matrix produced a reference dendrogram with 12 heterotic clusters (Figure 4.4). Again, tree robustness was tested with 1000 bootstrapping replications (Figure 4.5). Between figures 4.4 and 4.5 only 2 heterotic clusters (4 and 5) did not remain identical over 1000 replications. For the consensus SRAP tree and the GBS consensus tree (Figures 4.3 and 4.5), the GBS consensus tree showed remarkable robustness as many nodes have a 90% or higher clustering percent over 1000 bootstrapping replications (Felsenstein, 1985) and considerable confidence can be given to a tree that is supported by >80% of bootstrapping replicates (Zharkikh and Li, 1992). On the other hand, little confidence can be given to a tree that is supported by <75% of the replicates (Zharkikh and Li, 1992). This may apply to the SRAP consensus tree as it seemed to vary over 1000 replications and its node length ranged from approximately 1.3% to 99.5% intervals. Comparing all dendrograms, cluster 2 remained identical through the different methods and replicates. Cluster 2 is represented by genotypes all of European descent.

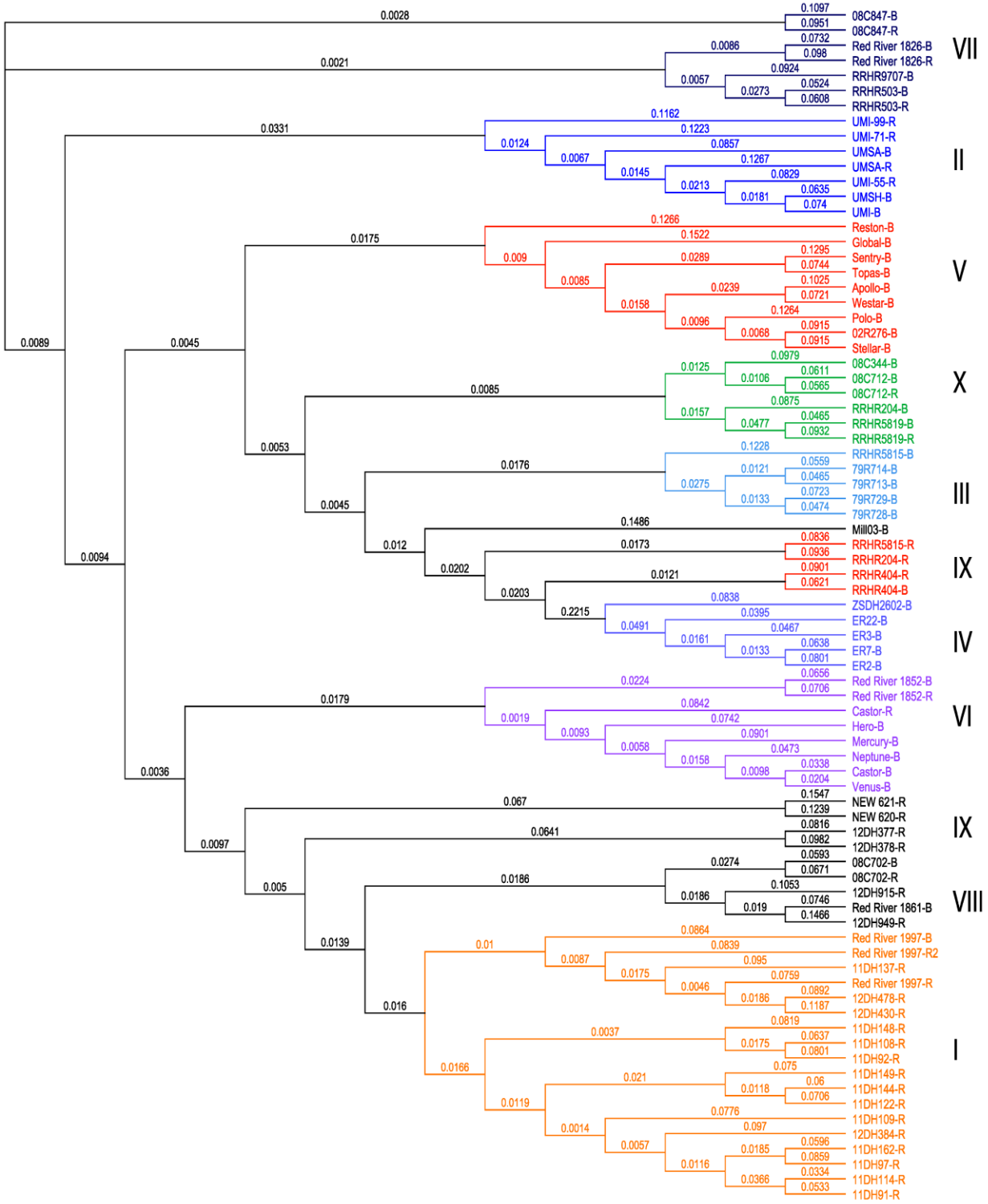

Figure 4.2: Neighbour joining dendrogram clustered using Nei's standard genetic distance based on 293 SRAP polymorphic bands obtained through sequence related amplified polymorphisms on 79 *Brassica napus* genotypes visualized in Geneious V.8.05. Distinct clusters have been numbered and color-coded for ease of viewing. Each genotype is either a maintainer (-B) or restorer (-R) in the *ogu*-INRA pollination control system.

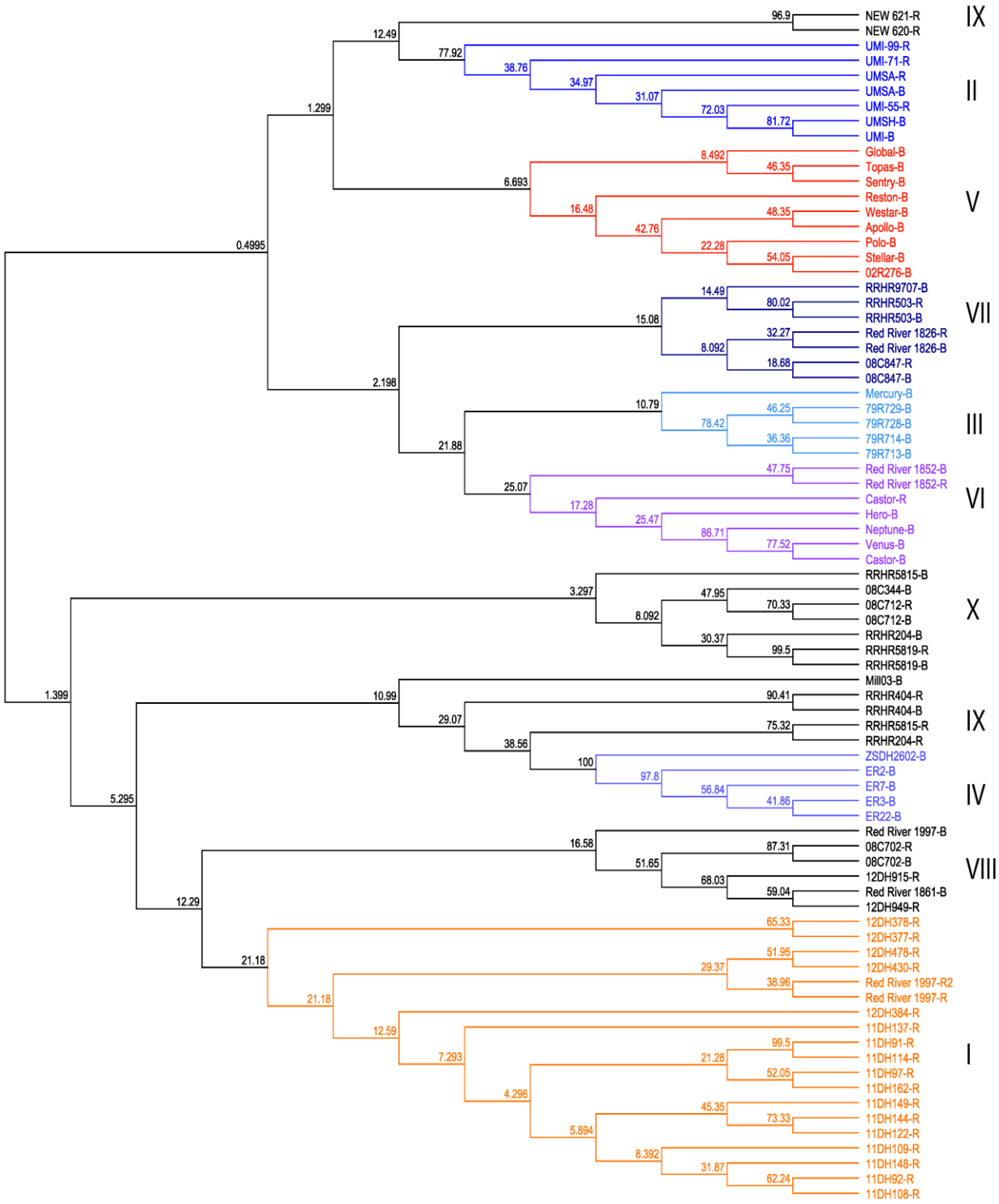

Figure 4.3: 1000 bootstrap replication of the neighbour joining cluster analysis based on 293 polymorphic bands obtained through sequence related amplified polymorphism. Consensus tree construction was implemented in Geneious V.8.05 over 1000 replicates with percent threshold set to 0. Node lengths equal percent commonality over 1000 trees. Numbers and colors have been added for ease of viewing. Each genotype is either a maintainer (-B) or restorer (-R) in the *ogu*-INRA pollination control system.

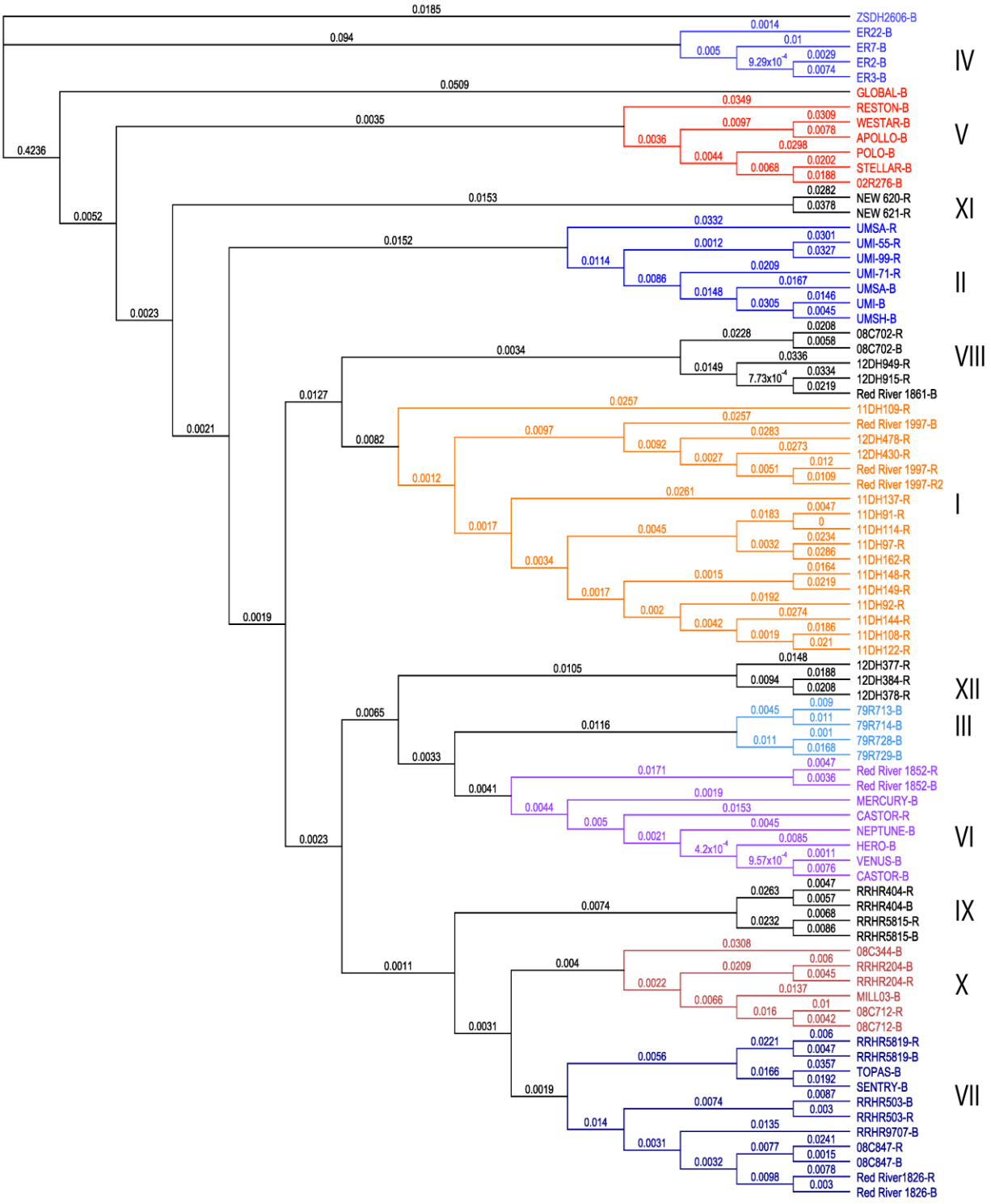

Figure 4.4: Neighbour joining dendogram based on Tamura-Nei's genetic distance calculated on 80,005 SNPs obtained from genotyping-by-sequencing on 79 *Brassica napus* genotypes visualized in Geneious (V.8.05). Distinct clusters have been numbered and color-coded for ease of viewing. Each genotype is either a maintainer (-B) or restorer (-R) in the *ogu*-INRA pollination control system.

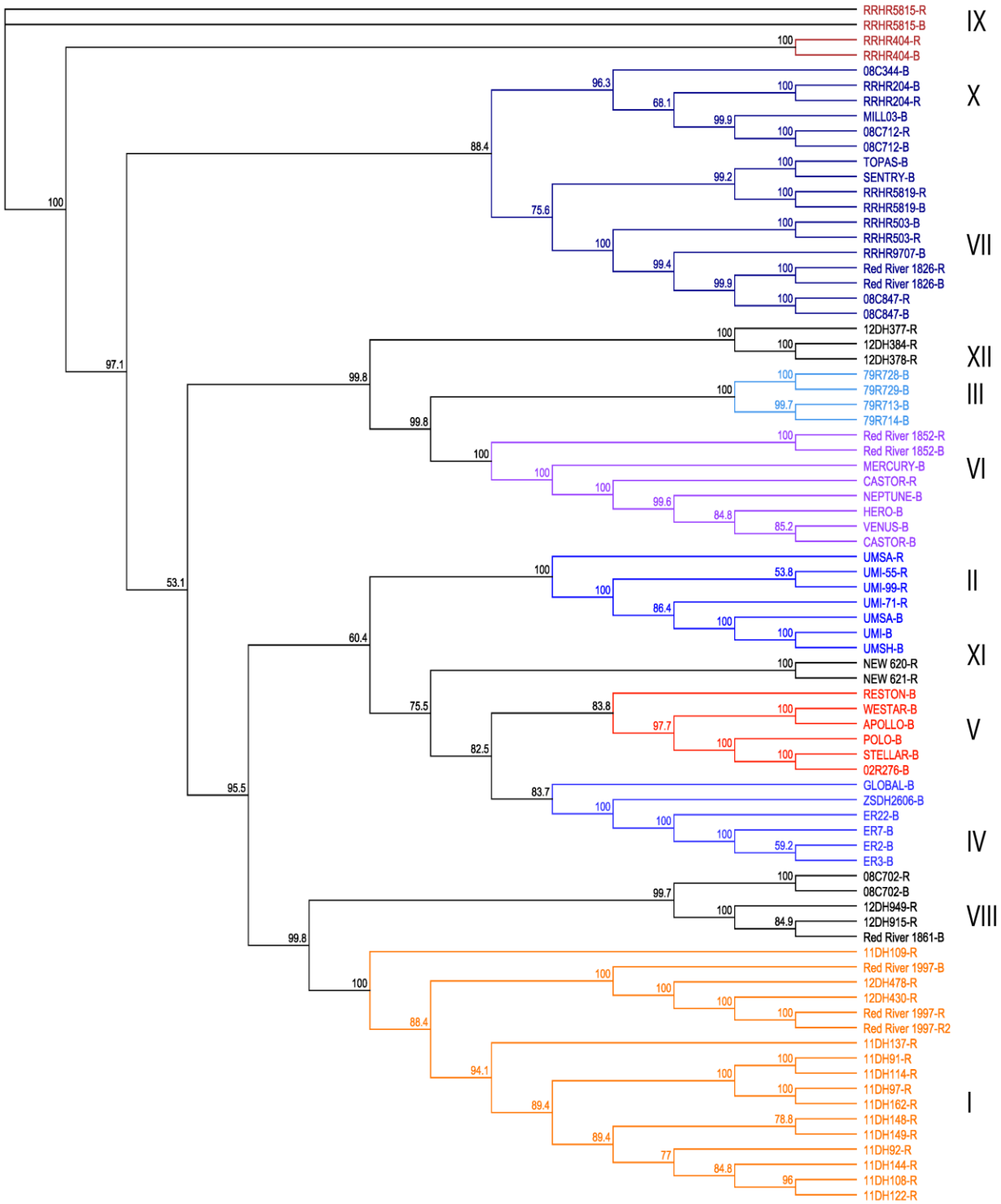

Figure 4.5: 1000 bootstrap replication of the neighbour joining cluster analysis based on genotyping-by-sequencing 80,005 SNPs. Node lengths equal percent commonality over 1000 replicates visualized in Geneious V.8.05. Distinct clusters have been color coded for ease of viewing. Each genotype is either a maintainer (-B) or restorer (-R) in the *ogu*-INRA pollination control system.

