Investigating the Link Between Genetic Distance and Seed Yield in Hybrid *Brassica napus* L. using Phenotypic and Genotypic Methods

by

Alexander Peter Cattini

A Thesis submitted to the Faculty of Graduate Studies of
The University of Manitoba
in partial fulfilment of the requirements of the degree of

MASTER OF SCIENCE

Department of Plant Science
University of Manitoba
Winnipeg

Acknowledgements

There is not enough room in this entire thesis to properly acknowledge and express my gratitude to the plethora of people who helped me achieve my goals and get to where I am, both academically and as a man.

To my family, I want to thank you for all the support and encouragement over my University career, helping me realize my goals, as well as providing me with the time and ability to even realize what my goals were.

To my supervisor, Dr. Robert Duncan, I would like to thank for the opportunity to advance my career within his program and benefit immeasurably from his guidance and experience, both in the field and in writing my thesis.

My project could not have been conducted without the faculty and support staff within the Department of Plant Science. Whether in the lab, the field, the growth rooms, the greenhouses, the quality lab or in the classroom you were all there to help make my project come to life.

To my committee members, Dr. Doug Cattani, Dr. Genyi Li and Dr. Mario Tenuta, I would like to express my thanks for their time and expertise in making my project what it is today.

I would like to thank all my fellow graduate students within the University of Manitoba's Department of Plant Science for their friendship and support throughout my M.Sc. A special thanks to the other members of Team Sea Onion for our impressive Escape Room record, I still think we were robbed during the "Nuclear Meltdown".

A special thanks to DL Seeds, Bunge Canada, NSERC Canada and the University of Manitoba for providing funding and support for my project, allowing me to conduct my research.

List of Tables

List of Figures

Abstract

Brassica napus L. is an economically important oilseed species widely cultivated across Western Canada. Hybrid *B. napus* cultivars compose the majority of the market due to their ability to improve seed yield and agronomic quality. When selecting hybrid parents in *B. napus*, it is important to attempt to predict high-yielding parental combinations in order to conserve resources during experimental hybrid evaluation. Genetic distance between parents has been implicated as a contributing factor to producing high-yielding hybrids and is often used as one criteria for determining parental combinations. Genetic distance can be calculated using phenotypic or genotypic data. In the current study, the genetic distance between high erucic acid rapeseed (HEAR) genotypes of *B. napus* was established using both phenotypic and genotypic data. Phenotypic criteria took the form of nine agronomic and seed quality traits gathered from 318 distinct *B. napus* genotypes over the 2013 and 2014 field seasons in Southern Manitoba. Genotypic criteria took the form of either 291,782 SNP markers identified in 231 distinct *B. napus* genotypes using genotyping-by-sequencing (GBS) or 230 polymorphic sequence-related amplified polymoprhism (SRAP) markers identified in 160 *B. napus* genotypes. The genetic distance between available pollinators and a single male-sterile female was established using each set of criteria in an attempt to correlate genetic distance between any given pollinator and the female with hybrid yield. Regression analysis was conducted with yield data from hybrid genotypes gathered from 37 field sites from 2011-2014. Using the phenotypic-derived genetic distance, a significant correlation between genetic distance and hybrid yield was uncovered explaining either 22 % or 42 % of the variation in hybrid yield depending upon whether hybrids were grown at three or more, or five or more sites in the analysis, respectively. No significant link was found between GBS or SRAP-derived genetic distance and hybrid yield. These results provide evidence that

genetic distance does have effect on hybrid yield, and specifically that phenotypic criteria can be used to establish genetic distance with utility to aid in the selection of high-yielding hybrid genotypes.

Table of Contents

1.0 General Introduction

The economically important oilseed crop *Brassica napus* L. (canola/rapeseed) is seeded to approximately 20 million acres across Canada, and contributes $19.3 billion annually to the Canadian economy (Canola Council of Canada, 2013). This production is mostly localized in the prairie provinces of Saskatchewan (10.6 million acres in 2014), Manitoba (3.0 million acres in 2014) and Alberta (6.7 million acres in 2014). The University of Manitoba has had a *B. napus* breeding program in place for over 40 years; releasing several commercial cultivars including: Hero (Scarth et al., 1991), MilleniUM 03 (McVetty et al., 2000), Red River 1826 (McVetty et al. 2006a), Red River 1852 (McVetty et al., 2006b), Red River 1997 (McVetty et al., 2010) and Red River 1861 (McVetty et al., 2011). To keep up with industry developments, and to meet increased demand for high performing *B. napus* cultivars, the University of Manitoba released their first Roundup Ready® high erucic acid hybrid rapeseed cultivar, HYHEAR 1, in 2014; yielding ~33 % more seed and 23 g kg^{-1} more seed oil than MilleniUM 03 (McVetty et al., 2014). The biological phenomenon of heterosis is one factor responsible for the high yield observed in HYHEAR 1 and other hybrid cultivars.

Heterotic gain is characterized by the superior performance of hybrid offspring over their parents as it relates to agronomic traits of interest; a process which is exploited for great effect in many important crop species (Ryder et al., 2014). In *B. napus*, high-parent heterosis (superior performance of hybrid offspring over parental material) from 20 % (McVetty et al., 1995), to as high as 43 % was reported by Cuthbert et al. (2009) at the University of Manitoba. Despite high parent heterosis values ranging from 20 – 50 %, high-parent heterosis values average around 30 % (McVetty and Scarth, 2012).These successes have attracted the interest of *B. napus* breeders

hoping to replicate this progress in their own breeding programs and by 2010, hybrid cultivars comprised the greatest acreages in nearly every major canola/rapeseed producing area worldwide (Basunanda et al., 2010).

To take proper advantage of heterosis in hybrid production, the selection of appropriate parents is essential. Falconer and McKay (1996) hypothesized that increased genetic distance between parental genotypes correlates with increased heterotic gain in the resulting hybrid. To this effect, Riaz et al. (2001) reported a correlation coefficient between genetic distance and hybrid yield of 0.64. This correlation is important in the development high performing hybrids, prompting the organization of available parents into heterotic groups from which diverse parents can be selected (Laude and Carena, 2015). Heterotic groups result from the clustering of parental genotypes into groups which display a similar heterotic response when crossed with genotypes from other groups (Ryder et al., 2014). Used to great effect in maize, where the stiff stalk (SS) vs. non-stiff stalk morphological groups resulted in increased hybrid agronomic performance when one parent from each group were crossed (Tracy and Chandler, 2006). However, the criteria that establishes a given population's diversity is not limited to morphological characteristics. Pedigree data (Badu-Apraku et al., 2006) and molecular marker data (Melchinger et al., 1991; Mohammadi and Prasanna 2003) have also been effective in heterotic group development. While phenotypic data is seen as the backbone of any germplasm evaluation (Violle et al 2014), phenotyping large germplasm collections can be logistically prohibitive, and subject to relatively large amounts of ambiguity (Langer et al., 2014). As such, for use in diversity analysis and heterotic group assignment, molecular characterization has become an important step in many crop breeding programs (Muranty et al., 2014).

Beyond the type of data set (phenotypic, pedigree, molecular etc.) that is chosen, there are several statistical methods available to hierarchically cluster data. The most commonly used methods are the unweighted pair group (UPGMA) and Ward's minimum variance method (Khodadadi et al., 2005; Teklewold, 2006; Fahid, 2014; Singh, 2015). Ward's method is increasingly cited as ideal, due to it not being subject to the "chaining effects" observed in UPGMA analysis; which results in higher internal affinity within clusters than is actually present (Lombard et al, 2000; Mohammadi and Prasanna, 2003; Khodadadi et al., 2005).

The overall goal of this research was to evaluate and organize *B. napus* genotypes from the University of Manitoba's germplasm collection into potential heterotic groups using morphological, agronomic, seed quality and molecular information. We attempted to fulfill this research goal in the following three projects.

1. Use two multivariate statistical methods (Ward's method and UPGMA) to separate 318 inbred *B. napus* genotypes, one male-sterile female genotypes and the two parents of the male-sterile female genotypes into heterotic groups based on nine agronomic and seed quality traits collected over two site/years.
2. Organize pollinators and the male-sterile genotype of *B. napus* into groups using molecular markers. To this end, 230 genotypes were subjected to genotyping-by-sequencing (GBS) analysis. Similarly, 160 *B. napus* genotypes were subjected to sequence related amplified polymorphism (SRAP) analysis at the University of Manitoba. This molecular data was used to hierarchically cluster genotypes into groups that were visualized using dendrograms.
3. To correlate genetic distance (calculated using agronomic and seed quality data, GBS derived SNPs, or SRAP markers) to hybrid yield performance.

2.0 Literature Review

2.1 History

2.1.1 Early Brassica Cultivation

Within the Eudicot family, Brassicaceae are an important group of cruciferous plants, the mustard family, also known as the Brassicas (*Brassica* spp.) (Gupta and Pratap, 2007; Franzke et al., 2011). Cultivation of *Brassica* species dates back to the earliest agricultural forays by humans, with evidence that early Neolithic groups domesticated members of the genus for both nutritional and medicinal purposes (Eskin, 2012). Domestication of a wild plant species broadly occurs in three stages: 1) the gathering of wild individuals based on a desirable trait, 2) the establishment of cultivated populations from the gathered individuals, increasing the proportion of favourable alleles in the population, 3) the active selection and breeding of desired individuals thereby refining suitability for agriculture (Meyer and Purugganan, 2013**)**. It was through this refined selection that the incredible amount of morphological diversity in cultivated *Brassica* species is observed today, for instance in the different morpho-types of *Brassica oleracea* L. (broccoli, brussel-sprouts, cauliflower, kale, cabbage) (Salmon et al., 2008).

Archaeological finds and ancient writings place *Brassica* species alongside humans throughout the ancient world. Therefore it is no surprise that the genus is thought to have originated in what Nikolai Vavilov referred to as a "center or origin" (1935**)**, in this case being the Fertile Crescent of the Middle East between 11,000, and 8300 B.C.E (Rai et al., 2007**)**.

Evidence in Asia, primarily India and China, where Sanskrit writings dating to ~2000 BC describe the use of *Brassica* species as a condiment; while in ~500 BC descriptions of cultivated Brassica appear in the writings of the Greek philosopher Pythagorus **(**Rai et al., 2007**)**. Writings by the

Roman historian Columella (35-65 A.D.) describe the roots of both *Brassica rapa* L. and *Brassica napus* L. and what it means for their cultivation (Mikíc, 2016).

2.1.2 Rapeseed (*Brassica napus*)

One of the most economically important members of the genus *Brassica* is the amphidiploid and allopolyploid *B. napus* (AACC, n = 19), more commonly referred to as rapeseed, or oilseed rape **(**McVetty and Zelmer, 2009**)**. *Brassica napus* is thought to have originated as the result of an interspecific cross between *B. rapa* L. (A Genome n = 10) and *B. oleracea* (C Genome, n = 9). The familial relationship between these members of the genus *Brassica*, as well as the interactions between other related species can be visualized in Figure 2.1 as first published by U in 1935.

Rapeseed, has been cultivated in Europe for its oil since the 13th century (Ramchiary et al., 2011), and in his 1570 AD works, the German priest Konrad Heresbach described rapeseed oil in Germany as an alternative lamp and cooking oil for the lower class (Eskin, 2012). Similarly, Appelqvist described the cultivation of rapeseed in Holland for use as a lamp oil in the 17th century (Appelqvist, 1972). The extensive use of Brassica oil as lamp oil continued until the end of the 19th century, when mineral and gas lighting as well as imports from colonial states became the norm for Western European nations (Baranyk and Fabry, 1999).

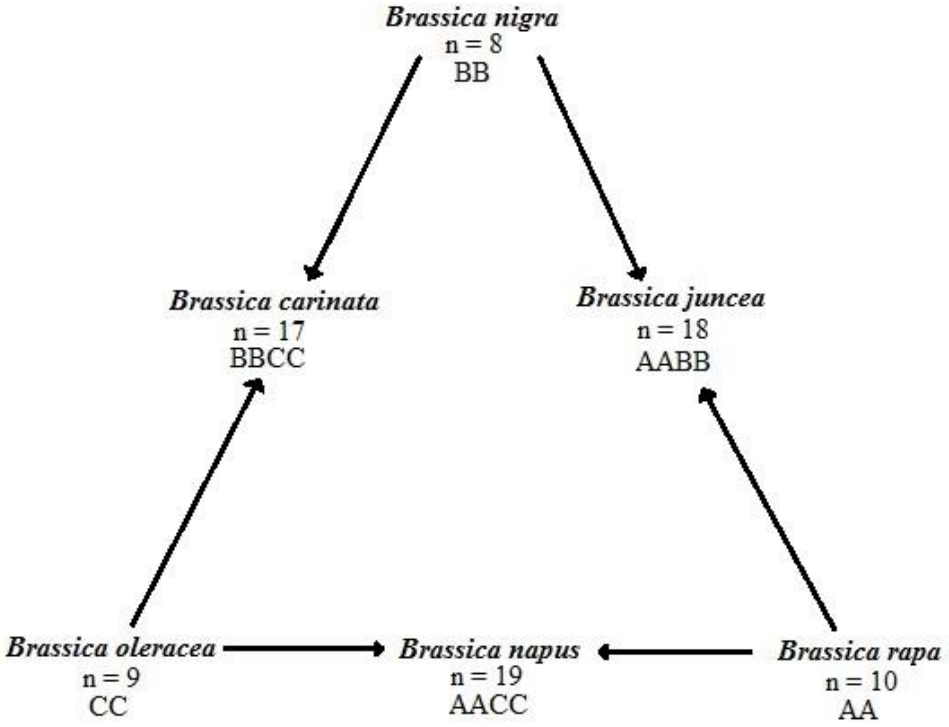

Figure 2.1. The "Triangle of U" first described in 1935 portraying theoretical relationships between important *Brassica* oilseed crops (U, 1935).

The jump of Brassica oil to North America occurred when a Canadian farmer named Fred Solvuniuk, received an envelope of seed from a friend in his native Poland, to be planted at his Shellbrook, Saskatchewan farm in 1936 (Gupta and Pratap, 2007). This seed turned out to be *B. campestris* and constitutes the first record of cultivation of Rapeseed in North America (Gupta and Pratap, 2007).

As a result of shipping embargoes between North America and Europe and Asia during the Second World War, Canadian agriculture faced two pressing problems: 1) the Canadian economy had taken a major hit, and economists and plant breeders quickly realized the danger of Canada's dependence on the production and export of wheat (Eskin, 2012); and 2) the Canadian Military needed a new and renewable source of lubricating engine oil for their burgeoning wartime navy, a

job which had previously been filled by imported petroleum based lubricants (Kneen, 1992). The seed oil from *B. napus* was billed as a renewable substitute for petroleum based lubricants, as it had the ability to maintain its lubricating ability and cling to metal under intense heat and pressure (Boulter, 1983). Beginning in 1942, T.A. Stevenson sent all available *B. napus* seed to Federal Experimental Stations across Canada, rapidly increasing national seed volume to meet Wartime demand, while at the same time 19 tonnes of *B. napus* seed was brought up from the United States for distribution to Western Canadian producers (McInnis, 2004). In 1943, the majority of rapeseed was crushed and processed at the W.R. Carpenter Plant in Hamilton Ontario, and in 1945 the first rapeseed crushing facility in the West was built by J. Gordon Ross at Moose Jaw, Saskatchewan (Kneen, 1992). Canadian rapeseed production attained a periodic high of 80,000 acres in 1948 (Bell, 1982), though as wartime policy ceased, rapeseed production was all but eliminated, and by 1950, barely 400 acres of rapeseed were being grown in Canada (Eskin, 2012). It is thought that rapeseed production could have disappeared from North America at this point were in not for the efforts of producers in western Canada, such as J. Ross, who tirelessly searched for new markets (Boulter, 1983).

2.1.3 A New Outlook on Rapeseed Oil

By the latter half of the 20th century in North America, rapeseed oil was quite limited in its utility as edible oil because of the poor taste, and health risks, associated with two major constituents of rapeseed oil; erucic acid, and glucosinolates (Khajali and Slominski, 2012). Removing these components was difficult due to an inaccurate means of evaluating seed oil content, paired with a general lack of interest in a new oil. Work done by A.T. James, J.P. Martin and Howard Smith changed that, when in 1952 they developed the gas-liquid partition chromatography (GLC) system which enabled for the separation and measurements of components in a mixture (James and

Martin, 1952). This was important to plant breeding because previously, the fatty acid analysis of oilseeds required a large volume of seed and multiple days of man-hours. Subsequently, a Swedish biochemist named Sixten Troëng (1955), used the new GLC technology to develop a method for a more rapid oilseed analysis. Keith Downey at the University of Saskatchewan and Baldur Stefansson at the University of Manitoba quickly recognized the possible utility of quick and efficient seed oil analysis for the improvement of rapeseed oil (Downey and Harvey, 1963; Stefansson and Hougan, 1963). Modifying the technique pioneered by Troëng, Keith Downey pioneered the "half-seed" method to evaluate rapeseed germplasm and characterize its erucic acid content; it was called the "half-seed" method because during development of the method half the seed was tested for oil content and later matched against the plant grown from the other half of the plant (Downey and Harvey, 1962). The half-seed method allowed Downey and his colleague B.L. Harvey to, in 1963, make the important discovery that the erucic acid content of a seed was controlled not by the maternal genotype, as had previously been thought, but by the genotype of the developing embryo within the seed (Downey and Harvey, 1963). At the same time, Baldur Stefansson in Manitoba was evaluating the previously described European forage rapeseed Liho (Vehov and Lebedeva, 1978), recording a low erucic acid content of 10 % (Stefansson and Hougen, 1964). Liho was a major contributor to the evolution of edible Brassica oil as after transferring Liho seed to Keith Downey, the first LEAR variety (Low Erucic Acid Rapeseed), Oro, was bred using Liho as the parental material (Downey et al., 1975). While Oro was a breakthrough in the history of rapeseed, the LEAR variety Oro was never a commercial success due to its relatively poor agronomic performance and its lackluster cost-benefit ratio compared to the far cheaper sunflower oil from the USSR (Eskin, 2012). The ability to properly measure and lower the eruric acid content of Brassica oil became essential due to legislation such as 76/621/EEC in 1976 by the

European Union requiring oils, fats, and blends for human consumption to have erucic acid levels no greater than 10 % by 1977 and no greater than 5 % by 1979 (Council of the European Union, 1976). Today, in the United States, Brassica oil for human consumption must have no greater than 2 % erucic acid content by weight (Food and Drug Administration, 2015).

2.1.4 Double-Low Rapeseed

In 1967, a Polish researcher named Dr. Jan Krzymanski characterized a variety of *B. napus* which was near free of glucosinolates (Baranyk and Fabry, 1999). This variety, called "Bronowski", was sent to both Baldur Stefansson and Keith Downey, and by 1973 both had succeeded in transferring the low glucosinolate trait into their elite LEAR lines (Downey et al., 1975; Kondra and Stefansson, 1970). From there, Dr. Baldur Stefansson was able to register "Tower", the first "double-low", low in both glucosinolates and erucic acid, variety of rapeseed (Stefansson, 1972). In order to distinguish the new double low rapeseed, the world "canola" (CANadian Oil Low Acid) was coined to describe any rapeseed where the concentration of erucic acid was below 2 % and glucosinolates were below 30 micromoles per gram of seed (Canola Council of Canada, 2014). In 1985, the FDA gave GRAS (generally regarded as safe) status to canola oil, paving the way for it to become one of the most important emerging oilseeds.

2.2 Canadian Rapeseed Production

Brassica napus contributes $19.3 billion annually to the Canadian economy (based on averages from 2009-2013); including 249,000 jobs and $12.5 billion in wages, making it one of the most important agricultural components of the Canadian economy (Canola council of Canada, 2013). Ten industry sectors are cited as "directly" benefitting from the Canadian canola/rapeseed industry, including: seed supply, seed development, farming, seed handling, elevation, ports and transportation, crushing, refining, livestock production and food processing (Canola council of

Canada, 2013). In 2014, nearly 20 million acres of rapeseed were planted to the Canadian Prairies (down from 22 million in 2012); the majority of which is localized in the prairie provinces of Saskatchewan, Manitoba, and Alberta. This is spectacular considering that in 1950, only 160 acres of rapeseed were planted in Canada (Ohlson, 1972), growing to 813,000 acres by 1969 and 6.4 million acres in 1986 (Canola council of Canada, 2013). This rapid growth has made Canada a world leader in oilseed production; of the 71,254 million metric tons of oilseed produced annually worldwide (averaged 2013-2015), Canada produced 16,761 million metric tons, equivalent to China, and only behind the combined contribution of European Union member states in raw oilseed production (Carré and Pouzét, 2014).

2.3 Economically Important Quality Characters of Rapeseed Oil

2.3.1 Erucic Acid

Erucic acid (C22:1 ω-9), or cis-13 docosenoic acid, is a long chain, unbranched, monounsaturated fatty acid, composed of 22 carbon molecules with a single double bond in the omega-9 position (Beare-Rogers and Nera, 1972) (Figure 2.2). Seed oil containing erucic acid is universal to members of the Brassicaceae, but rapeseed oil from *B. napus* contains particularly high levels of the molecule (Mcvetty and Scarth, 2012).

In the 1960s through 1980s, experiments were performed linking myocardial lipidosis with feeding animals (rats, pigs, chicks, monkeys, rabbits) a diet rich in high erucic acid oil (Beare-Rogers and Nera, 1972; de Wildt and Speijers, 1984; Hulsmannet al., 1979). A link between erucic acid metabolism and increased lipase activity around the heart and liver, results in increased lipid uptake by the heart (Hulsmann et al., 1979). This rapid lipid uptake causes pathological changes to heart muscle, which peaks 3-7 days after diet initiation; and can eventually result in cardiac

fibrosis, histocyte infiltration and necrosis. However, it is worth noting, that these pathological changes do not necessarily result in statistically significant alterations in myocardial contractility (de Wildt and Speijers, 1984). While these effects have never been experimentally confirmed in humans, recent mechanistic studies suggest that erucic acid may cause myocardial toxicity as a result of increased metabolism within myocardial peroxisomes and subsequent free radical formation (Imamura et al., 2013).

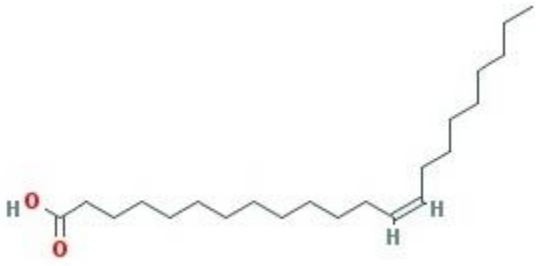

Figure 2.2. Two-dimensional structure of C22:1 ω-9 Erucic Acid (cis-13 docosenoic acid) showing position of double bond and oxygen molecules. Taken from the US National Library of Medicine (PubChem, 2004a).

It is in the industrial sector that erucic acid is most important. When mixed with ammonia, the biproduct erucamide is formed (Coonrod et al., 2008). Erucamide, and other erucic acid derivatives, are important feedstocks in the production of plastics, nylons 13-13, cosmetics, and most notably high temperature lubricants (Li et al., 2012). A potent surface lubricant, erucamide allows plastics, such as those found in sandwich or garbage bags, to not adhere to themselves or other bags (Coonrod et al., 2008). Beyond that, the increased biodegradability and sustainability of erucic acid oil make it an attractive alternative to petroleum based lubricants (Coles et al., 2008).

High Erucic Acid Rapeseed (HEAR) cultivars are varieties of *B. napus* unique in the high erucic acid concentration within their seed oil (McVetty et al, 2009). Where an unimproved rapeseed

cultivar will contain erucic acid concentrations of 30-45 % within their seed oil; a HEAR cultivar will boast seed oil erucic acid concentrations between 50 and 55 % (McVetty et al., 2009; Cuthbert et al., 2011). Erucic acid content within *B. napus* seed oil is largely controlled through the action of two additive loci, E^A and E^C, located on the A and C chromosomes, which together control 90 % of total existing variation in erucic acid content (Nesi et al., 2008). Modern HEAR breeding operations work to utilize and improve upon these loci, as well as identify or create new regulators on erucic acid production, to increase and optimize HEAR oil (Rahman et al., 2008).

2.3.2 Oleic Acid

Oleic acid (C18:1 ω-9), or 9-octadecanoic acid, is a polyunsaturated fatty acid belonging to the omega-9 family with 18 carbon molecules and a double bond at the omega-9 position (Snyder and Weselake, 2009). One of the major components of *B. napus* oil, standard canola quality oil is ~60 % oleic acid, alongside 20 % linoleic acid (C18:2), and 10 % linolenic acid (C18:3).

Canola oil high in oleic acid is most important to the cooking industry. Four major requirements must be fulfilled for a cooking oil to be desirable: (1) oil must have a clean and bland taste, (2) must have high oxidative stability, (3) must have a high smoke point, (4) must be low in saturated and trans fats (Syed, 2012). High oleic acid rapeseed oil fulfils all these requirements, with a bland taste, a 17H oxidative stability index, >450 ˚F smoke point, and 7 % saturated fatty acid composition (Syed, 2012).

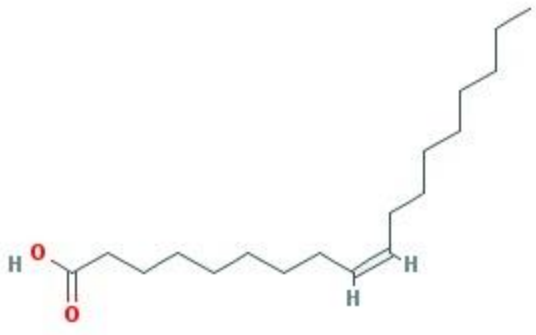

Figure 2.3. Two-dimensional structure of C18:1 ω-9 Oleic Acid (9-octadecanoic acid) showing position of double bond and oxygen molecules. Taken from the US National Library of Medicine (PubChem, 2004b).

Hu et al. (2006) discovered that a single nucleotide substitution in the *fad2* gene would result in a truncated, non-functional protein, and a sharp decrease in downstream products. However, this substitution results in levels of oleic acid > 85 % in varieties with the nucleotide substitution, and allows germplasm to be screened for high oleic potential (Snyder and Weselake, 2009).

Three varieties of specialty canola are available from Cargill under the trademark Victory® moniker which produce Clear Valley® oils and shortenings (Iassonova and Liu. 2014). Three Clear Valley® high oleic oils are available from Cargill, including Clear Valley® 65 (CV65), omega-9 Oilseed Rape, and Clear Valley® 75 (CV75). CV65 boasts a minimum oleic-acid content of 65 %, while CV80 has a minimum 80 % oleic acid (Debonte et al., 2012). In 2015, Cargill put the finishing touches on a 45,700 square foot research facility in Fort Collins, Colorado at which the next generation of Cargill's high-oleic canola lines are being developed using Victory® hybrid canola as a template.

In 2004, Dow Agroscience initiated a high oleic breeding program within their canola division to combine yield-competitive canola varieties with high oleic potential along with Roundup Ready® and Clearfield® herbicide tolerance (Shah et al., 1990; Schnable et al., 2013). This move was

implemented when legislation in the United States required the labelling of trans-fat containing products, and has been a boom to the canola industry. The Canola Council of Canada has set a goal of 33 % of total *B. napus* production (~22 million acres), or 7.25 million acres, to be of high-oleic or other specialty oil by 2025; which would be over double the ~3 million acres (~15 %) seeded in 2015 (Pratt, 2015)

2.3.3 Saturated Fatty Acids

Brassica napus seed oil contains various saturated fatty acids of agronomic importance, including stearic, lauric, palmitic and myristic acid (Hamam and Shahidi, 2005). Saturated fatty acids have generally been avoided in the food industry as they were thought to negatively affect plasma cholesterol levels in humans when ingested (Nesi et al., 2008). This way of thinking is slowly being altered as new studies are showing that short and medium chain fatty acids have no effect on plasma cholesterol levels (Legrand et al., 2010). Several new and unique Brassica oilseed profiles have made saturated fatty acids a new target in breeding programs.

Stearic acid, or octadecanoic acid (C18:0) is a saturated fatty acid found at 1-4 % in Brassica oil (Zarhloul et al., 2006). Oils rich in saturated fatty acids tend to be very oxidation resistant, which is desirable in cooking oils, and as such it was thought that artificially increasing stearic acid content in Brassica oil may increase oxidation resistance in vegetable oils (Zarhloul et al., 2006). Furthermore, recent medical studies have shown that stearic acid in vegetable oil has no increasing effect of plasma cholesterol in humans (Legrand et al., 2010). Two methods of increasing stearic acid content are; to downregulate the upstream endogenous stearoyl-ACP desaturase (SAD) activity using antisense DNA, or by insertion of the Garm *FatA1*gene from *Garcinia mangostana* (Nesi et al., 2008). These methods have been able to increase stearic acid concentration in single LEAR seeds by a factor of ten, from ~3.4 % to ~33 % (Zarhloul et al., 2006).

Lauric acid, or dodecanoic acid (C12:0) is also a medium chain fatty acid found in Brassica oil (Del Vecchio, 1997). A good source of medium chain fatty acids for use in structured lipids, this oil is found in the oil from many tropical plants, such as palm or coconut oil; and there has been industry attention at using modified Brassica oil as a substitute source of this oil (Hamam and Shahidi, 2005). Calgene Inc. of Davis, California, released Laurical 25, a Brassica oilseed variety with levels of lauric acid ~40 % (Del Vecchio, 1997). Further research is expected for upregulated lauric acid in rapeseed as medical applications are being investigated in the treatment of both maldigestion and cystic fibrosis (Hamam and Shahidi, 2005).

2.3.4 Glucosinolates

A group of agronomically important secondary compounds in the Brassicacaea family are the sulphur containing metabolites known collectively as the glucosinolates (GLS), found in all edible parts of the plant (Jeong et al., 2015) (Figure 2.4). Characterized by a β-thioglucoside-N-hydroxysulphate common structure (Fernández-Martínez and Velasco, 2012), there are 120 GLS variants based on side chain structure; the most common in canola meal being: gluconapin (3-butenyl), glucobrassiconapin (4-pentanyl), progoitrin (2-hydroxy-3-butenyl), glucanopoleiferin (2-hydroxy-4-pentenyl), glucobrassicin (3-indolymethyl) and 4-hydroxyglucobrassicin (4-hydroxy-3-indolymethyl) (Khajali and Slominski, 2012)

While non-toxic in their natural state, GLS molecules are almost always accompanied in-vivo by the enzyme myrosinase (thioglucoside glucohydrolase), which works to cleave the GLS thioglucoside linkage, yielding unstable aglycones such as isothiocyanate (ITC) which give Brassica plants their pungent mustard flavor (Kadir et al., 2015). When Brassica plant tissue is disrupted, the resulting glucosinolate-myrosinase action is referred to as the “mustard bomb”.

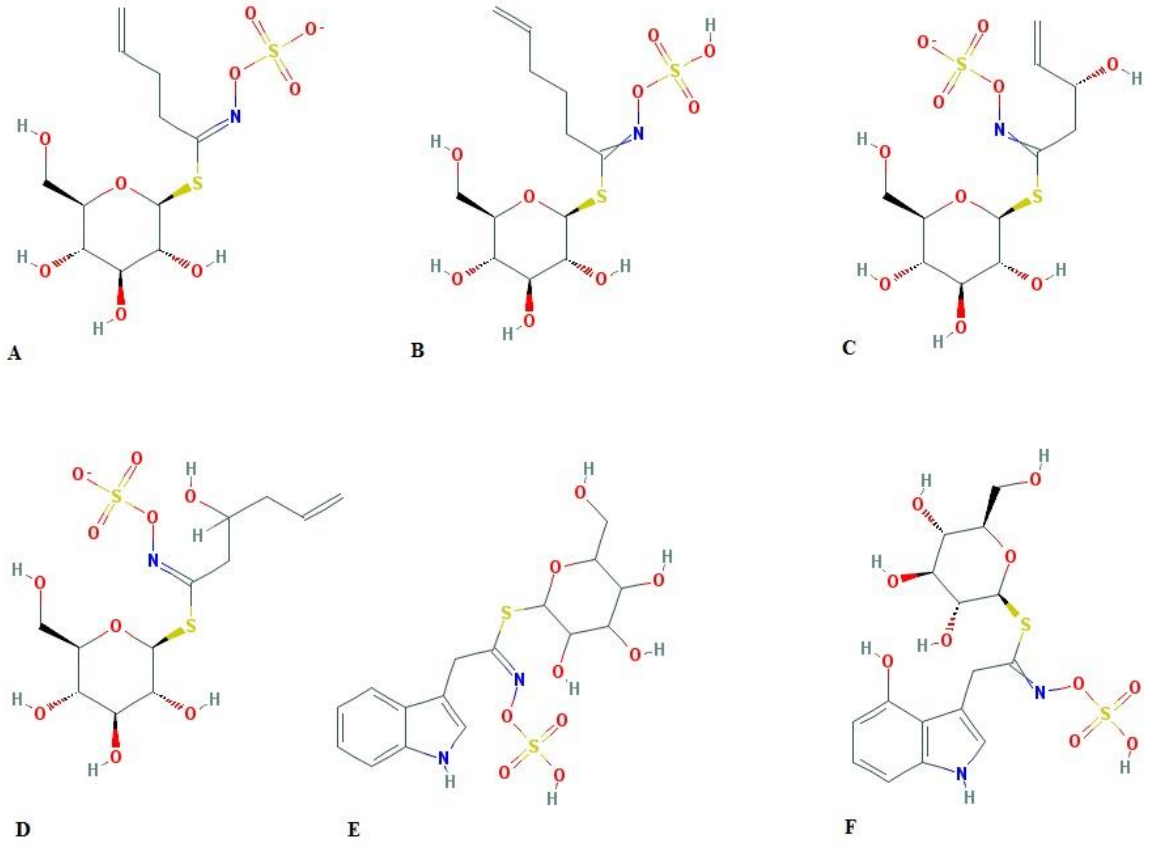

Figure 2.4. Two-dimensional chemical structure of common glucosinolate molecules, including: (A) Gluconapin, (B) Glucobrassiconapin, (C) Progoitrin, (D) Glucanopoleiferin, (E) Glucobrassicin, (F) 4-hydroxyglucobrassicin. Position of double bonds and active groups are shown. Taken from the US National Library of Medicine (PubChem, 2006).

In egg-laying poultry, high GLS diets (200g/kg) have been shown to cause decreased egg production and increased mortality due to hemorrhagic liver syndrome compared to poultry fed a diet of the low glucosinolate variety Tower (Ibrahim et al., 1980; Khajali and Slominski, 2012). Similar results were found in laying hens by Fenwick (1982), and McNeil et al. (2004) found that broiler chickens fed a high GLS diet exhibited lower feed intake, reduced growth, and higher mortality. Studies have since been conducted to determine a threshold level for acceptable

glucosinolate levels (which do not impair growth) in poultry, and found them to be ~10 µmol/g (Khajali and Slominski, 2012; Leeson and Summers, 2005; Tripathi and Mishra, 2007).

2.3.5 Meal Protein Content

As a bi-product of increased *B. napus* production for oil extraction, *B. napus* meal rich in protein is increasingly used as an alternative protein source in chicken, poultry, and cattle feed (Little et al., 2015). Brassica meal contains around 35-40 % protein with a favourable compliment of amino acids (Agnihotri et al., 2007), including lysine and cysteine which are not usually found in cereal meals (Zuo et al., 2015). While containing more methionine and cysteine than soybean meal, Brassica meal contains less raw crude protein and three times as much fibre as soybean yield (Little et al., 2015). However, pig diets rich in rapeseed meal have been shown to have fewer effects on nitrogen utilization and excretion than soybean meal (Zuo et al., 2015).

Due to bioenergetic constraints on *B. napus*, it is difficult for a rapeseed variety to produce both high protein and oil content, as the two components compete for the same energy supply (Agnihotri et al., 2007). Protein content can be increased by increasing nitrogen fertilizers, increasing relative levels of glutamine, proline, and arginine (Zuo et al., 2015). In the United States alone, 50 million tons of protein meal is consumed by animals in the livestock industry each year, with soybean meal comprising 26-30 million tons of that (Dow AgroSciences, 2016). To provide a new alternative protein meal, Dow AgroSciences have developed what they call ProPound™ quality meal, touted as having 20 % higher protein, 10 % higher energy, and 25 % less fibre than standard canola meal (Dow AgroSciences, 2016). This ProPound™ trait is usually paired with Nexera™ canola seed with an omega-9 oil profile making it a high protein, high energy, low fibre alternative to soybean meal with the added benefit of high oleic acid (The Dow Chemical Company, 2015).