4.55 Cluster Similarity

Cluster similarity was investigated between the two genotypic methods of SRAP and GBS in association with Nei's standard genetic distance and the Tamura-Nei distance model, respectively. Cluster topology differed between the genetic distance calculated on Nei's standard genetic distance based on SRAP and the genetic distance calculated on the Tamura-Nei model based on GBS despite using the same neighbour joining algorithm (Table 4.6). However, distinct clusters contained similar genotypes between the two methods. Clusters II, IV, VI and XI are all examples where all genotypes remained identical over both methods.

Table 4.6: Manual comparison of similarity for each conserved cluster between the genetic distances of sequence related amplified polymorphism and genotyping-by-sequencing dendrograms for individual genotypes (*n*).

Conserved Clusters Between Dendrograms	SRAP (*n*)[1]	GBS (*n*)[2]	Match %[3]
I	18	17	94.4
II	7	7	100
III	5	4	80
IV	5	5	100
V	9	7	78
VI	8	8	100
VII	7	11	64
VIII	7	5	71
IX	5	4	60
X	6	6	71
XI	2	2	100
XII	0	3	0
Total (*n*) and mean match percent (%)	79	79	76.5

[1]The total number of genotypes separated by neighbour joining cluster analysis based on SRAP genetic distance.
[2]The total number of genotypes separated by neighbour joining cluster analysis based on GBS genetic distance.
[3]Match percent obtained by dividing the smaller number of genotypes per cluster by the larger number of genotypes per cluster.

Despite the differences in cluster placement, each cluster between the two dendrograms shows highly conserved clusters for specific genotypes. There is an approximate homology of 77% between all genotypes in all clusters when manually compared. The Java applet Compare2trees was implemented for a branch-to-branch computational comparison (Nye *et al*., 2006). Compare2trees found an approximate homology of 68% between the SRAP and GBS trees.

4.6 DISCUSSION

Several primer combinations for SRAP have been previously been reported and found to be very successful for differentiating *B. napus* genotypes (Li and Quiros, 2001; Sun *et al*., 2007; Wen *et al*., 2006; Riaz *et al*., 2001). Sun *et al*., (2007) constructed an ultra-dense genetic map using 1,634 SRAP primer combinations to produce 13,551 mapped markers. Wen *et al*., (2006) discovered 123 polymorphic fragments using 25 SRAP primer combinations and Riaz *et al*., (2001) found 118 polymorphic bands based on 18 forward and reverse SRAP primer combinations. Here we report 293 polymorphic bands with 29 forward and reverse primer combinations for 79 *B. napus* genotypes.

Genotyping-by-sequencing is a relatively new protocol for high throughput SNP detection (Elshire *et al*., 2011). There is currently little GBS data published for *B. napus* diversity. We report that 285,736 tags or 17.5% of filtered tags could not be aligned to the *B. napus* reference genome (Chalhoub *et al.*, 2014). Elshire *et al*., (2011) reported that only 2% of parental maize line B73 filtered tags could not be aligned to the maize reference genome (B73 RefGen V.1). Elshire *et al*., (2011) found that this 2% of non-aligning reads were not present in the reference genome. In the current research, 17.5% could not be aligned and in conjunction with the Elshire *et al*., (2011) findings, these sequencing tags are probably not contained within the reference genome. Currently, the *B. napus* genome is of winter habit and is an open-pollinated genotype (Chalhoub *et al*., 2014). This may explain a vast majority of non-aligning reads as our material is considered to be spring habit and 38 of the 79 genotypes contain restorer introgressions from *Raphanus sativa* L. for use in the *ogu*-INRA pollination control system (Heyn, 1976; Delourme and Eber, 1992; Gourret *et al.*, 1992; Delourme *et al.*, 1994). These two differences may contribute to the unaligned sequences; however, this hypothesis warrants further investigation.

Genetic distance has been a well-cited mathematical tool for the determination of species and/or individual relatedness (Ali *et al*., 1995; Riaz *et al*., 2001; Yu *et al*., 2005; Jesske *et al*., 2013). However, very few studies present multiple genetic distance methods with the same population or genotypes with the same clustering method for the purpose of evaluating genetic distance measures. Here, the ultimate goal is to investigate which genetic distance method can produce highly accurate heterotic pools for the purpose of predicting heterosis. To pursue this endeavour intra- and inter-cluster hybrids from the current dendrograms need to be evaluated to gauge which genetic distance method has greater predictive power. However, despite these different methods, genotypes 11DH91-R and 11DH114-R produced the smallest genetic distance for both methods. In addition, the largest values obtained for both genetic distance methods involved Canadian *B. napus* genotypes compared to resynthesized *B. napus* and this is in agreement with Jesske *et al*. (2013) who presented evidence that resynthesized *B. napus* genotypes contain genetic diversity not seen in elite breeding lines. This experimental evidence lends credibility to both SRAP and GBS methods as each separate mathematical formula calculated the shortest genetic distance between the same pair of genotypes and also produced the largest genetic distance between Canadian *B. napus* genotypes and resynthesized genotypes.

The comparison between the genetic distance dendrograms derived from SRAP and GBS (Figures 4.2 and 4.4) were remarkably similar despite using different genetic distance formulas and different genotypic methods (77% homology based on manual match percent and 68% homology calculated by the Java applet Compare2trees). The close association in percentage shows that these trees are similar; however, when bootstrapping values were incorporated, the GBS bootstrapping tree was stronger and more robust as opposed to the SRAP bootstrapping dendrogram. Unfortunately, the program Compare2trees cannot compare bootstrapping trees

(Nye *et al*., 2006). However, across all dendrograms, a visual inspection shows cluster II remained identical. This supports the conclusions of Cuthbert *et al*. (2009) which showed that European-derived breeding material was distinct from Canadian high erucic acid rapeseed material based on heterotic performance and cluster II in this analysis retains only European-derived genotypes.

From a *B. napus* breeding standpoint, it is well established that inter-cluster hybrids exhibit higher heterosis than intra-cluster hybrids (Grant and Beversdorf, 1985; Riaz *et al.*, 2001; Riaz and Qiuros, 2011). This assumption is also well documented in maize hybrid breeding as many commercial hybrids are from complementary heterotic pools (e.g., Reid Yellow Dent and Lancaster Surecrop) (Lu *et al*., 2009; Chen *et al*., 2015). The conserved clusters between SRAP and GBS (II, IV, VI and XI) suggest that the genotypes within each cluster may be more important for heterotic gains than cluster placement in the overall topology of the dendrogram, since topologies shift between methods. Genotypic placement within clusters can be considered highly accurate given origin and pedigree information. For instances, cluster I for both methods contains mostly double haploid (DH) material and Red River 1997, a parental genotype for most of the DH material. Cluster II was all European sourced material, cluster IV was all resynthesized genotypes and cluster VI contains genotypes relseased by the University of Manitoba. Further investigation between inter- and intra-cluster hybrids as well as cluster placement and the genetic distance between each parental genotype would prove extremely useful for developing a breeding schematic based on genetic distance and cluster analysis for *B. napus* hybrids. Since cluster II was conserved across all methods and replicates, this European derived cluster is distinct and future inter-cluster hybrids should be explored using this cluster.

From a finacial standpoint, GBS was outsourced to Cornell University Institute of Biotechnology which (as of 2014) roughly had a price tag of $38.00 US per sample for one 96 well plate including bioinformatics with an additional cost of $4.56 US per sample for DNA extraction using a Qiagen® DNeasy Plant Mini Kit extraction kit and an additional $0.93 per sample for labour based on 20 samples. In relation, CTAB costs $0.80 US per genotype for DNA extraction, assessed by the Abarshi *et al*. (2010). CTAB is more laborious and takes longer giving it a higher labour cost per sample at $1.25 US based on 20 samples. However, PCR reagent costs were assessed by Duncan *et al*. (2012) at $1.22 US for a total cost of $3.27 US per sample for SRAP opposed to $43.49 US for GBS. Comparing the overall resolution between the two methods at ~77% and 68% similarity respectively, SRAP appears to be just as effective for separating diverse genotypes into distinct clusters for breeding purposes with a substantially lower cost. Despite the cost difference, the similarity in clustered genotypes between each method lends credibility to both methods. Since these methods are roughly a decade apart, yet still produce similar results we can infer that these genotypic methods are comparable when investigating heterotic pool placement based cluster analysis and genetic distance in *B. napus* genotypes.

These current heterotic clusters as defined by SRAP and GBS may prove useful for the development of hybrid *B. napus* cultivars based on genetic distance. Future investigations should concentrate on the accuracy of genotypic placement through inter- and intra- cluster hybrids with the concurrent measure of hybrid heterosis over parental values to gauge if genetic distance has any influence on heterotic gains. This will be examined in chapter 5 of the thesis. Overall tree topology and its effect on hybrid performance should also be investigated through the use of patristic distance based on the two different SRAP and GBS consensus trees.

5.0 Heterosis in *Brassica napus* L. Based on Genetic Distance and Cluster Analysis

5.1 ABSTRACT

Despite over a century of hybrid theory, research and breeding, the genetic control of heterotic gains are not fully understood. This has limited the ability to predict the hybrid performance for specific parental combinations. The predictive power of hybrid pairing in *Brassica napus* L. were investigated with the use of two different genotypic methods, sequence related amplified polymorphisms (SRAP) and genotyping-by-sequencing (GBS). Nei's standard genetic distance and Tamura-Nei genetic distance models were used with the two different genotypic methods, respectively. A consensus matrix was also constructed from the two genetic distance models to investigate if a consensus matrix would improve heterotic parental pairing predictions. Thirty-six parental inbred lines were selected from different clusters to create 44 inter-cluster hybrid combinations. GBS data produced the best prediction of yield (R = 0.47); however, this is only considered a weak correlation. Regression analysis showed that the SRAP method outperformed both GBS and consensus models with a moderately strong correlation (R = 0.61) for high-parent heterosis. All hybrid combinations beat their mid-parent values for seed yield and only three hybrid combinations did not yield greater than their high parent yield values. The best hybrid combination exhibited mid-parent heterosis of 145% along with high-parent heterosis of 89%. This study demonstrates that SRAP data along with genetic distance and cluster analysis can be a method for predicting high heterotic parental combinations.

5.2 INTRODUCTION

On a global scale, *Brassica napus* L. (N=19 AACC) is currently one of the most important oilseed crops grown today (Rahman, 2013). The EU, Canada and China are top producers at 20Mt, 15.5Mt, and 12Mt, respectively, as of 2011-12 (Carré and Pouzet, 2014). Over the past 20 years, *B. napus* production has increased globally by a factor of 2.4 (Carré and Pouzet, 2014). This massive expansion in recent years is primarily due to the development of canola quality cultivars (≤ 2% erucic acid, ≤ 30 μmol/g aliphatic glucosinolate content) from Canada, which provide one of the healthiest oil profiles for human nutrition and meal by-product for animal feed (Rahman, 2013; Wittkopp *et al*., 2009). Substantial research investment into *B. napus* hybrid development has occurred since the 1990s, creating a very competitive hybrid industry (Rahman, 2013). One of the first steps for a successful hybrid breeding program needs to be the classification of germplasm into genetically distinct heterotic pools (Girke *et al*., 2012). This classification can be accomplished using genetic distance and cluster analysis.

Genetic distance is based on a mathematical formula in which species divergence can be calculated (Nei, 1972). These calculations can be based on morphological differences or on polymorphism at the DNA level (Yu *et al*., 2005). A variety of genetic distance formulae exist, with each method addressing certain assumptions of molecular evolution (Nei, 1972; Nei, 1978; Tamura and Nei, 1993; Reynolds *et al*., 1983). Several publications have shown a positive correlation between genetic distance calculated on morphological phenotypic traits and hybrid heterosis (Ali *et al*., 1995; Geleta *et al*., 2004), as well as DNA polymorphisms and hybrid heterosis (Ahmad and Quiros, 2011; Tan *et al*., 2007). On the genetic side, Sequence Related Amplified Polymorphism [SRAP (Li and Quiros, 2001)] has been a popular method for genetic distance studies (Ahmad and Quiros, 2011; Tan *et al*., 2007). Next generation sequencing

technology has also been used for genetic distance studies in a variety of agriculturally important crops (Vieira *et al*., 2015; Crossa *et al*., 2013). However, there is currently little published research calculating genetic distance using genotyping-by-sequencing [GBS (Elshire *et al*., 2011)] in *B. napus*.

In addition to genetic distance, multivariate cluster analysis is a popular and frequently cited method throughout the plant breeding world for separating germplasm/accessions/genotypes into distinct heterotic clusters for breeding purposes (Odong *et al*., 2011). A variety of different clustering methods are available including Ward's method (Ward, 1963); unweighted pair group method with arithmetic mean [UPGMA, (Sokal and Michener, 1958)], and neighbour joining (Saitou and Nei, 1987). Despite decades of use with these clustering techniques, there is currently no scientific consensus on which clustering method provides the most accurate hierarchical clustering for breeding purposes (Mouchet *et al*., 2008). To alleviate this, Mouchet *et al*., (2008) suggests that a consensus tree (the combination of different trees into one dendrogram) may result in the most faithful representation of clustered data. The ultimate goal of genetic distance and cluster analysis is to separate breeding material into separate genetically distinct clusters. Once these clusters are defined as heterotic pools, breeding between pools can occur which may help maximise heterosis between parental parings (Hasan *et al*., 2006).

Hybrid heterosis, or more commonly called hybrid vigor, is the increase of performance traits over the parental phenotypes, and to this day, is still not fully understood (Fu *et al*., 2014; Kaeppler, 2012). Due to the uncertainty in how heterotic gains are fully derived, several competing theories have arisen to describe this observed phenomenon (Kaeppler, 2012). Dominance, first suggested by Bruce (1910) and Keeble and Pellew (1910), state that dominant alleles are positive and recessive alleles are negative, by stacking dominant alleles in the F_1

generation you increase performance traits of the hybrid. Currently, new evidence suggests that heterosis may be strongly linked to repulsion linkage and dominance (Li *et al*., 2015). This is counter to the overdominance theory, first suggested by Shull (1908) and East (1908), which suggested that the heterozygous combinations of diverse alleles has a greater impact on fitness than either of the homozygous states. Additionally, Falconer and McKay (1996) suggested that the genetic distance between two parental cultivars may influence the amount of expressed heterosis. In conjunction, numerous studies have implicated genetic distance as a predictor of high heterotic parental combinations in *B. napus* (Ali *et al*., 1995; Diers *et al.*, 1996; Riaz *et al.*, 2001; Radoev *et al.*, 2008; Cuthbert *et al.*, 2009; Ahmad and Quiros, 2011).

The objective of this study was to investigate the underpinnings of hybrid heterosis through the use of genetic distance calculated by the Nei and Tamura-Nei formulas, respectively, on two different genotypic methods of SRAP and GBS. The neighbour joining clustering method was used to cluster each data set and hybrid heterosis was evaluated based on genetic distance and calculated on hybrid yield over parental inbred lines. Each genetic distance data set was then combined into a consensus tree and evaluated to determine if the consensus method would outperform each genetic distance method independently.

5.3 MATERIAL AND METHODS

5.31 Greenhouse Production

Breeding material was provided by the University of Manitoba and consisted of mainly spring-habit oilseed rape. Initially, three plants for each genotype were planted at a depth of 1 cm into plastic 4 x 3 well cell packs (13 cm x 13 cm x 5 cm) containing Sunshine Metro Mix potting soil (Sungro® Horticulture, MA, USA) during spring of 2014. Plants were grown in a growth chamber (temperature: day: 22ºC, night: 18ºC, light cycle: 16 hours light, 8 hours dark) and watered daily. At the two-leaf stage [14 days after planting (DAP)] each plant was transferred to a plastic growers pot (14.5 cm x 15 cm) using Sunshine Metro Mix potting soil. Pots were transferred to an Argus controlled greenhouse (Argus Control Systems Ltd., Surrey, BC Canada) with the following specifications (temperature: high 25 ºC, low 22 ºC, relative humidity 40-50%, light cycle: 16 hours light, 8 hours dark) and watered daily. Fertilizer application was applied twice, once at the time of transplanting (14 DAP) and once during flowering (50 DAP) (Plant-Prod® water soluble fertilizer (10-52-10) at a concentration of 15 g / 3.8 L. Insect populations were controlled with Intercept™60 WP (imidacloprid, Bayer Environmental Science, Research Triangle Park, NC, USA) added several days after transplant (20 DAP) with a concentration of 4.1 g / 1000 seedlings. All mixing and application procedures were followed as per the manufacturer's instructions on label.

5.32 DNA Extraction

For the SRAP genotyping method, DNA was extracted from fresh leaves by the cetyltrimethylammonium bromide (CTAB) method (Doyle and Doyle, 1990) with minor modifications. A 500–600 mg leaf sample was crushed using a mortar and pestle with liquid

nitrogen. The ground leaf tissue was added to a 15 ml centrifuge tube. Six ml of preheated extraction buffer (500 ml of 2% CTAB, 100 mM Tris–HCl, pH 8.0, 1.4 M NaCl, 20 mM EDTA) was then added and the tube and was vortexed then incubated for 90 min in a 65ºC water bath. The tube was cooled to room temperature and 7 ml of chloroform–isoamyl alcohol (24:1) was added. The solution was mixed gently for 10 min and centrifuged at 4600 RPM for 16 min. The supernatant was transferred to a new 15 ml tube, 0.5 volumes of 2-propanol were added and mixed gently to precipitate the DNA. The mixture was centrifuged at 4600 RPM for 5 min, the supernatant was removed and the DNA pellet was washed with 8 ml 70% (v/v) ethanol. The pellet was then air-dried and resuspended in 3 ml of double distilled water.

Due to purity requirements for GBS, a Qiagen® DNeasy Plant Mini Kit (Qiagen, Valencia CA, USA) was utilized to ensure ultra-pure DNA. All procedures were followed as per manufacturer's instructions for fresh tissue DNA extraction and purification with the addition of 10 minutes total elusion time (2 x 5 min).

5.33 Sequence Related Amplified Polymorphism Procedure

Sequence related amplified polymorphisms (SRAP) is a polymerase chain reaction (PCR) method, which is designed to amplify open reading frames (ORF's) using a variable forward and reverse primer system (Li and Quiros, 2001). Each primer is 17-18 base pairs long with the forward primer containing a core sequence of CCGG. This forward core sequence targets ORF's due to the known distribution that exons are GC rich (Li and Quiros, 2001). The reverse primer has a core sequence of AATT near the 3′ region to target introns and promoter regions that are typically AT rich (Li and Quiros, 2001). Together, these primer combinations create polymorphic DNA bands that are separated through electrophoresis in polyacrylamide gels and

visualized through autoradiography (Li and Quiros, 2001). Presence/absence scoring is then applied to the visualized polymorphic bands and genotypes can be separated based on scoring (Li and Quiros, 2001). Briefly, PCR was completed with the following parameters: 10 µl aliquots of PCR master mix were allocated into a 384 well plate. The master mix was composed of 8.6 µl ddH_2O, 1 µl 10X PCR buffer (500 mM KCl, 100 mM Tris, 1% Triton, 1.5 mM $MgCl_2$, pH 9.3), 0.15 µl dNTP's, 0.15 µl forward primer (labelled, Table 4.2), 0.15 µl reverse primer (not labelled, Table 4.2) and 0.15 µl *Taq* polymerase (all reagents were from New England Biolabs, Ipswich, MA). DNA was transferred via a stainless steel 96 spike stamping plate and sealed with a PCR plate cover for the PCR procedure. PCR was completed using the following parameters: temperature cycle initiated at 94ºC for 3 min with cycle 2 at 94ºC for 55 s, cycle 3 was 35ºC for 55 s and cycle 4 was 72ºC for 55 s. Cycles 2-4 were repeated five times. Cycle 5 was set to 94ºC for 55 s, cycle 6 was 50ºC for 55 s and cycle 7 was set to 72ºC for 55 s. Cycles 5-7 were repeated thirty times. After the final cycle was completed the PCR products were separated by denaturing acrylamide gels and detected by autoradiography with an ABI Prism 3130XL in association with GenScan® software (V.3.7) (Applied Biosystems, Life Technologies, Carlsbad, CA, USA) (Li and Quiros, 2001).

5.34 Genotyping-by-Sequencing (GBS) Protocol

One 96-well plate (Eppendorf twintec PCR plate 96 well) (caps: thermo scientific PCR 8 strip flat caps) was submitted to Cornell University Institute of Biotechnology (http://www.biotech.cornell.edu/brc/genomic-diversity-facility) where GBS was completed as per the protocol defined by Elshire *et al*., (2011). *Ape*KI (GCWGC, where W is A or T) was the restriction enzyme chosen at a 95-plex level. Since GBS was performed on 95 genotypes and

SRAP only 79 genotypes, 16 genotypes were removed from the GBS analysis to standardize the analysis to 79 genotypes each.

5.35 Genotyping-by-Sequencing (GBS) Bioinformatics

All bioinformatics analysis (SNP calling) was completed by Cornell University Institute of Biotechnology using the Tassel computation pipeline V. 3.0.166 (Bradbury *et al*., 2007; Glaubitz *et al*, 2014) and the *B. napus* reference genome (Chalhoub *et al.*, 2014). Genome alignment was generated with Burrows-Wheeler transform algorithm (BWA) version: 0.7.8-r455 (Li and Durbin, 2010; Li and Homer, 2010).

5.36 Genetic Distance and Cluster Analysis

For SRAP, each presence/absence marker was scored using a binary system of 1 for present and 0 for absence creating a matrix. Genetic distance using Nei's standard genetic distance (Nei, 1972) formula was calculated using the software Powermarker (V3.25) (Liu and Muse, 2005, http://statgen.ncsu.edu/powermarker/) based upon the binary matrix. This calculation created a new matrix of 79 x 79 genotypes with the genetic distance between each genotype displayed. Neighbour joining cluster analysis (Saitou and Nei, 1987) was then applied to the genetic distance matrix using the software Powermarker (V3.25). This created a reference tree with branch lengths between genotypes approximately equal to the genetic distance between genotypes. Tree robustness was tested with the generation of 1000 bootstrapping replicates in Powermarker (V3.25) (Felsenstein, 1985). The reference tree file and the 1000 bootstrapping replicate trees were exported as Newick format into Mega 6 (V.6.06 [6140226] Tamura *et al*., 2013, http://www.megasoftware.net/) and exported as Newick file to be compatible with the software Geneious V.8.05 (Kearse *et al*., 2012, http://www.geneious.com/download). Consensus

tree construction based on the 1000 bootstrapping replicates was completed in Geneious V. 8.05 with the following parameters: support threshold set to 0, topology threshold set to 0, burn in set to 0, using a greedy clustering model.

For GBS, 80,005 filtered biallelic SNP's were imported into the software program Geneious V.8.05. To match SRAP, only 79 genotypes were analyzed. Genetic distance was calculated by Geneious based on the Tamura-Nei distance model (Tamura and Nei, 1993) and the subsequent distance matrix was clustered using neighbour joining method (Saitou and Nei, 1987). This created a reference tree with branch lengths between genotypes approximately equal to the genetic distance between genotypes. Tree robustness was tested with 1000 bootstrapping replicates with the following parameters: support threshold set to 0, topology threshold set to 0, burn in set to 0, using a greedy clustering model.

5.37 Consensus Tree Construction

The SRAP and GBS genetic distance matrixes were combined to create a consensus matrix. The consensus matrix was similar in design to Lapointe and Cucumel, (1997) with modification where $\bar{M} = \overline{d_{(i,J)}} = \frac{1}{2}\sum d_k$ (i,j) where i and j equal SRAP and GBS distance matrixes, respectively. This new consensus matrix was imported into Powermarker (V.3.25) (Liu and Muse, 2005) and the neighbour-joining clustering algorithm was applied. The new dendrogram was exported to MEGA6 (V.6.06 [6140226] Tamura *et al*., 2013) as a newick file and imported to Geneious (V.8.05) for visualization.

5.38 Inbred and Hybrid Selections

Seventy-nine breeding lines were available and thirty-six parental inbred *B. napus* lines (16 female and 20 male) were selected from different defined clusters to create forty-four hybrid combinations for this study (Table 5.1). Maintainer genotypes (-B) were converted to male sterile lines (female) using the *ogu*-INRA cytoplasmic male sterility (CMS) system to ensure proper cross-pollination (Delourme *et al*., 1998; Pelletier and Budar, 2015).

Table 5.1: Hybrid designation with parent cultivar combinations divided into two experiments with twenty-two hybrids in each grown in Winnipeg and Carman, MB, in 2014. Maintainer genotypes (-B) were converted to male-sterile lines using the *ogu*-INRA cytoplasmic male sterility system and were crossed with restorer genotypes (-R) to create fertile hybrids.

Hybrid Designation	Female Genotype	Male Genotype
Experiment 1		
1	08C702-B	11DH137-R
2	08C712-B	11DH137-R
3	08C712-B	12DH915-R
4	08C712-B	12DH949-R
5	08C847-B	11DH137-R
6	Mill03-B	RRHR503-R
7	Red River 1826-B	11DH137-R
8	Red River 1826-B2	11DH137-R
9	Red River 1861-B	11DH137-R
10	Red River 1861-B2	11DH137-R
11	Red River 1997-B	12DH915-R
12	Red River 1997-B	12DH949-R
13	Red River 1997-B2	12DH915-R
14	Red River 1997-B2	12DH949-R
15	RRHR204-B	11DH137-R
16	RRHR404-B	11DH137-R
17	RRHR404-B	12DH915-R
18	RRHR404-B	12DH949-R
19	RRHR5815-B	11DH137-R
20	RRHR9707-B	11DH137-R
21	UMI-B	12DH949-R
22	UMSA-B	11DH91-R

Hybrid Designation	Female Genotype	Male Genotype
Experiment 2		
23	UMI-B	Red River 1852-R
24	UMI-B	RRHR503-R
25	UMI-B	08C702-R
26	UMI-B	12DH915-R
27	UMSA-B	RRHR503-R
28	UMSA-B	08C702-R
29	UMSA-B	08C847-R
30	UMSA-B	Red River 1997-R1
31	UMSA-B	Red River 1826-R
32	UMSA-B	11DH92-R
33	UMSA-B	11DH97-R
34	UMSA-B	11DH108-R
35	UMSA-B	11DH109-R
36	UMSA-B	11DH114-R
37	UMSA-B	11DH122-R
38	UMSA-B	11DH137-R
39	UMSA-B	11DH144-R
40	UMSA-B	11DH148-R
41	UMSA-B	11DH149-R
42	UMSA-B	11DH162-R
43	UMSA-B	12DH915-R
44	UMSA-B	12DH949-R

5.39 Field Experimental Design

The 44 hybrid combinations along with the 36 parents and were planted in yield trials during the 2014 growing season (May – September) in Winnipeg, MB, Canada (Latitude: 49.80 | Longitude: -97.16) and at the Ian N. Morrison Research Farm, Carman, MB, Canada, (Latitude: 49.50 | Longitude -98.03). Each parental genotype and its hybrid combination were planted using 3.6 grams of seed planted in 1 m x 3 m, six row plot, with 30 cm spacing between each plot in a random complete block (RCB) design with 3 replicates. The experiment for the Winnipeg location was planted on May 28th, 2014, and the experiment for the Carman location was planted on May 23rd 2014. Soil type in southern Manitoba is generally considered to be slightly acid (pH

6.1-6.5) black chernozem fine clay (Li *et al*., 2012) with the Winnipeg location being classified Riverdale silty clay (Lewis and Gulden, 2014). The Carman, MB soil type is fine sandy loam soil (Lewis and Gulden, 2014; Kahimba and Ranjan, 2007).

Conventional management techniques were implemented at both locations that included fertilizer, herbicide, insecticide and fungicide use as needed. Specifically for the Winnipeg location, fertilizer was supplied in the late October, 2013 by Crop Production Services with a custom blend equivalent to 134.7 kg/ha of nitrogen broadcast (formula: 46-0-0 (42.7%), 11-52-0 (31%), 0-0-60 (7.5%) 20-0-0-24 (18.7%). At the same time in 2013, Edge™ granular herbicide (Dow AgroSciences Indianapolis, IN) was added to the site at 24.8 kg/ha and worked twice with a tandem disk. Roundup WeatherMAX® Herbicide (Monsanto, St. Louis, MO) was applied on May 20th and May 29th, 2014 (60gal @0.5 litres/gal), to kill emerging weeds. All mixing and handling procedures were followed as per the manufacturer's instructions. Decis® (group 3 insecticide) (Bayer CropScience, North Carolina, USA) was applied on June 10th to control flea beetle (*Phyllotreta cruciferae*) populations at 148 ml/ha. Poast-Muster-Lontrel herbicide mix (Muster 29.6 g/ha, Lontrel 0.84 L/ha, Poast Ultra 0.67 L/ha and Merge at 1% volume/volume (1 L merge in 100 L of water) was applied on June 14th, 2014. Lontrel 360/E (Group 4 herbicide) (Dow AgroSciences Indianapolis, IN) was applied June 20th with a backpack sprayer (15 litres @100mL/L) to spot control Canadian thistle (*Cirsium arvense* L.). All mixing procedures were followed as per the manufacturer's label. Lance® (BASF Canada Inc. Mississauga, Ont.) granular fungicide (Group 7 fungicide) was applied on July 21th, 2014 at 350 g/ha to control Sclerotinia stem rot (*Sclerotinia sclerotiorum* Lib.). The experiments were harvested with a Wintersteiger nursery harvester (Nursery master Classic; Wintersteiger, Salt Lake City, UT) on September 18 - 22, 2014.

For the Carman research site, fertilizer was applied in late October, 2013 by Crop Production Services with a custom blend equivalent to 134.7 kg/ha of nitrogen with a custom applicator (formula: 46-0-0 (42.7%), 11-52-0 (31%), 0-0-60 (7.5%) 20-0-0-24 (18.7%). Approximately during the same time in 2013, Edge™ granular herbicide (Dow AgroSciences Indianapolis, IN) was added to the site at 27.9 kg/ha and worked twice with a tandem disk. Roundup WeatherMAX® Herbicide (Monsanto, St. Louis, MO) was applied on May 16th, 2014 (2.47 L/ha), to kill emerging weeds. All mixing and handling procedures were followed as per the manufacturer's instructions. Decis® (group 3 insecticide) (Bayer CropScience, North Carolina, USA) was applied on June 10th to control flea beetle (*Phyllotreta cruciferae*) populations at 148 ml/ha. All mixing procedures followed the manufacturer's label. Lance® (BASF Canada Inc. Mississauga, Ont.) granular fungicide (Group 7 fungicide) was applied on July 17th, 2014 at 350 g/ha to control Sclerotinia stem rot. The Carman yield plots were swathed on August 18th – 19th. Unfortunately, the Carman location was hit with hail following swathing and this severely affected the yield, leaving only the Winnipeg site usable for heterosis calculations.

5.40 Hybrid Heterosis

Mid-parent hybrid heterosis was estimated on mid-parent values using the formula MPH = (F_1 - MP) / MP x 100% where MP = mean parent yield and F_1 = mean hybrid yield (Riaz *et al*., 2001). High-parent hybrid heterosis was estimated using the high parent value from each parental pairing using the following formula HPH = (F_1 - HP) / HP x 100% where HP = high parent yield and F_1 = hybrid yield (Riaz *et al*., 2001).

5.41 Statistical Analysis

Statistical analysis for mean (μ) and coeffiecent of variation (CV) for all data tables was completed with Excel 2010. Mean (μ) values were obtained with the Averge function in Excel and standard deviation (σ) was obtained with the STDEV.P function in Excel. CV was obtained as dictated by the formula $CV = \frac{\sigma}{\mu}$ x 100% in Excel. Least significant difference (LSD) for seed yield was completed in Agrobase Generation II® (Agronomix Software, Inc. Winnipeg, MB) with the analysis of complete blocks (ACB) function. All regression analyses were completed in Microsoft Excel 2010 (Microsoft, Mississauga, Ontario) by using the scatterplot function and selecting hybrid yield as the Y-axis and genetic distance as the X-axis. Once the scatterplot graph was displayed a linear regression trend line was added with the R^2 value displayed for each trend line. Regression analysis for heterosis displays genetic distance on the Y-axis and heterosis on the X-axis.

5.5 RESULTS

5.51 Genetic Distance Estimates

SRAP genotyping produced 293 polymorphic bands for the 79 genotypes. GBS genotyping produced 80,005 bi-allelic SNP's over 95 genotypes. Sixteen genotypes were removed from this analysis to standardize each method using the same 79 genotypes. SRAP genetic distance (GD) estimates for each parental combination ranged from 0.233 GD (hybrid 2) to 0.44 GD (hybrid 21). GBS genetic distance estimates for each parental combination ranged from 0.057 GD (hybrid 6) to 0.173 GD (hybrid 24). Table 5.2 summarizes all 44 hybrid combinations with mean yield data and the genetic distance between parents from both SRAP and GBS.

Table 5.2: Mean hybrid yield, hybrid yield over mid-parent yield, hybrid yield over high-parent yield, and the genetic distance for each hybrid combination using each method of sequence related amplified polymorphism (SRAP) and genotyping-by-sequencing (GBS) based on Nei and Tamura-Nei (TN93) distance formulas, respectively, using yield data from Winnipeg, 2014.