2.4 Commercial Breeding of Rapeseed

The goal of any successful plant breeding program is to utilize resources to reach defined agronomic objectives for a plant of interest. Defining resources in this case as the time, labour, germplasm, environment and finances a program has available to it. Since breeding objectives can be incredibly diverse, they often require varied experience and expertise to accomplish. To maximize these breeding objectives, the management of germplasm is usually geared towards hybrid production (Roy, 2012). Hybrids are F_1 individuals developed from crossing two inbred parents which are homozygous across their genomes; and as such, all F_1 hybrids from the same parental cross should be genetically identical (Banga and Raman, 1998).

2.4.1 Heterosis

Offspring showing vigorous growth and development compared to their parents is nothing new; in 1889 Charles Darwin described this occurrence as "..nature showing us that she abhors perpetual self-fertilization" (Chen and Liu, 2014). However it was not until 1908, that George Shull independently characterized this hybrid superiority in Maize, giving it the name "heterosis" (Shull, 1948). Heterosis describes a situation in which the F_1 shows superior performance to either parent used in creating the hybrid (Ryder et al., 2014). The impact heterosis has had on the agriculture industry is enormous (Lippman and Zamir, 2007). It has given rise to a viable and productive seed industry to meet increased demands for food, fibre and sustainability; which hinges on the successful exploitation of heterosis in potential parents and 15-50 % of all yield increases in the last century are directly attributed to heterotic benefit (Lippman and Zamir, 2007). Hybrid plant breeding involves developing inbred parental lines with the hope of utilizing heterosis (Shull, 1948) to market an agronomically superior F_1. It is imperative to select parental lines which are

genetically distinct to ensure good heterotic benefit and a competitive F_1 hybrid (Brandle and Mcvetty, 1989). Hybrids which result from diverse parental lines show more vigorous growth and yield than those from closer related parents (Li et al., 2015). Hybrid breeding programs often spend significant amounts of time and money categorizing their available germplasm into heterotic groups of distinct individuals from which to select potential parents for hybrid crosses. The understanding of the molecular mechanisms underlying heterotic performance is therefore of the utmost importance to agriculture (Liberatore et al., 2013). Three major international conferences have been held to discuss these mechanisms; including at the University of Iowa in 1950, a 1999 CIMMYT (International Maize and Wheat Improvement Center) conference in Mexico City, and most recently a 2009 conference in Stuttgart, Germany. From these conferences, and several independent studies, several theories have received support for explaining heterosis: the dominance theory (DOM) (Davenport, 1908; Keeble and Pellew, 1910; Bruce, 1910), overdominance/pseudo-overdominance (ODO) (Jones, 1917 ; East, 1936 ; Lippman and Zamir, 2007) (POD), and genetic epistasis (EPI) (Melchinger et al., 2007 ; Ni et al., 2009 ; Chen, 2013 ; Schnable and Springer, 2013).

2.4.1.1 The Dominance Hypothesis

Early research, conducted by researchers such as Davenport (1908), Keeble and Pellew (1910), and Bruce (1910), pointed to the masking of accumulated deleterious alleles as the causative mechanism of heterosis. This theory is now referred to as the dominance hypothesis. Inbreeding depression is a situation by which deleterious recessive alleles manifest themselves within inbred lines resulting in negative phenotypic effects for the plant (Liberatore et al., 2013). According to the dominance theory, by crossing two different inbred parents, negative recessive alleles from one parent are masked by superior alleles in the other (Lippman and Zamir, 2007). This

explanation is favored even today (Radoev et al., 2008), yet has major criticisms, notably that it should be possible to fix an inbred line with all the positive alleles, eliminating the need for hybridization (Chen and Liu, 2014)**,** a development which has yet to occur.

2.4.1.2 Overdominance and Pseudo-Overdominance

Another explanation for heterosis came from East (1936) in the form of the Overdominance Theory, which asserts that it is intra-locus allelic interactions (regardless of the type of interaction) result in the superior hybrid performance. In this model, the masking of unfavourable alleles by dominant ones is not implicated in increased performance, it is a unique allelic interaction at heterozygous loci which results in high-performing offspring (Liberatore et al., 2013). This effect is thought by supporters to be locus specific (Lippman and Zamir, 2007), and modern studies using QTL analysis have shown that this may indeed be true for a number of cases (Liberatore et al., 2013). A previous but related theory to East's Overdominance theory was proposed by Jones in 1917 describing a situation in which there are alleles at multiple closely linked locations with opposite effects that are linked in repulsion causing effects across multiple alleles to be mistaken for the effect of a single overdominant locus. (Li et al., 2015; Schnable and Springer, 2013) (Figure 2.4).

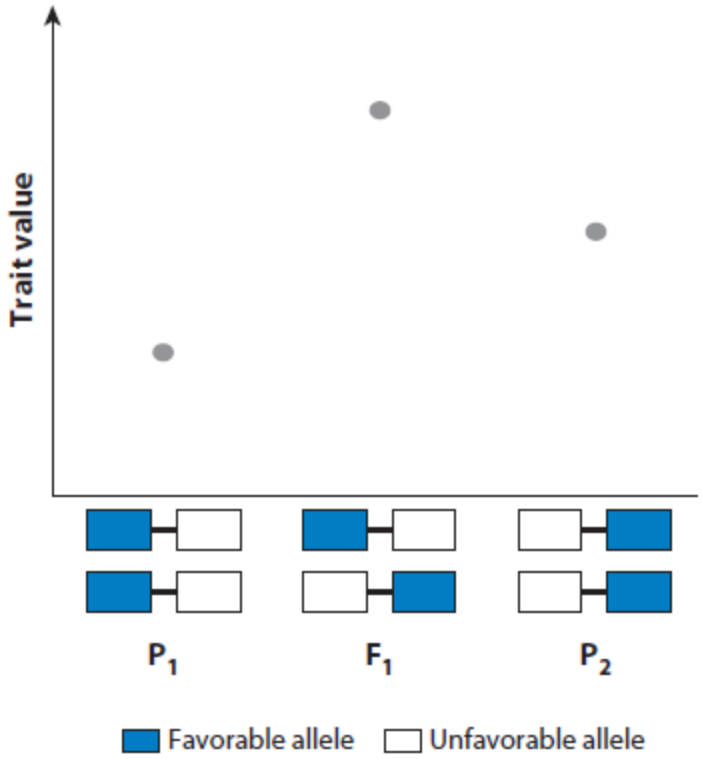

Figure 2.5. Pseudo-Overdominant effect from two linked loci in F_1 showing high parent heterosis in a trait value (Schnable and Springer, 2013).

2.4.1.3 Epistatic Effects on Heterosis

Epistasis, or the modification of a genetic locus by the action of another gene at another distinct locus, has been shown to have a small but significant role in the manifestation of heterosis (Schnable and Springer, 2013). The extent to which these epistatic interactions influence heterosis is notoriously difficult and expensive to test, due to the sheer number of crosses and tests you must perform to control for it (Melchinger et al., 2007). However, it is assumed that even small epistatic interactions contribute to the superior expression of QTLs during heterosis (Melchinger et al., 2007). Proponents of the epigenetic model cite the limitation of single gene studies in their inability to distinguish indirect regulatory effects from neighbouring genes from the effect of another

distinct QTL (Chen, 2013). Examples of this include epigenetic changes in protein expression resulting in altered flowering time and circadian rhythm in *Arabidopsis thaliana* (L.) Heynh. hybrids (Ni et al., 2009). The term epi-allele describes a number of stable linked loci that are inherited together which have been suggested to mimic more concrete genomic differences described by dominant and overdominant effects (Schnable and Springer, 2013).

2.4.1.4 Modern Theories on Heterosis

Since the first time it was described, evidence has been brought forward supporting each of the four hypotheses for heterosis. It is generally accepted that different mechanisms and factors causing heterosis are favored in different species and situations which makes a unified theory difficult to rationalize (Li et al., 2015). Although the dominance model has the most supporting evidence in maize, not all the dominant loci predicted to maximize heterosis can be fixed in inbred lines, implying other factors are at work (Chen, 2013). This could imply more about the necessity to detect more precise QTLs than a flaw in the dominance model.

Xin Li (2015) and his research team at Iowa State University investigated the mechanism of heterosis in sorghum (*Sorghum bicolor* (L.) Moench). To date, four important QTLs, *Dw1 – Dw4*, are linked to heterosis in plant height for sorghum, and of these, only *Dw3*, which codes for an auxin transporter, is mapped. Using recombinant inbred lines, they were able to determine that there was another QTL, *qHT7.1* that is linked in repulsion with *Dw3* mimicking the actions of an overdominant effect.

There are a few more unexplained factors described by Birchler et al. (2010) which need to be addressed in the creation of any unified heterotic theory. These include the fact that there isn't much of a decrease in heterosis when obvious negative alleles are purged from inbred genotypes. Other unexplained factors include; the tendency of polyploids to exhibit progressive heterosis, as

well as the dosage component of heterosis (situation where AABB shows higher heterosis than AAB in allopolyploids) (Kaeppler et al., 2012).

2.4.2 Germplasm Evaluation for Heterotic Benefit

For ease of parental selection in hybrid seed production, breeders find it useful to organize their available germplasm into distinct groups based on either shared phenotypic morphology, or on shared molecular identity (Melchinger and Gumber, 1998). To this end, it is common for most breeders to maintain at least two separate groups within their available genetic resources (Brown et al., 2014). Ana et al. (2009) used a combination of morphological (plant height, stem diameter, branch number, pod morphology, leaf morphology) and agronomic traits (yield, oil content, protein content) to evaluate and successfully organize *B. napus* genotypes into clusters based on diverse geographical origin (North America, Europe, Africa). In a similar study, Mahasi and Kamundia (2007) used agronomic data (yield, oil content, protein content) to cluster and differentiate between diverse genotypes of European and Canadian origin. Depending on the size of the population to be evaluated, morphological and agronomic traits can understandably create a logistical bottleneck due to time, space, and labor.

Over the past 40 years, the use molecular techniques to observe patterns of relatedness in germplasm have provided an efficient and effective use of a breeding programs time and resources (Xie et al., 2015). Genetic markers are single genes or sequences of DNA that have a myriad of uses in molecular technology; including identification, variation analysis, phylogenetic analysis, and plant breeding (Yang et al. 2015). Molecular technology such as genetic markers has allowed breeders to categorize the available germplasm allowing them to target their available time and space towards genetically distinct parents. Today, powerful next-generation sequencing (NGS)

platforms allow for heavily multiplexed analysis of genetic material screening hundreds of samples at a time (Brown et al., 2014). Molecular markers and NGS will be reviewed in section 2.7.

2.5 Pollination Control

Once parental genotypes have been selected and homozygosity ensured, pollen from one potential parent must fertilize the egg cell of the other potential parent thereby creating a hybrid. The problem is, *B. napus* is more likely self-pollinate (70-90 %) as it is to cross-pollinate (10-30 %) (Banga and Raman, 1998), therefore, steps must be taken the ensure cross-pollination and confidence in hybrid seed production. These steps work to ensure that one parent, does not contribute pollen of its own, and will functionally represent only the "female" plant. These individuals are referred to as "male-sterile". For the first half of the 20th century, the process of creating male-sterile individuals involved the hand emasculation of "female" plants, either through manual labour, machines or chemical treatments (Chen and Liu, 2014). All these options take time, money, and cannot ensure that pollen flow has been stopped. As such, the focus of plant breeding in self-pollinating species such as *B. napus* became the creation of high performing open-pollinating homozygous lines.

In 1794, a German botanist by the name of Joseph Gottieb Kölreuter described a phenomenon by which a subset of individuals within a species do not produce any pollen, whereas other individuals within the same species do; a phenomenon which has since been reported in as many as 610 other plant species (Chen and Liu, 2014). This phenomenon is known today as male sterility, and refers to a situation by which a "male" plant is either unable to produce functional pollen, or is unable to produce pollen at all (Brown, 1999). Male sterility would prove to be a crucial breeding tool in the production of many commercial crops, allowing breeders to exploit heterosis by reliably combining selected parents (Chen and Liu, 2014).

2.5.1 Cytoplasmic Male-Sterility

Cytoplasmic Male-Sterility (CMS) systems are the largest and most widely used male-sterility systems in agriculture today (Stiewe et al., 2008). CMS-T (Texas) was released for corn in the 1950s as the first commercially available male-sterility system (Chen and Liu, 2014), and since then various CMS systems have been released for essentially all agronomically important crops. Caused by widespread mitochondrial (or cytoplasmic) encoded genes, which prevent functional pollen formation in most plants, these genes are kept in check by the presence of nuclear "restorer of fertility" (Rf) genes (Brown, 1999). There are multiple different CMS genes capable of inducing male-sterility, each with it's own unique nuclear *Rf* gene, and if an individual carries a sterility gene without the corresponding *Rf* gene, no functional pollen is formed (Kausch et al., 2012).

2.5.1.1 *pol* and *nap* CMS

Polima CMS was first found in 1972 by Fu (Fu, 1981) as a spontaneous male-sterility system found in *B. napus*. The causative gene for *pol* male sterility is thought to be the presence of the mitochondrial encoded *orf224* in the absence of the corresponding "restorer" or *Rfp* gene in the nuclear genome (L'Homme and Brown 1993). This has been demonstrated both by an alteration in the *orf224* protein product in the presence of the *Rfp* (L'Homme and Brown, 1993) and by the injection of *orf224* into a normally fertile *B. oleracea* inducing sterility (Wang et al. 1995). Behaving in a similar fashion is Thompson-Shiga sterility, owing its name to the discovery by two different research groups; Thompson (1972) and Shiga and Baba (1973). Thompson and Shiga had originally thought they discovered separate systems, which were called TSCM and SCMS, respectively; but further analysis showed both TCMS and SCMS were caused by the same genes (Wang et al., 1995). Thompson-Shiga sterility is more commonly known today as *nap* sterility (L'Homme et al, 1997**)**. Instead of the *pol* sterility causing *orf224*, individuals displaying *nap*

sterility contain *orf222* and the corresponding restorer gene *Rfn* **(**L'Homme et al, 1997). Many naturally occurring populations of *B. napus* contain the cytoplasmic *pol* or *nap* CMS trait as well as the corresponding nuclear restorer *Rfp* or *Rfn*, as such male-sterility is not often seen (Brown, 1999). However, a plant with *nap* cytoplasm and the *Rfp* restorer gene will be male-sterile; and a plant with the *pol* cytoplasm and nuclear *Rfn* will also be male-sterile, and this is the fundamental principle between causing *nap* or *pol* male sterility (Fan and Stefansosn, 1986). Plants using the polima system are characterized by short petals, underdeveloped anthers, an increased carpel size, and suffer from limited restorer potential as they are also unstable at high or low temperatures (Stiewe et al., 2008). As such, the *pol* system is relatively limited in its utility for commercial breeding programs, though is used in China (Wang et al., 1995). As previously stated, *nap* male sterility is controlled by the *orf222/nad5c/orf139* region of mtDNA located in *nap* male-sterile cytoplasm (L'Homme et al., 1997) which is found in many lines of *B. napus,* though it is restored to fertility by relatively few genes (Fan and Stefansson, 1986). To this end, *nap* male-sterility is restored by inducing functional pollen production in male-sterile plants by growing them under high temperature conditions (Fan and Stefansson, 1986).

2.5.1.2 Ogura CMS

In 1968, Hiroshi Ogura described a system of male-sterility (Ogura, 1968) in the Japanese radish (*Raphanus sativus* L.), which was later transferred to *B. napus* through the introduction of *Ogura* containing cytoplasm and introgression of a nuclear restorer (*Rfo*) gene (Bannerot et al., 1974). The mitochondrial gene responsible for sterility in the Ogura cytoplasm is the open reading frame *orf138* which is expressed in sterile plants (Giancola et al., 2007) to arrest the development of microspores. However, the fragment of *Raphanus* DNA that contains *orf138* can cause negative effects in *B. napus* individuals carrying the introgression; such as: (1) low female fertility, (2) poor

homozygous plant performance, and (3) the expected fertile:sterile ratio in F_2 offspring is not observed (Primard-Brisset et al., 2005). Poor female fertility can result from malformed ovaries and underdeveloped nectaries which in turn lead to disfigured pods and poor seed set (Pelletier and Budar, 2015)

Because of the agronomic detriment of Ogura cytoplasm, considerable effort has been put on breeding Ogura CMS restorers that do not display these negative characteristics. The French Institute for Agricultural Research (INRA), has been a leader in this effort, and has employed the use of pedigree and backcross methodology to both decrease the size of the radish introgression containing *orf138* as well as decreasing the glucosinolate content and increasing overall vigor in restored females (Pelletier and Budar, 2015). INRA has developed the Ogu-INRA restorer lines P113 and subsequent R2000 via irradiation of heterozygous plants carrying the *orf138* introgression that have both displayed increased vigor in adult plants and fertility as it relates to seed number and quality (Primard-Brisset et al., 2005).

2.5.1.3 Seedlink® Sterility

In 1995, Bayer Cropscience released the SeedLink® cytoplasmic male-sterility system for use in hybrid crop development (De Both, 1995). SeedLink® is paired with Bayer Cropscience's Liberty® herbicide tolerance system and provides hybrid offspring with immunity to glufosinate-based herbicides (Singh, 2015). First pioneered in tobacco (*Nicotiana tobacum*) Seedlink® sterility is based on the chimaeric Barnase gene, which encodes for a cytotoxic ribonuclease expressed during anther formation, interrupting its progress (Mariani et al., 1990). Fertility is restored through the introduction of the Barstar gene that codes for a Barstar inhibitor allowing for normal anther development.

2.5.2 Nuclear Male-Sterility

After CMS systems, nuclear (or genic) male-sterility (NMS, GMS) systems, defined by complete control by nuclear genes, are the next most promising avenue in the development of pollination control systems. Unlike CMS, NMS systems have extensive sources of cytoplasm and restorer genes that make for more stable-sterility systems (Deng et al., 2016). Most of the genes involved are recessive, making fertility restoration possible by combination with most inbred lines (who carry the dominant form of the gene in question), giving many NMS systems a distinct advantage over CMS based systems (Dong et al., 2012). These systems however are difficult to handle, can be expensive to develop and maintain making them rarer than CMS systems in commercial breeding (Dong et al., 2012)

2.5.2.1 Recessive Genic Male-Sterility System

Recessive genic male-sterility (RGMS) describes a method of pollination control in that a homozygous recessive locus confers male-sterility on the individual (Dong et al., 2012). RGMS systems have been used extensively and to great effect in China for *B. napus* with the three lines; S45AB (Pan et al., 1988), 117AB (Hou et al., 1990), and 9012AB (Chen et al., 1993) acting as the female line in most crosses. S45AB and 117AB require that nearly 50 % of fertile plants be removed from the RGMS lines before the hybridization event can occur, which limited their utility outside of China. The more stable line, 9012AB does not require this step and has contributed to several elite lines coming out of China since its discovery (Dong et al. 2012). The gene *BnRf* has been implicated as the causative gene in 9012AB and was mapped by Deng et al. (2016) to a 13.8 kb fragment on the *B. napus* chromosome A7.

2.5.2.2 Male-Sterile Lembke

Described by one of its developers as an "alternative CMS system" (Stiewe et al., 2008), the Male-Sterile Lembke, or MSL system, is classed as a GMS system based on a spontaneous mutation at a German NPZ Lembke (Norddeutsche Pflanzenzucht Lembke) test facility in 1984. The system employs a fertile maintainer line to propagate sterile mothers and fully restored hybrids (Stiewe et al., 2008). However, the parents involved in the creation of MSL male-sterility as well as the process of hybrid seed formation are a trade secret of NPZ Lembke.

2.6 Statistical Estimation of Hybrid Performance

Once a desired trait has been identified, and a method of hybridization selected, F_1 hybrids can be produced and their performance evaluated for the trait of interest. This evaluation can be done in the greenhouse or via field trials of varying size and configuration. From the information gathered in these trials, the concept of general, and specific combining ability becomes important (Griffing, 1956). Griffing's method for statistically evaluating the suitability of parental lines in hybrid production was introduced, which has now been used successfully by hybrid plant breeders for over 50 years (Roy, 2012). Griffing's method uses crosses of all parental combinations to determine General Combining Ability (GCA), being the average performance of a given parental line in hybrid combinations; as well as Specific Combining Ability, being the difference in performance which cannot be attributed to GCA (Griffing, 1956). Using both the determined GCA and SCA, the performance of any cross between two parents for a trait of interest can be mathematically estimated.

However, in many breeding operations it is often not feasible to evaluate every single parental combination to determine combining ability; this can be due to the sheer number of parental combinations, or because of limitations in the pollination control system used. Thus, it is common

for breeders to evaluate parental performance against a common tester; and measure the performance from the resulting F_1's (Brown et al., 2014). Unfortunately, parent performance for a trait of interest has been a weak predictor of hybrid performance for the same trait of interest (Hallauer et al., 2009). For this reason, alternative methods of hybrid performance prediction have been developed, including the use of heterotic groups and whole genome prediction (WGP) (Reidelsheimer et al., 2012).

2.6.1 Heterotic Groups

The assumption that diverse inbred parents will produce high performing F_1 hybrids is central to the rationale for heterotic groups (Laude and Carena, 2015). The criteria that establishes a given dataset's diversity can vary including: pedigree data (Badu-Apraku et al., 2006), morphological data (Ana et al., 2009), and molecular marker data (Melchinger et al., 1991; Mohammadi and Prasanna 2003). These methods organize a dataset into distinct groups of genotypes that display a similar heterotic effect when crossed with individuals in other groups (Parentoni et al., 2001). Heterotic groups were first effectively used in maize, where germplasm was separated into stiff stalk (SS) vs. non-stiff stalk morphological groups resulting in increased agronomic performance in hybrid offspring with one parent from each group (Tracy and Chandler, 2006).

Cluster analysis refers to the establishment of groups with high internal homogeneity and high external heterogeneity for the trait of interest (Laude and Carena, 2015). Several statistical procedures exist for clustering a dataset, though the unweighted pair group method and Ward's minimum variance method are the two most widely used (Khodadadi et al., 2005; Teklewold, 2006; Fahid, 2014; Singh, 2015). Although the UPGMA is still more common, Ward's method is increasingly seen as ideal, as the UPGMA method often results in higher internal cluster affinity

than is actually present, a feature of the UGPMA method which is known as the “chaining effect” (Lombard et al, 2000; Mohammadi and Prasanna, 2003; Khodadadi et al., 2005).

2.7 Genetic Markers

2.7.1 Restriction Fragment Length Polymorphisms

The first widely used genetic markers were Restriction Fragment Length Polymorphisms (RFLPs), developed by Sir Alec Jeffreys to aid in a murder investigation (Jeffreys et al., 1985)**.** RFLP markers have since been used successfully in Brassica breeding. Sakamoto et al. (2000) used these markers to evaluate allelic variation in 62 *B. oleracea* (31 broccoli morphotype, 31 cabbage morphotype). They discovered a high degree of homogeneity in these lines as it related to the important S allele (91 % in broccoli morphotype, 97 % in cabbage morphotype) identifying the need to breed in new variation. Quantitative trait loci linked to important Brassica agronomic characteristics have been identified using RFLP markers. Mahmood et al. (2003) used 316 RFLP markers in *B. juncea* L. (Vassiliî Matveievitch Czernajew) to create 18 linkage groups over the A and B genomes aiding in the identification of QTL related to oleic, linoleic, linolenic and erucic acid content in the seed oil. However, because of its high cost and the large amount of high quality DNA required for its use, RFLP markers have since been replaced by the markers which use Polymerase Chain Reaction (PCR) to amplify fragments of DNA (Saiki et al., 1985). Today, genetic markers can generally be classified as either: a) those without genetic sequence specificity, and b) those making use of sequence specificity (Yang et al., 2015).

2.7.2 PCR-Based Without Specificity

2.7.2.1 Random Amplified Polymorphic DNA

Random Amplified Polymorphic DNA (RAPD) refers to a PCR-based system which uses small (8-12 nucleotides) and arbitrary primers to amplify random segments of DNA by several orders of magnitude (Caetano-Anollés and Gresshoff, 1994). These markers require no specificity and are therefore useful in species or biological systems which have received little research, as well as diversity studies (Kumar and Gurusubramanian, 2011). RAPD markers have been successfully used in Brassica species to great effect. Somers et al. (2001) was interested in *B. napus* seed with a yellow-pigmented seed coat, which had previously been linked to high protein, low fibre characteristics. Using RAPD markers, they were able to identify a gene, flanked by 8 RAPD markers, which explained 72 % of seed colour variation in their test population. RAPD markers have also been used to perform more broad diversity studies, such as conducted by Ahmad et al. (2007) in which RAPD fingerprinting aided in the characterization of *B. napus* and *B. rapa* breeding programs. Despite its successes, some researchers view the RAPD protocol negatively for plant breeding due to it being only cost and time effective in experiments with few genotypes (Ragot and Hoisington, 1993) as well as an unrepeatable and unreliable nature (Zuo et al., 2015).

2.7.2.2 Amplified Fragment Length Polymorphisms

Zabeau and Vos first described amplified Fragment Length Polymorphisms (AFLPs) in 1993 as a system of molecular markers with broad utility in research and DNA fingerprinting. A PCR-based system, AFLP uses restriction enzymes to digest a DNA sample, and then uses adaptor-restriction site specific sequence pairs in the digested DNA to create groups of DNA which can be visualized and isolated using gel-electrophoresis (Perry et al., 1999). Honsdorf et al. (2010) used 684 AFLP markers to test a population of *B. napus* for QTL related to 14 phenological, morphological and

quality traits of interest. Using these markers, they were able to identify 27 possible QTL candidates linked to at least two traits (more with the inclusion of phenotypically linked traits i.e. pods per main raceme and pod density etc.). Zeng et al. (2011) evaluated dwarf lines of *B. napus* containing the EMF generated recessive gene *bnaC.dwf* with the hope of identifying markers to speed up marker-assisted selection (MAS). They were able to identify two AFLP markers, 0.2 and 0.05 cM from *bnaC.dwf* respectfully, which they hoped, would speed up the development of dwarf varieties. Although the highly reproducible nature of AFLPs have made it an attractive marker choice, AFLP technology is a registered trademark of Keygene N.V, subject to patent, limiting its attractiveness (Van Eijk et al., 2009).

2.7.2.3 Microsatellite-Anchored Fragment Length Polymorphisms

Microsatellite-Anchored Fragment Length Polymorphisms (MFLP) were developed as a modification to AFLPs by Yang et al. (2002) with the addition of microsatellite-anchor primer techniques. Instead of adaptor-restriction site specific pairs to separate digested DNA, MFLP protocol uses microsatellite-AFLP primer pairs to develop polymorphisms which are SSR-*MseI* fragments (H. et al., 2002). MFLP markers have been extensively used in the breeding of *Lupinus angustifolius* L. (narrow leafed lupin) in Australia (Shahidul et al., 2013). Boersma et al. (2005) was successful in using MFLP technology to develop markers for reduced pod shatter, petal color and seed coat strength in lupin. Phomopsis Stem Blight (PSB) is an important disease for narrow leafed lupin in Australia. Yang et al. (2002) used MFLP protocol to search for markers associated with the PSB resistance gene *Phr1*, and were successful in identifying two markers that could be used to screen lupin populations for the presence of the gene.

2.7.2.4 Sequence Related Amplified Polymorphisms

Sequence Related Amplified Polymorphisms, or SRAP markers, are a simple, inexpensive way to amplify open reading frames using ambiguous primers to produce genome wide fragments (Li and Quiros, 2001). Forward primers designed to target GC rich exons and reverse primers targeting AT rich regions and amplify multiple fragments that are separated on an acrylamide gel and detected via autoradiography. These robust and diverse fragments have been very useful in MAS of *Brassica* species and the method is less technical than other forms of marker identification and allows for widespread utility.

Many diversity studies have been conducted with SRAP markers, with Wu et al. (2009) using SRAP markers to investigate genetic diversity in Eurasian landraces of *B. juncea*. Using eight primer combinations they were able to clearly separate 95 distinct genotypes of *B. juncea* into three clusters corresponding with vegetable, spring oil, and winter oil varieties. They were able to identify high levels of diversity between genotypes in the vegetable cluster and the two oil clusters, aiding in the creation of high heterosis crosses. In 2012, Raman et al. used SRAP markers as part of a method to discern QTL underlying flowering time in *B. napus*. These markers were aligned with the position of previously predicted QTLs in *B. rapa* and *A. thaliana*. The hope was to identify the chromosomal location of flowering time genes, speeding the breeding of early or late maturing varieties.

Ghanbarnia et al. (2012) at the University of Manitoba used SRAP markers to investigate *AvrLepR1*, an avirulence gene in the important *B. napus* pathogen *Leptosphaera maculans* (Sowerby (P. Karst)). Using 259 SRAP markers, they were able to map the *L. maculans* genome into 36 linkage groups, placing the *AvrLepR1* in linkage group 4. *AvrLepR1* corresponds to the

important *B. napus* resistance gene *Brassica LepR1*, and this study was conducted to learn if it was distinct from other previously identified avirulance genes on *L. maculans*.

2.7.2.5 Diversity Array Technology

Diversity Array Technology (DArT) is a microarray-based genotyping platform which can detect polymorphisms at restriction fragments and within indels (Jaccoud et al., 2001). This system is non-specific, does not require sequence-information and can be performed on parallel and automated platforms, significantly reducing costs (Yang et al., 2015). DArT has been successfully used in polyploidy species as well as successful diversity studies in crops such as sorghum (Mace et al., 2009) and rapeseed (Raman et al., 2012). Traini et al. (2013) used diversity array technology to investigate genome organization, gene function and genetic variation within wild potato species, *Solanum commersonii* (Dunal) and *S. bulbocastanum* (Dunal) which both lacked a reference genome of their own. Using 1423 DArT markers as well as the reference genomes from the related *S. tuberosum* L. and *S. lycopersicum* L. they were able to identify a large degree of microscale heterogeneity in the genetics of the related *Solanum* species. The variation discovered in *S. commersonii* and *S. bulocastanum* could potentially improve the genetic background of the heavily cultivated *S. tuberosum.*

Niedziela et al. (2015) converted 49 DArT markers into 24 polymorphic PCR assays to screen for aluminum tolerance genes in triticale (*x Triticosecale* W.). They cited aluminum stress as a characteristic of acidic soil that negatively affects yield, and hoped to design molecular markers to speed up a selection process that had previously been confined to physiological studies. Using their PCR assays, they were able to find a marker linking Al- tolerance to the 7R chromosome, providing a method for future MAS.

2.7.3 PCR-Based with Specificity

2.7.3.1 Sequence Tagged Sites

Sequence Tagged Sites (STS), are markers that take the form of short (200-500 nucleotides) fragments of DNA with a single occurrence in the genome and multiple allelic forms per species (Perry et al., 1999). First described in 1999, the STS system used the precise location and sequence at a particular locus in designing markers that are easily detectable, and produce a readily identifiable banding pattern on a polyacrylamide gel via electrophoresis.

Kar et al. (2012) used STS methodology to investigate Turmeric (*Curcuma longa* L.) germplasm and its susceptibility to the rhizome rot causing pathogen *Pythium aphanidermatum* (Edson (Fitzp)). Using bulk segregate analysis, they pooled DNA from susceptible and resistant genotypes and identified a 720 base pair segment that was found in all resistant genotypes, but none of the susceptible ones. Converting this 720 bp segment into an STS marker, they successfully used it to detect resistance. Kar and his team intended for it to be an early diagnostic tool in evaluating new turmeric germplasm for possible resistance, as the majority of turmeric is heavily susceptible to rhizome rot.

2.7.3.2 Simple Sequence Repeats

Microsatellites, or Simple Sequence Repeats (SSRs) as they are referred to in plant genetics, are simple chunks of DNA (2-5 base pairs) which repeat themselves over and over (Queller et al., 1993). SSR markers have an almost random distribution throughout genomes and have incredible degrees of polymorphism, making them a popular target for molecular markers as well as kinship analysis and genetic mapping (Schlötterer, 2000). When amplified by PCR, the large number of

different SSR markers available make them very robust and adaptable for several uses in plant breeding.

Guo et al. (2015) used 315 SSR markers in a diversity study on 96 resynthesized and 25 conventional genotypes of *B. napus*, hoping to identify variation that could be exploited by future breeding efforts. Clusters generated using SSR data displayed significant intra-cluster variation within the 96 resynthesized genotypes that were further separated from the conventional genotypes. This showed molecular diversity within resynthesized lines that Guo et al. (2015) planned to exploit in future *B. napus* breeding.

2.7.4 454-Pyrosequencing

454-pyrosequencing (454 Life Sciences, Branford, CT) was a sequencing platform released in 2005 which utilizes the pyrosequencing method described by Ronaghi et al. **(**1996**)** to luminometrically monitor the sequential addition of nucleotides to a template DNA strand and deduce the target DNA sequence through the order of these luminometric events (Metzker, 2010). In the 454-pyrosequencing method, a DNA sample is ligated into small fragments which are flanked with one of 454 available adaptors, and then bound individually to 28 µm beads before clonal amplification via PCR and attachment to an array for pyrosequencing (Shendure and Ji, 2009). Each bead could have one of 454 adaptor fragments giving the system its name. As the first commercially available next-generation sequencing platform, 454-pyrosequencing created a high throughput answer to library preparation, template preparation and sequencing. However, since the pyrosequencing method uses luminescence to identify nucleotides, the proper detection of homo-polymers, where the same base is added multiple times to a sequence, relied on the interpretation of signal strength, making insertion/deletion errors a common error in 454-pyrosequencing (Shendure and Ji, 2009). While an important step in next-generation sequencing

technology, the high cost, and error rate (Metzker, 2010) associated with 454-pyrosequencing resulted in the platforms discontinuation in 2013.

2.7.5 Illumina Sequencing

Born out of the merger of Solexa (Essex, UK), Lynx Therapeutics (Hayward, CA, USA), Manteia Predictive Medicine (Coinsins, Switzerland) and Illumina (San Diego, CA, USA) the illumine sequencing platform is a broadly used next-generation sequencing platform utilized today (Shendure and Ji, 2009). The Illumina platform uses an adaptor rich array, or flowcell, to bind one end of single-stranded DNA fragments which then subsequently bend and attach another adaptor with their free end allowing for selective solid-phase bridge amplification of complimentary strands bound by adaptors (Morozova and Marra, 2008**)**. Once amplified, these flowcell bound clusters of DNA can be sequenced and evaluated in a number of ways, two of which are 1) Restriction Site Association DNA sequencing and 2) Genotyping-by-sequencing.

2.7.5.1 Restriction Site Association DNA Sequencing

Using the Illumina sequencing platform, Restriction Site Association tags (RAD tags) provide a window into the genetic diversity of a species (Xiao et al, 2015). First developed using *Drosophila melanogaster*; RAD tags were used as a genetic marker system to isolate and identify single-nucleotide polymorphisms (SNPs) localized at the site which a particular restriction enzyme binds to a DNA fragment (Baird et al., 2008). This results in the genome wide revelation of the flanking SNP identity to a restriction enzyme binding site and are a useful way to evaluate species wide diversity at that location. Many different restriction enzymes can be used to increase resolution; however, since only the sites adjacent to restriction enzymes are evaluated, RAD sequencing can only offer a small fraction of segregating polymorphisms in a species (Baird et al., 2008).

RAD sequencing was using by Zhao et al. (2014) to construct the first high density genetic map in the Chinese date (*Ziziphus jujube* Mill). They were able to identify 47,784 polymorphisms in their test population, which they reduced to 2748 RAD markers for genetic mapping. These 2748 markers were grouped into 12 linkage groups spanning 913.89 cM with an average interval of 0.34 cM. Xiao et al. (2015) used a similar method to construct a RAD library for tobacco (*Nicotiana sp.*). Using a BC_1 population, a RAD-seq protocol was used to identify 2162 de-novo SNPs that they organized into 24 linkage groups spanning 2000.9 cM. To compare the efficacy of RAD-seq they constructed a second map using a reference genome, which resulted in the identification of 4138 SNPs organized into 24 linkage groups spanning 1944.74 cM. They concluded that RAD sequencing was suited for SNP identification, genetic map construction as well as for downstream investigation.

2.7.5.2 Genotyping-by-Sequencing

For plant species with large, complex, or redundant genomes, investigating genetic diversity can be challenging. To this end, the Genotyping-By-Sequencing (GBS) method, released in 2011 (Elshire et al., 2011) is a fast and cost-effective method, using the Illumina platform, to handle large diverse crop species such as maize or wheat (Xie et al., 2014). The key to this analysis is to reduce genome complexity using methylation sensitive restriction enzymes to avoid highly repetitive regions to ease the burden on computation (Elshire et al., 2011). The DNA to be read is affixed to a slide and subjected to polymerase amplification forming localized concentrations of DNA. Reversible terminators are added to these concentrations, affixing themselves to the free 3' end of the DNA, and non-incorporated nucleotides are removed. A camera takes images of fluorescent attached bases, showing which base is present. This fluorescent 3' base is then

chemically removed and the process starts all over again revealing nucleotide sequence order base by base.

Annacchiaro et al. (2015) used GBS generated SNP makers to evaluate two breeding populations of alfalfa (*Medicago sativa*), one adapted to the Po Valley in Italy (PV: 124 genotypes), and the other to the broader Mediterrannean region of Europe (Me: 154 genotypes). The rationale here was to identify molecular markers strongly linked to yield improving QTL to improve on the approximately 150 markers previously identified using RAPD, AFLP, SSR and RAPD methods, which Annacchiaro et al. (2015) found insufficient. Using GBS SNP identification, 68,972 were identified in the PV population, and 77,610 were identified in the Me population. However, when stringent thresholds for missing data were applied to the data GBS analysis still resulted in 7000 usable SNPs in the PV population and 11,000 in the Me population. Using support vector and pine ridge regression models, a predictive power of 0.35 was achieved in Me, and 0.32 in PV. Annacchiaro et al. (2015) concluded that the moderate predictive ability, the moderate cross population predictive ability and lack of subpopulation genetic structure make GBS an attractive method for molecular analysis in alfalfa.

Although there have been few diversity analyses in *B. napus* using GBS, Lees et al. (2016) used SNP markers gathered using GBS to characterize 79 *B. napus* genotypes into distinct groups. GBS analysis of the 79 genotypes yielded 80,005 bi-allelic SNP markers, by which 12 distinct groups were identified in a dendrogram generated by neighbour-joining cluster analysis. Genetic distance between the evaluated genotypes was also determined using the Tamura-Nei formula. Lees et al. (2016) comment on the robustness of the GBS generated SNP markers for diversity analyses, pointing to a > 90 % support of their dendrogram structure over 1000 bootstrapping replications. These results are positive for the use of GBS in *B. napus* diversity analyses and Lees et al. (2016)

concludes that GBS generated groups may have utility in parental selection for hybrid *B. napus* production.

There have been some concerns over the feasibility of GBS analysis in species with large and complex genomes such as barley (*Hordeum vulgare* L.) and wheat (*Triticum spp.*). In their 2012 study, Poland et al. sought to test this. They were able to identify 34,000 SNP markers in barley using the Wolfe Barley Reference Genome, and 20,000 SNP markers using the Synthetic W9784 x Opata85 Reference Genome. Both groups of SNP markers were placed in high density maps for their respective species, and were enough to identify chromosomal structural components such as probable centromere location. Within the SNP markers identified from the evaluated wheat, a large heterogenous deletion was found. A layover from the synthetic parental stock, which Poland et al. (2012) asserts would have distorted a genetic map generated using a presence/absence method such as DArT. This study highlighted the ability of GBS to develop high density genetic maps, which are a boon for genomic selection as well as in the mapping of QTL for MAS.

2.7.5 Markers and Breeding

2.7.5.1 Marker-Assisted Selection

Genome prediction hinges on the ability to split the genome of a target organism into manageable segments which can be tested for, and the effect of which to be estimated for any morphological, agronomic, or quality traits of interest (Hayes et al, 2009). Marker-assisted selection uses molecular data from the plethora of marker technologies available to detect allelic variation within a target population (Collard and Mackill, 2008). This variation can then be used to generate heterotic pools, or screen for QTL, both of which are central to successful plant breeding. Breseghello and Sorrells (2006) used 62 SSR markers on a population of elite wheat (*Triticum aestivum* L.) germplasm to successfully detect markers for kernel size across three chromosomes

(2D, 5A, 5B). In *B. napus*, Rahman et al. (2008) implicated two genes, *Bn-FAE1*, *Bn-FAE2*, on the A and C genomes respectively, as being responsible for the 0% erucic acid content of canola. These genes were then targeted for SRAP marker discovery, and they were able to associate 124 SRAP markers with *Bn-FAE1* and 42 markers with *Bn-FAE2*. Similarly, Hu et al. (2006) implicated two loci, corresponding to two fatty acid desaturase genes, as being partially responsible for high oleic (*fad2*) and low linolenic (*fad3c)* content in canola quality rapeseed oil. Using both SSR and AFLP marker analysis of differences in mutant and wildtype genotypes, they developed a marker for mutations in both *fad2* and *fad3c* which could be used to select for high oleic and low linolenic oil quality in *B. napus*. Also implicated in high oleic and low linolenic acid character in *B. napus* oil is the marker UBC_{830} identified by Javidfar et al. (2006), which was responsible for 43 % of variation in oleic acid content, and 13 % of low linolenic acid content in their test population. Markers such as these could potentially help with early generation selection for differences in oil quality in diverse or resynthesized lines of *B. napus*

However, MAS is not without it's challenges. While successful in screening populations for single alleles of large effect, it has thus far had difficulty with polygenic traits, where multiple small effect alleles segregate independently causing effects on a gradient (Jannink et al., 2010). Linear regression and least-squares models have had mixed results in addressing complex traits by selecting markers which have been shown to increase the trait of interest (Johnson, 2004).

2.7.5.2 Whole genome prediction.

Whole genome prediction (WGP) refers to the supplementation of a phenotypically characterized "training" population with molecular information to predict the performance of other genotypes (Riedelsheimer et al., 2012), giving them genomic estimated breeding values (GEBVs). This method has been highly successful in cattle breeding, where it was demonstrated that selection

decisions could be successfully made with just molecular data (Hayes et al., 2009). Where MAS struggles to find markers to attribute to polygenic traits, WGP does not test for marker significance, but takes all markers together to predict performance (Lorenzana et al. 2009). For polygenic traits in cattle, selection response using WGP has been up to 43 % more effective than with traditional linear regression models (Meuwissen et al., 2001). Promising results such as these have led to increased interest in WGP from the plant breeding industry (Reidelsheimer et al., 2012). Lorenzana et al. (2009) tested one maize (*Zea mays* L.) population, one arabidopsis population (*A. thaliana*) and two barley (*Hordeum vulgare* L.) populations for the efficacy of using WGP vs. MAS to predict polygenic traits in crop species. They found that while MAS managed to predict the highest performing individuals, its results were highly variable, producing a significant amount of inconsistency in predictive power. On the other hand, they found WGP was more effective overall, and its results did not significantly differ from genotype to genotype. In 2012, Reidelsheimer et al. (2012) tested the ability of WGP to create a prediction model for complex traits in 285 maize genotypes. Using 56,110 SNPs, a predictive model was created which was found to be 72-81 % effective. Results such as these support the idea of GWP as an important tool for the effective use of molecular data.

2.8 Objectives

The objectives of this study were to evaluate and organize 321 diverse genotypes of *B. napus* into heterotic groups using morphological, agronomic, quality and molecular information. This was accomplished in three projects: 1) the use of agronomic and quality traits to cluster genotypes into heterotic groups, 2) the use of SRAP and GBS molecular methods to cluster genotypes into heterotic groups, and 3) the evaluation of the clusters from projects 1 and 2 by using hybrid performance of each genotype with the same male-sterile female parent. This research will aid in

the selection of parental material for future hybrids in the University of Manitoba *B. napus* breeding program.

3.0 Characterization of *Brassica napus* Genotypes and Assignment to Phenotypic Clusters

3.1 Abstract

The morphological and agronomic diversity of 321 High Erucic Acid Rapeseed (HEAR) (*Brassica napus* L.) genotypes were evaluated for phenotypic characteristics with the goal of grouping the parents into heterotic clusters for hybrid rapeseed development. The evaluated traits included: days to flowering (FLR), days to maturity (MAT), plant height (HT), lodging (LOD), oil content (OIL), meal protein content (PROM), glucosinolate (GLUC) content, erucic acid content (ER) and saturated fatty acid content (SAT). Data for the above traits were collected during the 2013 and 2014 growing seasons near Winnipeg, Manitoba, Canada. Agronomic and quality data was subjected to principle component and hierarchical cluster analysis using the two most commonly used clustering algorithms for morphological data (Ward's Method and Unweighted Pair Group Method using Arithmetic means). Two principle components (PC1, PC2) were found to explain 44.7 % of the variation in the population, and were used to sort genotypes prior to cluster analysis. Both clustering methods sorted the 321 genotypes into four distinct groups of clusters.

3.2 Introduction

Heterosis is a natural process characterized by superior agronomic performance of hybrid offspring over their parents; a process which is exploited for great effect in many important crop species (Ryder et al., 2014). Heterotic improvements to yield, quality and disease resistance have been attained across many crops including maize (*Zea mays* L.) (Duvick, 2005), sorghum (*Sorghum bicolor* L.) (Chapman et al. 2000), tomato (*Solanum lycopersicum* L.) (Cuartero et al. 2005), rice (*Oryza sativa* L.) (Dan et al., 2016), cotton (*Gossypium hirsutum* L.) (Zhang et al., 2016) and sunflower (*Helianthus annuus* L.) (Rieseberg, 2000). These successes have attracted the interest of canola/rapeseed (*Brassica napus* L.) breeders hoping to replicate this progress in their own

breeding programs. Since the 1989 release of Hyola-40, the first *B. napus* hybrid cultivar, (Kneen, 1992); hybrid breeding in rapeseed has been largely successful, with the average high- parent heterosis in spring rapeseed estimated at 30 % (McVetty and Scarth, 2012). Hybrid canola/rapeseed accounts for 75 % of total rapeseed production in China (Fu et al., 2014) and 83 % of the rapeseed grown in France (Ruster, 2015). In Canada, hybrid acreages have grown exponentially since their introduction: accounting for 15 % in 2000, 35 % by 2003, 50 % by 2004 (Barnes, 2007) and by 2010 hybrid varieties made up the greatest acreages in nearly every major canola/rapeseed producing area worldwide (Basunanda et al., 2010).

The increased demand for hybrid cultivars means an increase in the demand for new or improved potential parents for these hybrids. However, parental selection can create a logistical bottleneck, as the number of potential parental combinations may take extensive time and resources to evaluate (Riaz et al., 2001). To address this bottlenecking of time and resources, heterotic groups can be applied to great effect, clustering genotypes into groups which display similar degrees of heterosis in crosses with individuals from other groups (Ryder et al., 2014).

First used in maize, heterotic grouping was used to separate germplasm into stiff stalk (SS) vs. non-stiff stalk groupings which were extremely successful in improving agronomic performance (Tracy and Chandler, 2006). Similarly, in Chinese three-line hybrid rice, two major heterotic groups (early season from China, and late season from Southeast Asia) were identified during the 1970s, which facilitated superior parent selection and increased yield (Xie et al., 2014). This work formed the foundation of hybrid rice breeding in China and sparked the identification of *Indica spp.* heterotic groups that continues today.

The success of different parental combinations can be highly variable, depending on the combining ability of the parents involved (Bechere et al., 2016). Combining ability being the breeding value

of a potential parent to create high performing hybrids (Cox and Frey, 1984). As such, methods for predicting heterosis are critical to a breeding program, and there is considerable evidence that heterosis is proportional to the genetic distance between the parents of the hybrid (Ali et al., 1995; Qian et al., 2007; Girke et al., 2012; Xing et al., 2014). Genetic distance can be defined by agronomic, morphological, pedigree, or genotypic traits; or a combination thereof (Abdel-Ghani and Lübberstedt, 2013). Once trait data is collected, statistical methods such as Euclidean distance (Goodman, 1972) and Principle Component Analysis can then be employed to analyze trait data and statistically predict hybrid performance (Ali et al., 1995).

Phenotypic traits are often used as the backbone of any germplasm evaluation; the word "phenotypic" describing any readily observable physiological, morphological, agronomic, or quality characteristic of an individual (Violle et al 2014). The phenotypic analysis of breeding populations can be complicated and time consuming, limiting its utility for heterotic group identification (Langer et al 2014). Traits identified using molecular markers instead of phenotypic criteria are often used to reduce the overall ambiguity of phenotypic data (Riaz et al., 2001). However, not all observed genetic variation is related to agronomic traits of interest, which can cause breeders to keep using the same heterotic loci again and again; plateauing or even reducing heterotic benefit (Phillips, 2010). This has been observed in maize and rice; where heterotic yield increases for rice averaged 3.1 % in the 80s, 1.4 % in the 90s and 0.8 % in the 2000s (Phillips, 2010). Conversely, a 2009 study by Troyer and Wellin (2009) showed that in many crop species, focusing on quality characteristics in breeding self-pollinating genotypes has increased yield by 1.9 - 3.5 over their hybrid cousins. Results such as these suggest that there could be significant gains made by establishing heterotic groups based on overall differences in agronomic or quality characteristics.

Several statistical procedures have been used to cluster agronomic and quality data in the past, though the unweighted pair group (UPGMA) and Ward's method are the two most widely used (Khodadadi et al., 2005; Teklewold, 2006; Fahid, 2014; Singh, 2015). Ward's method is increasingly cited as ideal, due to it not being subject to the "chaining effects" observed in UPGMA analysis, which results in higher internal affinity within clusters than is actually present (Lombard et al, 2000; Mohammadi and Prasanna, 2003; Khodadadi et al., 2005). It has been reported that although Ward's method may generate fewer clusters than the UPGMA method, the groups it does generate are far more homogenous, creating groups with defined characteristics (Padilla, 2007).

The objective of this study was to use two multivariate statistical methods (Ward's method and UPGMA) to separate: 318 inbred *B. napus* genotypes, one male-sterile female line and the two parents of the male-sterile female line into heterotic groups based on nine agronomic and quality traits collected over two field years.

3.3 Materials and Methods

3.3.1 Genetic Material

The *B. napus* genotypes selected for this study were drawn from 86 different pedigrees from the University of Manitoba's canola/rapeseed research program. Genotype names and designations are found in Appendix A. Although some genotypes may have the same pedigree, each genotype is unique. For evaluation and heterotic grouping, 303 diverse inbred HEAR genotypes were used in 2013, with 15 more were added in the 2014. These genotypes can potentially serve as pollen donors (pollinators) in hybrid crosses using the Male-Sterile Lembke (MSL) pollination control system. The male-sterile "female" (UM388) and its two parents (UM386, UM387) were obtained contractually between the University of Manitoba and DL Seeds (parent company, NPZ Lembke).

With the 318 genotypes from the University of Manitoba, as well as UM386, UM387, UM388, there were a total of 321 *B. napus* genotypes.

3.3.2 Field Evaluation

3.3.2.1 Field Management

In 2013, genotypes were evaluated in nursery rows at the University of Manitoba's Glenlea Research Station (49.648123, -97.13068), which is located on top of fine, imperfectly drained, Scanterbury and Red River clay soil. In October, 2012, in preparation for the 2013 growing season, a fertilizer was broadcast applied ((46-0-0, 43%), (11-52-0, 31%), (0-0-60, 7%), (20-0-0-24, 19%)) equivalent to ~137 kg/ha nitrogen. Alongside fertilizing in October 2012, an application of granular Edge™ herbicide (Dow Agrosciences, Indianapolis, IN, USA) was applied to the field at a rate of 30.9 kg per hectare. Fertilizer and herbicide were worked into the field twice with a tandem disk (Case International, Racine, WI, USA). Each genotype was grown in three-meter nursery rows, seeded at a rate of 1 g/row at a depth of 1.3 – 1.9 cm using a Nursery Master Classic Hege 3-point hitch double disc seeder from Wintersteiger (Salt Lake City, UT, USA) during the last week of May 2013. Corn grit treated with Helix™ insecticide (Syngenta Basel, Switzrland**)** was incorporated along with the seed for flea beetle control (*Phyllotreta cruciferae* G., *Phyllotreta striolata* F.). Nursery rows were treated with a mixture of the Group 1 herbicides in early June 2013: Poast® at 0.67 L/ha (BASF, Ludwigshafen, Germany), Lontrel® at 0.84 L/ha (Dow Agroscience, Midland, MI, USA); and the Group 2 herbicide Muster® (DuPont, Wilmington, DE, USA) at 29.6 g/ha in 1000 L/ha.

When nursery rows reached stage 66 (20% flowering, BBCH Scale as per Lancashire et al., 1991), the group 2 fungicide Rovral™ (Bayer Cropscience, Leverkusen, Germany) was applied at 62 ml/100 m^2 to manage sclerotinia stem rot (*Sclerotinia sclerotiorum* Lib. de Bary) infection. When

a given nursery row reached the growth stage of 97 (BBCH scale, Lancashire et al., 2001), they were straight cut using a Wintersteiger Nursery Master Elite plot combine (Salt Lake City, UT, USA).

In 2014, genotype evaluation occurred at the University of Manitoba's Bison Research Field (49.801217, -97.167434), on an orthic black soil. Field management was conducted using the same protocol as in 2013, however on June 10th, 2014 the Group 3 insecticide Decis® was applied at 148 ml/ha to control flea beetles.

3.3.2.2 Agronomic Characteristics

Each genotype was evaluated for nine agronomic and seed quality traits. Days to flowering (FLR) for any given genotype was measured as the number of days from seeding to flowering of at least 50 % of the individuals within a nursery row. Flowering refers to stage 60 on the BBCH scale when first flowers open (Lanchashire et al., 1991). Plant height (HT) was measured at early ripening, or stage 80. The height (cm) above the ground a group of plants extended was measured from the soil to the tip of the axial bud. Several groups of plants within a row were measured and heights were averaged to record a representative plant height for each genotype. Lodging (LOD) was measured according to the protocol of the Canola Performance Trials (CPT) on a 1-5 scale as the amount the individuals within a nursery row deviated from the vertical plane (Canola Council of Canada, 2016). Lodging score was measured at the ripening stage 80 (BBCH scale) once virtually all pods had reached their final size (Lancashire et al., 1991). For this study, the lodging scale was defined as: 1 = individuals making up row were completely vertical, 2 = slight lodging in all individuals, 3 = moderate lodging of all individuals, 4 = most individuals highly lodged, and 5 = all individuals excessively horizontal. Days to maturity (MAT) was determined at the end of the ripening stage (88-89 BBCH scale) of development and defined as the number of days between

the seeding date and when over 50 % of the plants were considered mature (Lancashire et al., 1991). Pod maturity occurs at stage 89 (BBCH scale), when individually sampled pods contain uniformly dark seeds and are considered "fully ripe" (Lancashire et al., 1991). Once plants senesced (97 BBCH scale), seed could be harvested and yield data was recorded in g per row for each genotype and converted to kg/ha.

3.3.2.2 Quality Characteristics

Seed quality characteristics were determined at the University of Manitoba, using Near Infrared Reflectance (NIR) for oil, protein, and glucosinolate content (Tkachuk, 1981) as per protocol laid out by Daun et al. (1994). NIR was conducted by running a 5 g bulk seed sample from each genotype through a Foss 6500 NIR system (Foss NIRSystems Inc., MD, USA) (Burns and Ciurzak, 2008). Oil content was determined using NIR adjusted to 0 % moisture, and expressed as a percentage of total collected dried seed mass. Glucosinolates were measured using NIR adjusted to the industry standard 8.5 % moisture and expressed as µmol/g of seed. Meal protein content was determined using NIR adjusted to 0 % moisture, and expressed as a percentage of total collected dried seed mass.

Gas-Chromatography was used to determine erucic acid and saturated fatty acid content by measuring the methyl ester production during base transesterification from a 3 g bulk seed sample. A Reference Standard 421 from Nu Check Prep (Elysian, MN, USA) was used to ensure proper GC readings. Saturated fatty acid levels were expressed as a percentage of the total dried seed mass (Hougen and Bodo, 1973) and measured using a Varian 3900 CP Wax 52 CB fitted with a capillary column and flame ionization detector (Hougen and Bodo, 1973; Liu, 1994; Kim et al. 2007**)**.

3.3.3 Statistical Analysis

Data from the agronomic and quality evaluation of each genotype in 2013 and 2014 were analyzed in isolation and together. When each growing season was analyzed separately, coefficient of variation (CV) was calculated as a relative measure of the populations variation for a given trait by dividing standard deviation by population mean; resulting values < 0.10 considered "low", 0.10 - 0.20 considered "medium", and values > 0.20 considered high (Deshmukh et al.,1986). These values for CV reflect trait variation across the whole population within each growing season. Traits for each growing season were examined in isolation, and clustered using the "dendextend" module within R-Studio 0.99.893 (R-Studio Inc., Boston, MA, USA), a program that utilizes the Rx64 3.2.4 (© The R Foundation for Statistical Computing) open sourced statistical software. Hierarchical clustering was performed in R-Studio according to Ward's method. Generated dendrograms for 2013 and 2014 were then compared using "dendextend" for data point merging, height, and order to determine mean-relative similarity. Mean relative similarity was then subjected to a statistical t-test to determine if similarity was statistically significant in order to combine the data from both years. If data was combined, analysis of variance was conducted on each of the nine agronomic and quality traits to highlight variation using PROC-GLM in SAS 9.3 software (SAS Institute, Cary, NC, USA). For statistical analysis, each morphological or quality trait was treated as a dependent variable and genotypes were treated as independent variables. If trait values for 2013 and 2014 were able to be combined; CV between the two year would reflect the degree to which a given genotype fluctuated from year to year; values < 0.10 were considered "low", values between 0.10 and 0.20 were considered "medium", and values > 0.20 were considered high (Deshmukh et al.,1986). Least significant distance (LSD) was calculated for each

trait using Tukey's test in SAS. Combined agronomic and quality data results were imported into JMP® 8 (SAS Institute Inc., Cary, NC, USA) for principle component analysis.

3.3.3.1 Principle Component Analysis

A principle component analysis (PCA) was run on the combined data from both growing seasons. This involved an orthogonal transformation of the data into a coordinate system of uncorrelated linear components so that the first coordinate explains the greatest amount of variance (principle component one or PC1), and the k^{th} greatest variance on the k^{th} coordinate (principle component *k*, or PC*k*) as per the procedure first described by Karl Pearson (1901). PCA was conducted in JMP® according to the protocol described by Jones and Sall (2011). This converted the numerical data set into principle components which explained the largest amount of the observed variation; principle component one (PC1) explaining the most variability, and each successive component explaining the remainder (Joliffe, 2002). The PC1 and PC2 value for each genotype was recorded and correlations between PC1 and PC2 for evaluated traits were calculated and tabulated. Significance of correlation values were examined using a student's t-test, with n = 321, df = 319, and $\alpha = 0.01$.

3.3.3.2 Hierarchical Clustering

All 321 genotypes were hierarchically clustered in R-Studio using Ward's Minimum Variance Method (Ward, 1963), weighing each phenotypic trait equally. This method treats each genotype as a separate cluster and uses the ANOVA (analysis of variance) sum of squares ($SSE=\sum^{(i=1)}_{n}(\chi^{i}-\bar{\chi})^{2}$) between clusters added over all variables to combine the genotypes into groups (Milligan and Cooper, 1988). At each step of the analysis, the program combines the genotypes from the previous step in such a way as to minimize the total sum of squares across all observations (SAS Institute Inc., 2014). The program uses this and the Euclidean-distance method to determine distance

between clusters. Prior to clustering, genotypes were ordered via scores for PC1 from the principle component analysis. After ordering, hierarchical cluster analysis was performed using Ward's Minimum Variance Method, following the equation $D_{KL}= (\|\bar{\chi}_K-\bar{\chi}^2\|^2)/((1/N^K - 1/N^L))$, where n = the number of observations, v = the number of variables, x_i = the value of the i^{th} observation, C_k = k^{th} cluster (subset of {1,2,…,n}), N_k = the number of observations in C_k, x = the sample mean vector, x_k = the mean vector for cluster C_k and $\|x\|$ = the square room of the sum of squares of the elements of x (Ward, 1963). Cluster distance was also determined using the unweighted pair group method with arithmetic mean (UPGMA) method as outlined by Sokal and Michener (1958).

For both Ward's and the UPGMA method, the number of clusters was determined by cubic clustering criterion outlined by the SAS Institute in Tech Report A-108 (1983) and Milligan and Cooper (1985). Using this method, the within cluster sum of squares is statistically tested using an ascending number of groups (eg. 1, 2, …, n). The ideal group number is determined to be the point at which sum of squares is at its maximum and then starts decreasing when you add more groups. Once set at this number, clusters can be coloured for easier visualization.

Once subjected to hierarchical clustering, genotypes were sorted by cluster membership, and mean trait values for each group were determined. Trait values were then averaged for all genotypes within clusters and subjected to another hierarchical analysis using Ward's Minimum Variance Method. When hierarchical clusters were generated using both Ward's Method and UPGMA, they were compared for near equality in R-Studio for dendrogram height, genotype merging and order to test if both methods produced statistically similar results **(**Chambers, 1998**)**.

3.4 Results

3.4.1 Agronomic and Quality Data

In 2013, genotypes had a mean of 45.4 (SD 2.6) days required for flowering, with a minimum of 39 (UM20) and a maximum of 50 days (UM364) (Table 3.1). In 2014, the mean days to flowering was 45.3 days (SD = 3.1), with UM256 requiring the fewest days at 39, and UM52 requiring the most at 53 days. Individually, for both 2013 and 2014, genotypes were not overly variable for number of days to flowering, producing CV values of 0.06 and 0.07 around the respective means for each year. When both years were taken together, genotypes averaged 45.4 days (SD = 1.62) required until flowering. A CV of 0.04 was calculated between the two growing seasons, meaning that for every genotype, there was a "low" degree of variation between data from both years.

In 2013, genotypes had an average plant height of 99.4 cm (SD = 8.9), with the shortest genotype being 80.0 cm (UM229), and the tallest genotype being 127.5 cm (UM392). In 2014 genotypes had a mean height of 99.3 cm (SD = 8.8 cm), with the shortest being UM120 at 75.0 cm, and UM333 being the tallest at 135.0 cm. When each growing season was analyzed individually, a CV = 0.09 was calculated. When data from both growing seasons was combined, genotypes had a mean of 99.3 cm in height (SD = 7.4 cm) with a CV of 0.05, meaning that there was a "low" degree of variation between height values for the same genotype over both growing seasons.

In 2013, genotypes had a mean lodging rating of 2.0 (SD = 0.9). Due to the nature of the trait, a distinct minimum and maximum was not available; although 103 genotypes received a classification of "1", indicating barely any lodging, while 17 received a rating of "4" indicating heavy lodging. The 2014 growing season yielded similar results, in which 95 genotypes were rated "1" and 15 genotypes were rated "4" with a mean value of 2.0 (SD = 0.8). When each growing

season was analyzed individually, each year yielded a CV between 0.40 and 0.45; indicating a "high" degree of variation around each year's respective mean for the lodging trait. When data from both growing seasons was combined, genotypes had a mean rating of 2 (SD = 0.6) and a CV of 0.30, meaning that there was a "high" degree of variation between lodging values for the same genotype over both growing seasons.

In 2013, genotypes required a mean of 100.2 days (SD = 2.5) to reach maturity; with UM392 maturing the fastest at 93 days, and UM186 maturing the slowest at 107 days. In 2014, genotypes required a mean of 102.9 days to reach maturity (SD = 2.2); with UM280 taking the shortest amount of time at 97.0 days, and UM195 taking the longest at 108.0 days. When both growing seasons were evaluated individually, each year yielded a CV of 0.02, indicating a "low" degree of variation around each year's respective mean. When data from both seasons was combined, genotypes required a mean of 101.6 days to maturity (SD = 2.1) with a CV of 0.02 indicating a "low" degree of variation between the values for the same genotype over both growing seasons.

In 2013, genotypes had a mean oil content of 45.6 % (SD = 3.1 %), with UM106 displaying the lowest oil content at 36.1 % and UM16 displaying the highest oil content at 52.7 %. Genotypes in 2014 averaged 46.5 %, (SD = 2.0 %), with UM386 having the lowest at 38.1 % and UM125 having the highest at 51.1 %. Individually, a CV (relative to the mean of each year) of 0.07 was calculated for 2013, and 0.05 for 2014 indicating a "low" degree of variation for oil content in both growing seasons. When the data from both growing seasons was combined, genotypes had a mean oil content of 46.1 % (SD = 1.7 %), with a CV of 0.03, meaning there was a "low" degree of variation between values for the same genotype over both growing seasons.

Table 3.1. Summary statistics for four agronomic traits measured for select *Brassica napus* L. genotypes in 2013[†] and 2014[‡].

		FLR[¶] (days)	HT[¶] (cm)	LOD[¶] (1-5)	MAT[¶] (days)
2013	**Mean**	45.4	99.4	2	100.1
	Min	39	75	1	93
	Max	50	135	4	107
	SD	2.6	8.9	0.9	2.5
	CV	0.06	0.09	0.45	0.02
2014	**Mean**	45.3	99.3	2	102.9
	Min	39	75	1	97
	Max	53	135	4	108
	SD	3.1	8.8	0.8	2.2
	CV	0.07	0.09	0.40	0.02
Combined	**Mean**	45.4	99.3	2	101.6
	SD	1.6	7.4	0.6	2.1
	CV	0.07	0.09	0.40	0.02
	LSD	4.9	15.5	1.3	4

[†]2013, n = 306, Glenlea Research Station, Manitoba, Canada
[‡] 2014, n = 321, Bison Research Field, Winnipeg, Manitoba, Canada.
[¶]FLR, number of days to first flowering; HT, plant height; LOD, lodging; MAT, days to maturity.
[§] SD, standard deviation; CV, population coefficient of variation (SD/mean).

In 2013, genotypes produced a mean of 48.2 % meal protein content (SD = 2.2), with UM67 producing the lowest protein content at 40.7 % and UM129 producing the highest at 53.6 %. Genotypes in 2014 yielded an average 50.2 % (SD = 1.7 %) for protein content, with UM67 yielding the lowest at 44.5 %, and UM106 yielding the highest at 54.4 %. When each growing season was evaluated individually, a CV of 0.05 was calculated for 2013, and a CV of 0.03 was calculated for 2014, indicating a "low" degree of variation around the mean for protein content in both 2013 and 2014. When growing seasons were taken together, genotypes averaged 49.2 % (SD = 1.1), with a CV of 0.02, indicating there was a "low" degree of variation between trait values for the same genotype over both growing seasons.

In 2013, genotypes had a mean of 12.6 µmol/g for glucosinolate content (SD = 5 µmol/g), with UM112 yielding the lowest glucosinolates at 1.2 µmol/g, and UM386 yielding the highest at 57.2 µmol/g. The male-sterile UM388, which uses UM386 as part of its parental material, also yielded numerically higher than the average glucosinolates of 26.1 µmol/g. In 2014, genotypes had a mean of 14.4 µmol/g (SD = 4.6) of glucosinolates, with UM233 yielding the lowest at 1 µmol/g, and UM386 yielding the highest at 57.2 µmol/g. When both growing seasons were evaluated individually, a CV of 0.40 was calculated for 2013, and a CV of 0.52 was calculated for 2014, indicating a "high" degree of variation around the mean in both 2013 and 2014. When both growing seasons were evaluated together, genotypes produced a mean of 13.6 µmol/g (SD = 2.0 µmol/g). A "moderate" CV of 0.15 was calculated for variation between traits over both growing seasons.

In 2013, genotypes had a mean of 5.2 % saturated fatty acid content, (SD = 0.5 %), with UM251 showing the lowest at 4.2 %, and UM362 showing the highest at 7.4 %. In 2014, genotypes averaged 5.2 % (SD = 0.4) with UM105 yielding the lowest at 3.2 %, and UM362 yielding the highest at 6.3 %. When both growing seasons were evaluated individually, a CV of 0.09 was calculated for 2013, and a CV of 0.03 was calculated for 2014, indicating a "low" degree of variation for saturated fatty acid content compared to the respective means of 2013 and 2014. When both growing seasons were evaluated together, genotypes showed a mean of 5.2 % (SD = 0.2 %) saturated fatty acid content, with a CV of 0.04, indicating a "low" amount of variation in glucosinolate trait values over both growing seasons.

In 2013, genotypes produced a mean of 47.6 % erucic acid (SD = 5.9 %), with UM387 yielding the lowest at 0 %, and UM251 yielding the highest at 59.3 %. In 2014, genotypes had a mean of 49.4 % (SD = 5.6 %) with UM387 yielding the lowest at 0 %, and both UM252 and UM105

yielding the highest at 59.5 %. When evaluated separately, a CV of 0.12 was calculated for 2013 and a CV of 0.11 was calculated for 2014, indicating a "medium" degree of variation for erucic acid content around the means of each respective growing season. When 2013 and 2014 were evaluated together, genotypes averaged 45.4 % (SD = 1.7 %) with a CV of 0.04.

Table 3.2. Summary statistics for five quality traits measured for select *Brassica napus* L. genotypes in 2013[†] and 2014[‡].

		OIL (%)	PROM (%)	GLUC (µmol/g)	SAT (%)	ER (%)
2013	**Mean**	45.6	48.2	12.6	5.2	47.6
	Min	36.1	40.7	1.2	4.2	0
	Max	52.7	53.6	57.2	7.4	59.3
	SD	3.1	2.2	5	0.47	5.9
	CV	0.07	0.05	0.40	0.09	0.12
2014	**Mean**	46.5	50.2	14.4	5.2	49.4
	Min	38.1	44.5	1	3.2	0
	Max	51.1	54	57.2	6.3	59.5
	SD	2	1.7	4.6	0.4	5.6
	CV	0.04	0.03	0.32	0.02	0.11
Combined	**Mean**	46.1	49	13.6	5.2	45.3
	SD	1.7	1.1	2	0.2	1.7
	CV	0.03	0.02	0.15	0.04	0.04
	LSD	3.5	2.5	8.6	0.7	7.8

[†]2013, n = 306, Glenlea Research Station, Manitoba, Canada
[‡] 2014, n = 321, Bison Research Field, Winnipeg, Manitoba, Canada.
[¶] OIL, oil content; PROM, meal protein content; GLUC, glucosinolates; SAT, saturated fatty acids; ER, erucic acid.
[§] SD, standard deviation; CV, population coefficient of variation (SD/mean).

3.4.2 Principle Component Analysis of Agronomic and Quality Data

Two principle components were identified which explained 44.7 % of observed variation. PC1 explaining 29.4 %, and PC2 explaining 15.3 %. When plotted against each other, PC1 and PC2 provided a snapshot of traits that were correlated with each other (Figure 3-1; Table 3-4). Positive correlations were found between days to flowering and days to maturity ($r = 0.5$, $P < 0.01$), lodging and days to flowering ($r = 0.2$, $P > 0.01$), glucosinolates and days to flowering, ($r = 0.2$, $P > 0.01$),

saturated fatty acid content and days to flowering, ($r = 0.3$, $P > 0.01$), glucosinolates and days to maturity ($r = 0.3$, $P > 0.01$), erucic acid and oil content ($r = 0.2$, $P > 0.01$), days to maturity and saturated fatty acid content ($r = 0.3$, $P > 0.01$), and meal protein content and glucosinolates ($r = 0.2$, $P > 0.01$). Significant negative correlations were found between days to flowering and oil content ($r = -0.4$, $P > 0.01$), days to flowering and meal protein ($r = -0.2$), days to flowering and erucic acid content ($r = -0.3$, $P > 0.01$), lodging and oil content ($r = -0.3$), lodging and erucic acid ($r = -0.2$), days to maturity and erucic acid ($r = -0.2$, $P > 0.01$), oil content and days to maturity ($r = -0.4$, $P > 0.01$), glucosinolates and oil content ($r = -0.5$, $P > 0.01$), and oil content and saturated fatty acids ($r = -0.4$. $P > 0.01$).