Hybrid designation	Hybrid yield (kg/ha)	Hybrid yield over mid-parent value (kg/ha)	Hybrid yield over high-parent value (kg/ha)	SRAP-GD (Nei 1972)	GBS-GD (TN93)
1	2400	467	58	0.250	0.078
2	2640	652	190	0.233	0.079
3	2855	1218	405	0.298	0.110
4	2623	659	173	0.364	0.107
5	2338	347	-119	0.292	0.100
6	2869	1023	488	0.315	0.057
7	2754	1154	1079	0.342	0.102
8	2458	858	782	0.342	0.102
9	2894	886	402	0.279	0.085
10	2668	660	176	0.279	0.085
11	2245	739	57	0.267	0.092
12	2608	883	201	0.297	0.092
13	2389	775	420	0.315	0.092
14	2762	929	574	0.276	0.092
15	2372	530	212	0.269	0.094
16	2014	271	52	0.316	0.078
17	2359	867	199	0.275	0.101
18	2564	844	602	0.384	0.099

Hybrid designation	Hybrid yield (kg/ha)	Hybrid yield over mid-parent value (kg/ha)	Hybrid yield over high-parent value (kg/ha)	SRAP-GD (Nei 1972)	GBS-GD (TN93)
19	2375	385	-80	0.290	0.086
20	2759	754	274	0.324	0.099
21	3739	2213	1768	0.440	0.169
22	3296	922	592	0.301	0.127
Mean	2635	820	387	0.307	0.097
CV%	13.94				
LSD	533				
23	2661	942	877	0.341	0.169
24	2563	1367	755	0.350	0.173
25	2977	603	552	0.314	0.144
26	2726	561	410	0.327	0.124
27	2788	1008	973	0.340	0.118
28	2621	639	460	0.316	0.129
29	3131	1159	981	0.271	0.120
30	2402	1121	988	0.298	0.122
31	2685	1036	841	0.353	0.137
32	2399	1068	948	0.332	0.132
33	3032	1038	1026	0.353	0.137
34	2807	908	828	0.273	0.117
35	3327	716	529	0.342	0.125
36	3243	1145	938	0.298	0.128
37	2830	799	406	0.320	0.147
38	2937	1145	1145	0.393	0.145
39	3039	1383	1001	0.345	0.128
40	2977	783	768	0.271	0.101
41	2893	282	-67	0.310	0.137
42	2927	738	335	0.284	0.128
43	2395	451	413	0.267	0.137
44	3105	1073	696	0.289	0.110
Mean	2839	907	718	0.318	0.132
CV%	14.86				
LSD	583				
Total Mean	2737	864	552	0.313	0.115

Hybrid yield had a mean value of 2737 kg/ha for all hybrids, which was greater than the mean mid-parent yield by a mean of 864 kg/ha. Hybrid yield was on average greater than high-parent yield with a mean of 552 kg/ha. The best hybrid combination for yield was hybrid 21 as its yield was the highest at 3739 kg/ha with a substantial gain of 2213 kg/ha over its mid-parent value and

1768 kg/ha over its high parent value. These values translate into 145% mid-parent heterosis and 89% high-parent heterosis. In addition, hybrid combination 21 had the largest genetic distance for SRAP and the second largest genetic distance for GBS at 0.440 GD and 0.169 GD, respectively. The lowest yielding hybrid was combination 16 at 2014 kg/ha, however hybrid 16 still outperformed its mid-parent value by 271 kg/ha and its high-parent value by 52 kg/ha. Each genetic distance method calculated the genetic distance differently for hybrid 16 as SRAP genetic distance was high at 0.316 GD yet GBS was a tie for second lowest with a distance of 0.078 GD. From the 44 hybrid combinations all hybrids had higher yield than their mid-parent values and only three hybrid combinations (5, 19, 41) did not yield greater than their high-parent values with a yield loss of -119 kg/ha, -80 kg/ha and -67 kg/ha, respectively. Genetic distance offered no insight as to why hybrids 5, 19 and 41 did not out yield their high-parent values as SRAP genetic distance values were 0.292, 0.290 and 0.310 GD and GBS genetic distance was 0.100, 0.086 and 0.137 GD, respectively.

5.52 Consensus Tree Dendrogram

Figure 5.1 displays the neighbour joining cluster analysis for all 79 genotypes that were available for hybrid creation based on a consensus distance matrix obtained from the combination of both SRAP and GBS genetic distance models. From the 79 genotypes, 36 parents were selected to create 44 hybrids and the parental genotypes are color coded with the females in green and males in red. All hybrid combinations are considered to be inter-cluster hybrids.

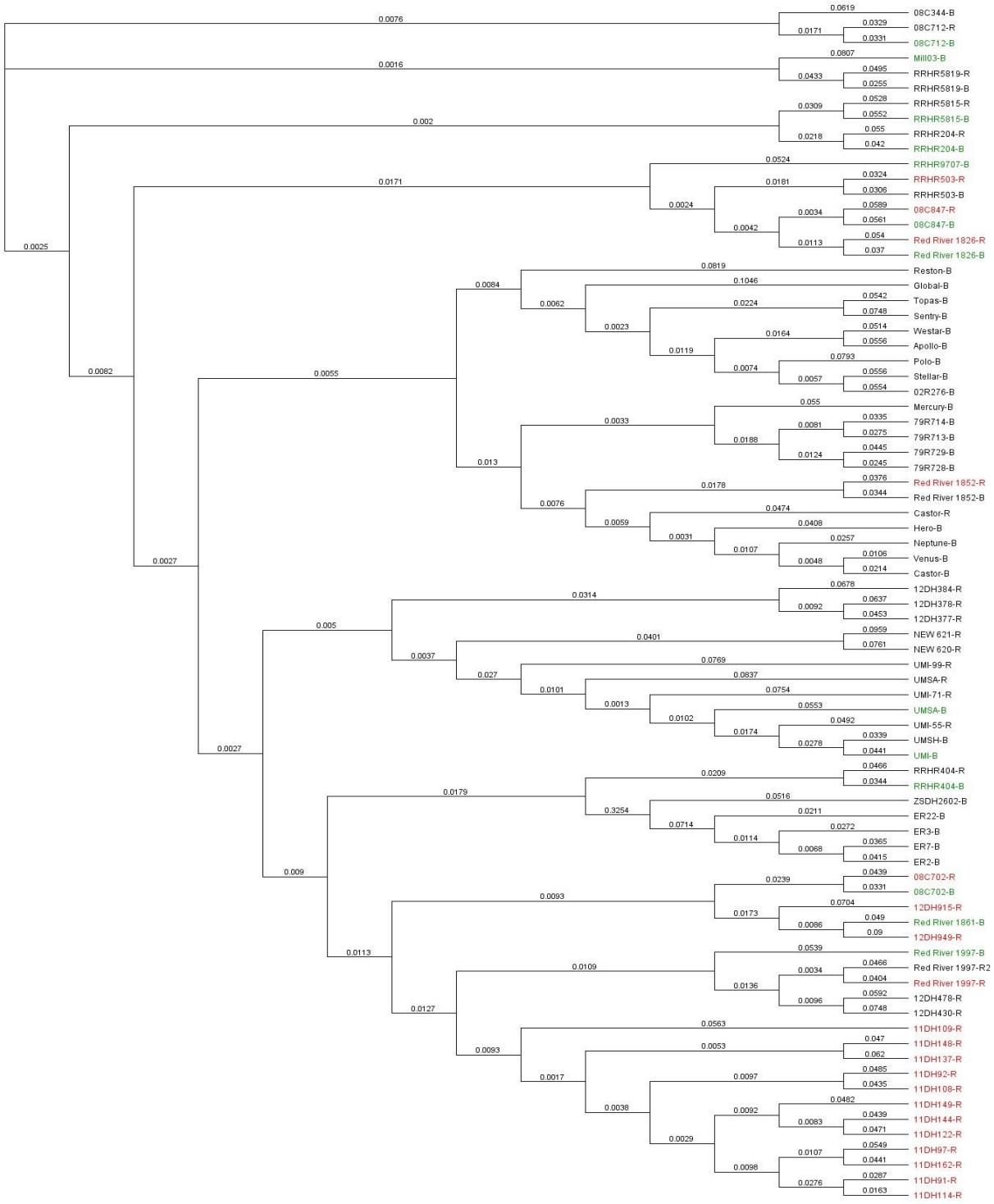

Figure 5.1: Neighbour joining dendrogram with 79 *Brassica napus* genotypes based on the combined mean distance matrixes of Nei and Tamura-Nei based on sequence related amplified polymorphisms SRAP and genotyping-by-sequencing (GBS) markers visualised in Geneious (V.8.03). Only parental genotypes chosen for hybrid creation are color coded with female genotypes (green) and male genotypes (red). The *ogu*-INRA cytoplasmic male sterility (CMS) system was used to ensure proper cross-pollination for hybrid production.

Following the completion of the consensus genetic distance matrix and neighbour joining dendrogram, each hybrid combination was measured for the consensus genetic distance between each parental genotype (Table 5.3). Hybrid combinations 11 and 12 were the shortest calculated distances at 0.159 GD each and both performed poorly in regard to yield at 2245 kg/ha and 2608 kg/ha, respectively. Whereas the largest genetic distance calculated was hybrid 21 at 0.304 GD, which was the highest yielding hybrid for this study at 3739 kg/ha.

Table 5.3: Consensus genetic distance (GD) obtained from the mean genetic distance of Nei and Tamura-Nei distance matrixes between each *Brassica napus* hybrid pairing sorted from highest to lowest yield.

Hybrid Designation	Female Genotype[1]	Male Genotype[2]	Consensus GD	Hybrid Yield (kg/ha)
21	UMI - B	12DH949-R	0.304	3739
35	UMSA -B	11DH109-R	0.202	3327
22	UMSA - B	11DH91-R	0.214	3296
36	UMSA -B	11DH114-R	0.204	3243
29	UMSA -B	08C847-R	0.179	3131
44	UMSA-B	12DH949-R	0.232	3105
39	UMSA-B	11DH144-R	0.237	3039
33	UMSA -B	11DH97-R	0.205	3032
25	UMI- B	08C702-R	0.229	2977
40	UMSA-B	11DH148-R	0.201	2977
38	UMSA -B	11DH137-R	0.266	2937
42	UMSA-B	11DH162-R	0.220	2927
9	Red River 1861-B	11DH137-R	0.181	2894
41	UMSA-B	11DH149-R	0.209	2893
6	Mill03-B	RRHR503-R	0.186	2869
3	08C712-B	12DH915-R	0.220	2855
37	UMSA -B	11DH122-R	0.231	2830
34	UMSA -B	11DH108-R	0.198	2807
27	UMSA -B	11DH97-R	0.219	2788
14	Red River 1997-B2	12DH949-R	0.194	2762
20	RRHR 9707-B	11DH137-R	0.211	2759
7	Red River 1826-B	11DH137-R	0.222	2754
26	UMI -B	12DH915-R	0.190	2726
31	UMSA -B	Red River 1826-R	0.211	2685
10	Red River 1861-B2	11DH137-R	0.181	2668
23	UMI - B	Red River 1852-R	0.255	2661

Hybrid Designation	Female Genotype[1]	Male Genotype[2]	Consensus GD	Hybrid Yield (kg/ha)
2	08C712-B	11DH137-R	0.192	2640
4	08C712-B	12DH949-R	0.258	2623
28	UMSA -B	08C702-R	0.181	2621
12	Red River 1997-B	12DH915-R	0.159	2608
18	RRHR 404-B	12DH949-R	0.256	2564
24	UMI - B	RRHR503-R	0.261	2563
8	Red River 1826-B2	11DH137-R	0.222	2458
30	UMSA -B	Red River 1997-R1	0.179	2402
1	08C702-B	11DH137-R	0.173	2400
32	UMSA -B	11DH92-R	0.207	2399
43	UMSA-B	12DH915-R	0.204	2395
13	Red River 1997-B2	12DH949-R	0.194	2389
19	RRHR 5815-B	11DH137-R	0.220	2375
15	RRHR 204-B	11DH137-R	0.215	2372
17	RRHR 404-B	12DH915-R	0.170	2359
5	08C847-B	11DH137-R	0.215	2338
11	Red River 1997-B	12DH915-R	0.159	2245
16	RRHR 404-B	11DH137-R	0.207	2014

[1]B=maintainer genotype, [2]R=restorer genotype for the *ogu*-INRA pollination control system.

5.53 Regression Analysis for Genetic Distance and Hybrid Yield

Following yield and genetic distance tabulation, a regression analysis was performed to investigate if genetic distance from either method of SRAP, GBS or consensus GD had any correlation to hybrid yield. Figures 5.2, 5.3 and 5.4 display the regression analysis of genetic distance against hybrid yield. SRAP regression analysis had a coefficient of determination (R^2) value of 0.17 or 17% of the variation in yield can be attributed to the SRAP genetic distance in this model (Figure 5.2). The Pearson product-moment correlation coefficient (r) value based on the regression analysis was 0.41 (significant at $df = 42$ $p = 0.01$), a positive weak correlation between SRAP genetic distance and hybrid yield (Figure 5.2).

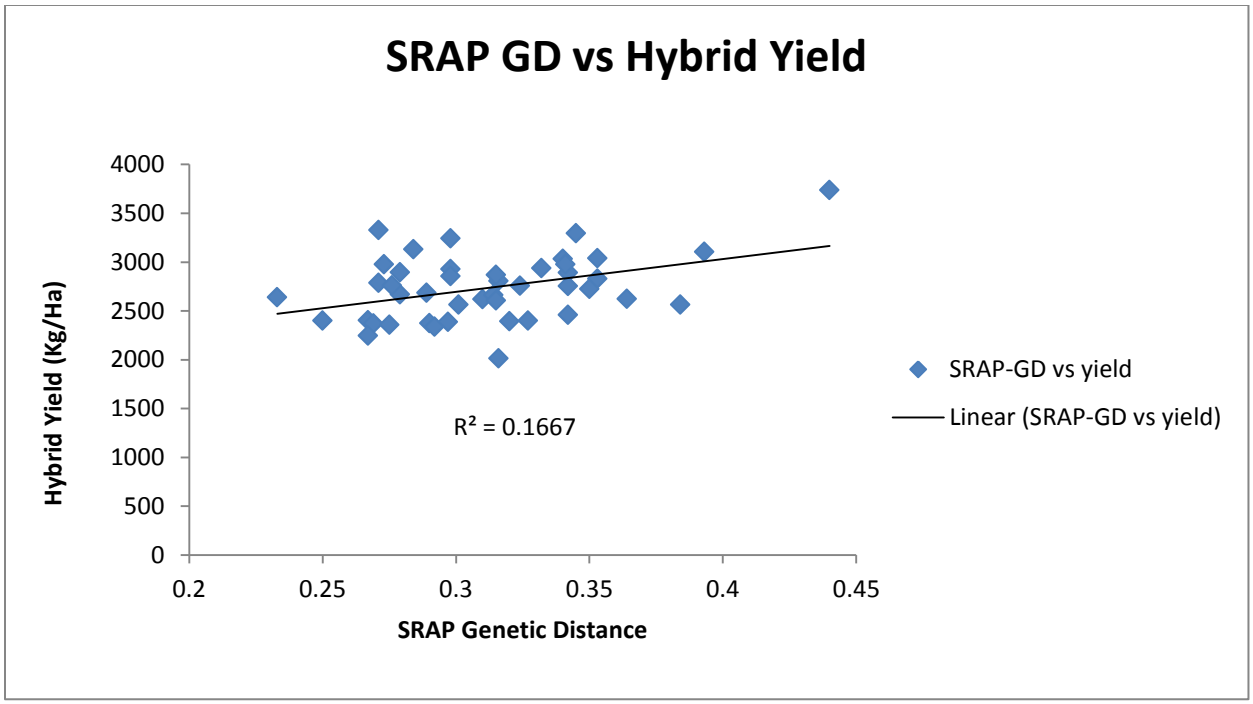

Figure 5.2: Forty-four *Brassica napus* hybrid combinations with hybrid yield (kg/ha) as a function of Nei's standard genetic distance calculated using 293 sequence related amplified polymorphism (SRAP) polymorphic bands.

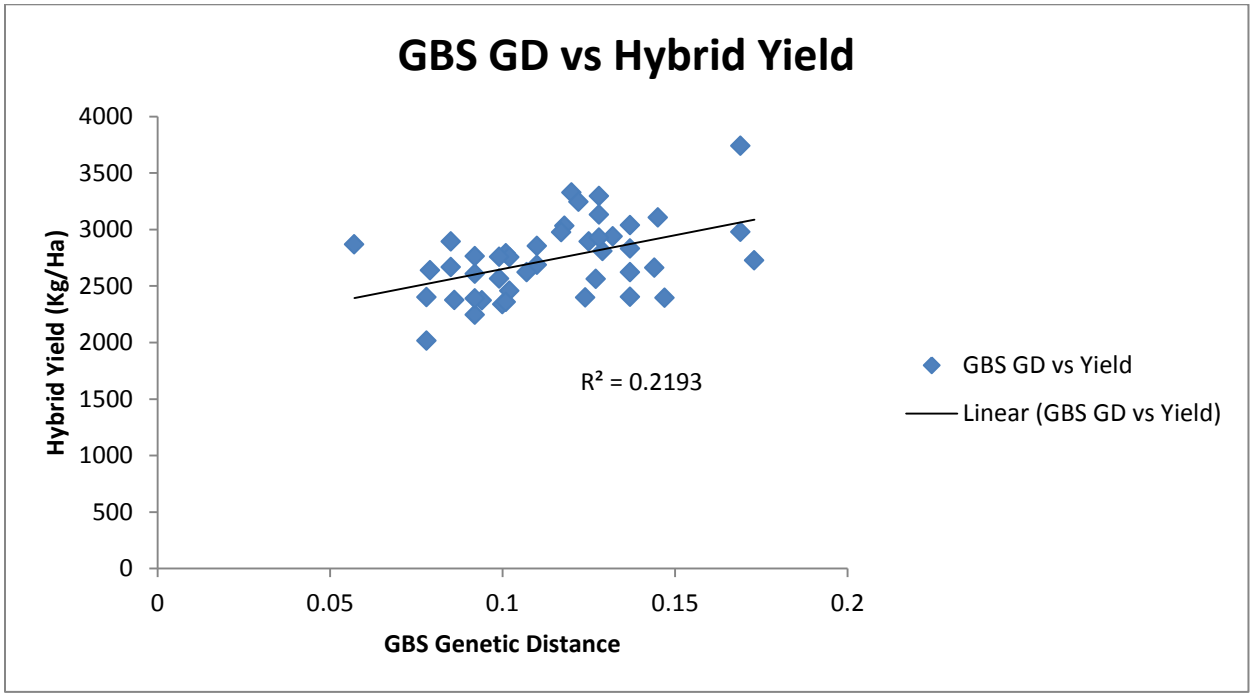

Figure 5.3: Forty-four *Brassica napus* hybrid combinations with hybrid yield (kg/ha) as a function of Tamura-Nei genetic distance calculated using 80,005 bi-allelic single nucleotide polymorphism's generated from genotyping-by-sequencing (GBS).