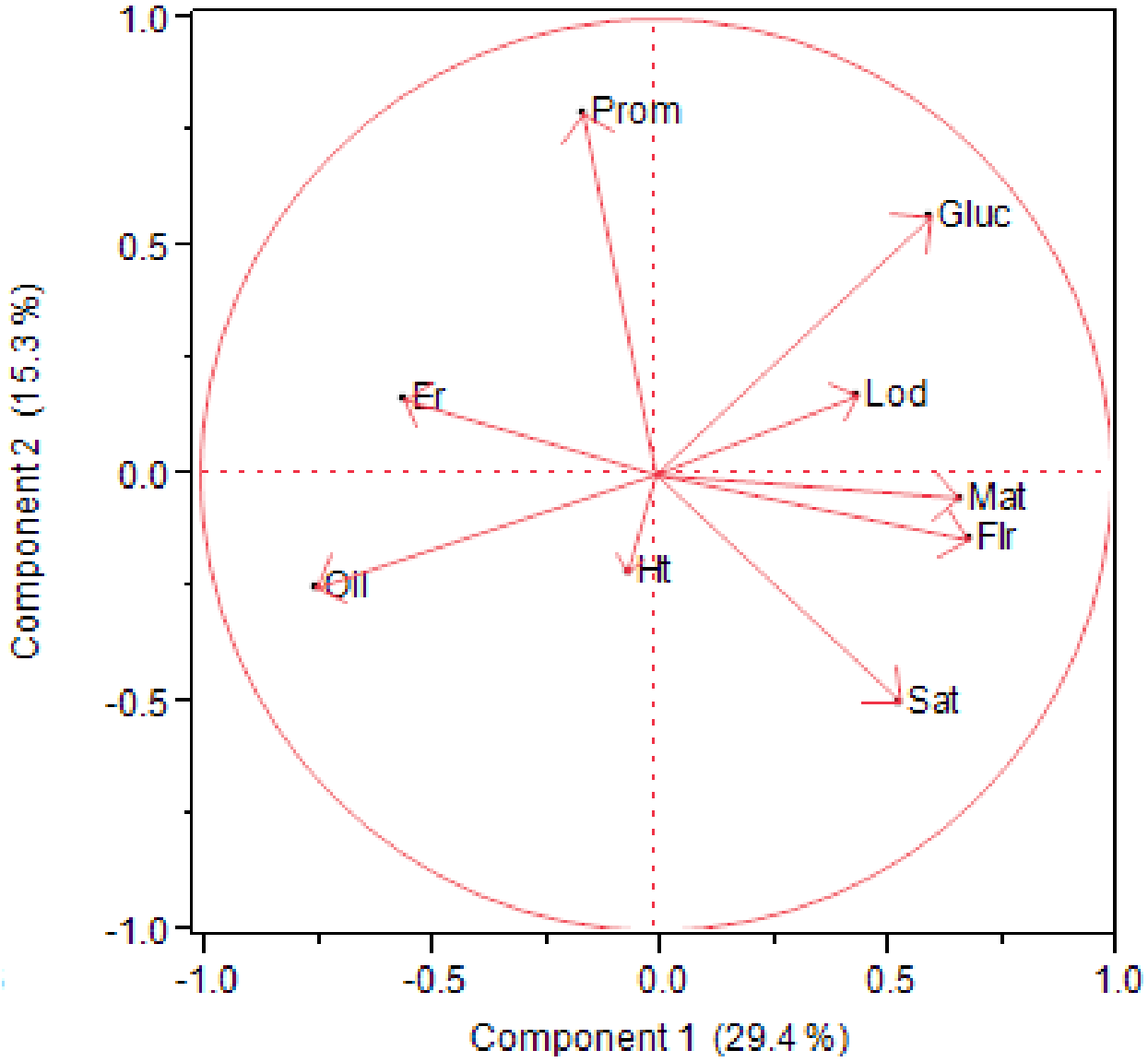

Figure 3.1. Principle component analysis from clustering 321 *Brassica napus* L. genotypes based on euclidian distance derived from nine traits including: glucosinolates (Gluc), lodging (Lod), days to maturity (Mat), days to flowering (Flr), saturated fatty acids (Sat), plant height (Ht), oil content (Oil), erucic acid (Er), and meal protein (Prom).

Table 3.3 Correlation matrix between nine traits for 321 *Brassica napus* L. genotypes collected over the 2013† and 2014† field seasons.

	FLR‡	HT	LOD	MAT	OIL	PROM	GLUC	SAT	ER
FLR	1.0	-	-	-	-	-	-	-	-
HT	-0.1*	1.0	-	-	-	-	-	-	-
LOD	0.2***	0.1*	1	-	-	-	-	-	-
MAT	0.5***	0.0	0.2***	1	-	-	-	-	-
OIL	-0.4***	0.1*	-0.3***	-0.4***	1	-	-	-	-
PROM	-0.2***	-0.1*	0.0	-0.1*	0.0	1	-	-	-
GLUC	0.2***	0.0	0.2***	0.3***	-0.5***	0.2***	1	-	-
SAT	0.3***	0.0	0.0	0.3***	-0.4***	-0.3***	0.0	-	-
ER	-0.3***	0.0	-0.2***	-0.1**	0.2***	0.1*	-0.3***	-0.3***	1

* significant at the 0.05 probability level.
** significant at the 0.01 probability level.
*** significant at the 0.001 probability level.
†2013, Glenlea, Manitoba, Canada; 2014, Winnipeg, Manitoba, Canada.
‡FLR, number of days to first flowering; HT, plant height; LOD, lodging; OIL, oil content; PROM, meal protein; GLUC, glucosinolates; SAT, saturated fatty acids; ER, erucic acid.

3.4.3 Hierarchical Clustering based on Agronomic and Seed Quality Data

3.4.3.1 Clustering using Ward's Method

Using Ward's Minimum Variance Method, the 321 *B. napus* genotypes were organized into four clusters, which are referred to as, Group A, Group B, Group C, and Group D (Figure 3.5). Cubic clustering criterion reached a high when genotypes were organized into four clusters and decreased as the number of clusters was increased; indicating four to be the ideal number of clusters.

Group sizes ranged from 38 genotypes in Group A, 122 genotypes in Group B, 5 genotypes in Group C, and 141 genotypes in Group D. Group A was the most distinct, as it split from Group(s) B, C and D at less than 140 Euclidean distance units. Group B split off from C and D at a height of ~105 distance units from the base and C and D separated from each other at ~60 distance units from the base. Group B and D displayed the most intra-cluster diversity, which is indicated by the

number of sub-clades that make up the clusters within these groups. Group C had the least intra-group diversity and was made up of only a small subset of one clade. The male-sterile female line UM388 was located in Group C, and its parents UM386 and UM387 were located in groups C and B, respectively. It is worth noting that UM388 and UM386 plotted closer together than UM388 and UM387, as UM388 and UM386 were more similar for observed agronomic and quality traits. Group C was made up of five genotypes all characterized by significantly lower erucic acid levels ($p < 0.0001$) than the population mean; managing an intragroup average of 17.6 %, versus the population average of 48.5 %.

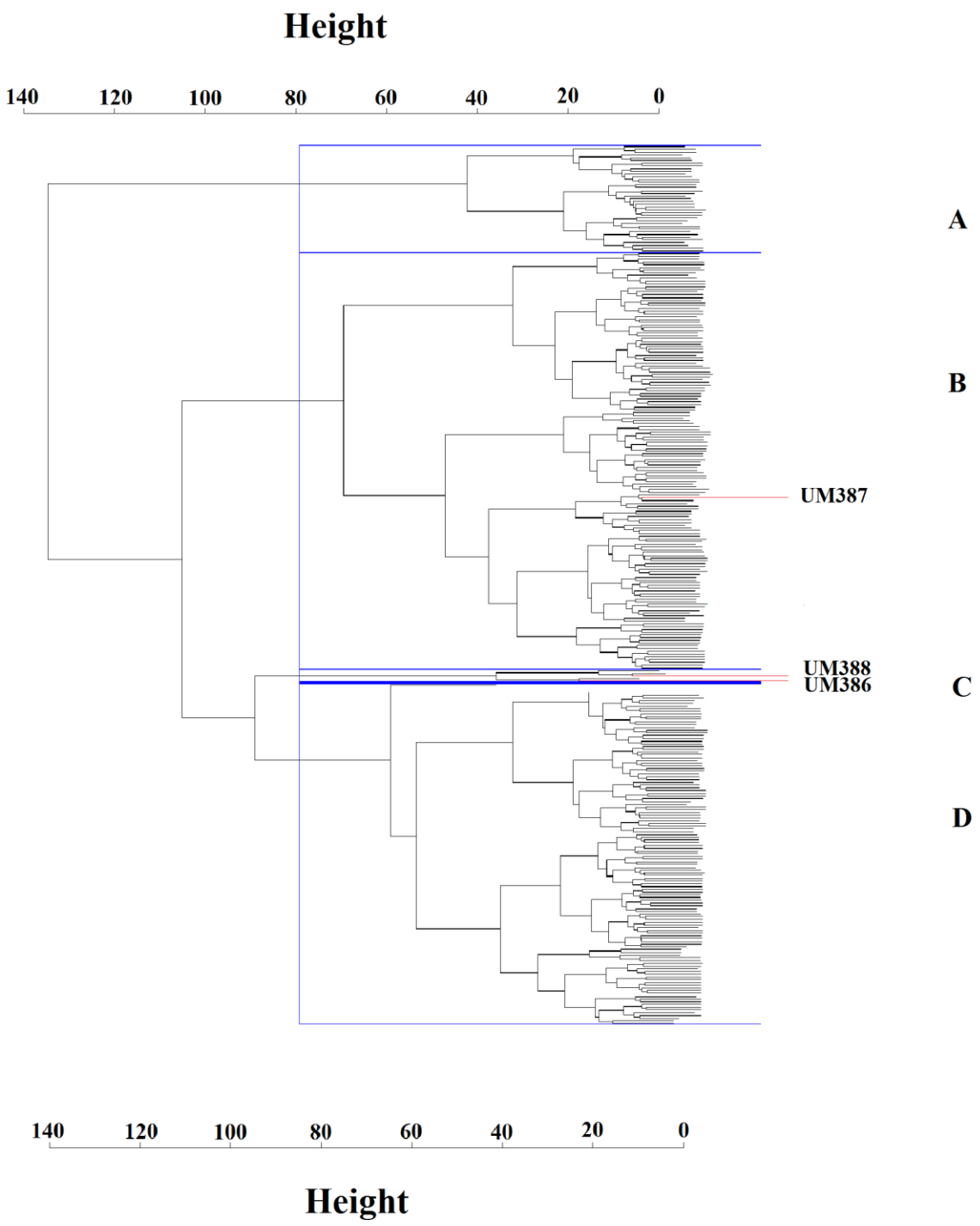

Figure 3.2. Hierarchical cluster analysis of 321 *Brassica napus* L. genotypes (*x*-axis) based on phenotypic data collected over in 2013 in Glenlea, Manitoba, and 2014 in Winnipeg, Manitoba. Clustering was conducted using Ward's Minimum Variance Method. Height (*y*-axis) refers to Euclidean distance units.

3.4.3.2 Clustering using UPGMA method

Using the UPGMA method, the 321 genotypes were also clustered into four groups (A, B, C, D) based on cubic clustering criterion (Figure 3.6). UPGMA created a bottom-heavy set of groups made up of: Group A with 2 genotypes, Group B with 1 genotype, Group C with 3 genotypes, and Group D with the remaining 315 genotypes. The male-sterile line UM388 was located in the large Group D along with 314 other genotypes. However, the parental material for UM388 was located outside Group D, with UM386 making up the monoecious group B, and UM387 being one of the two members of Group A, alongside UM34. These results are unsurprising due to the UPGMA method's sensitivity to outliers; UM386 had a glucosinolate content of 57.2 µmol/g, significantly higher than the overall mean of 13.52 µmol/g ($p < .0001$), and UM387 had a erucic acid content of 0 %, which was significantly ($p < .0001$) smaller than the overall average of 48.49 %. This would be enough to justify UPGMA placing it in its own group. Similarly, UM387 with a erucic acid content of 0 % and UM34, with a erucic acid content of 8.25 % both had erucic acid values significantly lower than the overall mean of 48.9 % ($p < 0.01$).

3.4.3.3 Comparison of Ward's and UPGMA Clusters

When UPGMA and Ward's method were compared, they had a mean relative similarity of 32 % for branch merging, 62 % for branch height, and 26 % for order. Consistency between values for merging ($t = 5.9 \geq 3.3$), height ($t = 13.8 \geq 3.3$) and order ($t = 4.7 \geq 3.3$) were determined to be significant ($P = 0.001$) when comparing UPGMA clusters and Ward's clusters. However, number of genotypes in each group was inconsistent between clustering algorithms ($t = (< 0.1) \leq 5.98$, $\alpha = .05$). Due to the of the unbalanced differences in group size of the data set identified by the UPGMA method, only the clusters generated by Ward's method were analyzed further.

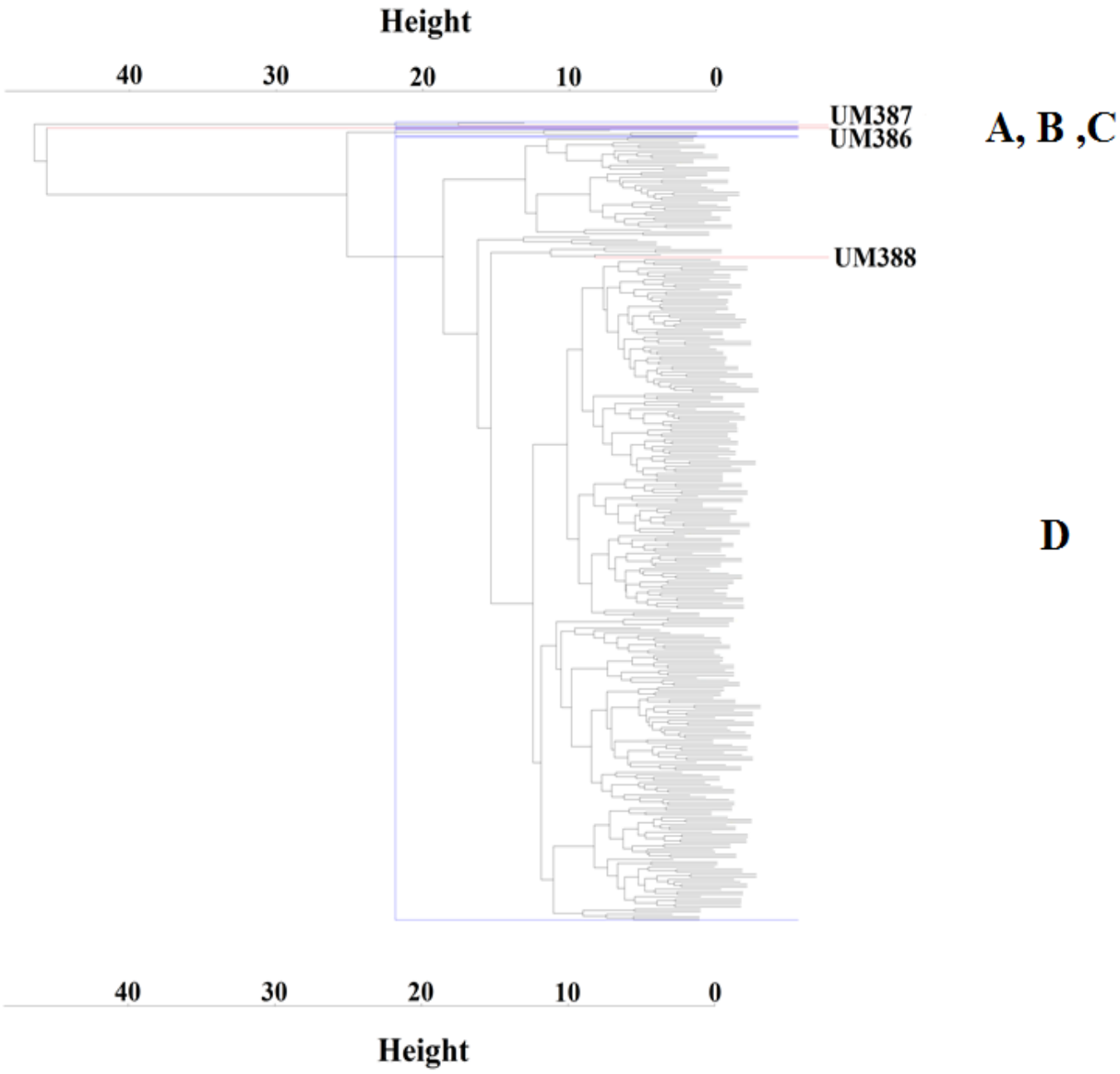

Figure 3.3. Hierarchical cluster analysis of 321 *Brassica napus* L. genotypes based on phenotypic data collected over in 2013 in Glenlea, Manitoba, and 2014 in Winnipeg, Manitoba. Clustering was conducted using Unweighted Pair Group Average Method with Arithmetic Means (UGPMA). Height (*y*-axis) refers to Euclidean distance units.

3.4.4 Group Composition using Ward's Method

Average values for each of the nine agronomic and quality traits from all genotypes within each of Ward's four clusters are presented in Table 3.5. Values calculated for CV for combined results (2013 + 2014 growing seasons together) refer to differences in trait values for the same genotype over the two years. When intra-group values were statistically compared, groups significantly differed for a number of traits (Table 3.5). For number of days before flowering (FLR), Group A (43.30) was distinct from each of Group B (45.50), Group C (45.90) and Group D (45.90) which did not significantly differ from each other based on a least significant distance (LSD) of 0.67. For plant height (HT), based on an LSD of 2.13, Group A (98.30) and Group B (96.50) did not significantly differ, but were significantly different from Group C (103.50) and Group D (101.80) who did not significantly differ from each other. For lodging (LOD) and number of days to maturity (MAT) based on LSD values of 0.45 and 2.80 respectively, there was no significant difference between any of the evaluated groups. For oil content (OIL), based on an LSD of 1.10, Group A (47.10) significantly differed from Group C (44.80) and D (45.80) which did not significantly differ from each other. For protein content (PROM), based on an LSD of 1.52, no group significantly differed from any other. For glucosinolate content (GLUC), based on an LSD of 1.47, Group A (11.8) differed significantly from Groups C (16.5) and D (14.2), but did not significantly differ from Group B (13.4). For saturated fatty acid content (SAT), based on an LSD of 0.12, Group B (5.1) and D (5.1) significantly differed from Group D (5.3), but did not significantly differ from Group A (5.2). For erucic acid content (ER), based on an LSD of 1.3, Group A (50.9) and Group C (50.7) significantly differed from Group B (48.3) and Group D (47.1), but not from each other.

Table 3.4. Mean intra-cluster values for each of nine morphological and quality traits[†] used to cluster 321 *Brassica napus* L. genotypes into four groups[‡], using Ward's Minimum Variance method. Data was collected in the 2013[§] and 2014[§] growing seasons.

	FLR[†] (days)	HT (cm)	LOD (1-5)	MAT (days)	OIL (%)	PROM (%)	GLUC (µmol/g)	SAT (%)	ER (%)
A[§]	43.30	98.30	1.80	99.90	47.10	49.50	11.80	5.20	50.90
B[§]	45.50	96.50	2.00	101.50	46.30	49.30	13.40	5.10	48.30
C[§]	45.90	103.50	2.00	102.30	44.80	49.50	16.50	5.10	50.70
D[§]	45.90	101.80	2.20	101.90	45.80	48.90	14.20	5.30	47.10
Mean	45.20	100.00	2.00	101.40	46	49.30	13.90	5.20	49.20
SD[¶]	1.09	2.77	0.17	0.91	0.82	0.24	1.68	0.07	1.60
CV[¶]	0.02	0.03	0.08	0.01	0.02	0.01	0.12	0.01	0.03
LSD[¶]	0.67	2.13	0.45	2.80	1.10	1.52	1.47	0.12	1.30

[†] FLR, number of days to first flowering; HT, plant height; LOD, lodging; OIL, oil content; PROM, meal protein; GLUC, glucosinolates; SAT, saturated fatty acids; ER, erucic acid.
[‡] A^n, 38; B^n, 122; C^n, 5; D^n = 41
[§] 2013, Glenlea, Manitoba, Canada; 2014, Winnipeg, Manitoba, Canada.
[¶] SD, standard deviation; CV, coefficient of variation between 2013 and 2014; LSD, least significant difference.

3.5 Discussion

Producing high quality F_1 hybrids has been a major goal of commercial *B. napus* breeding since reports in the 1970's and 1980's of increases in seed yield of 40 – 60 % compared to parental genotypes (Shiga, 1976; Schuster and Michael, 1976, Sernyk and Stefanssan, 1983, Grant and Beversdorf 1985). A prerequisite to the production of successful hybrids is the biological phenomenon of heterosis, which is thought to be higher in hybrids with genetically diverse parents in comparison to low heterosis hybrids (Hallaeur et al., 1988). To aid in parental selection for hybrids, breeding programs often employ the use of heterotic groups, which are a group of genotypes that produce similar levels of heterosis when hybridized with a genotype of a different genetic background (Debrueil et al. 1996)**.** The creation of heterotic groups involves the characterization of available germplasm by pedigree, morphological, or molecular means, or a combination thereof to identify genotypes that are distinct from each other (Esposito, 2013). Characterization of qualitative and quantitative phenotypic traits is essential to investigating

functional diversity within plant populations (Granier and Vile, 2014), and can be a successful way to assign heterotic pools to a breeding population **(**Geleta et al., 2004**)**.

In this study, 321 *B. napus* genotypes were evaluated for nine agronomic and quality traits collected over two growing seasons and clustered into four distinct groups based on variation in the evaluated traits. Hierarchical clustering using Ward's Minimum Variance method was found to be best suited to the population and a dendrogram was generated from Ward's Method that created clusters relating genotype clusters with Euclidean distance (Table 3.2). One trait, which was not used as a criterion for clustering was yield. As yield is controlled by multiple genes and susceptible to strong environmental influence many studies find it has limited use as a descriptive measure (Liu and Furnier, 1993; Magloire, 2005). Although, numerous studies have successfully used yield as a clustering criteria (Teklewold, 2006; Babic, 2010; Singh, 2015). In this research, yield was not utilized due to the fact that the genotypes were only grown in single nursery rows and consistent yield data is difficult to obtain in this format.

Despite significant similarity in genotype order while clustering ($t = 4.69 \geq 3.32$ $\alpha = 0.001$), when clusters were split into four groups using cubic clustering criterion, group sizes were inconsistent between the two algorithms ($t = 0.0017 \leq 5.98$, $\alpha = .05$). This is consistent with previous research showing that the UPGMA method can result in unbalanced clusters reflecting large intergroup differences (Mohammadi and Prasanna, 2003). However, the two groups containing one genotype each were the two most distinct genotypes in the population: UM388 and UM387. This sensitivity would make the UPGMA method ideal for evaluating micro-diversity in relatively homogenous populations, as clusters generated using this method tend to have high levels of intra-group similarity relative to other methods (Padilla et al., 2007**)**. Heterotic groups with so few members provide limited information for designing parental combinations. For this reason, the phenotypic

clusters created by Ward's Minimum Variance of the 321 genotypes were more practical than the UPGMA groups for further evaluation. In previous studies, the same resiliency was shown by Ward's Method for evaluating breeding populations. Not only in *B. napus* (Ana et al., 2009), but in *B. oleracea* L. (Dias et al., 1993) as well as *Triticum aestivum* L. (Khodadadi et al. 2011; Lombard, 2000; Fahid, 2014) and *Pisum sativum* L. (Esposito, 2013)

It was interesting to note where University of Manitoba cultivar releases, UM391 (Red River 1997, (McVetty et al., 1997)), UM389 (Red River 1826 (McVetty et al., 2006)), UM390 (Red River 1852 (McVetty et al., 2006)), and UM392 (Red River 1861 (McVetty et al., 2012)) were sorted using Ward's method. Ten genotypes used UM389 as a parent, and 8 of these clustered together in Group A. Also unique to Group A was UM392, which was the University of Manitoba's most recent open-pollinated Roundup Ready™ HEAR cultivar. This could reflect a similar genetic background due to their shared origin at the University of Manitoba. Of the twenty-two genotypes in this study that used UM390 as a parent, 17 sorted into Group B, and 5 others into Group D. Genotypes using UM392 as a parent were dispersed; with 1 in Group A, 14 in Group B, 1 in Group C, and 46 in Group D. Ninety-three genotypes used UM391 as a parent, and they were split between Groups B and D, with 46 in the former, and 45 in the latter and the remaining two genotypes were in Group C. Other pedigrees of note were the 12 genotypes (UM31, UM104, UM105, UM248, UM249, UM250, UM251, UM252, UM245, UM246, UM247) containing the resynthesized genotype UM421 in their parental material. Of these 12, 10 sorted into Group B, and 2 into Group D. The two that were sorted into Group D used UM390 as the other parent, and as previously mentioned, genotypes with UM390 tended to sort into Group D (17 of 22). Pedigree consistency in group assignment is one way to lend evidence to the accuracy of a clustering method in terms of relatedness.

The male-sterile female genotype, UM388 and its two parents (UM386 and UM387) sorted into Groups C, C and B respectively. Cluster placement of UM388 is of the utmost importance for the impact of this study as UM388 is the female for the creation of MSL hybrids, and hybrid heterosis is believed to be a function of parental diversity (Falconer, 1989). Heterotic groups designed using morphological characters have been successful in predicting mid-parent heterosis for days to flowering, and days to maturity in *P. sativum* (Geleta et al., 2014). A hybrid cross designed between UM388 (in Group C) and an individual in Group A, B, and D, should show more heterosis for earliness (days to flowering/maturity) compared to a cross designed between UM388 and another individual in Group C. Due to the epistatic nature of yield in *B. napus* (Radoev et al., 2008) it is reasonable to predict that hybrid crosses between UM388 and genotypes from other phenotypic clusters should display heterosis for yield; even though yield was not used as a trait for heterotic grouping. Basunanda et al. (2010) found unexpected pleiotropic effects in QTL studies on *B. napus*, such as co-localization of QTLs for seed yield and plant height and shoot weight, further highlighting the epistatic nature of yield in *B. napus*. The hope for this research is to aid in the design of future crosses maximizing Euclidean distance between UM388 and a high performing male parent.

In this chapter, a distinct group of 318 *B. napus* pollinator genotypes from the University of Manitoba's germplasm collection, along with a male-sterile female genotype (UM388) and its parents (UM386, UM387) were hierarchically clustered into four groups based on statistically significant differences in nine agronomic and seed quality traits. But more importantly, the Euclidean genetic distance between the 318 University of Manitoba pollinators and the male-sterile female UM388 was established, allowing for the testing of the hypothesis that pollinators that have greater genetic distance values between themselves and UM388 will produce higher

yielding hybrids (Chapter 5). However, phenotyping is not the only method of establishing genetic distance, and is not without problems. Depending upon the population size, evaluating morphological, agronomic and quality traits can create bottlenecks due to time, space, and labor (Riaz et al., 2001). As an alternative, molecular marker techniques have also been used to evaluate diversity in breeding populations (Mahmood et al., 2003; Ahmad et al., 2007; Zeng et al., 2011; Ghanbarnia et al., 2012). These methods have helped to facilitate efficient and effective use of time and resources (Xie et al., 2015). In the next chapter, *B. napus* genotypes from the University of Manitoba will be evaluated and clustered using sequencing and next-generation sequencing techniques, and these clusters will be compared to hierarchical clustering by Ward's Method using morphological and quality data as clustering criteria.

4.0 Using Genotyping-by-Sequencing and Sequence Related Amplified Polymorphisms to Assign *Brassica napus* Genotypes into Groups.

4.1 Abstract

The breeding of superior and competitive *Brassica napus* L. hybrids relies on the successful combining of available parental genotypes in such a way as to maximize the potential for yield gain. Heterotic groups are often used to generalize the heterotic response of similar genotypes in situations where there may be too many possible parental combinations to properly evaluate. The objective of this study was to hierarchically cluster select genotypes of from the University of Manitoba into groups using molecular markers. Molecular markers used were single-nucleotide polymorphisms (SNPs) from Genotyping-by-Sequencing (GBS) and Sequence Related Amplified Polymorphisms (SRAP), which were both used as differentiating molecular traits in the hierarchical cluster analysis. Using 291,782 SNPs gathered from GBS data, 231 *B. napus* genotypes were clustered into 12 distinct groups that appeared to be consistent with pedigree and origin of the genotypes. Using 24 forward and reverse primer pairs (SRAP), DNA fragments from 160 *B. napus* genotypes were selectively amplified and visualized, a process from which 230 polymorphisms were detected, which were then used to hierarchically cluster the genotypes into eight groups. The groupings generated using both GBS and SRAP can be further evaluated based on hybrid performance. If successful heterotic clusters can be produced using DNA sequencing information, the efficiency and effectiveness of future parent selection for hybrid production should increase significantly.

4.2 Introduction

The desirability of *Brassica napus* L. as an oilseed crop revolves around the ability of breeders to modify the fatty acid content to meet industrial and edible uses while at the same time improving agronomic performance (Cuthbert et al., 2009) and maintaining resistance to important diseases (Fernando et al., 2016). Rapeseed oil high in erucic acid (22:1) has many industrial applications, and is used in the development of plastics, lubricants, soaps, and surfactants, amongst other products (Sonntag, 1995; Piazza and Foglia, 2001). High erucic-acid rapeseed (HEAR) genotypes yield seed oil with erucic acid contents > 50 % (vs. 45-50 % in rapeseed oil) (Sanyal et al., 2015; McVetty and Scarth, 2012). The University of Manitoba has had a HEAR breeding program in place since the 1970s; releasing several open pollenated HEAR cultivars including: Hero (Scarth et al., 1991), MilleniUM 03 (McVetty et al., 2000), Red River 1826 (McVetty et al. 2006a), Red River 1852 (McVetty et al., 2006b), Red River 1997 (McVetty et al., 2010) and Red River 1861 (McVetty et al., 2011). Hybrid cultivars, resulting from the cross of two inbred parents, have been increasingly attractive in *B. napus* breeding; as high-parent heterosis (superior performance of hybrid offspring over parental values) of between 20-50 % has been reported (McVetty et al., 1995). The University of Manitoba released their first Roundup Ready^{TM} hybrid rapeseed cultivar, HYHEAR 1, in 2014, which yielded ~33 % more seed and 23 g kg^{-1} more seed oil than MilleniUM 03 (McVetty et al., 2014).

The development of new hybrids means the development of new parental combinations for these hybrids. To address this, heterotic groups can be generated, clustering genotypes into groups of genotypes displaying similar heterotic gains when crossed with genotypes from other groups (Ryder et al., 2014). As there is evidence that heterosis is proportional to the genetic distance between parents (Ali et al., 1995; Qian et al., 2007; Girke et al., 2012; Xing et al., 2014),

morphological, pedigree, or genotypic traits can be used to estimate genetic distance between parental material in a population to aid in parental selection.

Molecular markers are heritable segments of genetic variation reflecting differences between individuals of the same species at the genome level (Agrawal and Shrivastava, 2014), and have become a powerful tool in the efficient and precise use of genetic resources in modern breeding programs (Piquemal et al., 2005; Bertrand et al., 2008). Markers have been successful in selecting for traits with relatively simple inheritance (Blair et al., 2007; Jarquin et al., 2014; Ceballos et al., 2015), but have been less successful in affecting traits controlled by large numbers of polymorphic alleles such as seed yield (Heffner et al., 2009). Molecular markers have been useful in revealing underlying allelic variation within a population for diversity studies, (Ceballos et al., 2015). An advantage of molecular markers for diversity analysis is that the allelic variation and molecular identity of a genotype will not change under different growing conditions, which offers an advantage over phenotypic variation (only reflects a given trait value under the conditions in which the trait value was recorded) (Achtak et al., 2009). Thus, the use of molecular breeding could mean fewer field seasons required for evaluation, resulting in rapid breeding cycle. Additionally, molecular markers are detectable in all tissues at all stages of growth (Agrawal and Shrivastava, 2014). These two reasons make molecular markers more direct and reliable compared to phenotypic data for the development of genetic diversity analysis (Ni et al. 2002).

For use in marker-assisted selection, diversity analysis and heterotic group assignment, screening with molecular markers has become an important facet of many crop breeding programs (Muranty et al., 2014). Over the last 40 years, several methods have been developed and successfully used for the identification of molecular markers, including Random Amplified Polymorphic DNA (Williams et al., 1990), Restriction Fragment Length Polymorphisms (Jeffreys et al., 1985),

Amplified Fragment Length Polymorphisms (Zabeau and Vos, 1993), Simple Sequence Repeats (Queller et al., 1993), and Sequence Related Amplified Polymorphisms (SRAP) (Li and Quirous, 2001).

SRAP detection has been successful in evaluating genomic diversity and creating linkage maps within several crop species **(**Riaz et al. 2001; Ferriol et al. 2003; Fufa et al. 2005; Hu et al., 2015; Huang et al. 2016). SRAP analysis involves the selective amplification of the open reading frame (ORF) using synthetic forward and reverse primer pairs. This amplification allows for differential sequence amplification in analyzed genotypes, which can be quantified by capillary electrophoresis (Li and Quiros, 2001). SRAP specifically targets the ORF, which increases it's chances of selecting markers likely to affect phenotype (Zheng et al., 2015). Benefits of the SRAP method include high reproducibility and a high number of polymorphisms that have allowed it to gain popularity in lesser studied species as it does not require a reference genome (Peng et al., 2015). In the past, SRAP markers have been successfully used in diversity analyses of genera as varied as *Curcurbita* (Ferriol et al. 2003) *Pinus* (Xie et al., 2015) *Psidium* (Padmakar et al., 2015) *Tetrastigma* (Peng et al., 2015), *Morus* (Hu et al., 2015), *Triticum* (Fufa et al., 2005) and *Brassica* (Riaz et al., 2001; Ahmad et al., 2013; Huang et al. 2016.). Due to its highly polymorphic nature, its specific targeting of the ORF, and its simplicity, SRAP is considered a strong method for diversity analysis (Riaz et al., 2001).

Next-generation sequencing (NGS) methods have driven down the financial cost associated with genotyping and SNP discovery, to the point where whole genome analyses are now readily feasible (Xie et al., 2014). Low cost and high marker density make NGS the preferred method for genotyping in plant science (Beissinger et al., 2013) and allows most crop breeding programs to make full use of association mapping and gene-mining (Thudi, 2012).

Restriction-enzyme mediated methods of NGS have been very successful in SNP discovery from diverse germplasm collections (Hickey et al., 2014). One such method, Genotyping-by-sequencing (GBS), has been touted as requiring less DNA and fewer complex steps than other NGS methods such as RADseq (Elshire et al., 2011; Poland, 2012). These methods use methylation sensitive restriction enzymes to reduce genome complexity and target low copy regions of DNA and develop unambiguous SNPs outside of heavily repetitive (methylated) regions (Beissinger et al., 2013; Spindel et al., 2013). It's ability to preferentially target low coverage regions gives GBS generated SNPs added utility in diversity analysis, which has been successfully exploited in the diversity analysis of important crops (Xie et al., 2014). The unprecedented amount of SNPs generated across the whole genome using GBS paired with the relatively low cost make GBS an attractive option for modern plant breeders compared with traditional marker-discovery methods (Jarquín et al., 2014).

The objective of this study was to organize pollinator and male-sterile genotypes of *B. napus* into clusters based using molecular makers generated from SRAP and GBS. To evaluate the efficacy of GBS analysis, 230 individuals were submitted to Cornell University including 227 "pollinator" genotypes, and a male-sterile "female" genotype (UM388), and its two parents, the male-sterile mother line (UM386) and the fertile maintainer line (UM387). To evaluate the efficacy of SRAP analysis, 157 *B. napus* "pollinator" genotypes, as well as UM386, UM387, and UM388 were subjected to SRAP analysis at the University of Manitoba to gather molecular markers which could be used as differentiating criteria in hierarchical clustering. Heterotic groups were visualized using dendrograms to provide knowledge of genetic relationships within *B. napus* genotypes and facilitate parental selection for high-heterosis hybrids.