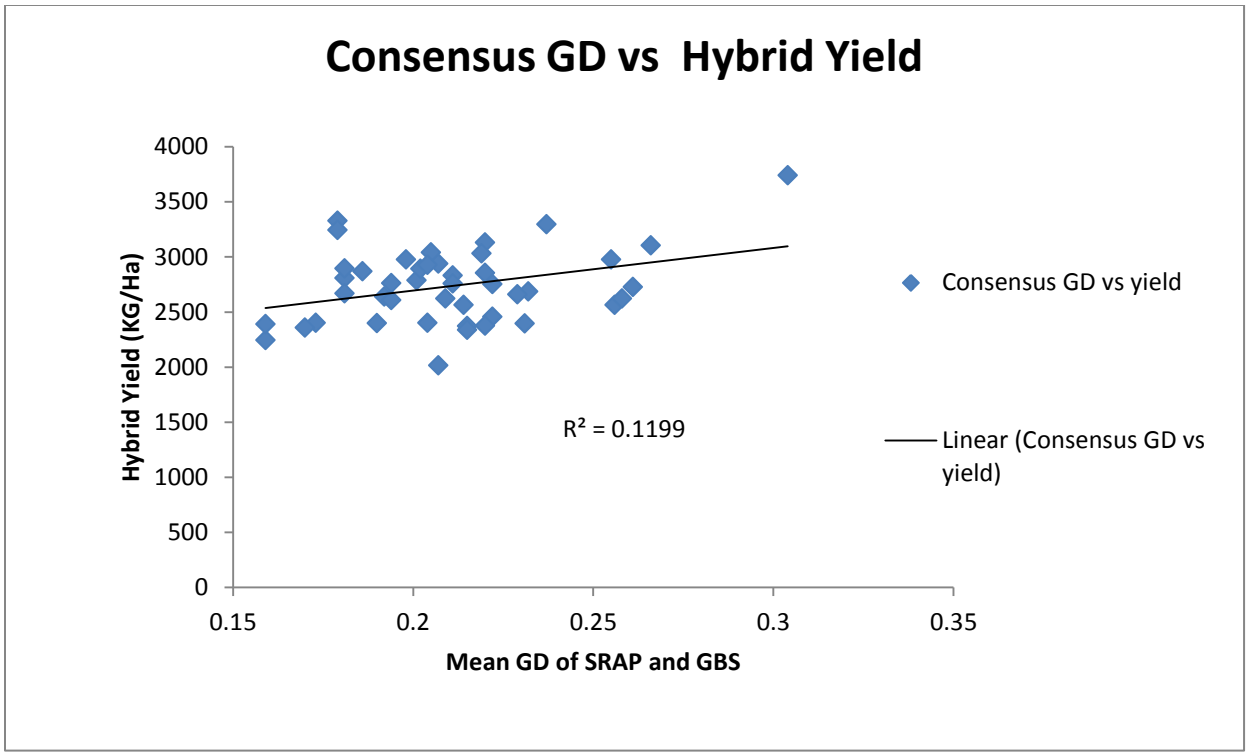

Figure 5.4: Forty-four *Brassica napus* hybrid combinations with hybrid yield as a function of consensus genetic distance obtained from the mean genetic distance matrixes from sequence related amplified polymorphisms (SRAP) and genotyping-by-sequencing (GBS) data.

Genetic distance regressed against hybrid yield as calculated by the Tamua-Nei model based on 80,005 SNP's obtained from GBS (Figure 5.3) had a slightly higher coefficient of determination (R^2) at 0.22 or 22% of the variation in yield could be attributed to the genetic distance between parental pairings. The correlation coefficient (r) was slightly higher for the Tamura-Nei model based on 80,005 SNP's at 0.47 (significant at *df* = 42 p = 0.01); however, this value is still considered a weak to moderate relationship between genetic distance and hybrid yield (Figure 5.3).

Following the determination of all genetic distances between parental pairings using the consensus genetic distance matrix, a regression analysis was performed to investigate whether a consensus genetic distance would prove more accurate for predicting hybrid yield, mid-parent

and high-parent heterosis. The consensus genetic distance was the worst predictor for hybrid yield (Figure 5.4) with an R^2 value of 0.12, and a correlation coefficient of r = 0.35

5.54 Regression Analysis for Genetic Distance and Mid-Parent Heterosis

Regression analysis was performed to investigate if genetic distance between different methods had any correlation to mid-parent heterosis. The SRAP model calculated with Nei's standard genetic distance had a coefficient of determination (R^2) of 0.28 and a correlation coefficient (r) of 0.53 (significant at *df* = 42 p = 0.01). The r value achieved is trending towards a moderate correlation.

The regression analysis for mid-parent heterosis against the Tamura-Nei genetics distance and the coefficient of determination (R^2) was 0.17 or 17% of the variation for mid-parent heterosis could be attributed to Tamura-Nei genetic distance between parental pairings. The correlation coefficient was 0.42 (significant at *df* = 42 p = 0.01). SRAP outperformed GBS in terms of mid-parent heterosis prediction based on Nei's standard genetic distance model over the Tamura-Nei genetic distance.

Regression analysis was performed to investigate if consensus genetic distance had any correlation to mid-parent heterosis. The coefficient of determination for mid-parent heterosis was R^2 = 0.21 and was higher than for the consensus genetic distance and hybrid yield. The correlation coefficient for consensus genetic distance and mid-parent heterosis was r = 0.46, a moderate correlation which was significant at P = 0.01. This value did not exceed the previous value of SRAP genetic distance for mid-parent heterosis; however, consensus regression analysis did have a higher correlation than the GBS model for mid-parent heterosis.

5.55 Regression Analysis for Genetic Distance and High-Parent Heterosis

Regression analysis was performed to investigate if genetic distance between the different methods had any correlation to high-parent heterosis. Again, SRAP outperformed the GBS distance model with the highest coefficient of determination for this study of $R^2 = 0.37$ and a correlation coefficient of r = 0.61 (significant at *df* = 42 p = 0.01). High-parent heterosis prediction for GBS had an R^2 value of 0.31 and an r value of 0.56 (significant at *df* = 42 p = 0.01). Both SRAP and GBS correlation coefficients are trending towards a moderately strong correlation.

Regression analysis was performed to investigate if consensus genetic distance had any correlation to high-parent heterosis. Again, consensus genetic distance was the worst predictor for high-parent heterosis with a coefficient of determination $R^2 = 0.24$ against the SRAP and GBS models of $R^2 = 0.37$ and $R^2 = 0.31$ respectively. The consensus genetic distance correlation coefficient was still significant at r = 0.49, P = 0.01.

Overall, GBS was the best predictor for high-yielding hybrid combinations however, SRAP genetic distance proved to have a higher r and R^2 values for mid-parent and high-parent heterosis over GBS and consensus methods, respectively (Table 5.4).

Table 5.4: Genetic distance (GD) comparison against hybrid yield, percent mid-parent and percent high-parent heterosis with the coefficient of determination (R^2) and the correlation coefficient (Pearson's r) based on sequence related amplified polymorphism (SRAP), genotyping-by-sequencing (GBS) and consensus models.

Genetic Distance Method Comparison	R^2	r
GBS GD vs Hybrid Yield	0.22	0.47*
SRAP GD vs Hybrid Yield	0.17	0.41*
Consensus GD vs Hybrid Yield	0.12	0.35
GBS GD vs % Mid-Parent Heterosis	0.17	0.42*
SRAP GD vs % Mid-Parent Heterosis	0.28	0.53*
Consensus GD vs % Mid-Parent Heterosis	0.21	0.46*
GBS GD vs % High-Parent Heterosis	0.31	0.56*
SRAP GD vs % High-Parent Heterosis	0.37	0.61*
Consensus GD vs % High-Parent Heterosis	0.24	0.49*

*Significant P = 0.01

5.6 DISCUSSION

Hybrid production in *B. napus* is currently a very competitive industry with nurmerous companies offering hybrid cultivars (Rahman, 2013). This study highlights some of the benefits that hybrid *B. napus* cultivars can offer as all hybrid combinations had higher yield when compared to their mid-parent yield values and only 3 hybrid combinations did not have higher yield than their high-parent values (hybrids 5, 19 and 41). This study supports previous research that *B. napus* hybrids can exhibit substantial heterosis over mid-parent and high-parent yield (Riaz *et al*., 2001; Ahmad and Quiros, 2011). Specifically, hybrid combination 21 showed 145% of mid-parent heterosis and 89% high-parent heterosis.

Numerous studies have implicated genetic distance as a predictor of high heterotic parental combinations in *B. napus* (Ali *et al*., 1995; Diers *et al.*, 1996; Riaz *et al.*, 2001; Radoev *et al.*, 2008; Cuthbert *et al.*, 2009; Ahmad and Quiros, 2011). Here, we discuss regression analyses for genetic distance calculated by Nei's standard genetic distance and Tamura-Nei's genetic distance model for SRAP and GBS, respectively. Regarding seed yield, the GBS method outperformed the SRAP method for predicting high-yield hybrid combinations with an R^2 of 0.22 compared to an R^2 of 0.17 (Table 5.4). The correlation coefficient was highest for GBS at r = 0.47 compared to SRAP at r = 0.41 (P = 0.01), and is generally considered a positive, weak correlation. This value for SRAP is smaller than previously reported SRAP values for yield (r = 0.64 and r = 0.60) (Riaz *et al*., 2001; Ahmad and Quiros, 2011). Currently, there is no data published to compare the GBS yield correlations.

Continuing with regression analysis for percent mid-parent and percent high-parent heterosis, it appears that SRAP outperformed GBS in both categories with R^2 values of 0.28 and 0.37 for

SRAP compared to R^2 values of 0.17 and 0.31 for GBS, respectively. The correlation coefficient was highest for SRAP with an r value of 0.61 for high-parent heterosis. This represents a moderate correlation for high-parent heterosis and is similar to previous results for SRAP (r = 0.66 and r = 0.67) (Riaz *et al*., 2001; Ahmad and Quiros, 2011). It is interesting that GBS seems to be a better predictor for yield, yet SRAP seems to be a better predictor for heterosis. The underlying mechanism for this discrepancy warrants further investigation.

Investigations into the hypothesis that a consensus method may help improve overall heterosis predictability has fallen short of its goal. The blending of genetic distance matrixes offered no improvement over either individual method of SRAP and GBS. Consensus regression values were lower than SRAP in every category and only provided a modest increase over GBS mid-parent heterosis association. Abeysundera *et al*., (2014) suggest that singular value decomposition (SVD) methods may improve the blending of distance matrixes through weighted averages between genes or markers; however, this requires further investigation.

The neighbour-joining clustering of genotypes based on a consensus matrix provided a scaffold for the investigations into inter-cluster hybrids. Figure 5.1 displays the cluster analysis for all seventy-nine genotypes available with select genotypes color-coded [male (red) and female (green)]. All hybrid combinations were considered inter-cluster hybrids for this study. Interestingly, the top twelve yielding hybrids (Table 5.3) were crosses from European-(EU) derived material with Canadian genotypes. These findings are directly in line with Cuthbert *et al*., (2009) in that the best yielding hybrid combinations have been crosses between EU and Canadian germplasm. Specifically, within the current cluster analysis, all EU-derived material formed a distinct cluster and future breeding strategies should focus on exploiting this genetically distinct high heterotic pool. This research demonstrates that SRAP can be just as

effective as and/or better than current next-generation sequencing technologies and procedures with a substantially smaller data set of 293 markers opposed to 80,005 SNP's. The moderate correlation for high-parent heterosis (r = 0.61) in this study is in line with previous reported values for SRAP (Riaz *et al*., 2001; Ahmad and Quiros, 2011) and therefore the SRAP procedure can be considered a tool for predicting high-heterotic parental pairings.

6.0 GENERAL DISCUSSION AND CONCLUSIONS

During the last twenty years, hybrid breeding has had a substantial role in shaping the agricultural industry. Maize, rice and canola are all crops where significant yield improvements have occurred impart through hybridization. However, continued hybrid improvement relies on the careful management of genetic diversity and germplasm resources.

Chapter three of this study explored the genetic diversity of 79 germplasm accessions through the phenotypic evaluation of 20 qualitative and quantitative traits. From these 20 traits it was demonstrated that resynthesized genotypes possess desirable traits such as high oil, high protein, variation in height, lodging resistance and high erucic acid content. Continued management of these genotypes should be a priority for the long-term integration of desirable traits into elite breeding material with the goal of broadening the genetic base of *B. napus*. Cluster analysis was also performed with the end goal of developing heterotic pool definitions based on phenotypic classification. These defined heterotic pools could be used to develop hybrid cultivars based on inter-cluster parental combinations. However, phenotypic pool definitions differ greatly from the genotypic definitions presented in chapter 4 and 5. Therefore, model optimization with a reduced set of traits should be undertaken before heterotic pairing is initiated. This may allow a reduction in the number of crosses required to find high-heterotic parental pairings.

Chapter 4 of this research investigated genotypic diversity using two independent methods, SRAP and GBS, in association with different genetic distance models. Neighbour-joining cluster analysis grouped all 79 genotypes into 11 and 12 clusters, respectively. Compare2trees found a 68% homology between each method and a manual comparison achieved 77% homology. Again these cluster definitions may prove useful in the development of heterotic pools based on genetic

distance and cluster analysis. SRAP was substantially more cost effective at $3.27 US per genotype opposed to GBS at approximately $43.49 US per genotype. Based on the similar results, SRAP would appear to be a more cost effective method for future investigations into genetic diversity; however GBS has many more downstream applications including SNP-trait association and copy number variation and may be worth the extra investment depending on the user's end goal.

Chapter 5 investigated the hypothesis of Falconer and McKay (1996) that genetic distance influences heterosis. Based on 44 hybrid combinations and genetic distance calculations achieved using SRAP and GBS data, a regression analysis was conducted. SRAP was the best predictor for mid-parent and high-parent heterosis with a correlation coefficient of r = 0.53 for mid-parent heterosis and r = 0.61 for high parent heterosis, moderate correlations. These values are similar to those achieved by Riaz *et al*. (2001), which increases the experimental evidence that SRAP data used for calculating genetic distance has a moderate correlation to hybrid heterosis in *B. napus*. Currently there is no published data detailing GBS genetic distance calculations in *B. napus*. Our correlation between GBS GD and hybrid yield was r = 0.47 and may be improved through SNP reduction and selection of optimal SNPs. Despite the overall success of SRAP and to a lesser extent GBS in hybrid prediction, developing a hybrid breeding program based solely on genetic distance may be inherently problematic as further research is needed to substantiate the exact influence genetic distance plays on heterosis. In addition, having only one site location with one year of data may contribute to the results. In conclusion, genetic distance does provide some degree of accuracy for the prediction of high-heterotic parental combinations. However, the predictive power is only moderate and should be considered as a single tool for the elucidation of proper parental pairings in the larger scale of a breeding

program. Multi-year and multi-location experiments have been conducted and need to be conducted in the future to truly gauge the interaction of genetic distance and hybrid heterosis.

7.0 FUTURE RESEARCH RECOMMENDATIONS

There is a general consensus within the plant breeding community that the broadening of genetic diversity along with the management of genetic resources should be a strategic priority for continual genetic improvement (Rahman *et al*., 2015). Although the use of resynthesized lines can be challenging, chapter three of this study demonstrates that these genotypes possess several desirable traits. The genotypic characterization in chapter 4 showed that the resynthesized lines possess genetic diversity not present in current Canadian germplasm as both phenotypic and genotypic methods clustered all resynthesized lines into distinct clusters. Future research recommendations include the use of resynthesized genotypes for novel trait development and their integration into elite *B. napus* inbred lines.

Continued development of heterotic pools with the inclusion of *B. napus* cultivars from a wider geographic origin, including Chinese and Australian genotypes, may provide a more complete view of genetic diversity present and may allow for a more refined cluster analysis. The inclusion of progenitor species, *B. rapa* and *B. oleracea* would also add more genetically divergent genotypes and aid in the assessment and accuracy of cluster placement.

Hybrid pairing based on genetic distance and cluster analysis should continue to aid in the verification and placement of genotypes through hybrid performance. Further investigation into SRAP and GBS correlations for yield and heterosis should be a future research priority concentrating on multi-year and multi-location experiments. Although the cost of GBS is substantially higher than SRAP, GBS has many more downstream applications and the extra investment may provide trait specific SNP's. However, markers from both genotypic methods may be linked to yield and yield components and therefore these determinations should be fully

explored through quantitative trait loci detection. Hybrids also need to be genotyped to correlate which markers from each parent contribute to hybrid heterosis in high yielding hybrids. This may allow for genome wide selection and prediction, greatly reducing the number of markers needed to discover high heterotic hybrids.

8.0 References

Abarshi, M. M., Mohammed, I. U., Wasswa, P., Hillocks, R. J., Holt, J., Legg, J. P., and Maruthi, M. N. (2010). Optimization of diagnostic RT-PCR protocols and sampling procedures for the reliable and cost-effective detection of Cassava brown streak virus. J. Virol. Methods. *163*(2): 353-359.

Abbadi, A., and Leckband, G. (2011). Rapeseed breeding for oil content, quality, and sustainability. Eur. J. Lipid Sci. Tech. *113*(10): 1198-1206.

Abdi, H., and Williams, L. J. (2010). Principal component analysis. Wiley Interdisciplinary Reviews: Computational Stat. *2*(4): 433-459.

Abeysundera, M., Kenney, T., Field, C., and Gu, H. (2014). Combining distance matrices on identical taxon sets for multi-gene analysis with singular value decomposition. PloS one *9*(4).

Ahmad, R., Potter, D., and Southwick, S. M. (2004). Genotyping of peach and nectarine cultivars with SSR and SRAP molecular markers. J. Am. Soc. Hortic. Sci. *129*(2): 204-210.

Ahmad, R., and Quiros, C. F. (2011). Inter-and intra-cluster heterosis in spring type oilseed rape (*Brassica napus* L.) hybrids and prediction of heterosis using SRAP molecular markers. SABRAO J. Breed. Genet. *43*(1): 27-43.

Ali, M., Copeland, L. O., Elias, S. G., and Kelly, J. D. (1995). Relationship between genetic distance and heterosis for yield and morphological traits in winter canola (*Brassica napus* L.). Theor. Appl. Genet. *91*(1): 118-121.

Allender, C. J., and King, G. J. (2010). Origins of the amphiploid species *Brassica napus* L. investigated by chloroplast and nuclear molecular markers. BMC Plant Boil. *10*(1): 54.

Al-Shehbaz, I. A., Beilstein, M. A., and Kellogg, E. A. (2006). Systematics and phylogeny of the *Brassicaceae* (*Cruciferae*): an overview. J. Syst. Evol. *259*(2-4): 89-120.

Amiri-Oghan, H., Fotokian, M. H., Javidfar, F., and Alizadeh, B. (2009). Genetic analysis of grain yield, days to flowering and maturity in oilseed rape (*Brassica napus* L.) using diallel crosses. Int. J. Plant Prod. *3*(2): 19-26.

Anderberg, M. R. (1973). Hierarchical clustering methods. Cluster Analysis for Applications. Academic Press (Ed) New York p.343-357.

Araus, J. L., and Cairns, J. E. (2014). Field high-throughput phenotyping: the new crop breeding frontier. Trends Plant Sci. *19*(1): 52-61.

Auldist, M. J., Marett, L. C., Greenwood, J. S., Wright, M. M., Hannah, M., Jacobs, J. L., and Wales, W. J. (2014). Replacing wheat with canola meal in a partial mixed ration increases the milk production of cows grazing at a restricted pasture allowance in spring. Anim. Prod. Sci. *54*(7): 869-878.

Baenziger, P. S., Graybosch, R. A., Dweikat, I. M., Wegulo, S. N., Hein, G. L., and Eskridge, K. M. (2008). Outstanding in their field: the phenotype of the 21st century plant breeder. Online article. URL: http://digitalcommons.unl.edu/entomologyfacpub/254.

Bannerot, H., Boulidard, L., Cauderon, Y., and Tempe, J. (1974). Cytoplasmic male sterility transfer from *Raphanus* to *Brassica*. Proc. EUCARPIA Meet. Crop Sect. Cruciferae *25*: 52-54.

Barbetti, M. J., Li, C. X., Banga, S. S., Banga, S. K., Singh, D., Sandhu, P. S., and You, M. P. (2015). New host resistances in Brassica napus and Brassica juncea from Australia, China and India: Key to managing Sclerotinia stem rot (*Sclerotinia sclerotiorum*) without fungicides. Crop Prot. *78*: 127-130.

Becker, H. C., Engqvist, G. M., and Karlsson, B. (1995). Comparison of rapeseed cultivars and resynthesized lines based on allozyme and RFLP markers. Theor. Appl. Genet. *91*(1): 62-67.

Bell, J. M. (1993). Factors affecting the nutritional value of canola meal: a review. Can. J. Anim. Sci. *73*(4): 689-697.

Bellenand, J. F., Baloutch, G., Ong, N., and Lecerf, J., (1980). Effects of coconut oil on heart lipids and on fatty acid utilization in rapeseed oil. Lipids. *15*: 938–943.

Berger, J., Suzuki, T., Senti, K. A., Stubbs, J., Schaffner, G., and Dickson, B. J. (2001). Genetic mapping with SNP markers in *Drosophila*. Nat. genet. *29*(4): 475-481.

Bhattacharjee, R., Khairwal, I. S., Bramel, P. J., and Reddy, K. N. (2007). Establishment of a pearl millet (*Pennisetum glaucum* L.) core collection based on geographical distribution and quantitative traits. Euphytica. *155*(1-2): 35-45.

Bonhomme, S., Budar, F., Lancelin, D., Small, I., Defrance, M. C., and Pelletier, G. (1992). Sequence and transcript analysis of the Nco2. 5 Ogura-specific fragment correlated with cytoplasmic male sterility in *Brassica* cybrids. Mol. Gen. Genet. *235*(2-3): 340-348.

Botstein, D., White, R. L., Skolnick, M., and Davis, R. W. (1980). Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am. J. Hum. Genet. *32*(3): 314.

Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y., and Buckler, E. S. (2007). TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics. *23*(19): 2633-2635.

Branca, F., (2008). Cauliflower and broccoli. In: Vegetables I: *Asteraceae*, *Brassicaceae*, *Chenopodiacea*,. *Cucurbitace.*, J. Prohens and F. Nuez (eds.), Springer, New York. pp. 151-186.

Brewin, D.G., and S. Malla (2012). The Consequence of Biotechnology: A Broad View of Changes in the Canadian Canola Sector, 1969 to 2012. AgBioForum. *15*:257-275.

Brown, G. G., Formanová, N., Jin, H., Wargachuk, R., Dendy, C., Patil, P., and Landry, B. S. (2003). The radish Rfo restorer gene of Ogura cytoplasmic male sterility encodes a protein with multiple pentatricopeptide repeats. Plant J. *35*(2): 262-272.

Bruce, A. B., 1910 The Mendelian theory of heredity and the augmentation of vigor. Science. *32*: 627–628.

Budar, F., and Pelletier, G. (2001). Male sterility in plants: occurrence, determinism, significance and use. Comptes Rendus de l'Académie des Sciences-Series III-Sciences de la Vie, *324*(6): 543-550.

Burns, D. R., Scarth, R., and McVetty, P. B. E. (1991). Temperature and genotypic effects on the expression of pol cytoplasmic male sterility in summer rape. Can. J. Plant Sci. *71*(3): 655-661.

Burstin, J., and Charcosset, A. (1997). Relationship between phenotypic and marker distances: theoretical and experimental investigations. Heredity. *79*(5): 477-483.

Bus, A., Körber, N., Snowdon, R. J., and Stich, B. (2011). Patterns of molecular variation in a species-wide germplasm set of *Brassica napus*. Theor. Appl. Genet. *123*(8): 1413-1423.

Canola Council, 2015: Internet database: URL: http://www.canolacouncil.org/markets-stats/statistics/harvest-acreage/

Carré, P., and Pouzet, A. (2014). Rapeseed market, worldwide and in Europe. OCL, *21*(1): D102.

Catchen, J. M., Amores, A., Hohenlohe, P., Cresko, W., and Postlethwait, J. H. (2011). Stacks: building and genotyping loci de novo from short-read sequences. G3: Genes, Genom., Genet. *1*(3): 171-182.

Cavalli-Sforza, L. L., and Edwards, A. W. (1967). Phylogenetic analysis. Models and estimation procedures. Am. J. Hum. Genet. *19*(3): 233.

Chalhoub, B., Denoeud, F., Liu, S., Parkin, I. A., Tang, H., Wang, X., and Sidebottom, C. H. (2014). Early allopolyploid evolution in the post-Neolithic *Brassica napus* oilseed genome. Science. *345*(6199): 950-953.

Chase, C. D. (2007). Cytoplasmic male sterility: a window to the world of plant mitochondrial–nuclear interactions. Trends Genet. *23*(2): 81-90.

Chen, C., Mitchell, S. E., Elshire, R. J., Buckler, E. S., & El-Kassaby, Y. A. (2013). Mining conifers' mega-genome using rapid and efficient multiplexed high-throughput genotyping-by-sequencing (GBS) SNP discovery platform. Tree Genet. Genomes. *9*(6): 1537-1544.

Chen, H. M., Zhang, Y. D., Jiang, F. Y., Huang, Y. X., Yao, W. H., Chen, X. H., and Fan, X. M. (2015). Genetic and Heterosis Analyses using Hayman's Six-Generation Model for Grain Yield and Yield Components in Maize. Crop Sci. *55*(3): 1006-1016.

Cheres, M. T., Miller, J. F., Crane, J. M., & Knapp, S. J. (2000). Genetic distance as a predictor of heterosis and hybrid performance within and between heterotic groups in sunflower. Theor. Appl. Genet. *100*(6): 889-894.

Cheung, F., Trick, M., Drou, N., Lim, Y. P., Park, J. Y., Kwon, S. J., and Bancroft, I. (2009). Comparative analysis between homoeologous genome segments of Brassica napus and its progenitor species reveals extensive sequence-level divergence. Plant Cell. *21*(7): 1912-1928.

Clarke, W. E., Parkin, I. A., Gajardo, H. A., Gerhardt, D. J., Higgins, E., Sidebottom, C., and Iniguez-Luy, F. L. (2013). Genomic DNA enrichment using sequence capture microarrays: a novel approach to discover sequence nucleotide polymorphisms (SNP) in *Brassica napus* L. PLoS ONE. *8*(12): e81992. doi:10.1371/journal.pone.0081992.

Crossa, J., Beyene, Y., Kassa, S., Pérez, P., Hickey, J. M., Chen, C., and Babu, R. (2013). Genomic prediction in maize breeding populations with genotyping-by-sequencing. G3: Genes, Genomes,Genet. *3*(11): 1903-1926.

Crow, J. F. (1948). Alternative hypotheses of hybrid vigor. Genetics. *33*(5): 477.

Cuevas, H. E., and Prom, L. K. (2013). Assessment of molecular diversity and population structure of the Ethiopian sorghum *Sorghum bicolor* (L.) *var* Moench germplasm collection maintained by the USDA–ARS National Plant Germplasm System using SSR markers. Genet. Resour. Crop Ev. *60*(6): 1817-1830.

Cuthbert, R. D., Crow, G., and McVetty, P. B. E. (2009). Assessment of agronomic performance and heterosis for agronomic traits in hybrid high erucic acid rapeseed (HEAR). Can. J. Plant Sci. *89*(2): 227-237.

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., and Durbin, R. (2011). The variant call format and VCFtools. Bioinformatics. *27*(15): 2156-2158.

Daun, J. K., Clear, K. M., and Williams, P. (1994). Comparison of three whole seed near-infrared analyzers for measuring quality components of canola seed. J. Am. Oil Chem. Soc. *71*(10): 1063-1068.

Dawkins, R. (1999). The extended phenotype: The long reach of the gene. Oxford University Press.

DeBonte, L., Iassonova, D., Liu, L., and Loh, W. (2012). Commercialization of high oleic canola oils. Lipid Technology. *24*(8): 175-177.

Delourme, R., Bouchereau, A., Hubert, N., Renard, M., and Landry, B. S. (1994). Identification of RAPD markers linked to a fertility restorer gene for the Ogura radish cytoplasmic male sterility of rapeseed (*Brassica napus* L.). Theor. Appl. Genet. *88*(6-7): 741-748.

Delourme, R., & Eber, F. (1992). Linkage between an isozyme marker and a restorer gene in radish cytoplasmic male sterility of rapeseed (*Brassica napus* L.). Theor. Appl. Genet. *85*(2-3): 222-228.

Delourme, R., Horvais, R., Barret, P., Renard, M., Foisset, N., Champagne, G., Cheung, W.Y. and Landry, B.S. (1998). Mapping the Radish Introgression Carrying the *RFO* Restorer Gene Used for the *OGU*-INRA Cytoplasmic Male Sterility in Rapeseed (*Brassica napus* L.). Acta Hortic. *459*: 135-138

Delucchi, C., Eyhérabide, G. H., Lorea, R. D., Presello, D. A., Otegui, M. E., and López, C. G. (2012). Classification of Argentine maize landraces in heterotic groups. Maydica. *57*(1): 26-33.

Diers, B. W., McVetty, P. B. E., and Osborn, T. C. (1996). Relationship between heterosis and genetic distance based on restriction fragment length polymorphism markers in oilseed rape (*Brassica napus* L.). Crop Sci. *36*(1): 79-83.

Diers, B. W., & Osborn, T. C. (1994). Genetic diversity of oilseed *Brassica napus* germplasm based on restriction fragment length polymorphisms. Theor. Appl. Genet. *88*(6-7): 662-668.

Dong, X., Kim, W. K., Lim, Y. P., Kim, Y. K., and Hur, Y. (2013). Ogura-CMS in Chinese cabbage *Brassica rapa* ssp. *pekinensis* causes delayed expression of many nuclear genes. Plant Sci. *199*: 7-17.

Downey R. K., and J. M. Bell. (1990). New developments in canola research. In: F. Shahidi, editor, Canola and rapeseed: Production, chemistry, nutrition and processing technology. Springer US, New York. p. 37-46.

Doyle, J. J., and Doyle, J. L. (1990). Isolation of DNA from small amounts of plant tissues. BRL focus. *12*: 13-15.

Duncan, R. W., Gilbertson, R. L., and Singh, S. P. (2012). Direct and marker-assisted selection for resistance to common bacterial blight in common bean. Crop Sci. *52*(4): 1511-1521.

East, E. M., (1908). Inbreeding in corn. Rep. Conn. Agric. Exp. Stn. p. 419–428.

Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., and Mitchell, S. E. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PloS one. *6*(5): e19379.

Engelke, T., Hirsche, J., and Roitsch, T. (2010). Anther-specific carbohydrate supply and restoration of metabolically engineered male sterility. J. Exp. Bot. *61*(10): 2693-2706.

Esposito, M. A., Gatti, I., Cravero, V. P., Anido, F. S. L., and Cointry, E. L. (2013). Combining abilities and heterotic groups in *Pisum sativum* L. Aust. J. Crop Sci. 7(11): 1634-1641

Falconer D. S., Mackay T. F. C. (1996). Introduction to quantitative genetics, 4th edn. Longman, London. p. 160-183.

Fan, Z., and Stefansson, B. R. (1986). Influence of temperature on sterility of two cytoplasmic male-sterility systems in rape (*Brassica napus* L.). Can. J. Plant Sci. *66*(2): 221-227.

Fang, G. H., and McVetty, P. B. E. (1989). Inheritance of male fertility restoration and allelism of restorer genes for the Polima cytoplasmic male sterility system in oilseed rape. Genome. *32*(6): 1044-1047.

Felsenstein, J. (1985). Confidence limits on phylogenies: an approach using the bootstrap. Evolution. p.783-791.

Ferriol, M., Pico, B., and Nuez, F. (2003). Genetic diversity of a germplasm collection of *Cucurbita pepo* using SRAP and AFLP markers. Theor. Appl. Genet. *107*(2): 271-282.

Fourmann, M., Barret, P., Renard, M., Pelletier, G., Delourme, R., and Brunel, D. (1998). The two genes homologous to Arabidopsis FAE1 co-segregate with the two loci governing erucic acid content in *Brassica napus*. Theor. Appl. Genet. *96*(6-7): 852-858.

Friedt, W. and Snowdon, R. (2010). Oilseed rape. In: Oil crops. Handbook of Plant Breeding, vol. 4. Vollmann, J. and Rajan, J. (eds.). Springer-Verlag, New York, p. 91–126.

Fu, D., Xiao, M., Hayward, A., Fu, Y., Liu, G., Jiang, G., and Zhang, H. (2014). Utilization of crop heterosis: a review. Euphytica. *197*(2): 161-173.

Fu, T. D. (1981). Production and research of rapeseed in the People's Republic of China. Eucarpia Cruciferae Newsl. *6*: 6-7.

Garg, H., Li, H., Sivasithamparam, K., and Barbetti, M. J. (2013). Differentially expressed proteins and associated histological and disease progression changes in cotyledon tissue of a resistant and susceptible genotype of *Brassica napus* infected with *Sclerotinia sclerotiorum*. PloS one. *8*(6): e65205.

Geleta, L. F., Labuschagne, M. T., and Viljoen, C. D. (2004). Relationship between heterosis and genetic distance based on morphological traits and AFLP markers in pepper. Plant Breeding. *123*(5): 467-473.

Gepts, P. (2006). Plant genetic resources conservation and utilization. Crop Sci. *46*(5): 2278-2292.

Girke, A., Schierholt, A., and Becker, H. C. (2012). Extending the rapeseed genepool with resynthesized *Brassica napus* L. I: Genetic diversity. Genet. Resour. Crop Ev. *59*(7): 1441-1447.

Gixhari B., Pavelková M., Ismaili H., Vrapi H., Jaupi A., and Smýkal P. (2014). Genetic diversity of Albanian pea (*Pisum sativum* L.) landraces assessed by morphological traits and molecular markers. Czech J. Genet. Plant Breed. 50: 177–184.

Ghani, M.A., Zhang, L., Li, J., Liu, B., and Chen, L., (2013). In: Biotechnology of Crucifers: Hybrid technology in cruciferous vegetables. Gupta, S.K. (eds.), Springer, New York, p. 209–220.

Glaubitz, J. C., Casstevens, T. M., Lu, F., Harriman, J., Elshire, R. J., Sun, Q., and Buckler, E. S. (2014). TASSEL-GBS: a high capacity genotyping by sequencing analysis pipeline. PloS one. *9*(2): e90346.

Gourret, J. P., Delourme, R., and Renard, M. (1992). Expression of *ogu* cytoplasmic male sterility in cybrids of *Brassica napus*. Theor. Appl. Genet. *83*(5): 549-556.

Grant, I., and Beversdorf, W. D. (1985). Heterosis and combining ability estimates in spring-planted oilseed rape (*Brassica napus* L.). Can. J. Genet. Cytol. *27*(4): 472-478.

Gül, M., Becker, H. C., and Ecke, W. (2003). QTL mapping and analysis of QTL x nitrogen interactions for protein and oil contents in *Brassica napus* L. In: Proceedings of the 11th international rapeseed congress, Copenhagen, Denmark, p. 91–93.