4.3 Materials and Methods

4.3.1 Plant Material

For GBS evaluation and heterotic grouping, 230 diverse inbred *B. napus* genotypes representing several distinct oil profiles were selected, including: 3 canola quality genotypes, 2 low linolenic quality genotypes, 1 high-oleic low-linolenic (HOLLi) quality genotype, 5 super-high erucic acid (SHEAR) genotypes, and 216 HEAR quality genotypes from the University of Manitoba. These genotypes were available to serve as pollen donors in hybrid crosses, becoming the male parent, or the "pollinator". A male-sterile "female" genotype, UM388, and its parental material (male-sterile mother line UM386 and fertile maintainer line UM387) were obtained contractually between the University of Manitoba and DL Seeds. UM388 utilizes the male-sterile Lembke (MSL) proprietary male-sterility system from Norddeutsche Pflanzenzucht (Holtsee, Germany). For SRAP analysis, 160 diverse inbred *B. napus* genotypes were selected, again including UM386, UM387, and UM388. A full listing of the 231 genotypes used in GBS analysis and the 160 used in SRAP analysis can be found in Appendix B and C.

4.3.2 Tissue Preparation

Single seeds from all genotypes evaluated by either GBS or SRAP were seeded into 4 x 3 trays at a depth of 2 cm in large aggregate professional growing mix from Sunshine® (SunGro, Agaram, MA, USA). Plants were grown in an Econair growth chamber (BioChambers, Winnipeg, MB, Canada). The growth chamber was set to a 16/8 (hour) day/night cycle; during which the temperature fluctuated between 22 °C during the 16-hour day cycle and 18 °C during the 8-hour night cycle. While in the growth rooms, plants were manually watered daily at 11 AM. At stage 12 BBCH scale (Lanchashire et al., 2001), seedlings were fertilized with Plant-Prod® 10-52-10

water-soluble fertilizer at a rate of 3.9 g / L (Master Plant-Prod Inc., Brampton, ON, Canada). To prevent insect damage on the seedlings, the imidacloprid insecticide Intercept™ 60 WP (Bayer Environmental Science, Calgary, AB, Canada) was applied at a rate of 4.1 g / 1000 seedlings. At stage 13 - 14 BBCH scale (3 – 4 first true leaves) (Lancashire et al., 1991), 1-2 g of true-leaf tissue was sampled using forceps and placed in sterile 4 oz Whirl Pak® sample bags with puncture proof tabs (Nasco, Fort Atkinson, WI, USA). Once sampled, tissue was either subjected to DNA extraction immediately or stored at -80˚C.

4.3.3 DNA Extraction

For use in GBS analysis, DNA was extracted following the DNEasy Plant Mini Kit® protocol from Qiagen® (Qiagen, Valencia, CA, USA). In brief, ~500 mg of fresh leaf tissue from the genotype being investigated was manually disrupted using a pestle within a DiaTec (DiaTec Canada, Kitchener, ON, Canada) 1.5 mL snap cap microtube. Once disrupted, 400 µL of buffer AP1 and 4 µL of RNAse A stock solution (100 mg / ml) was added to the previously disrupted lysate, and then incubated for 10 minutes at 65 ˚C in a Julabo ShakeTemp SW22 water bath (Julabo USA, Allentown, PA, USA), mixing 2-3 times during incubation. After incubation, 130 µL of buffer P3 was added to the lysate, which was then mixed and put on ice for 5 minutes. At this point, the lysate was spun for 5 minutes at 14,000 RPM in an Eppendorf 5424 Centrifuge (Eppendorf AG, Mississauga, ON, Canada). The lysate was then pipetted into the QIAshredder mini spin column Qiagen® (Qiagen, Valencia, CA, USA) and centrifuged again for 2 minutes at 14,000 RPM. The flow through (~450 µL) from the QIAshredder column was then transferred to a new DiaTEC 1.5 mL microtube (DiaTec Canada, Kitchener, ON, Canada) and 1.5 volumes of Buffer AW1 were added to the lysate and mixed thoroughly by pipetting the solution up and down several times. At this point, 650 µL of lysate solution was pipetted into the DNEasy Mini Spin column (Qiagen,

Valencia, CA, USA), which was then centrifuged for 2 minutes at 8000 RPM. Flow through in the collection tube under the spin column was then discarded. This step was then repeated, adding another 650 µL to the spin column and centrifuged for another 2 minutes at 8000 RPM, and the flow through was once again discarded. Buffer AW2 (500 µL) was then added to the spin column and the column was centrifuged for 2 minutes at 14,000 RPM. The spin column was then removed from the collection tube and placed in a new 1.5 mL DiaTec microtube. Between 75-100 µL of AE buffer was then pipetted onto the spin column and allowed to sit for 5-10 minutes. The spin column was then centrifuged for 1 minute at 8000 RPM to elute the DNA from the membrane into the AE buffer in the 1.5 mL tube. This step was then repeated with another 75-100 µL of AE buffer through the spin column. Buffer(s) AP1, AW1, AW2 and AE; as well as RNAse were all supplied as concentrates within the DNEasy Mini Kit®. DNA purity was measured using a Nano Drop 2000 spectrophotometer/associated software following the manufacturer protocol (Thermo Fisher Scientific, Waltham, MS, USA).

For SRAP analysis, DNA was extracted using a modified cetyltrimethylammonium bromide (CTAB) protocol (Doyle and Doyle, 1990). In short, 500 mg of fresh leaf tissue was placed in a DiaTec (DiaTec Canada, Kitchener, ON, Canada) 1.5 mL microtube and physically disrupted within the microtube using a mini-pestle placed in a Craftsman™ drill (Sears Canada, Toronto, ON, Canada). Extraction buffer (500 mL of 2% CTAB, 100 mM Tris–HCl, pH 8.0, 1.4 M NaCl, 20 mM EDTA) was added to the lysate, and the mixture was incubated in a ShakeTemp SW22 water bath for 90 minutes. Once removed from the water bath, 7 mL of chloroform–isoamyl alcohol (24:1) was added to the lysate followed by gentle mixing in a Vortex Mixer (Thermo Fisher Scientific, Waltham, MA, USA). Lysate was then placed in an Eppendorf 5424 Centrifuge (Eppendorf AG, Mississauga, ON, Canada) and centrifuged for 16 minutes at 4600 RPM. After

centrifuging, the supernatant was pipetted out and transferred to a new 1.5 mL microtube. At this point, 0.5 volumes of isopropanol were added to the supernatant and the mixture was mixed gently to precipitate the DNA in the solution. The mixture was then centrifuged at 4600 RPM for 5 min, causing the DNA to congeal in a pellet at the base of the microtube. The supernatant above the pellet was pipetted off and the pellet was washed twice with 8 mL of 70 % ethanol which was then poured off. After washing, the pellet was air-dried and suspended in 300-500 µL distilled H_20. At the conclusion of DNA extraction, 70 µL of each sample was placed in its own well on one half of a 384 well plate (Thermo Fisher Scientific, Waltham, MA, USA), and then replicated on the other half of the plate which would later be used as a template during SRAP analysis. DNA purity from analysis was measured using a Nano Drop 2000 spectrophotometer/associated software following the manufacturer protocol (Thermo Fisher Scientific, Waltham, MS, USA).

4.3.4 Molecular Analysis

Each of the 231 samples analyzed using genotyping-by-sequencing were shipped to the Genomic Diversity Facility at Cornell University in Ithaca, New York for analysis. A semi-skirted 96 well twin-tec PCR plate (Eppendorf AG, Mississauga, ON, Canada) covered by 8 Strip PCR Flat Caps (Thermo Fisher Scientific, Waltham, MS, USA) was used to ship the samples on dry ice. GBS analysis at Cornell University followed the protocol described by Elshire et al. (2011). Extracted DNA was treated with methylation sensitive restriction enzyme ApeKI at a 95-plex level (5'…GCWGC…3', 3'…CGWCG…5') to reduce genomic complexity. A total of three plates, titled Plate A, Plate B, and Plate C were sent to Cornell for GBS analysis. A proprietary *B. napus* reference genome obtained from the University of Saskatchewan (Chalhoub et al., 2014) was used to filter and align genotypic data using the Barrows-Wheeler transforming algorithm 0.7.8 – r455 (Li and Durbin, 2010). Bioinformatic analysis was conducted at Cornell University using the

Tassel Computational Pipeline V.3.0.166 (Bradbury et al., 2007; Glaubitz et al., 2014). Results were received in the form of Variant Call Format (VCF) files from the 1000 Genome Project and were converted into "hapmap" files, developed by the International HapMap Project using the program Tassel 9.0 (Cornell University, Ithaca, NY, USA).

SRAP analysis was performed at the University of Manitoba using a modification of the procedure outlined by (Li and Quiros in 2001). PCR cocktail was mixed using 8.6 µL ddH_20, 1 µL 10x PCR Buffer (500 mM KCl, 100 mM Tris Buffer, 1% Triton X-100 surfactant, 1.5 mM $MgCl_2$), 0.15 Taq-polymerase, 0.15 µL dNTP, 0. 15 µL labelled forward primer, 0.15 µL reverse primer), Taq-polymerase, and dNTPase. Each of these ingredients was obtained individually from Invitrogen (Thermo Fisher Scientific, Carlsbad, CA, USA) and diluted with distilled water before mixing into the cocktail. DNA amplification was conducted in 384 well PCR plates from Fisher Scientific (Fisher Scientific, Waltham, MS, USA), into which DNA was added using a stainless steel 384 spike stamping plate. Twenty-five forward and reverse primer pairs with three different attached fluorescent dyes were selected to screen the 160 *B. napus* genotypes. Both forward and reverse primers (Table 4.1) were approximately 17-18 base pairs in length; with forward primers containing a core of CCGG in order to target GC rich regions in the ORF, and reverse primers containing an AATT near the 3' end with the hope of binding to AT rich regions in introns and promoters. DNA and primer pair mixes were amplified and subsequently denatured via PCR in a Mastercycler 384 (Eppendorf AG, Mississauga, ON, Canada) using temperature cycles of: (1) 94 ºC for 3 min, (2) 94 ºC for 55 seconds, (3) 35 ºC for 55 seconds, and (4) 72 ºC for 55 seconds. Steps 2-4 were then repeated five times before continuing to: (5) 94 ºC for 55 seconds, (6) 50 ºC for 55 seconds, (7) 72 ºC for 55 seconds. At this point, steps 5-7 were cycled thirty times. Once the PCR cycle was completed, amplified DNA was separated within acrylamide gels and run

through ABI Prism 3130XL using GenScan® software (V.3.7) (Applied Biosystems, Carlsbad, CA, USA) (Li and Quiros, 2001). PCR products were separated within acrylamide gels and run through ABI Prism 3130XL using GenScan® software (V.3.7) (Applied Biosystems, Carlsbad, CA, USA) (Li and Quiros, 2001). Chromatographs for SRAP analysis were developed from this data using ABI Sequencing Analysis to quantify the presence or absence of primer pairs within each DNA sample (Hagemann and Kwan, 1999). The amount of amplified DNA in each sample was quantified with gel analysis using the open sourced Genographer 2.1 software (University of Montana, Missoula, MT, USA) and the presence/absence of a specific primer pair could be used as a marker. The goal was to identify ~5-10 markers for each primer pair that could be used for cluster analysis. PCR products were placed in a Hitachi 3130xl Genetic Analyzer to undergo ABI DNA sequencing and detection of fluorescently dyed primer-pairs present in the analyzed DNA.

4.3.5 Hierarchical Cluster Analysis

For GBS clustering, hapmap files containing GBS data was analyzed in Tassel 9.0 (Ithaca, NY, USA) software. In Tassel, dendrograms were produced using the unweighted pair-group method with arithmetic means (UPGMA) to organize the *B. napus* genotypes into groups based on presence-absence of SNPs collected in GBS analysis. In Tassel, an "identical by site" (IBS) distance matrix was generated, in which the genetic distance between two genotypes is the proportion of shared SNP similarity. Dendrograms were converted into Archaeopteryx trees within Tassel 9.0 (Ithaca, NY, USA) for ease of viewing and analysis. Archaeopteryx is a java program developed in 2009 for editing and analyzing phylogenetic trees (Han and Zmasek, 2009). In hierarchical cluster analysis there is a point where every entry will make up its own cluster, this is either at the beginning of the analysis (agglomerative method) or at the end (divisive method) (Yim and Ramdeen, 2015). In agglomerative analysis, at each step of clustering, entries will be joined

based on an objective criterion, in this case, IBS variation. Large volumes of data and a high number of possible patterns therein can cause difficulty when imposing an objective number of groups onto a hierarchical cluster analysis, considering the huge number of possible configurations; especially when it relates to potentially hundreds of thousands of molecular markers (Anderberg, 1973; Jung et al., 2008). Methods for imposing groups onto large hierarchical clusters in past have been heuristic or ad-hoc, as hierarchical clustering is performed with no preconceived group numbers, moreover there is no formal method for determining the correct "cut-off" point for hierarchical cluster analysis (Bratchell, 1989; Jain and Dubes., 1988; Jain et al., 1999, Jung et al., 2008). For ease of visualization, after hierarchical clustering was performed via the UPGMA method, labelled groups were assigned by setting a cut-off point on the dendrogram based on the agglomerative measure, in this case % similarity (Yim and Ramdeen, 2015). To separate genotypes into a manageable number of clusters, the threshold for group membership was set at ~ 10 %. This means that members of any given group share the same identity at ~ 10% of evaluated sites (~29,000 sites).

For SRAP cluster analysis, the presence/absence of polymorphic markers in each genotype evaluated was scored using a binary system (1 = present, 0 = absent) to create a matrix and saved as a comma-separated value (CSV) file in Microsoft Excel (Redmond, WA, USA). This CSV file was uploaded into R-Studio 0.99.893 (R-Studio Inc., Boston, MA, USA), a program that utilizes the Rx64 3.2.4 (© The R Foundation for Statistical Computing) open sourced statistical software.

Table 4.1. Twenty-five forward and reverse primer pairs chosen for SRAP analysis of genomic DNA from 160 *Brassica napus* genotypes. Names of each forward and reverse primer are included as well as the color of the fluorescent dye present within the primer for detection in ABI analysis.

Forward Primer[†]	**Reverse Primer**[‡]	**Colour**[§]	**Polymorphisms**[¶]
ME2	BG11	Blue	5
ME2	BG68	Blue	10
ME2	BG62	Blue	9
FC1	BG62	Blue	13
FC1	BG68	Blue	10
FC1	BG73	Blue	9
ODD3	PM32	Blue	5
ODD3	PM34	Blue	5
EM1	BG11	Green	6
EM1	BG32	Green	8
EM1	BG45	Green	7
EM1	BG72	Green	6
SA7	BG11	Green	9
SA7	BG37	Green	15
SA7	BG39	Green	8
SA7	BG41	Green	7
SA7	PM18	Green	13
SA7	PM29	Green	8
BG23	BG37	Yellow	12
BG23	BG34	Yellow	11
BG23	BG41	Yellow	13
BG23	PM18	Yellow	7
BG23	PM29	Yellow	10
BG23	PM32	Yellow	8
BG23	PM34	Yellow	11

[†] Labeled 17-18 base pair CCGG rich forward primers.

[‡] Unlabeled 17-18 base pair AATT rich reverse primers.

[§] Fluorescent dye, attached to primer for each of visualization.

[¶] Number of polymorphisms detected using Genographer 2.1 (Missoula, MT, USA)

Euclidean hierarchical clustering was performed in R-Studio according to Ward's method. A dendrogram was generated using R-Studio and exported as a bitmap file so groups could be visualized. For ease of visualization, after hierarchical clustering was performed via the UPGMA method, labeled groups were assigned by setting a cut-off point on the dendrogram based on the agglomerative measure, in this case % similarity (Yim and Ramdeen, 2015). To separate genotypes into a manageable number of clusters, the cut-off for group membership was set at Euclidean distance of .30 – .35 between constituent genotypes. Dendrogram branches which separated within this zone were deemed their own groups.

4.4 Results

4.4.1 GBS Analysis

Using GBS analysis across the whole genome, 291,782 filtered SNPs were identified in the 231 individuals submitted for genotypic analysis. The most polymorphisms were detected on chromosome C3, with 25,795 polymorphisms identified. The number of polymorphisms detected on each chromosome are visible in Table 4.2.

Hierarchical cluster analysis produced a dendrogram of the examined *B. napus* genotypes (Figure 4.1). According to the IBS distance matrix, the majority of genotypes in this study (all except the five genotypes of Group N) are identical at ~7 % of evaluated sites. The genotypes range from sharing 8 – 12 % similarity depending on group membership. A full listing of the group assignment for each of the 231 genotypes in the GBS hierarchical cluster analysis can be found in Appendix B.

Group A was the largest group identified using GBS hierarchical analysis, containing 58 genotypes. Membership of Group A contained many of the University of Manitoba's important *B.*

napus releases: including a closely related cluster containing Red River 1852 (McVetty et al., 2006b), Venus (McVetty et al., 1996a), Neptune (McVetty et al., 1996b), Castor (McVetty et al., 1998), Hero (Scarth et al., 1989), and Mercury (Scarth et al, 1994). Also in Group A were the the *B. napus* cultivar(s) Reston, which was the second HEAR variety released by the University of Manitoba in 1982, Stellar (Scarth et al., 1987), Apollo (Scarth et al, 1995), Westar (Klassen et al., 1996), Polo (Rahman et al., 2001), Topas (released by Svalof A.B. of Sweden in 1987), Sentry (Rimmer et al., 1998**)**, MilleniUM 03 (McVetty et al. 2000), Red River 1826 (McVetty et al., 2006a), and the commercially available cultivar Global. An important member of Group B was UM391, which was the University of Manitoba commercial release Red River 1997 (McVetty et al., 2010). Clustered closely with UM391 were several genotyppes wth UM391 in their pedigrees including the genotypes UM47, UM48, UM49, UM81, UM82, UM50, UM61, UM58, UM60, UM70, UM54, UM57, UM68, UM72, UM52, UM53, UM46 and UM62 (listed in the order they clustered). Also in Group B was UM395, which was present along with another 14 genotypes (UM21, UM16, UM23, UM34 UM22, U242, UM19, UM20, UM15, UM18, UM 232, UM26, UM239, UM25) (listed in the order in which they cluster), which are all full or half-siblings derived from UM395. Group B also contained UM5, and a number of genotypes that have UM5 in their pedigree (UM33, UM40, UM34, UM37, UM38, UM39, UM76, UM80, UM76 and UM41).

Table 4.2. Number and chromosomal location of Single Nucleotide Polymorhisms (SNPs) detected during Genotyping-by-Sequencing analysis of 231 *Brassica napus* genotypes.

Chromosome	Polymorphisms
A1	11991
A2	10991
A3	17760
A4	8743
A5	12296
A6	12084
A7	11943
A8	8519
A9	17128
A10	10168
C1	18645
C2	17059
C3	25795
C4	19952
C5	19479
C6	14079
C7	18400
C8	16629
C9	20121
Total	291782

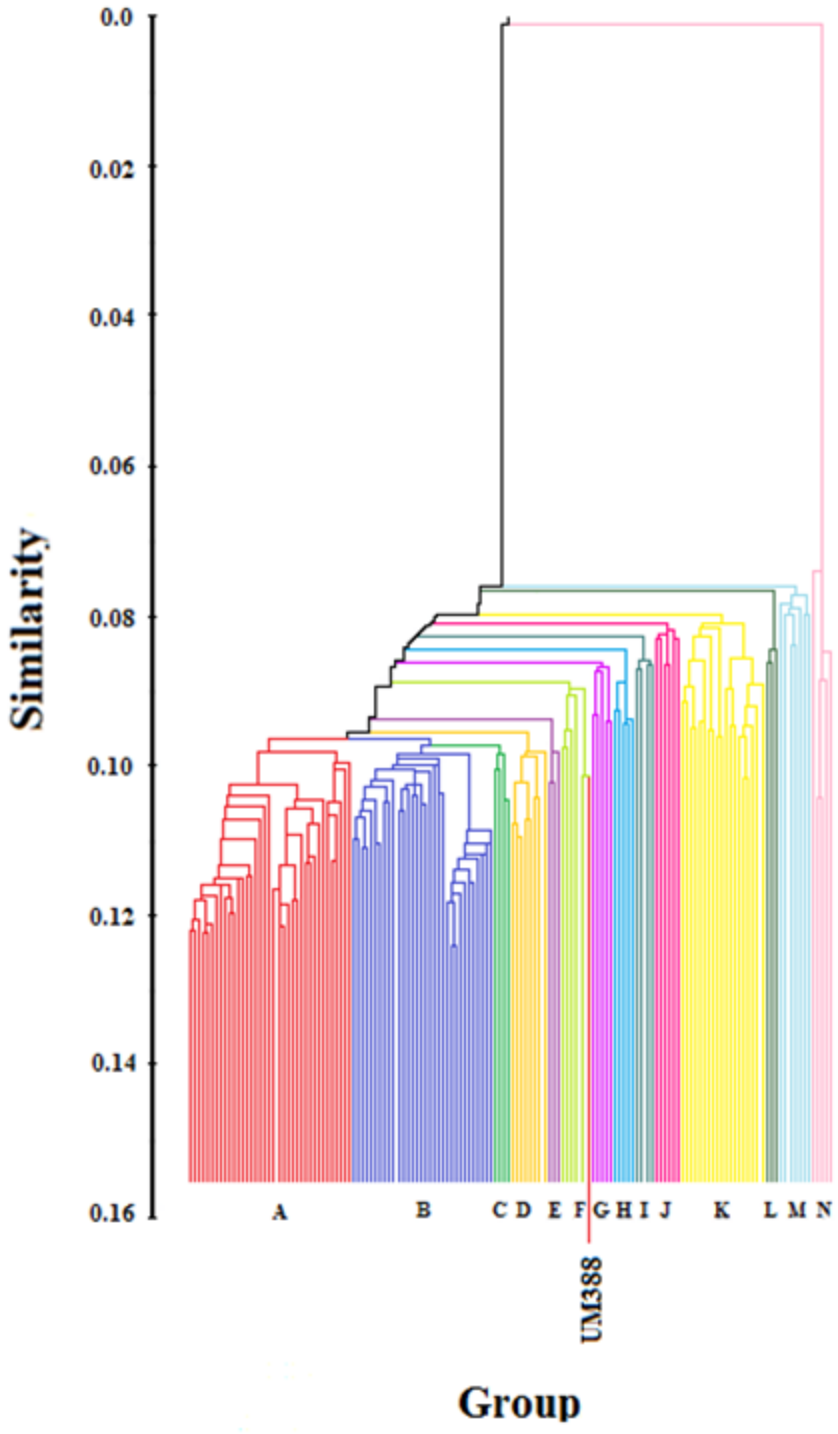

Figure 4.1. UPGMA hierarchical clustering of 231 *Brassica napus* L. genotypes using identical-by-state variation at 291,782 GBS identified sites as differentiating criteria. Visualized in Tassel 9.0, fourteen groups (A-N) are coloured for ease of visualization. Group is labeled along the x-axis as is the position of the male-sterile female UM388.

The University of Manitoba cultivar release Red River 1861 (UM392) (McVetty et al., 2012) was present in Group D alongside two full sibling groups; UM63, UM64, UM65 and UM66; and UM77, UM78, UM79 and UM75, which all contain UM392 in their pedigrees and are listed in the order they clustered.

Group F contained 11 genotypes and was home to the male-sterile female UM388 (Figure 4.1). Clustered closely with UM388, was UM386 which was one half of UM388's pedigree, and the source of it's male-sterility. Along with UM386 and UM388 were the genotypes UM402, UM403, and UM404. These three genotypes are European-derived cultivars from the Danish company Danisco (Copenhagen, Denmark). This is interesting as UM388 (and UM386) clustered with the only other European-derived genotypes in the study, implying a genetic affinity. Group N contained 5 genotypes and branched off from every other genotype in the study at the dendrogram base, making Group N the most genetically distinct group in the GBS hierarchical cluster analysis. This group was made up of resynthesized *B. napus* genotypes, each with SHEAR oil profiles. They included UM417, UM418, UM419, UM420, and UM421.

There were situations where genotypes were organized into several different groups despite having the same pedigree. There were genotypes in Group D (UM71), Group H (UM115, UM119) and Group I (UM118) which were derived from the same pedigree and would have been placed in the same group if the organization were based on such. Similarly, there were genotypes in Group D (UM63, UM64, UM65, UM66), Group I (UM117), Group J (UM113) and Group L (UM114) which all had the same pedigree. There were even genotypes derived from the same backcross organized into different groups, such as UM211, UM11, UM12, UM101 in Group A, when UM99 and UM100 were sorted into Group F.

4.4.2 SRAP Analysis

To organize the SRAP dendrogram into groups that explain the largest amount of variation (based Euclidean genetic distance) the cut-off for group membership was set between 30 – 35 Euclidean distance units. Dendrogram branches which separated within this zone were deemed their own groups, resulting in the identification of 8 (A, B, C, D, E, F, G, H) groups out of the hierarchical clustering of the 160 *B. napus* genotypes using Ward's method in R-Studio (Figure 4.2). Group sizes were 13, 13, 15, 30, 45, 17, 13, and 14 genotypes for groups A-H, respectively. At approximately 10 Euclidean distance units the dendrogram splits into two sub-groups. Subgroup I includes groups A-E and subgroup II groups F-H. Within subgroup I, groups A and B split from C, D and E at 32.5 Euclidean distance units. Within subgroup II, groups G and H split off from group F at 30 Euclidean distance units. Clustering based on SRAP analysis appeared consistent with reported pedigree, and according to their pedigrees, groups of related genotypes clustered together all over the dendrogram. Many of the genotypes evaluated with SRAP analysis contained either UM391 (Red River 1997), or UM392 (Red River 1861) in their pedigrees, 64 and 67 genotypes, respectively. Genotypes containing UM391 or UM392 were found from one end of the dendrogram to the other (Groups A-H), which suggests that the overall homogeneity of the evaluated genotypes is quite high, at least as it related to the evaluated SRAP markers.

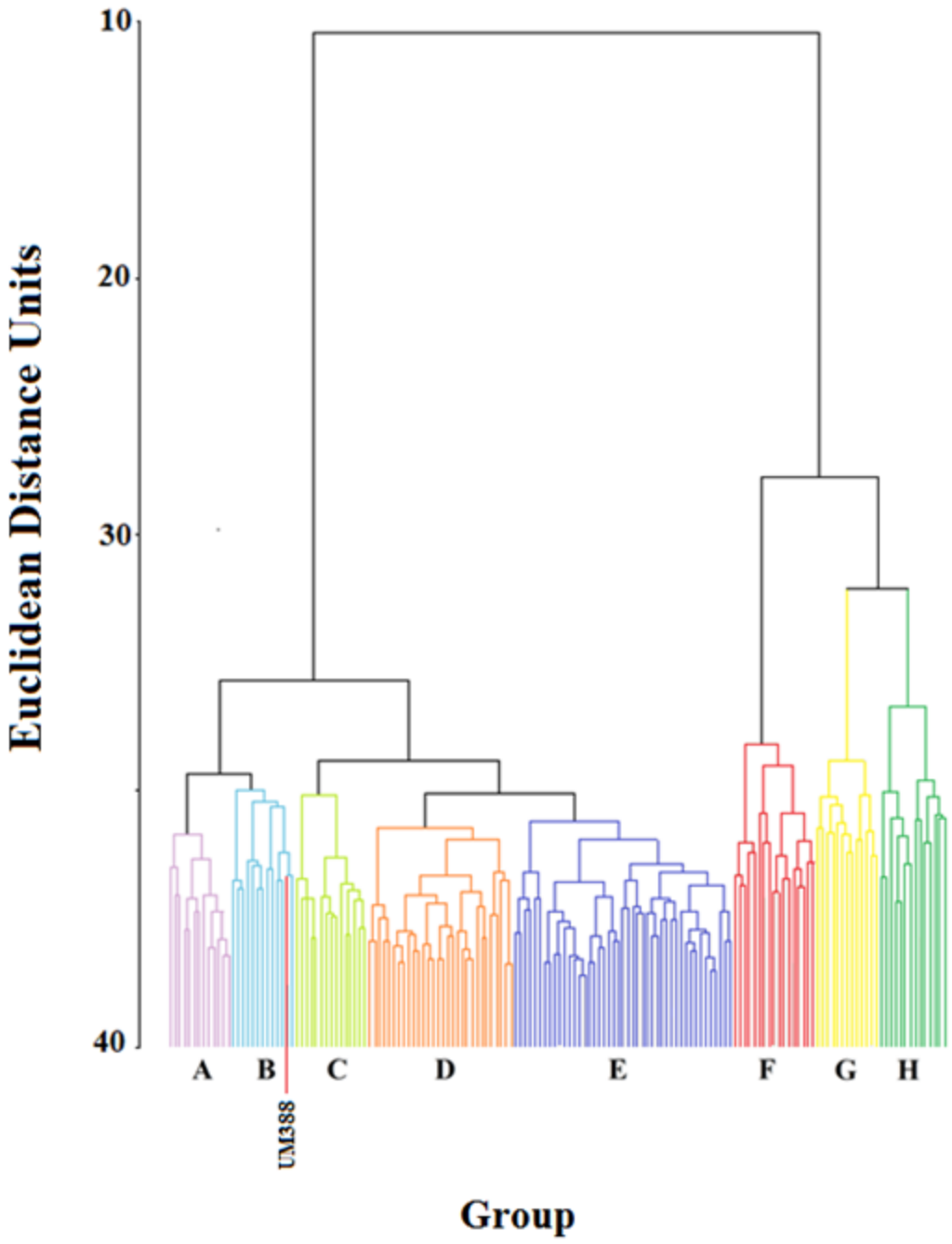

Figure 4.2. Hierarchical clustering of 160 *Brassica napus* L. genotypes via Ward's minimum variance method based on Euclidean distance established using 230 SRAP markers. Visualised in RStudio v. 0.99.893. Eight groups (A-H) are colored for ease of differentiation. Group is labeled along x-axis as is the position of important male-sterile female UM388.

Within subgroup I, Group A contained UM337, UM359, UM360, UM372, UM373, UM374 and UM376 which originated as crosses between full-siblings and UM392 (Red River 1861). These genotypes were half-siblings with other members of Group A. Group B contained 13 entries, including the male-sterile female UM388 and one half of its parental material, UM387. Group C contained 15 entries, the majority of which was made up by half-siblings sharing UM392 in their parental makeup (UM113, UM114, UM177, UM313, UM320, UM65, UM76, UM79 and UM80).

Group D contained 30 genotypes, including several genotypes containing UM391 in their parental material. These were UM127 and UM128 (full-siblings); UM301, UM304 and UM305 (full-siblings); UM324, UM325 and UM346 (full-siblings); UM57, UM58, UM59, UM60, and UM62 (full-siblings); UM81 and UM82 (full-siblings); as well as UM380 and UM67. Also present in Group D were several genotypes containing UM392 in their pedigrees, including UM316, UM322 and UM323 (full-siblings); UM63, UM64 and UM66 (full-siblings); as well as UM335 and UM371. Group E was the largest group generated using SRAP analysis, made up of 45 genotypes, including UM386, which was one half of the parental material, and the male-sterility donor to the female genotype UM388. Nearly one half of the genotypes in Group E contained UM392 (Red River 1861) as part of their parental material. These were UM117, UM119, UM331, UM332, UM333, UM334, UM338, UM342, UM343 and UM71 (full-siblings); UM124, UM355, UM356, UM357 and UM358 (full-siblings); UM184 and UM186 (full-siblings); as well as UM312, UM314, UM315 and UM321 (full-siblings). Also in Group E were 16 genotypes containing UM391 (Red River 1997) in their parental material including UM120, UM121, UM347, UM353, UM354 and UM72 (full-siblings); UM299, UM302 and UM311 (full-siblings); UM328, UM329 and UM330 (full-siblings); and UM367, UM370, UM379 and UM73 (full-siblings).

Within subgroup II, Group F contained 4 full-sibling genotypes with UM392 in their pedigrees. These were UM188, UM183, UM317 and UM318. Four genotypes in Group F had UM391 in their pedigree, these were UM325, UM327 (full-siblings); UM68 and UM69. Another pair of full-siblings in Group F was made up of UM22 and UM23, which each included UM395 in their pedigree. Group G contained 13 entries, including 4 genotypes that shared UM392 (Red River 1861) in their parental material, these being UM112, UM340 and UM341 and UM75. Another 6 genotypes were present in Group G, each sharing UM391 in their parental material. These were UM125, and UM365 (full-siblings); UM326, and UM70 (full-siblings); as well as UM351 and UM352 (full-siblings). Group H also contained 6 genotypes with UM392 in their parental material, these were UM319, UM339, UM362, UM377, UM378 and UM77. Another four genotypes were also present in Group H, each sharing UM391 in their parental material, these were UM308, UM349, UM356 and UM385.

4.4.3 GBS and SRAP Hierarchical Clusters

Of the 160 genotypes analyzed using SRAP markers, and the 231 genotypes analyzed using GBS markers, there were 56 genotypes shared in common between the two methods. Those 56 genotypes were made up from 36 pedigrees, four of which had three or more representative genotypes. When the genotypes from these four pedigrees were arrayed in both the GBS an SRAP clusters, GBS seemed to be more successful than SRAP at clustering them together. To illustrate this, the clustering results from all four pedigrees with greater than three representative genotypes will be outlined. The genotypes UM112, UM113, UM114 and UM117 were all from the same pedigree, and using GBS marker data they were all present in Group E of the GBS hierarchical cluster, whereas using SRAP markers, UM113 and UM114 were clustered together in Group C, while UM112 was in Group D and UM117 was clustered into Group B. A similar situation was

observed with the genotypes UM63, UM64, UM65 and UM66, while using GBS markers all four genotypes clustered together, while with SRAP markers UM63 and UM66 were clustered together, while UM64 and UM65 were each clustered into distinct groups. For the genotypes UM75, UM76, UM77 and UM79 the GBS marker data clustered UM75, UM77 and UM79 in the same cluster, with UM76 in a separate one. This juxtaposes with the use of SRAP markers for clustering which placed UM76 and UM79 in the same cluster and UM75 and UM75 in separate clusters. UM186, UM188, and UM189 were all clustered into the same group using GBS markers, while with SRAP markers UM186 and UM188 were clustered together and UM189 was placed in a separate cluster. It is interesting to note that using SRAP analysis, UM388 clustered closer with UM387 than with the other half of its parental material UM386, while using GBS analysis UM388 clustered closer with UM386 than with UM387.

4.5. Discussion

In this study, we report the use of molecular data gathered using GBS and SRAP analyses to evaluate the genetic diversity within select genotypes of *B. napus* and the hierarchically clustering of these genotypes into groups based on shared marker compliment. Due to their heritability and relative abundance (Hayward et al., 2012), SNP markers have been an invaluable tool for trait improvement in breeding populations populations of *B. napus* (Batley and Edwards, 2007), as well as the fine mapping of superior agronomic genes (Chagné et al. 2007**)**. Next-generation sequencing has been billed as an cost-effecive and efficient means for high-volume genome-wide mapping in breeding populations (Bayer et al., 2015**)**.