Hanson, M. R., and Bentolila, S. (2004). Interactions of mitochondrial and nuclear genes that affect male gametophyte development. Plant Cell. *16*(1): S154-S169.

Harbered, D. J. (1976). Cytotaxonomic studies of Brassica and related genera. In: The biology and chemistry of the Cruciferae. Vaughan J. G., MacLeod A. J., Jones M. G., (eds.). Academic, London. p. 47–68.

Harlan, J. R., & de Wet, J. M. (1971). Toward a rational classification of cultivated plants. Taxon. *20*(4): 509-517.

Harper, A. L., Trick, M., Higgins, J., Fraser, F., Clissold, L., Wells, R., and Bancroft, I. (2012). Associative transcriptomics of traits in the polyploid crop species *Brassica napus*. Nature Biotechnology. *30*(8): 798-802.

Harvey, B. L., and Downey, R. K. (1964). The inheritance of erucic acid content in rapeseed (*Brassica napus*). Can. J. Plant Sci. *44*(1): 104-111.

Hasan, M., Seyis, F., Badani, A. G., Pons-Kühnemann, J., Friedt, W., Lühs, W., and Snowdon, R. J. (2006). Analysis of genetic diversity in the *Brassica napus* L. gene pool using SSR markers. Genet. Resour. Crop Ev. *53*(4): 793-802.

Havlíčková, L., Klíma, M., Přibylová, M., Hilgert-Delgado, A. A., Kučera, V., and Čurn, V. (2014). Non-destructive *in vitro* selection of microspore-derived embryos with the fertility restorer gene for CMS *Ogu*-INRA in winter oilseed rape *Brassica napus*L.). Electron. J. Biotechn. *18*(1): 58-60.

Hayward, A., Morgan, J. D., and Edwards, D. (2012). Special Issue: Reviews; SNP discovery and applications in *Brassica napus*. J. Plant Biotechn. *39*(1): 49-61.

Heyn, F. W. (1976). Transfer of restorer genes from Raphanus to cytoplasmic male sterile *Brassica napus*. Cruciferae Newsl. *1*: 15-16.

Hougen, F. W., and Bodo, V. (1973). Extraction and methanolysis of oil from whole or crushed rapeseed for fatty acid analysis. J. Am. Oil Chem. Soc. *50*(7): 230-234.

Hu, X., Sullivan-Gilbert, M., Kubik, T., Danielson, J., Hnatiuk, N., Marchione, W., and Thompson, S. A. (2008). Mapping of the Ogura fertility restorer gene Rfo and development of Rfo allele-specific markers in canola (*Brassica napus* L.). Mol. Breed. *22*(4): 663-674.

Hua, W., Li, R. J., Zhan, G. M., Liu, J., Li, J., Wang, X. F., and Wang, H. Z. (2012). Maternal control of seed oil content in *Brassica napus*: the role of silique wall photosynthesis. The Plant Journal. *69*(3): 432-444.

Huang, Z., Chen, Y., Yi, B., Xiao, L., Ma, C., Tu, J., and Fu, T. (2007). Fine mapping of the recessive genic male sterility gene (Bnms3) in *Brassica napus* L. Theor. Appl. Genet. *115*(1): 113-118.

Hufford, M. B., Xu, X., Van Heerwaarden, J., Pyhäjärvi, T., Chia, J. M., Cartwright, R. A., and Ross-Ibarra, J. (2012). Comparative population genomics of maize domestication and improvement. Nat. Genet. *44*(7): 808-811.

Hwang, S. F., Howard, R. J., Strelkov, S. E., Gossen, B. D., & Peng, G. (2014). Management of clubroot (*Plasmodiophora brassicae*) on canola (*Brassica napus*) in western Canada. Can. J. Plant Pathol. *36*(sup1): 49-65.

Hwang, S. F., Strelkov, S. E., Feng, J. I. E., Gossen, B. D., and Howard, R. J. (2012). Plasmodiophora brassicae: a review of an emerging pathogen of the Canadian canola (*Brassica napus*) crop. Mol. Plant. Pathol. *13*(2): 105-113.

Jain, A., Bhatia, S., Banga, S. S., Prakash, S., and Lakshmikumaran, M. (1994). Potential use of random amplified polymorphic DNA (RAPD) technique to study the genetic diversity in Indian mustard (*Brassica juncea*) and its relationship to heterosis. Theor. Appl. Genet. *88*(1): 116-122.

Jean, M., Brown, G. G., and Landry, B. S. (1997). Genetic mapping of nuclear fertility restorer genes for the 'Polima'cytoplasmic male sterility in canola (*Brassica napus* L.) using DNA markers. Theor. Appl. Genet. *95*(3): 321-328.

Jesske, T., Olberg, B., Schierholt, A., and Becker, H. C. (2013). Resynthesized lines from domesticated and wild *Brassica* taxa and their hybrids with *B. napus* L.: genetic diversity and hybrid yield. Theor. Appl. Genet. *126*(4): 1053-1065.

Kaeppler, S. (2012). Heterosis: many genes, many mechanisms—end the search for an undiscovered unifying theory. ISRN Botany. Doi:10.5402/2012/682824

Kahimba, F. C., and Ranjan, R. S. (2007). Soil temperature correction of field TDR readings obtained under near freezing conditions. Can. Biosys. Eng. *49*: 1.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Mentjies, P., and Drummond, A. (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics. *28*(12): 1647-1649.

Keeble, F., and C. Pellew, 1910 The mode of inheritance of stature and of time of flowering in peas (*Pisum sativum*). J. Genet. (1): 47–56.

Khajali, F., and Slominski, B. A. (2012). Factors that affect the nutritive value of canola meal for poultry. Poul. Sci. *91*(10): 2564-2575.

Kim, Y., Himmelsbach, D. S., and Kays, S. E. (2007). ATR-Fourier transform mid-infrared spectroscopy for determination of trans fatty acids in ground cereal products without oil extraction. J. Agric. Food Chem. *55*(11): 4327-4333.

Koizuka, N., Imai, R., Fujimoto, H., Hayakawa, T., Kimura, Y., Kohno-Murase, J., and Imamura, J. (2003). Genetic characterization of a pentatricopeptide repeat protein gene, orf687, that restores fertility in the cytoplasmic male-sterile Kosena radish. The plant journal. *34*(4): 407-415.

Kondra, Z. P., and Stefansson, B. R. (1970). Inheritance of the major glucosinolates of rapeseed (*Brassica napus*) meal. Can. J. Plant Sci. *50*(6): 643-647.

Krzymanski, J., and Downey, R. K. (1969). Inheritance of fatty acid composition in winter forms of rapeseed, Brassica napus. Can. J. Plant Sci. *49*(3): 313-319.

Lapointe, F.J., and Cucumel, G. (1997). The Average Consensus Procedure: Combination of Weighted Trees Containing Identical or Overlapping Sets of Taxa. Syst. Bio. *46*(2): 306–312. http://doi.org/10.2307/2413625

Li, G., and Quiros, C. F. (2001). Sequence-related amplified polymorphism (SRAP), a new marker system based on a simple PCR reaction: its application to mapping and gene tagging in Brassica. Theor. Appl. Genet. *103*(2-3): 455-461.

Li, H., and Durbin, R. (2010). Fast and accurate long-read alignment with Burrows–Wheeler transform. Bioinformatics. *26*(5): 589-595.

Li, H., and Homer, N. (2010). A survey of sequence alignment algorithms for next-generation sequencing. Briefings in bioinformatics. *11*(5): 473-483.

Li, H., Li, J., Zhao, B., Wang, J., Yi, L., Liu, C., and Liu, K. (2014). Generation and characterization of tribenuron-methyl herbicide-resistant rapeseed (*Brasscia napus*) for hybrid seed production using chemically induced male sterility. Theor. Appl. Genet. *128*(1): 107-118.

Li, J., Zhao, Z., Hayward, A., Cheng, H., and Fu, D. (2015). Integration analysis of quantitative trait loci for resistance to *Sclerotinia sclerotiorum* in *Brassica napus*. Euphytica. *205*(2): 483-489.

Li, R., Khafipour, E., Krause, D. O., Entz, M. H., de Kievit, T. R., and Fernando, W. D. (2012). Pyrosequencing reveals the influence of organic and conventional farming systems on bacterial communities. PloS one. *7*(12): e51897.

Li, X., Li, X., Fridman, E., Tesso, T. T., and Yu, J. (2015). Dissecting repulsion linkage in the dwarfing gene Dw3 region for sorghum plant height provides insights into heterosis. Proc. Natl. Acad. Sci. *112*(38): 11823-11828.

Li, Y., Beisson, F., Pollard, M., and Ohlrogge, J. (2006). Oil content of Arabidopsis seeds: the influence of seed anatomy, light and plant-to-plant variation. Phytochemistry. *67*(9): 904-915.

Lin, L., Allemekinders, H., Dansby, A., Campbell, L., Durance-Tod, S., Berger, A., and Jones, P. J. (2013). Evidence of health benefits of canola oil. Nutrition reviews. *71*(6): 370-385.

Liu, H., Bayer, M., Druka, A., Russell, J. R., Hackett, C. A., Poland, J., and Waugh, R. (2014). An evaluation of genotyping by sequencing (GBS) to map the Breviaristatum-e (ari-e) locus in cultivated barley. BMC genomics. *15*(1): 104.

Liu, K. S. (1994). Preparation of fatty acid methyl esters for gas-chromatographic analysis of lipids in biological materials. J. Am. Oil Chem. Soc. *71*(11): 1179-1187.

Liu, K., and Muse, S. V. (2005). PowerMarker: an integrated analysis environment for genetic marker analysis. Bioinformatics. *21*(9): 2128-2129.

Liu, Q., Wu, L., Pu, H., Li, C., and Hu, Q. (2012). Profile and distribution of soluble and insoluble phenolics in Chinese rapeseed (*Brassica napus*). Food chemistry. *135*(2): 616-622.

Lee, A., and Willcox, B. (2014). Minkowski Generalizations of Ward's Method in Hierarchical Clustering. J. Classif. *31*(2): 194-218.

Lewis, D. W., and Gulden, R. H. (2014). Effect of Kochia (*Kochia scoparia*) Interference on Sunflower (*Helianthus annuus*) Yield. Weed Sci. *62*(1): 158-165.

Lombard, V., Baril, C. P., Dubreuil, P., Blouet, F., and Zhang, D. (2000). Genetic relationships and fingerprinting of rapeseed cultivars by AFLP: consequences for varietal registration. Crop Sci. *40*(5): 1417-1425.

Lowhorn, G. L. (2007). Qualitative and quantitative research: How to choose the best design. In: Academic Business World International Conference. Nashville, Tennessee.

Lu, G. Y., Yang, G. S., and Fu, T. D. (2004). Molecular mapping of a dominant genic male sterility gene Ms in rapeseed (*Brassica napus*). Plant Breeding. *123*(3): 262-265.

Lu, Y., Yan, J., Guimaraes, C. T., Taba, S., Hao, Z., Gao, S., and Xu, Y. (2009). Molecular characterization of global maize breeding germplasm based on genome-wide single nucleotide polymorphisms. Theor. Appl. Genet. *120*(1): 93-115.

Mahner, M., and Kary, M. (1997). What exactly are genomes, genotypes and phenotypes? And what about phenomes?. J. Theor. Bio. *186*(1): 55-63.

Mailer, R. J., Scarth, R., and Fristensky, B. (1994). Discrimination among cultivars of rapeseed (*Brassica napus* L.) using DNA polymorphisms amplified from arbitrary primers. Theor. Appl. Genet. *87*(6): 697-704.

Marcroft, S. J., Van de Wouw, A. P., Salisbury, P. A., Potter, T. D., and Howlett, B. J. (2012). Effect of rotation of canola (*Brassica napus*) cultivars with different complements of blackleg resistance genes on disease severity. Plant Pathology. *61*(5): 934-944.

Mason, A. S., Batley, J., Bayer, P. E., Hayward, A., Cowling, W. A., and Nelson, M. N. (2014). High-resolution molecular karyotyping uncovers pairing between ancestrally related *Brassica* chromosomes. New Phytol. *202*(3): 964-974.

Mason, A. S., Nelson, M. N., Takahira, J., Cowling, W. A., Alves, G. M., Chaudhuri, A., and Batley, J. (2014). The Fate of Chromosomes and Alleles in an Allohexaploid *Brassica* Population. Genetics. *197*(1): 273-283.

Matthäus, B. (2006). Utilization of high-oleic rapeseed oil for deep-fat frying of French fries compared to other commonly used edible oils. Eur. J. Lipid Sci. Tech. *108*(3): 200-211.

McKeown, P. C., Fort, A., Duszynska, D., Sulpice, R., and Spillane, C. (2013). Emerging molecular mechanisms for biotechnological harnessing of heterosis in crops. Trends Biotechnol. *31*(10): 549-551.

McVetty, P. B. E., Cuthbert, J. L., Marwede, V., Paulmann, W., Sass, O., Duncan, R. W., and Zelmer, C. D. (2014). HYHEAR 1 hybrid Roundup Ready™ high erucic acid, low glucosinolate summer rape. Can. J. Plant Sci. *94*(2): 453-455.

McVetty, P. B. E., Duncan, R. W., Fernando, W. G. D., Li, G., and Zelmer, C. D. (2012). Red River 1861 Roundup Ready™ high erucic acid, low glucosinolate summer rape. Can. J. Plant Sci. *92*(7): 1407-1409.

McVetty, P. B., and Duncan, R. W. (2015). Canola, Rapeseed, and Mustard: For Biofuels and Bioproducts. In: Industrial Crops. Springer New York. p. 133-156.

McVetty, P. B., Fernando, D., Li, G., Tahir, M., and Zelmer, C. (2009). High-erucic acid, low-glucosinolate rapeseed (HEAR) cultivar development in Canada. Biocatal. Agric. Biotechnol. p. 43-61.

McVetty, P. B., and Scarth, R. (2012). Breeding Oil Quality for Edible and Industrial Applications in Oilseed Rape and Linseed. In: Technological Innovations in Major World Oil Crops, Volume 2. Springer New York. p. 105-122.

Melchinger, A. E. (2010). The International Conference on Heterosis in Plants. Theor. Appl. Genet. *120*(2): 201-203.

Mohammadi, S. A., and Prasanna, B. M. (2003). Analysis of genetic diversity in crop plants—salient statistical tools and considerations. Crop Sci. *43*(4): 1235-1248.

Mouchet, M., Guilhaumon, F., Villéger, S., Mason, N. W., Tomasini, J. A., and Mouillot, D. (2008). Towards a consensus for calculating dendrogram-based functional diversity indices. Oikos. *117*(5): 794-800.

Moradi, M., Hoveize, M. S., and Ebrahimi, A. (2015). Evaluation of drought tolerance some of spring canola cultivars (*Brassica napus* L.). In. J. Biosci. *6*(5): 103-108.

Morota, G., Abdollahi-Arpanahi, R., Kranis, A., and Gianola, D. (2014). Genome-enabled prediction of quantitative traits in chickens using genomic annotation. BMC genomics. *15*(1): 109.

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

Nath, U. K., Wilmer, J. A., Wallington, E. J., Becker, H. C., and Möllers, C. (2009). Increasing erucic acid content through combination of endogenous low polyunsaturated fatty acids alleles with Ld-LPAAT+ Bn-fae1 transgenes in rapeseed (*Brassica napus* L.). Theor. Appl. Genet. *118*(4): 765-773.

Nei, M. (1972). Genetic distance between populations. American naturalist, p.283-292.

Nei, M. (1978). Estimation of average heterozygosity and genetic distance from a small number of individuals. Genetics. *89*(3): 583-590.

Nei, M. (1987). Molecular evolutionary genetics. Columbia University Press.

Nielsen, R., Paul, J. S., Albrechtsen, A., and Song, Y. S. (2011). Genotype and SNP calling from next-generation sequencing data. Nat. Rev. Genet. *12*(6): 443-451.

Nye, T. M., Lio, P., and Gilks, W. R. (2006). A novel algorithm and web-based tool for comparing two alternative phylogenetic trees. Bioinformatics. *22*(1): 117-119.

Odong, T. L., van Heerwaarden, J., Jansen, J., van Hintum, T. J., and van Eeuwijk, F. A. (2011). Determination of genetic structure of germplasm collections: are traditional hierarchical clustering methods appropriate for molecular marker data?. Theor. Appl. Genet. *123*(2): 195-205.

Ogura, H. (1968). Studies on the new male-sterility in Japanese radish, with special reference to the utilization of this sterility towards the practical raising of hybrid seeds. Mem. Fac. Agric. Kagoshima Univ. *6*(2): 39-78.

Oil World, 2014, Internet access: http://www.businessweek.com/news/2014-04-29/global-soybean-production-forecast-raised-by-oil-world-on-brazil

Østergaard, L., and King, G. J. (2008). Standardized gene nomenclature for the *Brassica* genus. Plant methods. *4*(10): 1-4.

Pelletier, G., and Budar, F. (2007). The molecular biology of cytoplasmically inherited male sterility and prospects for its engineering. Curr. Opin. Biotechnol. *18*(2): 121-125.

Pelletier, G., and Budar, F. (2015). *Brassica Ogu*-INRA cytoplasmic male sterility: an example of successful plant somatic fusion for hybrid seed production. In: Somatic Genome Manipulation. Springer New York. p. 199-216.

Pelletier, G., Primard, C., Vedel, F., Chetrit, P., Remy, R., and Renard, M. (1983). Intergeneric cytoplasmic hybridization in *Cruciferae* by protoplast fusion. Mol. Gen. Genet. *191*(2): 244-250.

Peng, P. F., Li, Y. C., Mei, D. S., Colasanti, J., Fu, L., Liu, J., and Hu, Q. (2015). Expression divergence of FRUITFULL homeologs enhanced pod shatter resistance in *Brassica napus*. Genet. Mol. Res. *14*(1): 871.

Primard-Brisset, C., Poupard, J. P., Horvais, R., Eber, F., Pelletier, G., Renard, M., and Delourme, R. (2005). A new recombined double low restorer line for the *Ogu*-INRA cms in rapeseed (*Brassica napus* L.). Theor. Appl. Genet. *111*(4): 736-746.

Quiniou, N., Quinsac, A., Crépon, K., Evrard, J., Peyronnet, C., Bourdillon, A., and Etienne, M. (2012). Effects of feeding 10% rapeseed meal (*Brassica napus*) during gestation and lactation over three reproductive cycles on the performance of hyperprolific sows and their litters. Can. J. Anim. Sci. *92*(4): 513-524.

Rabbani, M. A., Iwabuchi, A., Murakami, Y., Suzuki, T., and Takayanagi, K. (1998). Phenotypic variation and the relationships among mustard (*Brassica juncea* L.) germplasm from Pakistan. Euphytica. *101*(3): 357-366.

Radoev, M., Becker, H. C., and Ecke, W. (2008). Genetic analysis of heterosis for yield and yield components in rapeseed (*Brassica napus* L.) by quantitative trait locus mapping. Genetics. *179*(3): 1547-1558.

Rahman, H. (2013). Review: Breeding spring canola (*Brassica napus* L.) by the use of exotic germplasm. Can. J. Plant Sci. *93*(3): 363-373.

Rahman, H., Bennett, R. A., and Séguin-Swartz, G. (2015). Broadening genetic diversity in *Brassica napus* canola: Development of canola-quality spring *B. napus* from *B. napus*× *B. oleracea* var. *alboglabra* interspecific crosses. Can. J. Plant Sci. *95*(1): 29-41.

Rahman, H., Peng, G., Yu, F., Falk, KC., Kulkarni, M., and Selvaraj, G., (2014) Genetics and breeding for clubroot resistance in Canadian spring canola (*Brassica napus* L.), Can. J. Plant Path. 36:122-134, DOI: 10.1080/07060661.2013.862571.

Rahman, M., Sun, Z., McVetty, P. B., and Li, G. (2008). High throughput genome-specific and gene-specific molecular markers for erucic acid genes in *Brassica napus* (L.) for marker-assisted selection in plant breeding. Theor. Appl. Genet. *117*(6): 895-904.

Raman, R., Taylor, B., Marcroft, S., Stiller, J., Eckermann, P., Coombes, N., and Raman, H. (2012). Molecular mapping of qualitative and quantitative loci for resistance to *Leptosphaeria maculans* causing blackleg disease in canola (*Brassica napus* L.). Theor. Appl. Genet. *125*(2): 405-418.

Rask, L., Andréasson, E., Ekbom, B., Eriksson, S., Pontoppidan, B., and Meijer, J. (2000). Myrosinase: gene family evolution and herbivore defense in *Brassicaceae*. In Plant Molecular Evolution. Springer Netherlands. p. 93-113.

Reynolds J., Weir B. S., and Cockerham C.C. (1983). Estimation of the coancestry coefficient: Basis for a short-term genetic distance. Genetics. *105*: 767–779.

Rhoades, M. M. (1933). The cytoplasmic inheritance of male sterility in *Zea mays*. J. Genet. *27*:71-93

Riaz, A., Li, G., Quresh, Z., Swati, M. S., and Quiros, C. F. (2001). Genetic diversity of oilseed *Brassica napus* inbred lines based on sequence-related amplified polymorphism and its relation to hybrid performance. Plant breeding. *120*(5): 411-415.

Ruiz, J. J., García-Martínez, S., Picó, B., Gao, M., and Quiros, C. F. (2005). Genetic variability and relationship of closely related Spanish traditional cultivars of tomato as detected by SRAP and SSR markers. J. Am. Soc. Hort. Sci. *130*(1): 88-94.

Sabaghnia, N., Dehghani, H., Alizadeh, B., and Mohghaddam, M. (2010). Heterosis and combining ability analysis for oil yield and its components in rapeseed. Aust. J. Crop Sci. *4*(6): 390-397.

Saitou, N., and Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Boil. Evol. *4*(4): 406-425.

Sasongko, N. D., and Möllers, C. (2005). Toward increasing erucic acid content in oilseed rape (*Brassica napus* L.) through the combination with genes for high oleic acid. J. Am. Oil Chem. Soc. *82*(6): 445-449.

Scarth, R., and McVetty, P. B. (1999). Designer oil canola–a review of new food-grade *Brassica* oils with focus on high oleic, low linolenic types. In: 10th International Rapeseed Congress (p. 26-29).

Schnable, P. S., and Wise, R. P. (1998). The molecular basis of cytoplasmic male sterility and fertility restoration. Trends Plant Sci. *3*(5): 175-180.

Sernyk, J. L. (1983). Heterosis and cytoplasmic male sterility in summer rape (*Brassica napus* L.). Dissertation Abstracts International, B: *43*(8).

Sernyk, J. L., and Stefansson, B. R. (1983). Heterosis in summer rape (*Brassica napus* L.). Can. J. Plant Sci. *63*(2): 407-413.

Seyis, F., Snowdon, R. J., Luhs, W., and Friedt, W. (2003). Molecular characterization of novel resynthesized rapeseed (*Brassica napus*) lines and analysis of their genetic diversity in comparison with spring rapeseed cultivars. Plant Breeding. *122*(6): 473-478.

Shahidi, F. (1990). Canola and rapeseed: production, chemistry, nutrition, and processing technology. Springer. p.1-24.

Sherazi, S. T. H., Arain, S., Mahesar, S. A., Bhanger, M. I., and Khaskheli, A. R. (2013). Erucic acid evaluation in rapeseed and canola oil by Fourier transform-infrared spectroscopy. Eur. J. Lipid Sci. Technol. *115*(5): 535-540.

Shull, G. H., 1908 The composition of a field of maize. Am. Breeders Assoc. Rep. *4*: 296–301.

Singh, M., and Brown, G. G. (1991). Suppression of cytoplasmic male sterility by nuclear genes alters expression of a novel mitochondrial gene region. The Plant Cell Online. *3*(12): 1349-1362.

Sleper, D. A., and Poehlman, J. M. (2006). Methods in plant breeding. In: Breeding field crops. Blackwell Publishing, p.155-182.

Smith, J.S.C. and Smith O.S. (1989). The description and assessment of distances between inbred lines of maize: The utility of morphological, biochemical, and genetic descriptors and a scheme for the testing of distinctiveness between inbred lines. Maydica. 34: 151-161.

Snowdon, R., Lühs, W., and Friedt, W. (2007). Oilseed rape. In: Oilseeds. Springer Berlin Heidelberg. p. 55-114.

Snowdon, R. J., Friedrich, T., Friedt, W., and Köhler, W. (2002). Identifying the chromosomes of the A-and C-genome diploid Brassica species *B. rapa* (syn. *campestris*) and *B. oleracea* in their amphidiploid *B. napus*. Theor. Appl. Genet. *104*(4): 533-538.

Sokal, R. and Michener, C. (1958). A statistical method for evaluating systematic relationships. University of Kansas Science Bulletin. 38: 1409–1438.

Sonah, H., Bastien, M., Iquira, E., Tardivel, A., Légaré, G., Boyle, B., and Belzile, F. (2013). An improved genotyping by sequencing (GBS) approach offering increased versatility and efficiency of SNP discovery and genotyping. PloS one. *8*(1): e54603.

Song, L. Q., Fu, T. D., Tu, J. X., Ma, C. Z., and Yang, G. S. (2006). Molecular validation of multiple allele inheritance for dominant genic male sterility gene in *Brassica napus* L. Theor. Appl. Genet. *113*(1): 55-62.

Spindel, J., Wright, M., Chen, C., Cobb, J., Gage, J., Harrington, S., and McCouch, S. (2013). Bridging the genotyping gap: using genotyping by sequencing (GBS) to add high-density SNP markers and new value to traditional bi-parental mapping and breeding populations. Theor. Appl. Genet. *126*(11): 2699-2716.

Stefansson, B. R. (1983). The development of improved rapeseed cultivars. In: High and Low Erucic Acid Rapeseed Oils. p.143-159.

Stefansson, B. R., Hougen, F. W., and Downey, R. K. (1961). Note on the isolation of rape plants with seed oil free from erucic acid. Can. J. Plant Sci. *41*(1): 218-219.

Stefansson, B. R., and Hougen, F. W. (1964). Selection of rape plants (*Brassica napus*) with seed oil practically free from erucic acid. Can. J. Plant Sci. *44*(4): 359-364.

Stefansson, B. R., and Kondra, Z. P. (1975). Tower summer rape. Can. J. Plant Sci. *55*(1): 343-344.

Sultan, S. E. (2000). Phenotypic plasticity for plant development, function and life history. Trends Plant Sci. *5*(12): 537-542.

Sun, Z., Wang, Z., Tu, J., Zhang, J., Yu, F., McVetty, P. B., and Li, G. (2007). An ultradense genetic recombination map for *Brassica napus*, consisting of 13551 SRAP markers. Theor. Appl. Genet. *114*(8): 1305-1317.

Tamura, K., and Nei, M. (1993). Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol. Biol. Evol. *10*(3): 512-526.

Tamura, K., Nei, M., and Kumar, S. (2004). Prospects for inferring very large phylogenies by using the neighbor-joining method. Proc. Natl. Acad. Sci. U. S. A. *101*(30): 11030-11035.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. (2013). MEGA6: molecular evolutionary genetics analysis version 6.0. Mol. Biol. Evol. *30*(12): 2725-2729.

Tan, Z. M., Li, Y. C., Hu, Q., Mei, D. S., Li, Y. D., and Xu, Y. S. (2007). Heterosis prediction based on genetic distance estimated by molecular markers in rapeseed. Chin. J. Oil Crop Sci. *29*(2): 126-134.

Taylor, A., Coventry, E., Jones, J. E., and Clarkson, J. P. (2015). Resistance to a highly aggressive isolate of *Sclerotinia sclerotiorum* in a *Brassica napus* diversity set. Plant Pathology. *64*: 932–940. doi: 10.1111/ppa.12327.

Teklewold, A., and Becker, H. C. (2005). Heterosis and combining ability in a diallel cross of Ethiopian mustard inbred lines. Crop Sci. *45*(6): 2629-2635.

Thiemann, A., Meyer, S., and Scholten, S. (2009). Heterosis in plants: Manifestation in early seed development and prediction approaches to assist hybrid breeding. Chinese Science Bulletin. *54*(14): 2363-2375.

Town, C. D., Cheung, F., Maiti, R., Crabtree, J., Haas, B. J., Wortman, J. R., and Bancroft, I. (2006). Comparative genomics of *Brassica oleracea* and *Arabidopsis thaliana* reveal gene loss, fragmentation, and dispersal after polyploidy. The Plant Cell Online. *18*(6): 1348-1359.

Tripathi, M. K., and Mishra, A. S. (2007). Glucosinolates in animal nutrition: A review. Anim. Feed Sci. Technol. *132*(1): 1-27.

Udall, J.A., Quijada, P.A., Lambert, B., and Osborn, T. C. (2006). Quantitative trait analysis of seed yield and other complex traits in hybrid spring rapeseed (*Brassica napus* L.): Identification of alleles from unadapted germplasm. Theor. Appl. Genet. 113: 597–609.

Upadhyaya, H. D., Bramel, P. J., and Singh, S. (2001). Development of a chickpea core subset using geographic distribution and quantitative traits. Crop Sci. *41*(1): 206-210.

Uyttewaal, M., Arnal, N., Quadrado, M., Martin-Canadell, A., Vrielynck, N., Hiard, S., and Mireau, H. (2008). Characterization of Raphanus sativus pentatricopeptide repeat proteins encoded by the fertility restorer locus for Ogura cytoplasmic male sterility. The Plant Cell Online. *20*(12): 3331-3345.

Van de Wouw, A. P., Marcroft, S. J., Ware, A., Lindbeck, K., Khangura, R., and Howlett, B. J. (2014). Breakdown of resistance to the fungal disease, blackleg, is averted in commercial canola (*Brassica napus*) crops in Australia. Field Crop Res. *166*: 144-151.

Vieira, F. G., Lassalle, F., Korneliussen, T. S., and Fumagalli, M. (2015). Improving the estimation of genetic distances from Next-Generation Sequencing data. Biol. J. Lin. Soc. (pre-published online preview).

Vincourt, P. (2014). Research fields, challenges and opportunities in European oilseed crops breeding. OCL. *21*(6): D602.

Vinod, K. K. (2005). Cytoplasmic genetic male sterility in plants. A molecular perspective. In: Proceedings of the training programme on advances and accomplishments in heterosis breeding. Tamil Nadu Agricultural University, Coimbotore, India. p.147-162.

Vos, P., Hogers, R., Bleeker, M., Reijans, M., Van de Lee, T., Hornes, M., and Zabeau, M. (1995). AFLP: a new technique for DNA fingerprinting. Nucleic Acid Res. *23*(21): 4407-4414.

Walsh, B. (2001). Quantitative genetics in the age of genomics. Theor. Popul. Biol. *59*(3): 175-184.

Walter, A., Studer, B., and Kölliker, R. (2012). Advanced phenotyping offers opportunities for improved breeding of forage and turf species. Ann. Bot. mcs026.

Wang, X., Liu, G., Yang, Q., Hua, W., Liu, J., and Wang, H. (2010). Genetic analysis on oil content in rapeseed (*Brassica napus* L.). Euphytica. *173*(1): 17-24.

Wang, X., Wang, H., Long, Y., Liu, L., Zhao, Y., Tian, J., and Li, M. (2015). Dynamic and comparative QTL analysis for plant height in different developmental stages of *Brassica napus* L. Theor. Appl. Genet. *128*(6): 1175-1192.

Ward Jr, J. H. (1963). Hierarchical grouping to optimize an objective function. J. Am. Stat. Assoc. *58*(301): 236-244.

Weber, J. L., and May, P. E. (1989). Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. Am. J. Hum. Genet. *44*(3): 388.

Wei, M., Song, M., Fan, S., and Yu, S. (2013). Transcriptomic analysis of differentially expressed genes during anther development in genetic male sterile and wild type cotton by digital gene-expression profiling. BMC genomics. *14*(1): 97.

Wen, Y. C., Wang, H. Z., Shen, J. X., Liu, G. H., and Zhang, S. F. (2006). Analysis of genetic diversity and genetic basis of Chinese rapeseed cultivars (Brassica napus L.) by sequence-related amplified polymorphism markers. Sci. Agric. Sinica. *39*(2): 246-256.

Wittkop, B., RJ Snowdon, and W. Friedt. (2009). Status and perspectives of breeding for enhanced yield and quality of oilseed crops for Europe. Euphytica. *170*(1-2): 131-40.

Wu, J. G., Shi, C. H., and Zhang, H. Z. (2006). Partitioning genetic effects due to embryo, cytoplasm and maternal parent for oil content in oilseed rape (*Brassica napus* L.). Genet. Mol. Biol. *29*(3): 533-538.

Xing, N., Fan, C., and Zhou, Y. (2014). Parental Selection of Hybrid Breeding Based On Maternal and Paternal Inheritance of Traits in Rapeseed (*Brassica napus* L.). PLoS ONE *9*(7): e103165. doi:10.1371/journal.pone.0103165.

Yamagishi, H. and Bhat, S.R., 2014. Cytoplasmic male sterility in Brassicaceae crops. Breeding Sci. *64*(1): 38-47.

Yang, J., and Zhang, M. (2013). Molecular Basis of Cytoplasmic Male Sterility. In: Biotechnology of Crucifers. Springer New York. p. 173-185.

Yang, P., Shu, C., Chen, L., Xu, J., Wu, J., and Liu, K. (2012). Identification of a major QTL for silique length and seed weight in oilseed rape (*Brassica napus* L.). Theor. Appl. Genet. *125*(2): 285-296.

Yu, C. Y., Hu, S. W., Zhao, H. X., Guo, A. G., and Sun, G. L. (2005). Genetic distances revealed by morphological characters, isozymes, proteins and RAPD markers and their relationships with hybrid performance in oilseed rape (*Brassica napus* L.). Theor. Appl. Genet. *110*(3): 511-518.

Zanetti, F., Vamerali, T., and Mosca, G. (2009). Yield and oil variability in modern varieties of high-erucic winter oilseed rape (*Brassica napus* L. var. *oleifera*) and Ethiopian mustard (*Brassica carinata* A. Braun) under reduced agricultural inputs. Ind. Crop Prod. *30*(2): 265-270.

Zhang, Y., De, S., Garner, J. R., Smith, K., Wang, S. A., and Becker, K. G. (2010). Systematic analysis, comparison, and integration of disease based human genetic association data and mouse genetic phenotypic information. BMC Med. Genomics. *3*(1): 1.

Zhao, J., Becker, H. C., Zhang, D., Zhang, Y., and Ecke, W. (2006). Conditional QTL mapping of oil content in rapeseed with respect to protein content and traits related to plant development and grain yield. Theor. Appl. Genet. *113*(1): 33-38.

Zharkikh, A., and Li, W. H. (1992). Statistical properties of bootstrap estimation of phylogenetic variability from nucleotide sequences. I. Four taxa with a molecular clock. Mol. Biol. Evol. *9*(6): 1119-1147.