Diversity between parental lines in breeding populations is essential for the creation of high performing hybrids due to the effect of heterosis (Melchinger and Gumber, 1998**)**, and many breeders find it essential to sort their germplasm into, and maintain at least two distinct heterotic

groups (van Heerwaarden et al., 2012**)** within their breeding population**.** Next-generation methods such as GBS, have been succesfully used in diversity studies for crop species like *Arabidopsis thaliana* L. (Annachiaro et al., 2015), *Zea mays* L. (Elshire et al. 2011), *Triticum aestivum* L. (Poland et al. 2012), and *Brassica oleracea* L. (Lee et al., 2015). In the *B.* napus genome there are estimated to be approximately 1.7 million SNPs (Hayward et al., 2012), some of which could be invaluable in organizing of breeding populations of *B. napus*. In this study, 291,782 SNP markers were identified and used to evaluate the genetic similarity between 231 distinct *B. napus* genotypes. To this effect, GBS was successful in picking out and clustering the distinct and exotic genotypes based on origin. All the European-derived genotypes in the study sorted into Group C, including the genotypes UM402, UM403, UM404 and UM423 from the Danish company Danisco (Copenhagen, Denmark). Similarly, the canola-quality genotypes UM422, UM424, UM426 grouped together; as did the super high erucic acid rapeseed quality genotypes (UM417, UM418, UM419 and UM420). It is also worth noting how closely together previous HEAR cultivar releases from the University of Manitoba grouped together; Red River 1852 (UM390), Venus (UM405), Neptune (UM406), Castor (UM407), Hero (UM410), Mercury (UM409), Stellar (UM428), Apollo (UM427), MilleniUM03 (UM408), and Red River 1826 (UM389) all grouped within Group A. As hybrids often perform better based on larger genetic distances between parental genotypes (Ali et al., 1995), this study provides evidence of the need for the University of Manitoba to broaden the genetic basis of its commercial releases. That being said, Red River 1997 (UM391), and Red River 1861 (UM392), two other commercial cultivars from the University of Manitoba did cluster elsewhere. Another example of group consistency was where UM395 grouped closely with the genotypes UM15, UM16, UM18, UM19, UM20, UM21, UM22, UM23, UM24, UM25, UM26 and UM232; all of which contain UM395 within their pedigree. These results lend credibility to

the hypothesis that GBS is an effective way to organize genotypes into groups. However, from a financial standpoint, GBS is not without its drawbacks. With a price approximately $38.00/sample (2015), and an added $4.56 per sample for DNA extraction via Qiagen® DNeasy Mini Kit the price can become prohibitive if large volumes of samples are to be evaluated.

The use of SRAP markers to evaluate or map populations of *B. napus* has been succesful in the past. Sun et al. (2007) used approximately 13,551 SRAP markers to develop a linkage map for *B. napus*. The utility of SRAP markers do not end with mapping; Javed et al. (2016) used SRAP markers to aid in the identification of *B. napus* QTL for seed oil content, fatty acid content, as well as time to flower. Similarly, Raman et al. (2012) succesfully used SRAP markers to identify QTLs associated with flowering time in *B. napus*, and Ghanbarnia et al. (2012) used SRAP markers to investigate avirulence genes in *B. napus* to the pathogen *Leptosphaera maculans* (Desm.). For heterotic group evaluation, Riaz et al. (2001) used 118 polymorphisms generated from 18 forward-reverse primer combinations to organize 22 distinct *B. napus* genotypes into three heterotic groups. In the current study, 230 polymorphisms, generated from 25 forward-reverse primer combinations and were used to organize 160 *B. napus* genotypes into eight groups. These eight groups were comprised of discrete clusters of individuals that shared a genetic background. The majority of the genetic material examined using SRAP analysis either had Red River 1997 (UM391) or Red River 1861 (UM392) as part of their pedigree. Of the total 160 genotypes analyzed using SRAP, 64 had UM391 as a parent, and 67 had UM392 as a parent. However, these genotypes did not all cluster together; in fact, each SRAP Group (A-H) contained genotypes with UM391 and UM392 present in their pedigree. The SRAP analysis did not include any canola, HOLLi, or HEAR genotypes that were not from the University of Manitoba. Thus, the SRAP groups could reflect the narrow genetic base of the evaluated material.

Distinct groups were successfully created using GBS data from 231 *B. napus* genotypes including HEAR, canola quality, HOLLi, SHEAR, and a male-sterile female line and its parents. A hierarchical cluster analysis using SRAP molecular markers was also conducted. Both of these methods groups could aid in the creation of future *B. napus* hybrids from the University of Manitoba by way of using genetic distance during parental selection for future hybrid genotypes. In the following chapter, the "accuracy" of the GBS and SRAP generated groups will be evaluated by way of hybrid analysis of select *B. napus* genotypes and the male-sterile line UM388, operating under the hypothesis that larger genetic distances between potential parents within the heterotic groupings should result in better performing hybrids.

5.0 Investigating a link between genetic distance and hybrid performance in *Brassica napus*

5.1. Abstract

Genetic distance between parental genotypes has been implicated in the superior performance of hybrids for many economically important crop species. Genetic distance can be defined by morphological, agronomic or molecular traits. The objective of this research was to investigate the possible link between genetic distance and hybrid yield in *Brassica napus* L. when combining University of Manitoba pollinator genotypes with a single male-sterile female. Three sets of criteria for establishing genetic distance between the pollinators and female parent were utilized, including 321 genotypes organized using phenotypic trait evaluation, 231 genotypes organized using genotyping-by-sequencing (GBS), and 160 genotypes organized using sequence-related amplified polymorphism (SRAP) markers. The hypothesis was that hybrid yield would correlate positively with the genetic distance between the pollinators and male-sterile female. Regression analysis was performed to test the hypothesis utilizing hybrid yield data from 2011-2014. Using phenotypic criteria, the analysis produced a significant correlation between genetic distance and hybrid yield explaining either 5 % or 42 % of the variation in hybrid yield depending whether hybrids were grown at three or more, or five or more sites included in the analysis, respectively. No significant link was found between GBS or SRAP-derived genetic distance and hybrid yield. However, the phenotypic results provide evidence for the contribution of the genetic distance between parental genotypes and resulting performance of the hybrids. This could limit the number of possible parental combinations to be tested in a hybrid breeding program and become a tool in the production of superior *B. napus* hybrids.

5.2. Introduction

The breeding of high erucic-acid rapeseed (HEAR) genotypes has been a successful focus for the University of Manitoba's *Brassica napus* L. breeding program for decades, resulting in several commercial releases (Scarth et al., 1991; McVetty et al., 2000; McVetty et al. 2006a; McVetty et al., 2006b, McVetty et al., 2010; McVetty et al., 2011). As in any crop, the effective breeding of superior hybrids depends on the correct utilization of available germplasm (Qian et al., 2005). This involves maximizing the combining ability of individual parents while at the same time maximizing their heterotic potential through genetic or Euclidean distance (Grant and Beversdorf, 1985; Ali et al., 1995; Riaz et al., 2001; Qian et al. 2005; Memon et al., 2015). To this effect, the biological phenomenon of heterosis has been utilized to create more productive, more resilient and faster developing cultivars in nearly all major crop species and is a staple of modern plant breeding (Klimenko et al., 2014). While not completely understood, heterosis refers to the superior performance of F_1 offspring over both of the inbred parental genotypes in major agronomic characteristics (Shull, 1948). Observational evidence of superior performance in F_1 hybrids derived from distinct parental genotypes led to a hypothesis asserting a positive correlation between genetic distance of the inbred parents and the resulting performance of the hybrid (Grand and Beversdorf, 1985; Diers et al., 1996; Falconer and MacKay, 1996; Riaz et al., 2001; Singh, 2015). Although conflicting results have been found (Girke et al., 2012; Jesske et al., 2013.), there is a large body of evidence demonstrating that high-yielding hybrids are often the result of crossing genetically divergent parental genotypes (Diers et al., 1996; Riaz et al., 2001; Klimenko et al., 2014; Singh, 2015; Xie et al., 2015). To aid in the selection of superior parental material, breeding programs often find it useful to organize their germplasm into distinct groups of genotypes that

display a similar heterotic effect when crossed with individuals in other groups (Parentoni et al., 2001).

Genetic distance (GD) is often used to estimate the genetic divergence between species, or between individuals of the same species and assign each into groups (Abdel-Ghani and Lübberstedt, 2013). The criteria for group organization can include pedigree data (Badu-Apraku et al., 2006), morphological data (Ana et al., 2009), molecular marker data (Melchinger et al., 1991) or a combination thereof (Abdel-Ghani and Lübberstedt, 2013; Mohammadi and Prasanna 2003). However, regardless of the data set used, the distance is still referred to as genetic distance. Using data from a given set of criteria, genetic distance is then estimated mathematically **(**Cavalli-Sforza and Edwards, 1967; Nei, 1972).

Phenotypic traits are often used as the backbone of any germplasm evaluation, the word "phenotypic" describing any readily observable physiological, morphological, agronomic, or quality characteristic of an individual (Violle et al., 2014). However, the phenotyping procedures needed to properly evaluate large germplasm collections can be time consuming, difficult and subject to relatively large amounts of ambiguity (Langer, Longin, and Würschum, 2014). Due to these issues with phenotyping, molecular markers have been employed to evaluate and organize germplasm collections that are too large for phenotypic evaluation (Badu-Apraku et al., 2013; Wu et al., 2014; Obiadalla-Ali et al., 2014; Singh et al., 2015).

Maximizing heterotic potential in *B. napus* is different than in maize, where established heterotic groups make it easier to assume heterotic potential between potential parents based solely on genetic distance (Tracy and Chandler, 2006). Although per-se performance of potential parents can be useful in developing potential heterotic groups, the use of combining ability can be used to augment genetic distance in the selection of hybrid parents (Krystkowjak et al., 2009). Combining

ability can be applied in situations with a large number of potential parental combinations. In situations where male-sterile female genotypes are a factor; potential male parents can be crossed to a tester as a general measure of performance (Kempthorne, 1957).

The effective use of male-sterile female genotypes to ensure efficient hybridization has become central to hybrid seed production (Schnable and Wise, 1999; Budar and Pelletier, 2001). In *B. napus*, genic male-sterility systems such as Male-Sterile Lembke (MSL) are available to the University of Manitoba, which benefit in that all known *B. napus* genotypes act as fertility restorers, producing fertile F_1 (Stiewe et al., 2008; Jan et al., 2016). These females, while limited in number, are central to hybrid breeding at the University of Manitoba, and any potential "male" pollinator genotype should be tested against such females to aid in the production of high-performing hybrids. However, due to logistical constraints such as time, space, labour and money, the field evaluation of every possible hybrid combination, even against a tester, is not always feasible. As such, methods of predicting hybrid performance are critical for the proper allocation of time and resources by ensuring only favourable parental combinations are made and tested.

The objective of this experiment was to evaluate *B. napus* genotypes that have been hierarchically clustered from agronomic and seed quality data, genotyping-by-sequencing- (GBS) derived SNPs, or sequence related amplified polymorphism (SRAP) markers for their ability to produce high-yielding hybrids with a common male-sterile female. As each evaluated hybrid would be created using the male-sterile UM388, the genetic distances between each pollinator genotype and UM388 established in Chapter 3 and Chapter 4 were evaluated against the resulting performance of various hybrids in yield trials. It was hypothesized that a link would be found for genetic distance between parental genotypes and yield. Such a link could aid in parental selection and limit the number of hybrids that required field evaluation.

5.3 Methods and Materials

5.3.1 Parental Genotypes

For this study, hybrids were produced by combining a pollen producing "pollinator" genotype with a male-sterile "female" genotype. Pollinators were drawn from 318 HEAR quality *B. napus* genotypes from the University of Manitoba, the names of which are found in Appendix A, and their agronomic and quality characteristics are discussed in Chapter 3. The male-sterile female genotype with which all evaluated hybrids were produced was contractually obtained from DL Seeds (Morden, MB, Canada) and was designated as UM388.

5.3.2 Hybrid Genotypes

The process of combining University of Manitoba pollinators with UM388 was conducted annually near Temuco, Chile. Hybrid seed was produced on an annual basis in three-row mini-cages (approximately 1 m^2 tents), with planting in mid-October and harvest in late May or early April. The mini-cages consisted of the pollinator in the middle row and the male-sterile female in the two outside rows. In total, 119 distinct *B. napus* hybrids were produced and evaluated at 37 field sites in 2011-2014. If a hybrid did not perform well in the first year of yield evaluation it was not tested further. This led to testing at an unbalanced number of sites for each hybrid.

5.3.3 Hybrid Evaluation

Hybrid yield trials were designed as randomized complete blocks, with constituent genotypes replicated three times and managed according to Canola Co-operative Trial protocols, administered by the Canola Council of Canada. Briefly, in late May, all hybrid genotypes to be tested in a particular trial were seeded into 1.5 – 2.1 m x 5 – 6 m plots, each with six rows and a rate of 1200 - 1400 seeds/plot. Trials were designed as a randomized complete block with

constituent genotypes replicated three times in each trial. Every plot at a given field location was fertilized in accordance with local soil test recommendations. Herbicide was applied at stage 13 on the BBCH scale, or the three-leaf stage. Roundup Ready® genotypes were treated with Roundup® (Monsanto, Winnipeg, MB, Canada) at 1.25 L/ha in 111 L of H_2O. Stand agronomic practices were used on an as-needed basis throughout the rest of the growing season and differed between sites. Once plants senesced (97 BBCH scale), seed from each plot was harvested separately, collecting yield values for each plot in kg/ha adjusted for 10 % moisture. Yield data from four years (2011-2014) of yield trials in Western Canada was utilized for evaluating hybrid performance.

In 2011, eight hybrids derived from University of Manitoba pollinators crossed with the male-sterile UM388 were evaluated. The pollinators for these eight hybrids were UM1, UM5, UM6, UM8, UM13, UM97, UM205 and UM211. In 2012, another eight hybrids derived from University of Manitoba pollinators crossed with UM388 were included for evaluation. The pollinators were UM5, UM6, UM9, UM23, UM32, UM39, UM97 and UM211. Q2 from the University of Alberta, 46A65 from Pioneer Hi-Bred Production LP (Chatham, ON, Canada), and MilleniUM03 from the University of Manitoba were used as control genotypes in 2011 and 2012. Locations for 2011 and 2012 hybrid analysis were Carman, MB, Canada (orthic black soil); Glenboro, MB, Canada (gleyed black chernozem soil); Lake Lenore, SK, Canada (orthic black chernozem); Rosthern. SK, Canada (orthic black soil); Vegreville, AB, Canada (solozonetzic soil), Rosebank, MB, Canada (gleyed rego black chernozem); Valparaiso, Chile (alluvial soil).

In 2013, hybrid yield data was gathered from multi-location yield trials, as well as a single location preliminary yield trial. The first multi-location yield trial (5 sites) contained 16 hybrids, the pollinators for which were UM5, UM9, UM23, UM32, UM33, UM34, UM35, UM36, UM37,

UM39, UM40, UM43, UM57, UM97, UM98 and UM259. The control genotypes included 5540 from Bayer Crop Science (Calgary, AB, Canada), 46A65 from Pioneer Hi-Bred Production LP (Chatham, ON, Canada) and Red River 1861 from the University of Manitoba. Field locations for this trial were Glenboro, MB, Canada (gleyed black chernozem soil); Lake Lenore, SK, Canada (orthic black chernozem soil); Rosebank, MB, Canada (gleyed rego black chernozem soil); Watrous, SK, Canada (orthic dark brown chernozem soil) and Winkler, MB, Canada (orthic dark brown chernozem). Another multi-location advanced yield trial (3 sites) was conducted in 2013, testing 14 hybrids, which used UM2, UM4, UM7, UM11, UM14, UM16, UM17, UM44, UM256, UM264, UM267, UM268, UM269 and UM271 as pollinator genotypes. The University of Manitoba cultivars Red River 1861 and HYHEAR 1 were used as control genotypes for this trial and the field sites were Portage la Prairie, MB, Canada (gleyed rego black soil); Rosebank, MB, Canada (orthic black chernozem soil); Watrous, SK, Canada (orthic dark brown chernozem soil).

Yield data was gathered from several trials in 2014 including three multi-location yield trials, and three single-location yield trials. In the first multi-location yield trial (8 sites), 16 hybrids using the pollinators UM4, UM9, UM14, UM15, UM17, UM22, UM33, UM34, UM36, UM40, UM43, UM44, UM57, UM98, UM203 and UM259 were evaluated. The control genotypes included 5540 from Bayer Crop Science (Calgary, AB, Canada), 46A65 from Pioneer Hi-Bred Production LP (Chatham, ON, Canada) and Red River 1861 from the University of Manitoba. Field locations for this trial were at Glenboro, MB, Canada (orthic black chernozem soil); Killam, AB, Canada (orthic black chernozem soil); Lake Lenore, SK, Canada (orthic black chernozem soil); Rosthern, SK, Canada (orthic black chernozem soil); Saint Albert, AB, Canada (silty clay loam soil); Thornhill, MB, Canada (orthic black chernozem clay soil); Warren, MB, Canada (rego black chernozem soil); Watrous, SK, Canada (orthic dark brown chernozem soil). The next multi-location yield trial in

2014 (3 sites) tested 19 hybrids using the pollinators UM18, UM24, UM25, UM26, UM29, UM30, UM47, UM48, UM50, UM53, UM54, UM58, UM60, UM126, UM129, UM183, UM238, UM246, and UM248. The University of Manitoba cultivars Red River 1861 and HYHEAR 1 were used as control genotypes for this trial, which was conducted at Portage la Prairie, MB, Canada (gleyed rego black soil); Thornhill, MB, Canada (orthic black chernozem clay soil); Watrous, SK, Canada (orthic dark brown chernozem soil). The third multi-location trial in 2014 (2 sites) tested 18 hybrids using UM23, UM31, UM46, UM49, UM51, UM52, UM59, UM61, UM64, UM67, UM73, UM109, UM180, UM182, UM184, UM243, UM245 and UM309 as pollinators against the control genotypes Red River 1861 and HYHEAR 1. The two field sites for this trial were at Portage la Prairie, MB, Canada (gleyed rego black soil) and Carman, MB, Canada (orthic black chernozem soil). A single-location yield trial was located at Portage la Prairie, MB, Canada (gleyed rego black soil), testing hybrids using UM27, UM29 UM45, UM56, UM63, UM65, UM70, UM71, UM77, UM181, UM189, UM236, UM247, UM295, UM311 and UM368 as pollinators. Two more single location yield trials were conducted in 2014, both at Carman, Manitoba, Canada (orthic black chernozem soil). The first of these trials tested 16 hybrids, using UM18, UM68, UM178, UM185, UM191, UM241, UM250, UM251, UM296, UM300, UM301, UM303, UM304, UM308, UM317 and UM346 as pollinators. The second tested 15 hybrids, the pollinators for which were UM21, UM55, UM62, UM104, UM112, UM177, UM187, UM188, UM190, UM298, UM324, UM327, UM355, UM357 and UM367. The University of Manitoba cultivars HYHEAR 1 and Red River 1861 were used as control genotypes for all single-location trials.

5.3.4 Determining Genetic Distance

Three sets of criteria including phenotypic data (Chapter 3), GBS SNP data (Chapter 4), and SRAP molecular data (Chapter 4) were used to establish genetic distance between *B. napus* genotypes

and group them into hierarchical clusters. The phenotypic groups (Chapter 3) contained 321 *B. napus* genotypes that were clustered via Ward's minimum variance method using the "dendextend" module within R-Studio 0.99.893 (R-Studio Inc., Boston, MA, USA), a program that utilizes the Rx64 3.2.4 (© The R Foundation for Statistical Computing) statistical software. Euclidean distance between genotypes calculated using Ward's method served as genetic distance in this case. Nine agronomic and quality traits were used as criteria for Ward's clustering, including days to flowering, height, lodging, days to maturity, oil content, erucic acid content, glucosinolate content, saturated fatty acid content, and meal protein content. Details on clustering these criteria are described in more detail within Chapter 3. This clustering resulted in four distinct groups of genotypes (See Chapter 3, Figure 3.5). Of the 321 possible genotypes within the phenotypic groups, 62 were used as the male parent in hybrid combinations with UM388 to determine their hybrid yield.

The genotypic groups created using GBS data (Chapter 4) contained 231 *B. napus* genotypes that were evaluated for single-nucleotide polymorphisms (SNP). Using GBS, 291,782 filtered SNP were identified in the 231 genotypes. The variation in these SNP was used to hierarchically cluster the genotypes using Tassel 9.0 (Ithaca, NY, USA) via the unweighted pair-group method with arithmetic means (UPGMA) into fourteen distinct groups of genotypes (A-N) based on identical-by-state % similarity of all constituent genotypes (See Chapter 4, Figure 4.1). A full listing of the 231 genotypes included in the GBS analysis and which groups they were included in is visible in Appendix B. Of the 231 possible genotypes within the GBS groups, 55 were used as the male parent in hybrid combinations with UM388.

The genotypic groups created with SRAP data (Chapter 4) contained 160 *B. napus* genotypes that were evaluated for sequence-related amplified polymorphisms (SRAP) resulting in the

identification of 225 polymorphic markers (See Chapter 4, Figure 4.2). These markers were used to establish Euclidean genetic distance between genotypes to hierarchically cluster the 160 genotypes into 8 distinct groups (A-H) using the "dendextend" module within R-Studio 0.99.893 (R-Studio Inc., Boston, MA, USA). A full listing of the 160 genotypes included in the GBS analysis and which groups they were included in is visible in Appendix C. Of the 160 possible genotypes within the phenotypic groups, 16 were used as the male parent in hybrid combinations with UM388 over the course of the hybrid trials evaluated in this study.

5.3.5 Statistical Analysis

Statistical analysis was performed in Microsoft® Excel® 2016 (16.0.6828.1019) (Redmond, Washington, United States), with mean and standard deviation calculated using the "average" and "stdev" functions, respectively. Yield data from each site was analyzed individually, calculating the mean, LSD and CV (data not shown). For a site to be considered as valid, a coefficient of variation needed to be less than 15 %. Due to the unbalanced number of sites for each hybrid, the data was standardized by expressing hybrid yield as a % of Red River 1861's yield for each trial. Thus, each site/year or environment contributed equally to the mean for each hybrid (Yau et al., 1991; Bull et al., 1992).

For each clustering method, genetic distance between any given pollinator and UM388 was used for predictive regression analysis. Regression analysis investigating the relationship between genetic distance and hybrid yield for each of the three hierarchical clusters was performed in Microsoft® (Mississauga, Ontario, Canada) Excel® 2016 (16.0.6828.1019) using the scatterplot function. Once the scatterplot had been visualized, the Pearson correlation coefficient was calculated using the "correlation" function in Excel, and the coefficient of determination (r^2) value was calculated by squaring the calculated r-value for % of check yield.

5.4 Results

5.4.1 Hybrid Yield

Mean hybrid yield results for every pollinator-female combination included in the study are in Table 5.1. The highest yielding hybrid (percentage of Red River 1861) used UM61 as a pollinator and yielded a conservative 1569.6 kg/ha which represented a yield of 140.0 % of Red River 1861. The numerically highest yielding hybrid used UM6 as a pollinator yielding a mean 2946.6 kg/ha across six locations; however, this only translated to 106.6 % of the check. For every hybrid, the mean yield value was 1885 kg/ha, with a standard deviation of 584.76 and a range of 1875 kg/ha. When represented as a percentage of Red River 1861, the mean yield value was 80.0 %, with a standard deviation of 24.8 % and a range of 79.6 %. Descriptive statistics of the pooled hybrid yield data revealed a variance of 341922.6 and showed the data to be relatively symmetrical with a skewness of 0.20 and a kurtosis value of -1.52 relative to the height of a standard bell curve.

Table 5.1. Yield for *B. napus* hybrids produced using University of Manitoba pollinators crossed with a single male-sterile female (UM388) across 38 sites in Manitoba, Saskatchewan and Alberta, Canada, within the 2011-2014 growing seasons listed in ascending order of mean hybrid yield.

Pollinator †	Sites ‡	Mean Yield (kg/ha)	% Check¶
UM67	2	1071.6	45.5
UM65	1	1081.1	45.9
UM355	1	1084.9	46.1
UM182	2	1116	47.4
UM178	1	1153.6	48.9
UM245	2	1167.6	49.6
UM49	2	1172.3	49.8
UM177	1	1174.7	49.9
UM311	2	1176.2	49.9
UM191	1	1179	50.1
UM180	2	1185.2	50.3
UM45	1	1202.4	51
UM51	2	1202.9	51.1
UM261	1	1207.3	51.3
UM185	1	1212.1	51.5
UM317	1	1213.9	51.5
UM250	1	1215.4	51.6
UM46	2	1219.3	51.8
UM184	2	1240.8	52.7
UM63	1	1244.4	52.8
UM52	2	1255.2	53.3
UM104	1	1256.4	53.3
UM243	2	1275.3	54.1
UM23	2	1284.4	54.5
UM55	1	1285.5	54.6
UM64	2	1290.6	54.8
UM73	2	1293.6	54.9
UM346	1	1295.3	55
UM109	2	1301.1	55.2
UM309	2	1303.2	55.3
UM62	1	1319.3	56
UM187	1	1320.4	56.1
UM186	1	1323	56.7
UM190	1	1337.8	56.8
UM112	1	1346.5	57.2

Pollinator †	Sites ‡	Mean Yield (kg/ha)	% Check¶
UM59	2	1357	57.6
UM303	1	1370.1	58.2
UM324	1	1371.9	58.2
UM308	1	1406.8	59.7
UM241	1	1410.8	59.9
UM300	1	1418.5	60.2
UM77	1	1429.4	60.7
UM367	1	1449	61.5
UM440	1	1456.7	61.8
UM188	1	1477.7	62.7
UM295	1	1487.1	63.1
UM327	1	1506	63.9
UM311	1	1514.9	64.3
UM189	1	1527	64.8
UM357	1	1553.6	66
UM296	1	1568.9	66.6
UM61	2	1569.6	66.6
UM56	1	1570.5	66.7
UM236	1	1573.7	66.8
UM247	1	1577.2	67
UM368	1	1587	67.4
UM20	1	1621.3	68.8
UM304	1	1647	69.9
UM181	1	1665.1	70.7
UM70	1	1735.5	73.6
UM297	1	1746.9	74.2
UM71	1	1771.4	75.2
UM301	1	1786.5	75.8
UM68	1	1798.2	76.3
UM271	3	1897.1	80.5
UM25	3	2065.8	87.7
UM248	3	2083.2	88.4
UM246	3	2086.6	88.6
UM269	3	2112.1	89.7
UM238	3	2143.8	91
UM7	3	2167	92
UM60	3	2189.4	92.9
UM24	3	2211.2	93.8
UM50	3	2215.1	94

Pollinator †	Sites ‡	Mean Yield (kg/ha)	% Check¶
UM47	3	2222.7	94.4
UM16	3	2230.8	94.7
UM39	11	2264.8	96.2
UM211	9	2269.3	96.3
UM183	3	2277.6	96.7
UM129	3	2287.3	97.1
UM32	12	2290.2	97.2
UM26	3	2294.6	97.4
UM11	3	2314.4	98.3
UM54	3	2341.4	99.4
UM30	3	2356.4	100
UM268	3	2372.2	100.7
UM267	3	2373.1	100.8
UM126	3	2393.8	101.6
UM18	3	2398.5	101.8
UM48	3	2414.8	102.5
UM97	18	2430.8	103.2
UM203	21	2449.8	104
UM9	21	2458.7	104.4
UM5	18	2460.1	104.4
UM29	3	2462.2	104.5
UM53	3	2471.8	104.9
UM22	9	2533.4	107.6
UM17	12	2539.5	107.8
UM14	12	2540.1	107.8
UM58	3	2547.9	108.2
UM35	8	2573.5	109.3
UM256	3	2573.9	109.3
UM4	12	2580.1	109.5
UM43	17	2588.9	109.9
UM57	17	2597.9	110.3
UM13	6	2604.5	110.6
UM98	17	2606.9	110.7
UM44	12	2626.4	111.5
UM34	17	2631.8	111.7
UM15	9	2642.9	112.2
UM2	3	2719.8	115.5
UM1	6	2729.5	115.9
UM40	9	2742.8	116.4
UM8	6	2771.5	117.7

Pollinator †	Sites ‡	Mean Yield (kg/ha)	% Check¶
UM3	6	2785.1	118.2
UM264	3	2790	118.5
UM33	8	2850.8	121
UM37	8	2867.6	121.7
UM259	17	2868.9	121.8
UM36	17	2873.6	122
UM6	6	2946.6	125.1

† Pollen donating male-parent in hybrid combination with male-sterile female genotype UM388.
‡ Number of sites at which hybrid was grown at from 2011-2014.
¶ Check, Red River 1861.

5.4.2 Regression Analysis for Hybrid Yield as a Function of Genetic Distance

Regression analysis was performed to investigate whether genetic distance between parental genotypes estimated using the Ward's method on phenotypic traits correlated with an increase in seed yield (Table 5.2). When all hybrids were analyzed, a Pearson product-moment correlation coefficient (r) of -0.14 was found to be not significant. When only hybrids grown at 3 or more sites were included in the analysis, an r-value of -0.47 was found to be significant at $\alpha = 0.05$ ($p = 0.05$). When only hybrids grown at 5 or more sites were analyzed, an r-value of -0.65 was determined, which was significant at $\alpha = 0.01$ ($p = 0.005$). This result would suggest that as phenotypic-derived genetic distance between pollinator and male-sterile female genotypes increased, the resulting hybrid yield decreased. Based on the significant results for hybrids grown at 3 or more sites, 22 % of the variation visible in these hybrids (r^2) could be attributed to the genetic distance and with the results from five or more combined sites, 42 % of yield variation could be attributed to genetic distance.

Regression analysis was performed to investigate whether genetic distance between the pollinator and UM388 (calculated using GBS) correlated with seed yield (Table 5.2). Genetic distance using GBS marker data was estimated using identical-by-site percent similarity. When all available

hybrids were included in the analysis, an r-value of 0.15 was determined, which was not significant at $\alpha = 0.05$ ($p = 0.13$). When hybrids from three or more sites, were included in the analysis, an r-value of 0.22 was calculated, which was also not significant at $\alpha = 0.05$ ($p = 0.09$). When hybrids from 5 or more sites were included in the analysis, an r-value of 0.36 was determined which was not significant at $\alpha = 0.05$.

Regression analysis was performed to investigate whether genetic distance correlated with seed yield in hybrid combinations between pollinators from the SRAP analysis and the female genotype UM388 (Table 5.2). The analysis yielded a Pearson product-moment correlation coefficient (r) of -0.37, suggesting a negative correlation between genetic distance and yield; however, the r value was not significant due to the small number of tested hybrids (n = 16). Using only hybrid yield combined from 3 or more sites, or 5 or more sites did not yield significant results, with r-values of -0.03 and 0.34, respectively.

Table 5.2. Results of regression analyses comparing genetic distance to hybrid yield data gathered in *Brassica napus* L. Genetic distance was determined using one of three sets of differentiating criteria (phenotypic, GBS, SRAP) and hybrid yield was gathered from yield trials across western Canada in 2011-2014.

	Minimum Sites	**n**	**r**	**r^2**
Phenotypic	1	62	-0.14	0.01
	3	44	-0.47*	0.22*
	5	22	-0.65**	0.42**
GBS	1	54	0.15	0.02
	3	38	0.22	0.05
	5	20	0.36	0.15
SRAP	1	16	-0.37	0.14
	3	9	-0.03	0.00
	5	6	0.34	0.11

* Significant at $\alpha = 0.05$.
* Significant at $\alpha = 0.01$
n, number of hybrids analyzed.
r, correlation coefficient.
r^2, coefficient of determination.

5.5. Discussion

The superior performance of a hybrid over the parental genotypes used in the creation of the hybrid is central to modern plant breeding, and is estimated to be responsible for 15-50 % of yield increases in the last century (Lippman and Zamir, 2007). Explaining this performance advantage is difficult, though there is considerable evidence showing a link between the performance of hybrids and the distinctness, or genetic distance of the parental genotypes (Jesske et al., 2013; Klimenko et al., 2014; Singh, 2015; Xie et al., 2015; Singh, 2015). Multiple mathematical methods exist to estimate genetic distance defining a population by using phenotypic or molecular traits as variables in their analysis (Melchinger and Gumber, 1998). In the present study, three methods of determining genetic distance were tested for their ability to correlate with yield when evaluated individuals were hybridized with the male-sterile UM388. These methods included, 1) a phenotypic approach, using morphological and quality traits to determine the genetic distance between 321 *B. napus* genotypes; 2) a genotypic approach, using 291,782 filtered SNPs gathered via GBS to determine the genetic distance between 231 *Brassica napus* genotypes; 3) another genotypic approach, using 225 SRAP markers to determine genetic distance between 160 *Brassica napus* genotypes.

Regression analysis evaluated the impact of genetic distance and its relationship to hybrid yield data collected in yield trials from 2011 – 2014. A significant correlation between genetic distance and hybrid yield was reported with the use of phenotypic data as differentiating criteria with genetic distance being implicated in as high as 42 % of the variation in observed hybrid yield.

Although there is significant evidence showing a link between genetic distance and hybrid yield (Klimenko et al., 2014; Singh, 2015; Xie et al., 2015; Singh, 2015); there is also a body of evidence showing no, weak, or even negative correlations between the two (Qian et al., 2007; Pankovic et

al., 2009; Yao et al., 2015). For example, Girke et al. (2012) tested 44 lines of *B. napus* against two male-sterile females and found no significant correlation between hybrid yield and genetic distance between the parents. In this study, using phenotypic trait-derived genetic distance, negative correlations of r = -0.47, and r = -0.65 were found. Jesske et al. (2013) reported a negative correlation of r = - 0.29 between genetic distance and hybrid yield when they tested 64 *B. napus* hybrids in Germany and France, showing that a negative correlation between genetic distance and hybrid yield is not without precedent. Using SRAP derived genetic distance, no significant correlation was found between genetic distance and hybrid yield, which does not agree with previous work done by Riaz et al. (2001) and Ahmad and Quiros (2011) which report positive correlations (r = 0.64 and r = .60, respectively) between hybrid yield and SRAP-derived genetic distance. The results from the regression analysis between GBS-derived genetic distance and hybrid yield did not result in a significant correlation when all hybrid yield data were used (r = 0.15, $P < 0.05$), or when hybrid yield data combined from three or more sites (r = 0.22, P = 0.06) were analyzed. While there are no published results comparing GBS-derived genetic distance to hybrid yield in *B. napus*, the results using GBS marker data were numerically similar to previously reported work done using other marker types. Yu et al. (2005) reported a weak positive correlation (r = 0.26, $P < 0.05$) between RAPD marker-derived genetic distance and hybrid yield when they examined yield data from 30 F_1 *B. napus* hybrids. The results from the present study however, support that genetic distance between parental genotypes does have an effect on the yield performance of the F_1; however, are inconsistent in the nature or strength of this effect.

Molecular techniques are able to detect phenotypically silent allelic variation and are detectable in all tissues at all stages of growth (Agrawal and Shrivastava, 2014); this aids in creating the view that molecular markers are more direct and reliable than morphological methods (Ni et al. 2002).

However, usefulness in diversity studies and practical applications in predicting heterosis are not always linked. It has been suggested that not all polymorphisms are created equal contributing to yield increases, as most are located in non-coding regions, or in regions not linked to yield increases (Yu et al., 2005). This linkage disequilibrium between markers used to establish genetic distance and actual trait loci controlling yield may be at fault for weak/no correlations between genetic distance and hybrid yield (Jesske et al. 2013). The sheer volume of SNP markers identified via GBS and used to determine genetic distance may have led to this issue. Filtering the SNP markers for effect on yield may help refine the GBS-developed hierarchical clusters for more accurate correlation with yield. Although its results were not statistically significant in this study, SRAP has the unique quality of specifically targeting the open-reading frame, or DNA fragments which have the potential to code for protein or peptides (Li and Quiros, 2001). Although this does not mean SRAP markers contribute to yield, it does mean they may have a greater potential than random SNPs. With the smallest number of hybrids tested, a larger correlation coefficient was needed to show significance using SRAP markers. With more hybrids tested, correlation values would most likely have been increased or decreased further.

Furthermore, the variability itself in the tested germplasm may have contributed to the lack of large significant correlations. Melchinger (1999) suggested that parents with diverse pedigrees result in inconsistent predictive ability for yield, and weak or no correlations should be expected. While most of the germplasm used in this study was from the University of Manitoba, many of the pollinators were produced with distinct pedigrees. In the GBS analysis, resynthesized and European material was also included; whose exotic nature, relative to the North American genotypes, may have influenced genetic distance. These genotypes may have reduced adaptation to local conditions, which could act to confound evaluation of their yield potential.

In multi-environment trials, where genotypes are tested at multiple different locations and/or different years, a common realization is the appearance of significant genotype x environment interactions (GxE) (Oakey et al., 2016). GxE refers to the variation that will affect the growth and testing of different genotypes in how they respond to different biotic and abiotic environments (Crossa et al., 2013). These different environments can lead to differences in the performance of certain genotypes among the different locations or years of the analysis (Asfaw et al., 2013). Seed yield is particularly sensitive to GxE as it is a complex trait controlled by many loci (Heffner et al., 2009). GxE can potentially complicate the selection for yield and reduce the overall efficiency of the breeding effort, requiring further selection for different production environments (Ebdon and Gauch, 2002). In this study, yield data from 37 possible sites over 4 years was included in analysis, making growth conditions very diverse. To illustrate this, the check mean of each experiment ranged from a minimum of 926.1 kg/ha at Vegreville, SK, in 2013, to a maximum of 3852.0 kg/ha at Rosebank, MB, in 2013, reflecting a range of 2925.9 kg/ha for the same genotype (Red River 1861). This inconsistency in yield data may have resulted in an under or over estimation of the correlation between genetic distance and yield.

In conclusion, imputed genetic distance between parental genotypes was shown to be a significant contributor to hybrid yield when phenotypic traits were used as criteria. However, the proportion of hybrid yield that could be statistically attributed to genetic distance was variable, at 5 % to 42 % depending upon the number of locations utilized. Neither molecular criteria, GBS or SRAP, resulted in a statistically significant correlation between genetic distance and yield. Due to the inconsistent nature of measuring and predicting hybrid yield, the genetic distance between *B. napus* genotypes could contribute to the selection of parental genotypes for future hybrids, but should not be the only consideration.

6.0. General Discussion

High Erucic Acid Rapeseed (HEAR) genotypes of *Brassica napus* L. boast seed oil erucic acid concentrations between 50 and 55 %. HEAR genotypes have been developed at the University of Manitoba since the 1970s, releasing several open-pollinated, and more recently hybrid HEAR cultivars (McVetty et al., 2009; Cuthbert et al., 2011; McVetty et al., 2012; McVetty et al., 2014; Duncan et al., 2016). The desirability of *B. napus* hybrid cultivars lies in the large potential for heterotic yield gain (20 – 50 %) achieved through the combination of diverse parental lines (McVetty et al., 1995). As such, characterizing genetic diversity within the available parental population is essential to superior hybrid production, and this diversity can be established through phenotypic or genotypic traits.

In Chapter 3, 321 genotypes from the University of Manitoba's HEAR germplasm collection, including a male-sterile female and its parental material, were evaluated over the 2013 and 2014 field seasons for variation in nine morphological or seed quality characteristics. The variation within these nine traits was analyzed using principle component analysis and was used to establish Euclidean distance between genotypes. This distance was used to hierarchically cluster the genotypes into groups using the two most widely used statistical methods (Ward's minimum variance method and the unweighted pair group method with arithmetic means) (Khodadadi et al., 2005; Teklewold, 2006; Fahid, 2014; Singh, 2015). In previous research, Ward's method has gained prominence, as the UPGMA method was found to be subject to "chaining" effects, resulting in the overestimation of intra-group affinity and unbalanced group generation (Lombard et al, 2000; Mohammadi and Prasanna, 2003; Khodadadi et al., 2005). In this study the use of the UPGMA method did result in unbalanced group sizes (2, 1, 3 and 315 genotypes respectively), pulling out diverse genotypes, such as the male-sterile female early on in the clustering procedure.

The use of Ward's method however, did result in more balanced group sizes, more suited to being investigated as possible heterotic groups. However, if one was investigating diversity within a closely related population, the UPGMA method may be preferable, as overestimating relatedness and highlighting diversity would be an asset in that case. The PCA plot revealed that two components (PC1, PC2) accounted for 29.4 % and 15.3 % of the variation in the 321 genotypes, respectively and the variation examined in the principle component analysis was then used to successfully clustered into four groups using both methods of hierarchical clustering. Much of this diversity was found in the lodging trait, which boasted a high CV of 0.40 over the 2013 and 2014 growth seasons. This high amount of variation could reflect strong GxE pressure and effect on the lodging score, or could reflect inconsistencies in evaluation. With such a high degree of variation, this study provides evidence that for a two-year research window, lodging is not suited as a differentiating criterion. Lodging is however an important agronomic trait, which may be better suited to evaluation in a controlled experiment with increased fertility and irrigation to promote expression. The successful clustering of the *B. napus* genotypes in this study agrees with similar studies which suggest clustering imputed genotypes based on phenotypic characteristics is sufficient to reveal geographic origin and pedigree diversity within a population (Ana et al., 2009; Ebrahimi et al., 2011). Phenotypic traits are often used as the backbone of any germplasm evaluation (Violle et al 2014), and knowledge of the morphological and quality characteristics of the evaluated genotypes could be valuable in further breeding efforts at the University of Manitoba to exploit and interpret complex traits such as yield, disease resistance and quality traits.

In Chapter 4, two molecular methods, genotyping-by-sequencing (GBS), and sequence related amplified polymorphisms (SRAP), were used to investigate the diversity within 231 and 160 genotypes of *B. napus*, respectively. A male-sterile female and its parental material was included

in both the 231 genotypes evaluated using GBS and the 160 evaluated with SRAP. The next-generation GBS method identified 291,782 SNPs that were used to cluster the 231 genotypes into 12 distinct groups based on identical-by-site variation. GBS markers as differentiating criteria were highly successful in clustering together genotypes of similar origin, such as the HEAR releases Red River 1852, Venus, Neptune, Castor, Hero, Mercury, Stellar, Apollo, MilleniUM 03, and Red River 1826. This success was echoed with the grouping together of the SHEAR genotypes, as well as the genotypes of European origin included in the study. The ability shown in this study to accurately cluster genotypes based on pedigree makes GBS a useful tool in the precision diversity analysis of a breeding population. Similarly, using 24 primer pairs, 230 SRAP markers were used to characterize 160 *B. napus*, which were then used to hierarchically cluster the genotypes into eight groups. In the past, SRAP markers have been shown to be an effective tool in the creation of groupings within populations of *B. napus* (Riaz et al. 2001; Ahmad and Quiros, 2006) and the results of chapter 4 support that conclusion. The results of chapter 4 identified eight potential heterotic groups within the 160 analyzed genotypes, which could be tested for heterotic potential. As such, like the groups defined using morphological and seed quality traits in Chapter 3 the goal of molecular diversity analysis was to use the established groups and group position to predict which genotypes would make desirable pollinators for the male-sterile female genotype, under the assumption that genotypes with greater genetic distance from the female would produce higher performing hybrids.

In Chapter 5, the genetic distances established via morphological and seed quality traits in Chapter 3, and the GBS and SRAP molecular methods in Chapter 4 were tested for their ability to predict hybrid performance when genotypes were crossed with the male-sterile female included in each analysis. In theory, hybrid yield should correlate positively with genetic distance between the

pollinator and female genotypes (Riaz et al., 2001). Using phenotypic data as differentiating criteria, when 45 hybrid genotypes grown at 3 or more sites were analyzed, a statistically significant link was found between genetic distance and hybrid yield, producing an r^2 of 0.22. These results suggest that the genetic distance between the pollinator genotype and the male-sterile line UM388 effects 22 % of the resulting hybrid yield. More significant were the results considering genotypes grown at 5 or more sites. Although belied by a negative correlation between distance and yield, an r^2 of 0.42 was calculated, implying 42 % of difference in hybrid yield for that analysis could be attributed to phenotypic derived genetic distance between parent genotypes. Based on the large increase in r^2 from hybrids evaluated at three to hybrids evaluated five sites, a recommendation could be made when designing future experiments to replicate evaluated genotypes at a minimum of five sites. When GBS or SRAP markers were used to establish genetic distance, no statistically significant link between genetic distanced and hybrid yield was detected. Despite this, the significance of the phenotypic analysis provides evidence that genetic distance between parental genotypes is just one of the factors affecting yield, though the strength of this discovered link was inconsistent. These results were similar to several previous studies in which genetic distance alone was not enough to predict favorable hybrid yield (Qian et al., 2007; Pankovic et al., 2009; Girke et al. 2012; Jesske et al 2013; Yao et al., 2015).

The use of only one male-sterile female in the production of hybrids may have contributed to reduced ability of genetic distance to be an effective predictor of hybrid performance. In previous studies that were unable to link genetic distance with hybrid yield, combining ability proved to be a more effective predictor of performance (Qian et al., 2007; Rameeh, 2010; Xing et al., 2014). In their 2014 study, when looking at the GCA contribution of both parents to agronomic traits in *B. napus,* Xing et al. showed that the maternal genotype had a 55.2 % greater contribution to thousand

seed weight than the male parent. These studies concluded that hybrid performance would be better predicted using combining ability, supplemented by genetic distance. As it relates to the current study, the results of Qian et al., 2007, Rameeh, 2010 and Xing et al., 2014 suggest that it may have been specific combining ability of the female parent, and general combining ability of individual male-parents, not genetic distance which was effecting yield the most. Based on the results from this study, where phenotypically distinct parents produced poor yielding hybrids, a maximum and minimum genetic distance threshold should be defined to limit the number of potential parental combinations.

Based upon this research, genetic distance did not prove to be a consistent predictor of hybrid performance. In this study, 97 hybrids were evaluated and the results may be modified if a greater number of hybrids are evaluated. Moreover, the data for these 97 hybrids were collected from multiple locations over multiple (2011-2014) growing seasons. Even with yield data being standardized, this could have provided a major source of error. To better evaluate the ability of each method of genetic distance's ability to predict hybrid performance, multi-year and multi-location yield trials should be conducted with the same genotypes in each trial with proportionately equal numbers of genotypes in each group to be evaluated. This would perhaps provide more insight into the possible link between genetic distance and hybrid yield in available genotypes.

7.0. Future Research

Methods for predicting performance are critical to an efficient hybrid breeding program, and there is evidence that hybrid performance is proportional to the genetic distance between the parental genotypes (Ali et al., 1995; Qian et al., 2007; Girke et al., 2012; Xing et al., 2014). In this study, significant, yet inconsistent correlations were found between genetic distance and hybrid performance, which agrees with previous evidence showing that genetic distance between parental genotypes does play a role in the creation of superior hybrids (Ali et al., 1995; Riaz et al., 2001; Qian et al., 2007; Girke et al., 2012; Xing et al., 2014) it may be beneficial to take the placement of individual genotypes within the hierarchical cluster analyses conducted in this study into account when selecting parents, to minimize the crossing of related genotypes.

An easy, consistent, and productive method of procuring hybrid seed is imperative for hybrid production, and for many commercial hybrid operations this method is male-sterility, allowing for the exploitation of heterosis by reliably combining parent genotypes (Chen and Liu, 2014). In this study, one male-sterile female genotype was used in the production of every hybrid due to the nature of the male-sterility system chosen. It may be beneficial to include and test other female parents in the hierarchical clusters developed in this study, to broaden the genetic base of crosses and provide more depth to the analyses.

Using genotyping-by-sequencing (GBS) analysis, 291,782 SNP markers were identified in the 230 evaluated genotypes. These markers should have multiple downstream effects in quantitative trait loci (QTL) analysis. Some of these markers could be linked to agronomically important traits, and screening the highest performing hybrids in the study could provide a platform with which to screen available SNP markers for potential QTL associated with yield. This could also be done for other important traits such as oil profile, protein content, or disease resistance. This procedure

could also be performed with SRAP markers, though with only 230 of them, there are less potential opportunites for QTL discovery. For further use in diversity analsysis on the genotypes evaluated in this study, a smaller subset of the available SNP markers could provide a more accurate picture of the molecular diversity within the population. Quality control for SNP markers is often performed through complex filtering algorithms designed to limit extremes in the data. These extremes can refer to Hardy-Weinberg equilibrium (HWE), missing proportion errors, and minor allele frequency, where a minor allele is present in a small subset of your population (Pongpanich et al., 2010). Typically, extremes in HWE reflect gross genotyping error (Teo et al., 2007), missing proportion is indicative of poor genotyping performance (Neale and Purcell, 2008), and minor allele frequency within a population makes them more prone to error (Teo et al., 2007; Neale and Purcell, 2008; Pongpanich et al., 2010). To limit these errors, filters typically use arbitrary thresholds to determine if a SNP is of high enough quality to be reported. Of the 291,782 SNP markers reported using GBS analysis, a smaller subset of those analyzed with stricter standards for HWE extremes, missing proportion or minor alleles could provide a more accurate differentiating criteria. Since, with the 291,782 SNPs, a significant link was not found between genetic distance and hybrid yield, it may be beneficial to re-run the regression analysis multiple times with increasingly stringent levels of quality control for SNP filtering. Once you find the level which results in the highest correlation, evidence could be collected for the proper number of GBS-derived SNPs required to conduct a diversity analysis. For this study, while using GBS markers in hierarchical cluster analysis did not prove significant at $\alpha = 0.05$, they were significant at $\alpha = 0.10$, which suggests with optimization and reducing the number of SNPs, the results could potentially be improved.

In conclusion, the results from Chapter 5 have provided evidence for a link between genetic distance and hybrid yield for the University of Manitoba pollinators investigated using phenotypic criteria. These results will aid in parental selection and identify which pollinators should be selected for producing new hybrids. This should save time and resources in allocating field resources. Furthermore, a huge population of SNP markers were generated using GBS. Paired with agronomic and seed quality data, this research provides the potential for QTL discovery in oil quality, protein content, and yield. Genetic distance has long been hypothesized to be associated with hybrid yield, and the phenotypic results of this study provide evidence to support that conclusion. However, based on inability to produce significant predictive results using molecular marker defined genetic distance, further research is needed to confirm this possible link.

References

Abdel-Ghani, A.H., Lübberstedt, T. 2013. Parent selection – usefulness and prediction of hybrid performance. Chapter 17 in Diagnostics in Plant Breeding. Lübberstedt, T., Varshney, R.K. (eds.). Springer Science, New York, N.Y., USA.

Achtak, H., Oukabli, A., Ater, M., Santoni, S., Kjellberg, F., Khadari, B. 2009. Microsatellite markers as reliable tools for fig cultivar identification. Journal of the American Society for Horticultural Science, 134:624-631.

Agnihotri, A., Prem, D., and Gupta, S. K. 2007. The chronicles of oil and meal quality improvement in rapeseed. In S. K. Gupta (Ed.), Advances in Botanical Research: Rapeseed Breeding (pp. 50–59). San Diego, California, USA: Academic Press.

Agrawal, P.K., Shrivastava, R. 2014. Chapter 4: Molecular markers. pp. 25-39 IN Advances in Biotechnology. Ravi, I., Baunthiyal, M., Saxena, J. (Eds.). New Delhi, India: Springer India.

Ahmad, R., Quiros, C. F., Rahman, H., and Swati, Z. A. 2013. Genetic diversity analyses of *Brassica napus* accessions using SRAP molecular markers. Plant Genetic Resources, 12:14-21.

Ali, M., Copeland, L. O., Elias, S. G., and Kelly, J. D. 1995. Relationship between genetic distance and heterosis for morphological traits in winter canola (*Brassica napus* L.). Theoretical and Applied Genetics, 95:118–121.

Ana, M.J., Ankica, K.S., Dejana, S.P., Radovan, M., Nikola, H. 2009. Phenotypic and molecular evaluation of genetic diversity of rapeseed (*Brassica napus* L.) genotypes. African Journal of Biotechnology, 8:4835-4844.

Anderberg, M.R. 1973. Cluster analysis for applications. Academic Press. New York, NY, USA.

Appelqvist, L. A., Ohlson, R. 1972. Rapeseed: Cultivation, Composition, Processing, and Utilization. Appelqvist, L.A., Ohlson, R (Eds.). Amsterdam, Netherlands: Elsevier Publishing Company.

Asfaw, A., Gurum, F., Alemayehu, F., Rezene, Y. 2012. Analysis of multi-environment grain yield trials in mung bean *Vigna radiate* L. based on GCE biplot in Southern Ethiopia. Journal of Agricultural Science and Technology, 14: 389-398.

Babic, V., Pajic, Z., Prodanovic, S., Babic, M., Filipovic, M. 2010. Visual assessment of sweet maize lines phenotype, according to upov descriptor, as indictor of heterosis. Genetika, 42:313-322.

Badu-Apraku, B., Oyekunle, M., Fakorede, M. a B., Vroh, I., O Akinwale, R., and Aderounmu, M. 2013. Combining ability, heterotic patterns and genetic diversity of extra-early yellow inbreds under contrasting environments. Euphytica, 192:13-433.

Baird, N. A., Etter, P. D., Atwood, T. S., Currey, M. C., Shiver, A. L., Lewis, Z. A., Johnson, E. A. 2008. Rapid SNP discovery and genetic mapping using sequenced RAD markers. PLoS ONE, 3:e3376.

Banga, S. S., and Raman, R. 1998. Biotechnology and Hybrid Breeding. In S. S. Banga and R. Raman (Eds.), Hybrid cultivar development. New Delhi, India: Narosa Publishing House.

Bannerot, T., Boulichard, L., Couderon, Y., and Tempe, J. 1974. Transfer of cytoplasmic male sterility from *Raphanus sativus* to *Brassica oleracea*. pp 52-54 In Eucarpia Meeting on Cruciferae: Vegetable Crops Section. Scottish Horticultural Institute.

Baranyk, D., and Fabry, A. 1999. History of rapeseed (*Brassica napus* L.) growing from middle ages Europe to Canberra. pp. 374 In 10th Annual Rapeseed Conference. Canberra, Australia.

Barnes, P.J. 2007. News review. Lipid Technology, 19:3-7.

Bayer, P.E., Ruperao, P. Mason, A.S., Stiller, J., Chan, C.K.K., Hayashi, S., Long, Y., Meng, J., Sutton, T., Visendi, P., Varshney, R.K., Batley, J., Edwards, D. 2015. High-resolution skim genotyping by sequencing reveals the distribution of crossovers and gene conversions in *Cicer arietinum* and *Brassica napus*. Theoretical and Applied Genetics, 128:1039-1047.

Batley, J., Jewell, E., Edwards, D. 2007. Automated discovery of single nucleotide polymorphism (SNP) and simple sequence repeat (SSR) molecular genetic markers. pp. 473-494 In: Edwards D (ed). Plant bioinformatics. Humana Press, New York.

Beare-Rogers, J. L., and Nera, E. A. 1972. Cardiac fatty acids and histopathology of rats, pigs, monkeys and gerbils fed rapeseed oil. Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology, 41: 793–800.

Beissinger, T. M., Hirsch, C. N., Sekhon, R. S., Foerster, J. M., Johnson, J. M., Muttoni, G., de Leon, N. 2013. Marker density and read depth for genotyping populations using genotyping-by-sequencing. Genetics, 193:1073–1081.

Bell, J. M. 1982. From rapeseed to canola: a brief history of research for superior meal and edible oil. Poultry Science, 61:613-622.

Birchler, J. A., Yao, H., Chudalayandi, S., Vaiman, D., and Veitia, R. A. 2010. Heterosis. The Plant Cell Online, 22:2105-2112.

Blair, M. W., Fregene, M. A., Beebe, S. E., and Ceballos, H. 2007. Marker-assisted selection in common beans and cassava. pp. 81-115 In Marker-assisted selection (MAS) in crops, livestock, forestry and fish: Current status and the way forward. Rome, Italy: FAO.

Boersma, J.G., Pallota, M., Chengdao, L., Buirchell, B., Sivasithamparam, K.P., Yang, H. 2005. Construction of a genetic linkage map using MFLP and identification of molecular markers linked to domestication genes in narrow-leafed lupin (*Lupinus angustifolius* L.). Cellular and Molecular Biology Letters, 10:331-344.

Boulter, G. S. 1983. The history and marketing of rapeseed oil in Canada. Kramer, J.K.G., Saeuer, F.D., and Pidgen, W.J. (Eds.). Toronto, Ontario, Canada: Academic Press.

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: 2633-2635.

Brandle, J. E., and Mcvetty, P. B. E. 1989. Heterosis and combining ability in hybrids derived from Oilseed Rape Cultivars and Inbred Lines. Crop Science, 29:1191-1195.

Breseghello, F., Sorrells, M.E. 2006. Association mapping of kernel size and milling quality in wheat (*Triticum aestivum* L.) cultivars. Genetics, 172:1165-1177.

Brown, G. G. 1999. Unique aspects of cytoplasmic male sterility and fertility restoration in *Brassica napus*. Journal of Heredity, 90:351-356.

Brown, T. B., Cheng, R., Sirault, X. R. R., Rungrat, T., Murray, K. D., Trtilek, M., Borevitz, J. O. 2014. TraitCapture: genomic and environment modeling of plant phenomic data. Current Opinion in Plant Biology, 18:73–79.

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

Budar, P., Pelletier, G. 2001. Male sterility in plants: occurance, determinism, significance and use. Comptes Redus de l'Academie des Sciences – Series III – Science de la Vie, 324:543-550.

Burns, D.A., Ciurczak, E.W. 2008. Handbook of near-infrared analysis. 3rd Edition. Burns, D.A., Ciurczak, E.W. (Eds.). CRC Press: Boca Raton, FL, USA.

Caetano-Anollésm, G. and Gresshoff, P. M. 1994. DNA amplification fingerprinting using arbitrary mini-hairpin oligonucleotide primers. Nature Biotechnology, 12:619-623.

Canola Council of Canada. 2013. Industry overview. Taken from http://www.canolacouncil.org/markets-stats/industry-overview/ on 2/10/2016.

Canola Council of Canada 2014. What is canola? Taken from http://www.canolacouncil.org/oil-and-meal/what-is-canola/ on 2/10/2016.

Canola Council of Canada 2016. Canola variety trials – small plot protocol. Retrieved April 14, 2016 from http://www.canolaperformancetrials.ca/trial-protocol.

Carré, P., Pouzét, A. 2014 rapeseed market, worldwide and in Europe. Oilseeds and Fats, Crops and Lipids, 21:D102.

Ceballos, H., Kawuki, R. S., Gracen, V. E., Yencho, G. C., and Hershey, C. H. 2015. Conventional breeding, marker-assisted selection, genomic selection and inbreeding in clonally propagated crops: a case study for cassava. Theoretical and Applied Genetics, 128:1647-1667.

Chagné, D., Batley, J., Edwards, D., Forster, J.W. 2007. Single nucleotide polymorphism genotyping in plants. pp. 77-94 In: Association mapping in plants. Oraguzie, N., Rikkerink E., Gardiner, S., De Silva, H. (eds.). Springer. New York.

Chalhoub, B., Denoeud, F., Liu, S., Parkin, I. A., Tang, H., Wang, X., Chiquet, J., Belcram, H., Tong, C., Samans, B., Corre, M., da Silva, C., Just, J., Falentin, C., Koh, C.S., Le Clainche, I., Bernard, M., Bento, P., Noel, B., Labadie, K., Alberti, A., Charles, M., Arnaud, D., Guo, H., Daviaud, C., Alamery, S., Jabbari, K., Zhao, M., Edger, P.P., Chelaifa, H., Tack, D., Lassalle, G., Mestiri, I., Schnel, N., Le Paslier, M.C., Fan, G., Renault, V., Bayer, P.E., Golicz, A.A., Manoli, S., Lee, T.H., Thi, V.H., Chalabi, S., Hu, Q., Fan, C., Tollenaere, R., Lu, Y., Battail, C., Shen, J., Sidebottom, C.H., Wang, X., Canaguier, Chauveau, A., Berard, A., Deniot, G., Guan, M., Liu, Z., Sun, F., Lim, Y.P., Lyons, E., Town, C.D., Bancroft, I., Wang, X., Meng, J., Ma, J., Pires, J.C., King, G.J., Brunel, D., Delourme, R., Renard, M., Aury, J.M., Adamns, K.L., Batley, J., Snowdon, R.J., Tost, J., Edwards, D., Zhou, Y., Hua, W., Sharpe, A.G., Paterson, A.H., Guan, C., Wincker, P. 2014. Early allopolyploid evolution in the post-Neolithic *Brassica napus* oilseed genome. Science, 345: 950-953.

Chambers, J.M. 1998. Programming with Data: A Guide to the S Language. Chambers, J.M. (Ed.). Springer Science and Business Media: Berlin, Germany.

Chapman, S.C., Cooper, M., Hammer, G.L., Butler, D.G. 2000. Genotype by environment interactions affecting grain sorghum II. Frequencies of different seasonal patterns of drought stress are related to location effects on hybrid yield. Australian Journal of Agricultural Research, 51:209-222.

Chen, L., and Liu, Y.-G. 2014. Male sterility and fertility restoration in crops. Annual Review of Plant Biology, 65:579-606.

Chen, Z. J. 2013. Genomic and epigenetic insights into the molecular bases of heterosis. Nature Reviews Genetics, 14:471-482.

Coles, S. R., Barker, G., Clark, A. J., Kirwan, K., Jacobs, D., Makenji, K., and Pink, D. 2008. Synthetic mimicking of plant oils and comparison with naturally grown products in polyurethane synthesis. Macromolecular Bioscience, 8:526-532.

Coonrod, D., Brick, M. a., Byrne, P. F., DeBonte, L., and Chen, Z. (2008). Inheritance of long chain fatty acid content in rapeseed (*Brassica napus* L.). Euphytica, 164:583–592.

Cox, D.J., Frey, K.J. 1984. Combining ability and the selection of parents for interspecific oat mating. Crop Science, 24:963-967.

Crossa, J., Beyene, Y., Kassa, S., Perez, P., Hickey, J.M., Chen, C., de los Campos, G., Burgueno, J., Windhousen, V.S., Buckler, E., Jannink, J.L., Lopez-Cruz, M.A., Bobu, R. 2013. Genomic prediction in maize breeding populations with genotyping-by-sequencing. G3 Genes Genomes Genetics, 3:1903-1926.

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

Cuthbert, R. D., Crow, G., and McVetty, P. B. E. 2011. Assessment of seed quality performance and heterosis for seed quality traits in hybrid high erucic acid rapeseed (HEAR). Canadian Journal of Plant Science, 91:837–846.

Dan, Z., Hu, J., Zhou, W., Yao, G., Zhu, R., Zhu, Y., Huang, W. 2016. Metabolic prediction of important agronomic traits in hybrid rice (*Oryza sativa* L.). Scientific Reports, 6:21732.

Davenport, C. B. (1908). Degeneration, albinism and inbreeding. Science, 28:454-455.

Daun. J.K., Clear, J.K., Williams, P.1994. Comparison of three whole seed near-infrared analyzers for measuring quality components of canola seed. Journal of the American Oil Chemists Society, 71:1063-1068.

de Wildt, D. J., and Speijers, G. J. 1984. Influence of dietary rapeseed oil and erucic acid upon myocardial performance and hemodynamics in rats. Toxicology and Applied Pharmacology, 74:99–108.

Debonte, L., Iassonova, D., Liu, L., and Loh, W. 2012. Commercialization of high oleic canola oils, 24:175-177.

De Both, G. 1995. Seedlink^{TM} Technology. pp.64-69 In Proc. 9th Int. Rapeseed Congress, Cambridge, UK.

Debrueil, P.P., Dufour, P., Krejci, E., Causse, M., DeVienne, D., Gallais, A., Charcosset, A. 1996. Organization of RFLP diversity among inbred lines of maize representing the most significant heterotic groups. Crop Science, 36:790-799.

Del Vecchio, A. J. 1997. Food Products Containing Structured Triglycerides. Calgene Inc, assignee. Patent WO1997038588A1. 23 Oct. 1997.

Deng, Z., Li, X., Wang, Z., Jiang, Y., Wan, L., Dong, F., Chen, F., Hong, D., Yang, G. 2016. Map-based cloning reveals the complex organization of the *BnRf* locus and leads to the identification of *BnRf*^{b}, a male sterility gene, in *Brassica napus*. Theoretical and Applied Genetics, 129:53-64.

Deshmukh, S.N., Basu, M.N., Reddy, P.S. 1986. Genetic variation, character association and path coefficient analysis of quantitative traits in Virginia bunch varieties of ground nut. Indian Journal of Agricultural Science, 56:816-821.

Dias, J.S., Monteiro, A.A., Lima, M.B. 1993. Numerical taxonomy of Portuguese tronchuda cabbage and Galega kale landraces using morphological characters. Euphytica, 69:51-68.

Dong, F., Hong, D., Xie, Y., Wen, Y., Dong, L., Liu, P.,Yang, G. 2012. Molecular validation of a multiple-allele recessive genic male sterility locus (*BnRf*) in *Brassica napus* L. Molecular Breeding, 30:1193-1205.

Dow AgroSciences. 2014. Edge™ Herbicide. Edge Herbicide E 20980. Taken from http://msdssearch.dow.com/PublishedLiteratureDAS/dh_0955/0901b80380955a3e.pdf?filepath=ca/pdfs/noreg/010-22258.pdf&fromPage=GetDoc on 21/3/2016.

Dow AgroSciences. 2016. About ProPound™. Taken from http://www.propoundmeal.com/en/about on 2/11/2016.

Downey, R. K., and Harvey, B. L. 1963. Methods of breeding for oil quality in rape. Canadian Journal of Plant Science, 43:271-275.

Downey, R.K., Stefansson, B.R., Stringam, D.I. 1975. Breeding rapeseed and mustard crops. P. 157-178. In: Oilseed and Pulse Crops in Western Canada. Harapiak J.T. (Ed.).

Duncan, R.W., McVetty, P.B.E., Duguid, J., Fox, S., Paulmann, W., Sass, O., Fernando, D.G., Li, G. 2016. HYHEAR 2 Roundup Ready™ high erucic acid, low glucosinolate, hybrid summer rape. Canadian Journal of Plant Science, 10.1139/CJPS-2015-0327.

Duvick, D.N. 2005. The contribution of breeding to yield advances in maize (*Zea mays* L.). Advances in Agronomy, 86:83-145.

East, E.M. 1936. Heterosis. Genetics, 21:375-397.

Ebdon, J.S., Gauch, H.G. 2002. Additive main effect and multiplicative interaction analysis of national turfgrass performance trials. Interpretation of genotype x environment interaction. Crop Science, 42:489-496.

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:e19379.

Eskin, N. A. 2012. Canola Research. In U. Thiyam-Hollander, N. A. Michael Eskin, and B. Mathaus (Eds.), Canola and rapeseed: production, processing, and utilization (pp. 1–19). CRC Press.

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

Fahid, M.G. 2014. Grouping of promising agronomic traits of genotypes and lines of bread wheat using cluster analysis and detection function. Indian Journal of Fundamental and Applied Life Sciences 4:182-185.

Falconer, D. S., 1989: Introduction to quantitative genetics, 3rd edn. Longman Group Ltd, Longman and John Wiley & Sons, New York.

Fan, Z., and Stefansson, B. R. 1986. Influence of temperature on sterility of two cytoplasmic male-sterility-inducing cytoplasms in rape. (*Brassica napus* L.). Canadian Journal of Plant Science, 66:229-234.

Fenwick, G.R. 1982. The assessment of a new protein-source- Rapeseed. Proceedings of the Nutrition Society, 41:277-288.

Fernando, W.G.D., Zhang, X., Amarasinghe, C. 2016. Detection of *Leptosphaeria maculans* and *Leptosphaeria biglobosa* causing blackleg disease in canola from Canadian seed lots and dockage. Plants, 5:12; doi:10.3390/plants5010012.

Ferriol, M., Pico, B., Nuez, F. 2003. Genetic diversity of a germplasm collection of *Cucurbita pepo* using SRAP and AFLP markers. Theoretical and Applied Genetics, 107:271-282.

Fernández-Martínez, J. M., and Velasco, L. 2012. Technological innovations in major world oil crops. World Oil, 1:237–265.

Franzke, A., Lysak, M.A., Al-Shehbaz, I.A., Koch, M.A., Mummenhoff, K. 2011. Cabbage family affairs: the evolutionary history of Brassicaceae. Trends in Plant Science, 16:108-116.Fu,

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

Fu, T.D. 1981. Production and research on rapeseed in the People's Republic of China. Cruciferae Newsletter, 6:6-7.

Fufa, H., Baenziger P.S., Beecher, B.S., Dweikat I., Graybosch R.A., Eskridge K.M. 2005. Comparison of phenotypic and molecular marker-based classifications of hard red winter wheat cultivars. Euphytica, 145:133-146.Geleta, L.F., Labuschange, M.T., Viljoen, C.D. 2004. Relationship between heterosis and genetic distance based on morphological traits and AFLP markers in pepper. Plant Breeding, 123:467-473.

Ghanbarnia, K., Lydiate, D.J., Rimmer, S.R., Li, G., Kutcher, H., Larkan, N. 2012. Genetic mapping of the *Leptosphaeria maculans* avirulence gene corresponding to the LepR1 resistance gene of *Brassica napus*. Theoretical and Applied Genetics, 124:505-513.

Giancola, S., Rao, Y., Chaillou, S., Martin-canadell, S. H. A., Pelletier, G., and Budar, F. 2007. Cytoplasmic suppression of ogura cytoplasmic male sterility in European natural populations of *Raphanus raphanistrum*. Theoretical and Applied Genetics, 114:1333–1343.

Girke, A., Schierholt, A., Becker, H.C. (2012). Extending the rapeseed gene pool with resynthesized *Brassica napus* II: heterosis. Theoretical and Applied Genetics, 124:1017–1026.

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*: e90346.

Goodman, M. M. 1972. Distance analysis in biology. Systemic Zoology, 21:174–186.

Grant, I., Beversdorf, W.D. 1985. Heterosis and combining ability estimates in spring-planted oilseed rape (*Brassica napus* L.). Canadian Journal of Genetics and Cytology, 27:472-478.

Granier, C., Vile, D. 2014. Phenotyping and beyond: modeling the relationships between traits. Current Opinion in Plant Biology, 18:96-102.

Griffing, B. 1956. Concept of general and specific combining ability in relation to diallel crossing systems. Australian Journal of Biological Sciences, 9:463–493.

Guo, F., Yu, H., Tang, Z., Jiang, X., Wang, L., Wang, X., Xu, Q., Deng, X. 2015. Construction of a SNP-based high density genetic map for pummelo using RAD sequencing. Tree Genetics and Genomes, 11:2.

Gupta, S. K., and Pratap, A. 2007. History, origin, and evolution, 45:1–20.

Hagemann, T. L., and Kwan, S. P. 1999. ABI sequencing analysis. Manipulation of sequence data from the ABI DNA sequencer. Molecular Biotechnology, 13:137–52.

Hallaeur A.R., Carena, M., Miranda Filho, J.B. 2010. Quantitative genetics in maize breeding for handbook of plant breeding vol. 6. Springer-Verlag. New York, U.S.A.

Hamam, F., and Shahidi, F. 2005. Structured lipids from high-laurate canola oil and long-chain omega-3 fatty acids. Journal of the American Oil Chemists' Society, 82:731–736.

Han, M. V, and Zmasek, C. M. 2009. phyloXML: XML for evolutionary biology and comparative genomics. BMC Bioinformatics, 10:356.

Hayes, B.J., Bowman, P.J., Chamberlain, A.J., Goddard, M.E. 2009. Invited review: genomic selection in dairy cattle: progress and challenges. Journal of Dairy Science, 92:433-443.

Hayward, A., Mason, A., Dalton-Morgan, J., Zander, M., Edwards, D., Batley, J. 2012. SNP discovery and applications in *Brassica napus.* Journal of Plant Biotechnology, 39:1-12.

Heffner, E. L., Sorrells, M. E., and Jannink, J. L. 2009. Genomic selection for crop improvement. Crop Science, 49:1–12.

Hickey, J. M., Dreisigacker, S., Crossa, J., Hearne, S., Babu, R., Prasanna, B. M.,Gorjanc, G. 2014. Evaluation of genomic selection training population designs and genotyping strategies in plant breeding programs using simulation. Crop Science, 54:1476–1488.

Hougen, F., Bodo, V. 1973. Extraction and methanalysis of oil from whole or crushed rapeseed for fatty acid analysis. Journal of the American Oil Chemists' Society, 50:230-234.

Hu, D., Zhang, P., Hao, W., Sun, Y., Wang, Z., and Chen, C. 2015. SRAP analysis on the genetic relationships of 24 mulberry (*Morus* sp.) accessions in the lower regions of the yellow river. Genetic Resources and Crop Evolution, 62:13–19.

Hu, G., Ablett, G. R., Pauls, K. P., and Rajcan, I. 2006. Environment Effects on Fatty Acid Levels in Soybean Oil. Journal of American Oil Chemists' Society, 83:759–763.

Hu, X., Sullivan-Gilbert, M., Gupta, M., Thompson, S.A. 2006. Mapping of the loci controlling oleic and linolenic acid contents and development of *fad2* and *fad3* allele-specific markers in canola (*Brassica napus* L.). Theoretical and Applied Genetics, 113:497-507.

Huang, X., Huang, T., Hou, G., Li Li. Hou, Y., Lu, Y. 2016. Identification of QTLs for seed quality traits in rapeseed (*Brassica napus* L.) using recombinant inbred lines (RILs). Euphytica, DOI 10.1007/s10681-016-1675-5.

Hulsmann, W. C., Geelhoes-Mierus, M. M., and Houtsmuller, U. M. T. 1979. Alteration of the lipase activities of muscle, adipose tissue, and liver by rapeseed oil feeding of rats. Biochemica et Biophysica Acta, 572:183–187.

Iassonova, Diliara, and Linsen Liu. 2012. Oxidatively-stabilized Fats Containing Very Long-chain Omega-3 Polyunsaturated Fatty Acids. Cargill Incorporated, assignee. Patent EP2713725A1. 9 Apr. 2014.

Ibrahim, I. K., Burdett, B. M., and Hill, R. 1980. The effects of rapeseed meals including those from new low glucosinolate varieties on broiler production. In 6th European Poultry Conference. Poznan, Poland.

Imamura, F., Lemaitre, R. N., King, I. B., Song, X., Steffen, L. M., Folsom, A. R., Mozaffarian, D. 2013. Long-chain monounsaturated fatty acids and incidence of congestive heart failure in 2 prospective cohorts. Circulation, 127:1512–1521.

Jaccoud, D., Peng, K., Feinstein, D., and Kilian, A. 2001. Diversity arrays: a solid state technology for sequence information independent genotyping. Nucleic Acids Research, 29: E25.

Javidfar, F., Ripley, V.L., Roslinsky, V., Zeinali, H., Abdmishani, C. Identification of molecular markers associated with oleic and linolenic acid in spring oilseed rape (*Brassica napus* L.). Plant Breeding, 125:65-71.

James, T., and Martin, J. P. 1952. Gas-liquid partition chromatography: the separation and micro-estimation of ammonia and the methylamines. The Biochemical Journal, 52:238–242.

Jain, A.K., Dubes, R.C. 1988. Algorithms for Clustering Data. Prentice Hall. Englewood, NJ, USA.

Jain, A.K., Murty, M.N., Flynn, P.J. 1999. Data clustering: A review. ACM computing survey, 31:264-323.

Jannink, J.L., Lorenz, A.J., Iwata, H. 2010. Genomic selection in plant breeding: from theory to practice. Briefings in Functional Genomics, 9:166-177.

Jarquín, D., Crossa, J., Lacaze, X., Du Cheyron, P., Daucourt, J., Lorgeou, J., de los Campos, G. 2014. A reaction norm model for genomic selection using high-dimensional genomic and environmental data. Theoretical and Applied Genetics, 127:595–607.

Javed, N., Geng, J., Tahir, M., McVetty, P.B.E., Li, G., Duncan, R.W. 2016. Identification of QTL influencing seed oil content, fatty acid profile and days to flowering in *Brassica napus*. Euphytica, 207:191-211.

Jeffreys, A., Wilson, V., and Thein, S. L. 1985. Individual-specific "fingerprints" of human DNA. Nature, 316, 76–79.

Jeong, J., Park, H., Hyun, H., Kim, J., Kim, H., Oh, H. Il, Kim, H.H. 2015. Effects of glucosinolates from turnip (*Brassica rapa* L.) Root on Bone Formation by Human Osteoblast-Like MG-63 Cells and in Normal Young Rats. Phytotherapy Research, 29:902–909.

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. Theoretical and Applied Genetics, 126:1053–1065.

Jin, L., and Chakraborty, R. 1994. Estimation of genetic distance and coefficient of gene diversity from single-probe multilocus DNA fingerprinting data. Molecular Biology and Evolution, 11:120–127.

Jones, D.F. 1917. Dominance of linked factors as a means of accounting for heterosis. Genetics, 2:466-479.

Jones, B., Sall, J. 2011. JMP Statistical Discovery Software. Wiley Interdisciplinary Reviews: Computational Statistics Volume 3.

Jung, W.S., Moon, H.T., Stanley, H.E. 2008. Dynamics of clustered opinions in complex networks. Journal of Economic Interaction and Coordination, 3:81-88.

Kadir, N. H. A., David, R., Rossiter, J. T., and Gooderham, N. J. 2015. The selective cytotoxicity of the alkenyl glucosinolate hydrolysis products and their presence in Brassica vegetables. Toxicology, 334:59–71.

Kar, B., Nayak, S., Joshi, R.K. 2012 Classification and comparative analysis of *Curcuma longa* L. expressed sequences tags (ESTs) encoding glycine-rich proteins (GRPs). Bioinformation, 8:142-146

Kausch, A. P., Hague, J., Deresienski, A., Tilelli, M., and Nelson, K. 2012. Male sterility and hybrid plant systems for gene confinement. Plant Gene Containment, 1:85–100.

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

Kempthorne, O. 1957. An Introduction to genetic statistics. John Wiley and Sons, Inc. New York, USA pp 468-473.

Khajali, F., and Slominski, B. a. 2012. Factors that affect the nutritive value of canola meal for poultry. Poultry Science, 91:2564–2575.

Kim, Y., Himmelsbach, D.S., Kays, S.E. 2007. ATR-fourier transorm mid-infrared spectroscopy for determination of trans fatty acids in ground cereal products without oil extraction. Journal of Agricultural and Food Chemistry, 55: 4327-4333.

Klimenko, A. I., Azarin, K. V, Gorbachenko, O. F., Markin, N. V, Tikhobaeva, V. E., Kolosov, Y. A., Getmantseva, L. 2014. The relationship between heterosis and genetic distances based on SSR markers in *Helianthus* annus. American Journal of Agricultural and Biological Sciences, 9:270–276.

Kneen, B. 1992. The Rape of Canola. Toronto, Ontario, Canada: Dundurn.

Kondra, Z.P. and Stefansson, B.R. 1970. Inheritance of the major glucosinolates of rapeseed (*Brassica napus*) Meal. Canadian Journal of Plant Science, 50:643-647.

Kumar, N. S., and Gurusubramanian, G. 2011. Random amplified polymorphic DNA (RAPD) markers and its applications. Science Vision, 11:116–124.

Lancashire, P.D., Bleiholder, H., Langeluddecke, P., Stauss, R., van den Boom, T., Weber, E., Witzen-Berger, A. 1991. A uniform decimal code for growth stages of crops and weeds. Annals of Applied Biology, 119: 561–601.

Langer, S. M., Longin, C. F. H., and Würschum, T. 2014. Phenotypic evaluation of floral and flowering traits with relevance for hybrid breeding in wheat (*Triticum aestivum* L.). Plant Breeding, 133:433–441.

Lee, J., Izzah, N.K., Choi, B., Joh, H.J., Lee, S.C., Perumal, S., Seo, J., Ahn, K., Ju. E., Choi, G.J., Nou, I.S., Yu, Y., Yang, T.J. 2015. Genotyping-by-sequencing map permits identification of clubroot resistance QTLs and revision of the reference genome assembly in cabbage (*Brassica oleracea L.*). DNA Research 2015,1-13, doi: 10.1093/dnares/dsv034.

Leeson, S., and Summers, J. D. S. 2005. Commercial Poultry Nutrition (3rd ed.). Guelph, Ontario, Canada: University Books.

Legrand, P., Beauchamp, E., Catheline, D., Pédrono, F., and Rioux, V. 2010. Short chain saturated fatty acids decrease circulating cholesterol and increase tissue PUFA content in the rat. Lipids, 45:975–86.

L'Homme, Y., Stahl, R.J., Li, X., Hameed, A., Brown, G.G. 1997. Brassica nap cytoplasmic male sterility is associated with expression of mtDNA region containing a chimeric gene similar to the pol CMS-associated orf224 gene. Current Genetics, 31:325-335.

L'Homme, Y., Brown, G.G. 1993. Organizational differences between cytoplasmic male sterile and male fertile Brassica mitochondrial genomes are confined to a single transposed locus. Nucleic Acids Research, 21:1903-1909.

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. Theoretical and Applied Genetics, 103:455-461.

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. Proceedings of the National Academy of Sciences, 112:11823–11828.

Li, X., van Loo, E. N., Gruber, J., Fan, J., Guan, R., Frentzen, M., Zhu, L. (2012). Development of ultra-high erucic acid oil in the industrial oil crop *Crambe abyssinica*. Plant Biotechnology Journal, 10:862–870.

Liberatore, K. L., Jiang, K., and Zamir, D. 2013. Heterosis: the case for single-gene overdominance BT - polyploid and hybrid Genomics. Polyploid and Hybrid Genomics, 8:137–152.

Lippman, Z. B., and Zamir, D. 2007. Heterosis: revisiting the magic. Trends in Genetics, 23: 60–66.

Little, K. L., Bohrer, B. M., Stein, H. H., and Boler, D. D. 2015. Effects of feeding high protein or conventional canola meal on dry cured and conventionally cured bacon. Meat Science, 103:28–38.

Liu, Z., Furnier, G.R. 1993. Comparison of allozyme, RFLP, and RAPD markers for revealing genetic variation within and between trembling aspen and bigtooth aspen. Theoretical and Applied Genetics, 87:97-105.

Liu, K.S. 1994. Preparation of fatty acid methyl esters for gas chromatographic analysis of lipids in biological materials. Journal of the American Oil Chemists' Society, 71:1179-1187.

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

Lorenzana, R.E., Bernardo, R. 2009. Accuracy of genotypic value predictions for marker-based selection in biparental plant populations. Theoretical and Applied Genetics, 120:151-161.

Mace, E. S., Rami, J.-F., Bouchet, S., Klein, P. E., Klein, R. R., Kilian, A., Jordan, D. R. 2009. A consensus genetic map of sorghum that integrates multiple component maps and high-throughput Diversity Array Technology (DArT) markers. BMC Plant Biology, 9:13.

Magloire, N. 2005. The genetic, morphological and physiological evaluation of African cowpea genotypes. Thesis submitted to University of the Free State, Bloemfontein under the supervision of Prof. R.L. Verhoeven and Prof. M.T. Labuschagne.

Mahasi, M.J., Kamundia, J.W. 2007. Cluster analysis in rapeseed (*Brassica napus* L.) African Journal of Agricultural Research, 2:409-411.

Mahmood, T., Ekuere, U., Yeh, F., Good, A.G., and Stringam, G.R. 2003. RFLP linkage analysis and mapping genes controlling the fatty acid profile of *Brassica juncea* using reciprocal DH populations. Theoretical and Applied Genetics, 107:283-290.

Mariani, C., De Beuckeleer, M., Treuttner, J., Leemans, J., and Goldberg, R. 1990. A chimeric ribonuclease gene induces male sterility in plants. Nature, 347:737 – 741.

McInnis, A. 2004. From Rapeseed to Canola: The Billion Dollar Success Story. In Winning the prairie gamble: the Saskatchewan story (pp. 1–16). North Battleford, Saskatchewan.

McNeil, L., Bernard, K., MacLeod, M.G. 2004, Food intake, growth rate, food conversion and food choice in broilers fed on diets high in rapeseed meal and pea meal with observations of the resulting poultry meat. British Poultry Science, 45:519-523.

McVetty, P.B.E., Cuthbert, R., Marwede, V., Paulmann, W., Sass, O., Duncan, R.W., Fernando, W.G.D., Li, G., Zelmer, C.D. 2014. HYHEAR 1 hybrid Roundup Ready™ high erucic acid, low glucosinolate summer rape. Canadian Journal of Plant Science, 94:453-455.

McVetty, P.B.E., Fernando, W.G.D., Scarth, R., Li, G. 2006a. Red River 1826 Roundup-Ready™ high erucic acid, low glucosinolate summer rape. Canadian Journal of Plant Science, 86:1179-1180.

McVetty, P.B.E., Fernando, W.G.D., Scarth, R., Li, G. 2006b. Red River 1852 Roundup-Ready™ high erucic acid, low glucosinolate summer rape. Canadian Journal of Plant Science, 86:1181-1182.

McVetty, P.B.E., Fernando, W.G.D., Li, G., Tahir, M., Zelmar, D. 2010. Red River 1997 Roundup Ready™ high erucic acid, low glucosinolate, summer rape. Canadian Journal of Plant Science, 90:711-713.

McVetty, P.B.E., Scarth, R., Rimmer, S.R. 2000. Millenium 03 high erucic acid low glucosinolate summer rape. Canadian Journal of Plant Science, 80:611-612.

McVetty, P. B. E., Scarth, R. 2012. Breeding oil quality for edible and industrial applications in oilseed rape and linseed. In S. K. Gupta (Ed.), Technological innovations in major world oil crops Volume 2 (pp. 105–122). Springer New York.

McVetty, P.B.E., Scarth, R. 2007 Breeding Herbicide-Tolerant Oilseed Rape Cultivars. Advances in Botanical Research, 45:234-270

McVetty, P.B.E., Duncan, R.W., Fernando, W.G.D., Li, G., Zelmer, C.D. 2012. Red River 1861™ high erucic acid, low glucosinolate summer rape. Canadian Journal of Plant Science, 92:1407-1409.

McVetty, P.B.E., Fernando, W.G.D., Tahir, M., Zelmer, C. 2009. High erucic, low glucosinolate rapeseed (HEAR) cultivar development in Canada. In: Hou, C.T., Shaw, J.F., editors. Biocatalysis and agricultural biotechnology. Boca Raton: Taylor and Francis; 2009. p. 43-62.

Melchinger, A.E., Gumber, R.K. 1998. Overview of Heterosis and Heterotic Groups in Agronomic Crops. pp. 29-44. In: Concepts and Breeding of Heterosis in Crop Plants. Volenec, J.J., Kral, D.M., Viney, M.K. (eds.) Crop Science Society of America, USA,

Melchinger, A.E., Messmer, M.M., Woodman, W.L., Lamkey, K.R., 1991. Diversity and relationships among US maize inbreds revealed by restriction length fragment polymorphisms. Crop Science, 31:669-678.

Melchinger, A. E., Utz, H. F., Piepho, H. P., Zeng, Z. B., and Schön, C. C. 2007. The role of epistasis in the manifestation of heterosis: A systems-oriented approach. Genetics, 177:1815–1825.

Memon, S., Baloch, M. J., Baloch, G. M., and Jatoi, W. A. 2015. Combining ability through line × tester analysis for phenological, seed yield, and oil traits in sunflower (*Helianthus annuus* L.). Euphytica, 204:199–209.

Metzker, M.L. 2010. Sequencing technologies – the next generation. Nature Reviews, 11:31-46.

Meuwissen, T.H.E., Hayes, B.J., Goddard, M.E. 2001. Prediction of total genetic value using genome-wide dense marker maps. Genetics, 157:1819-1829.

Meyer, R.S., Purugganan, M.D. 2013. Evolution of crop species: genetics of domestication and diversification. Nature Review Genetics, 14:840-852.

Milligan, G.W., Cooper, M. 1985. An examination of procedures for determining the number of clusters in a data set. Psychometrika, 50:159-179.

Mikíc, A. 2016. Reminiscences of the cultivated plants early days as treasured by ancient religious traditions: the mustard crop (*Brassica* spp. and *Sinapis* spp.) in earliest Christian and Islamic texts. Genetic Resources and Crop Evolution, 63:1-6.

Mohammadi S.A., Prasanna, B.M. 2003. Analysis of genetic diversity in crop plants – salient statistical tools and considerations. Crop Science, 43:1235-1248.

Morozova, O., Marra, M.A. 2008. Applications of next-generation sequencing technologies in functional genomics. Genomics, 92:255-264.

Muranty, H., Jorge, V., Bastien, C., Lepoittevin, C., Bouffier, L., and Sanchez, L. 2014. Potential for marker-assisted selection for forest tree breeding: Lessons from 20 years of MAS in crops. Tree Genetics and Genomes, 10:1491–1510.

Neale, B.M., Pucell, S. 2008. The positives, protocols, and perils of genome-wide association. American Journal of Medical Genetics, 147:1288-1294.

Nesi, N., Delourme, R., Brégeon, M., Falentin, C., and Renard, M. 2008. Genetic and molecular approaches to improve nutritional value of Brassica napus L. seed. Comptes Rendus - Biologies, 331:763–771.

Ni, Z., Kim, E.-D., Ha, M., Lackey, E., Liu, J., Zhang, Y., Chen, Z. J. 2009. Altered circadian rhythms regulate growth vigour in hybrids and allopolyploids. Nature, 457:327–331.

Ni, J., Colowit, P.M., Mackill, D.J. 2002. Evaluation of genetic diversity in rice subspecies using microsatellite markers. Crop Science, 42:601-607.

Niedziela, A., Mankowski, D., Bednarek, P.T. 2015. Diversity Arrays Technology-based PCR markers for marker assisted selection of aluminum tolerance in triticale (*x Triticosecale* Wittmack). Molecular Breeding, 35:209.

Oakey, H., Cullis, B., Thompson, R., Comadran, J., Halpin, C., Waught, R. 2016. Genomic selection in multi environment crop trials. G3 Genes Genomes Genetics, 6:1313-1326.

Obiadalla-Ali, H. A., Mohamed, N. E. M., and Khaled, A. G. A. 2014. Inbreeding, outbreeding and (RAPD) markers studies of faba bean (*Vicia Faba* L.) Crop Journal of Advanced Research, 6:859-686.

Ogura, H. 1968. Memoirs of the faculty of agriculture, Kagoshima University, 6:39-78.

Ohlson, R. 1972. Projection and prospects for rapeseed and mustard seed. Journal of the American Oil Chemists' Society, 49:522-526.

Padilla, G., Cartea, M.E., Ordas, A. 2007. Comparison of several clustering methods in grouping kale landraces. Journal of the American Society for Horticultural Science, 132:387-395.

Padmakar, B., Kanupriya, C., Latha, P. M., Prashant, K. S., Dinesh, M. R., Sailaja, D., and Aswath, C. 2015. Development of SRAP and SSR marker-based genetic linkage maps of guava (*Psidium guajava* L.). Scientia Horticulturae, 192:158–165.

Pan, T. Zeng, F., Wu, S. 1988. A study on breeding and application GMS line of low erucic acid in rapeseed (*B. napus*). Chinese Journal of Oil Crop Sciences, 3:5-8.

Parentoni, S.N., Magalhae, J.V., Pacheco, C.A.P., Santos, M.X., Abadie, T., Gama, E.E.G., Guimaraes, P.E.O., Meirelles, W.F., Lopes, M.A., Vasconcelos, M.J.V., Paiva, E. (2001). Heterotic groups based on yield-specific combining ability data and phylogenetic relationship determined by RAPD markers for 28 tropical maize open pollinated varieties. Euphytica, 121:197-208.

Pearson, K. 1901. On lines and planes of closest fit to systems of points in space. Philosophical Magazine, 2:559-572.

Pelletier, G., and Budar, F. 2015. Brassica ogu-INRA cytoplasmic male sterility: an example of successful plant somatic fusion for hybrid seed production. pp.199-215 In somatic genome manipulation. Li, Q. (Ed.). Springer Scientific: Berlin, Germany.

Peng, X., Ji, Q., Fan, S., Zhang, Y., and Zhang, J. 2015. Genetic diversity in populations of the endangered medicinal plant *Tetrastigma hemsleyanum* revealed by ISSR and SRAP markers: implications for conservation. Genetic Resources and Crop Evolution, 62:1069–1078.

Perry, D. J., Isabel, N., and Bousquet, J. 1999. Sequence-tagged-site (STS) markers of arbitrary genes: the amount and nature of variation revealed in Norway spruce. Heredity, 83:239-48.

Phillips, R. L. 2010. Mobilizing Science to Break Yield Barriers. Crop Science, 50:99-108.

Piazza, G.J., Foglia, T.A. 2001. Rapeseed oil for oleochemical uses. European Journal of Lipid Science and Technology, 103:405–454.

Poland, J., Rife, T. W. 2012. Genotyping-by-sequencing for plant breeding and genetics. Plant Genome, 5:92–102.

Pongpanich, M., Sullivan, P.F., Tzeng J.Y. 2010. A quality control algorithm for filtering SNPs in genome-wide association studies. Bioinformatics, 26:1731-1737.

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*). Theoretical and Applied Genetics, 111:736–746.

PubChem. 2004a. PubChem: Erucic Acid. http://doi.org/5281116.

PubChem. 2004b. PubChem: Oleic Acid. http://doi.org/445639.

PubChem. 2006. PubChem: Glucosinolates. http://doi.org/9576416.

Qian, W., Sass, O., Meng, J., Li, M., Frauen, M., Jung, C. 2007. Heterotic patterns in rapeseed (*Brassica napus* L.): crosses between spring and Chinese semi-winter lines. Theoretical and Applied Genetics, 115:27-34.

Qian, W., Chen, X., Fu, D., Zou, J., Meng, J. 2005. Intersubgenomic heterosis in seed yield potential observed in a new type of *Brassica napus* introgressed with partial *Brassica rapa* genome. Theoretical and Applied Genetics, 110:1187-1194.

Queller, D. C., Strassmann, J. E., and Hughes, C. R. 1993. Microsatellites and kinship, 8:285–288.

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, 179:1547–1558.

Ragot, M., Hoisington, D.A. 1993. Molecular markers for plant breeding: comparisons of RFLP and RAPD genotyping costs. Theoretical and Applied Genetics, 86:975-984.

Rahman, H., Rahman, R., Nelson, M. N., Aslam, M. N., Rajasekaran, R., Wratten, N., Schondelmaier, J. (2012). Diversity array technology markers: genetic diversity analyses and linkage map construction in rapeseed (*Brassica napus* L.). DNA Research: An International Journal for Rapid Publication of Reports on Genes and Genomes, 19:51–65.

Rahman, M., Sun, Z., McVetty, P.B.E., 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. Theoretical and Applied Genetics, 117:895-904.

Rai, B., Gupta, S. K., and Pratap, A. 2007. Breeding methods. Advances in Botanical Research, 45:21–48.

Ramchiary, N., Suhyoung, P., and Yang, P. L. 2011. Classical breeding and genetic analysis of vegetable Brassicas. In J. Sadowski and K. Chittaranjan (Eds.). CRC Press.

Rameeh, V. 2010. Combining ability and factor analysis in F_2 diallel crosses of rapeseed varieties. Plant Breeding and Seed Sciences, 62:73-83.

Reidelsheimer C., Czedik-Eysenberg, A., Grieder, C., Lisec, K., Technow, F., Sulpice R., Altmann, T., Stitt, M., Willmitzer, L., Melchinger, A.E. 2012. Genomic and metabolic prediction of complex heterotic traits in hybrid maize. Nature Genetics, 44:217-222.

Riaz, A., Li, G., Queresh, Z., Swati, M.S., 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, 120:411-415.

Rieseberg, L.H. 2000. Crossing relationships among ancient and experimental sunflower hybrid lineages. Evolution, 54:859-65

Ronaghi, M., Karamohamed, S., Pettersson, B., Uhlen, M., Nyren, P. 1996. Real-time DNA sequencing using detection of pyrophosphate release. Analytical biochemistry, 242:84-89.

Roy, D. 2012. Plant Breeding: A Biometrical Approach (1st ed.). Pangbourne, UK.: Alpha Science.

Ruster, W. 2015. Who benefits from intellectual property rights for agricultural innovation? The case of ogura oilseed rape in France. WTO – Geneva 10/6/2015.

Ryder, P., McKeown, P. C., and Spillane, C. 2014. Epigenetics and heterosis in crop plants. In Epigenetics in plants of agronomic importance: fundamentals and applications (pp. 13–31). Springer International.

Saiki, R. K., Scharf, S., Faloona, F., Mullis, K. B., Horn, G. T., Erlich, H. A., and Arnheim, N. 1985. Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. Science, 1350–1354.

Sakamoto, K., Kusaba, M., Nishio, T. 2000. Single-seed PCR-RFLP analysis for the identification of S haplotypes in commercial F_1 hybrid cultivars of broccoli and cabbage. Plant Cell Reports, 19:400-406.

Salmon, A., Clotault, J., Jenczewski, E., Chable, V., Manzanares-Dauleux, M. 2008. *Brassica oleracea* displays a high level of DNA methylation polymorphism. Plant Science, 174:61-70.

SAS Institute Inc. 1983. SAS® Technical Report A-108, Cubic Clustering Criterion. 56pp. SAS Institute Inc. Cary, North Carolina, USA.

SAS Institute Inc. 2014. JMP® 11 Multivariate Methods. Cary, NC: SAS Institute Inc.

Schlötterer, C. 2000. Evolutionary dynamics of microsatellite DNA. Chromosoma, 109:365-371.

Schnable, P. S., Springer, N. M. 2013. Progress toward understanding heterosis in crop plants. Annual Review of Plant Biology, 64:71–88.

Schnable, P. S., Wise, R.P. 1998. The molecular basis of cytoplasmic male sterility and fertility restoration. Trends in Plant Science, 3:175-180.

Schuster, W., Michael, J. 1976. Investigation of inbreeding depression and heterosis effects in rape (*B. napus*). Z Pflanzenzuecht, 77:56-65.

Sernyk, J.L., Stefansson B.R. 1983. Heterosis in summer rape (*Brassica napus* L.). Canadian Journal of Plant Science, 63:407-413.

Shah, Dilip M., Stephen G. Rogers, Robert B. Horsch, and Robert T. Fraley. 1990. Glyphosate-resistant plants. Monsanto Company, assignee. Patent US4940835A.

Shahidul, I., Yang, H., Yan, G. 2013. Molecular markers for genetics and plant breeding: the MFLP marker system and its application in narrow-leafed lupin (*Lupinus angustifolius*). Methods in Molecular Biology, 1069:179-201.

Shendure, J., Hanlee, J. 2008. Next-generation DNA sequencing. Nature Biotechnology, 26:1135-1145.

Shiga, T., Baba, S. 1973. Cytoplasmic male sterility in oil seed rape, *Brassica napus* L. and its utilization to breeding. Japanese Journal of Breeding, 23:187-197.

Shiga, T. 1976. Cytoplasmic male sterility and its utilization in rapeseed (*Brassica napus* L.). Japan Agricultural Research Quarterly, 10:178-182

Shull, G. 1948. What is "heterosis"? Genetics, 33, 439. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1209417/.

Singh, N., Yadava, D. K., Vasudev, S., Singh, R., Giri, S. C., Dass, B.,Prabhu, K. V. 2015. Combining ability and heterobeltiosis for yield and yield contributing traits in high quality oil Indian mustard (*Brassica juncea*) genotypes. Indian Journal of Agricultural Sciences, 85:498–503.

Singh, P. 2015. Genetic distance, heterosis and combing ability studies in maize for predicting F1 hybrid performance. Society for the Advancement of Breeding Research in Asia and Oceania Journal of Breeding and Genetics, 47:21-28.

Sokal, R.R., Michener, C.D. 1958. Statistical Method for Evaluating Systemic Relationships. The University of Kansas Science Bulletin, 38:1409-1438.

Sonntag, N.O.V. 1995. Industrial utilization of long-chain fatty acids and their derivatives. pp. 339-352 In: Kimber, D.S., McGregor, D.I (Eds.) Brassica oilseeds. CAB International, Wallingford, Oxon, UK.

Spindel J., Wright M., Chen C., Cobb J., Gage J., Harrington S., et al. 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. Theoretical and Applied Genetics, 126:2699–2716.

Snyder, C. L., and Weselake, R. J. (2009). Modification of Seed Oil formation in *Brassica* oilseed species. pp. 101–124. In: Canola and rapeseed: production, processing, food quality and nutrition. Thiyam-Hollander, U., Eskin, N.A.M., Matthaus, B. (Eds.). CRC Press: Boca Raton, FL, USA.

Stefansson, B.R., Hougen, F.W. 1964. Selection of rape plants (*Brassica napus*) with seed oil practically free from erucic Acid. Canadian Journal of Plant Science, 44:359-364.

Stefansson, B.R. 1972. Tower summer rape. Canadian Journal of Plant Science, 55:343-344.

Stiewe, G., Pleines, S., Coque, S., and Gielen, J. 2008. New hybrid system for *Brassica napus*. Germany: European Patent Office.

Syed, A. 2012. Future of omega-9 oils. pp. 79–100 In: Canola and rapeseed: production, processing, and utilization. Thiyam-Hollander, U., Eskin, N.A.M., Matthaus, B. (Eds.) CRC Press: Boca Raton, FL, USA.

Teklewold, A., and Becker, H. C. 2006. Comparison of phenotypic and molecular distances to predict heterosis and F1 performance in Ethiopian mustard (*Brassica carinata* A. Braun). Theoretical and Applied Genetics, 112:752–759.

Teo, Y.Y., Fry, A.E., Clark, T.G., Tai, E.S., Seielstad, M. 2007. On the usage of HWE for identifying genotyping errors. Annals of Human Genetics, 71:701-703.

Thompson, K.F. (1972). Cytoplasmic male sterility in oil-seed rape plants. Heredity, 29:253-257

Tkachuk, R. (1981). Oil and protein analysis of whole rapeseed kernels by near infrared reflectance spectroscopy. Journal of the American Oil Chemists Society', 58:819-822.

Tracy, W. I., and Chandler, M. A. 2006. The historical and biological basis of the concept of heterotic patterns in corn belt dent maize. pp. 219–233 In: Plant breeding: The Arnel R. Hallauer International Symposium. K. R. Lamkey and M. Lee (Eds.). Blackwell Publishing: Hoboken, NJ, USA.

Tripathi, M. K., and Mishra, A. S. 2007. Glucosinolates in animal nutriton. A review. Animal Feed Science and Technology, 132:1-27.

Troyer, F., and Wellin, E. J. 2009. Heterosis decreasing in hybrids: Yield test inbreds. Crop Science, 49:1969–1976.

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

USDA. 2016. Classification for kingdom plantae down to genus Brassica L. Retrieved from http://plants.usda.gov/java/ClassificationServlet?source=displayandclassid=BRASS2 on February 10, 2016.

Van Eijk, T., Van Schriek, G. M. V., Josephus, M., and Sorenson, A. P. 2009. Method for high-throughput AFLP-based polymorphism detection. United States: United States Patent Office.

Vavilov, N.I. 1951. The origin, variation, immunity, and breeding of cultivated plants. Selected writings translated from Russian by K. Starr Chester, Chronica Botanica, 13:1-6.

Vehov V.N., Gubanov I.A., Lebedeva G.F. 1978. Cultural plants of the USSR. Moscow.

Violle, C., Reich, P.R., Pacala, S.W., Enquist, B., Kattge, J. 2014. The emergence and promise of functional biogeography. PNAS 111:13690-13696

Wu, X., Li, Y., Shi, Y., Song, Y., Wang, T., Huang, Y., and Li, Y. 2014. Fine genetic characterization of elite maize germplasm using high-throughput SNP genotyping. Theoretical and Applied Genetics, 127:621–631.

Wu, X., Chen, B., Lu, G., Wang, H., Xu, K., Guizhan, G., Song, Y. 2009. Genetic diversity in oil and vegetable mustard (*Brassica juncea*) landraces by SRAP markers. Genetic Resources and Crop Evaluation, 56:1011-1022.

Xiao, B., Tan, Y., Long, N., Tong, Tong, Z., Dong, Y., Li, Y. 2015. SNP-based genetic linkage map of tobacco (*Nicotiana tabacum* L.) using next-generation RAD sequencing. Journal of Biological Research, 22:11

Xie, F., He, Z., Esguerra, M. Q., Qiu, F., and Ramanathan, V. 2014. Determination of heterotic groups for tropical Indica hybrid rice germplasm. Theoretical and Applied Genetics, 127:407-417.

Xie, Q., Liu, Z., Wang, S., and Li, Z. 2015. Genetic diversity and phylogenetic relationships among five endemic Pinus taxa (Pinaceae) of China as revealed by SRAP markers. Biochemical Systematics and Ecology, 62:115–120.

Wang, H.M., Ketela, T., Keller, W.A., Gleddie, S.C., Brown, G.G. 1995. Genetic correlation of the orf224/atp6 gene region with Polima CMS in Brassica somatic hybrids. Plant Molecular Biology, 24:801-807.

Williams, J.G., Kubelik, A.R., Livak, K.J., Rafalski, J.A., Tingey, S.V. 1990. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucleic Acids Research 18: 6531-6535.

Yang, H., Shankar, M., Buirchell, B.J., Sweetingham, M.W., Caminero, C., Smith, P.M.C. 2002. Development of molecular markers using MFLP linked to a gene conferring resistance to diaporthe toxica in narrow-leafed lupin (*Lupinus angustifolius* L.). Theoretical and Applied Genetics, 105:265–270.

Yang, H., Li, C., Lam, H.-M., Clements, J., Yan, G., and Zhao, S. 2015. Sequencing consolidates molecular markers with plant breeding practice. Theoretical and Applied Genetics 128(5):779–795.

Yao, J., Li, H., Ye, J., and Shi, L. 2015. Relationship between parental genetic distance and offspring's heterosis for early growth traits in Liriodendron: implication for parent pair selection in cross breeding. New Forests, 47:163-177.

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

Zabeau, M., Vos, P. 1993. Selective restriction fragment amplification: a general method for DNA fingerprinting. European Patent Office. Publication 0534858A1 bulletin 93/13.

Zarhloul, M.K., Stoll, C., Lühs, W., Syring-Ehemann, A., Hausmann, L., Töpfer, R., Friedt, W. 2006, Breeding high-stearic oilseed rape (*Brassica napus*) with high- and low- erucic background using optimized promotor-gene constructs. Molecular Breeding, 3:241-251.

Zeng, X., Zhu, L., Chen, Y., Pu, Y., Wen, J., Yi, B., Shen, J., Ma, C., Tu, J., Fu, T. (2011) Identification, fine mapping and characterization of a dwarf mutant (bnaC.dwf) in *Brassica napus*. Theoretical and Applied Genetics, 122:421-428.

Zhang, J., Wu, M., Yu, J., Li, X., Pei, W. 2016. Breeding potential of introgression lines developed from interspecific crossing between uplant cotton (*Gossypium hirsutum*) and *Gossypium barbadense*: Heterosis, combining ability and genetic effects. PLoS One:10.1371/journal.pone.0143646.

Zhao, J., Jian, J., Liu, G., Wang, J., Lin, M., Ming, Y., Liu, Z., Chen, Y., Liu, X., Liu, M. 2014 Rapid SNP discovery and a RAD-based high density linkage map in jujube (*Ziziphus* Mill.) PLoS One, 10:e109850.

Zheng, H., Duan, H., Hu, D., Wei, R., and Li, Y. 2015. Sequence-related amplified polymorphism primer screening on Chinese fir (*Cunninghamia lanceolata* (Lamb.) Hook). Journal of Forestry Research, 26:101–106.

Zuo, Q. S., Zhou, G. S., Yang, S. F., Yang, Y., Wu, L. R., Leng, S. H., Wu, J. S. 2015. Effects of nitrogen rate and genotype on seed protein and amino acid content in canola. The Journal of Agricultural Science, 154:438-455.

Appendix A. Complete list of *Brassica napus* genotypes used in phenotypic[†] and molecular[‡] hierarchical cluster analysis to create heterotic groups.

Identity	Quality[§]	Origin[¶]
UM1	HEAR	U of M
UM10	HEAR	U of M
UM100	HEAR	U of M
UM101	HEAR	U of M
UM102	HEAR	U of M
UM103	HEAR	U of M
UM104	HEAR	U of M
UM105	HEAR	U of M
UM106	HEAR	U of M
UM107	HEAR	U of M
UM108	HEAR	U of M
UM109	HEAR	U of M
UM11	HEAR	U of M
UM110	HEAR	U of M
UM111	HEAR	U of M
UM112	HEAR	U of M
UM113	HEAR	U of M
UM114	HEAR	U of M
UM115	HEAR	U of M
UM117	HEAR	U of M
UM118	HEAR	U of M
UM119	HEAR	U of M
UM12	HEAR	U of M
UM120	HEAR	U of M
UM121	HEAR	U of M
UM122	HEAR	U of M
UM123	HEAR	U of M
UM124	HEAR	U of M
UM125	HEAR	U of M
UM126	HEAR	U of M
UM127	HEAR	U of M
UM128	HEAR	U of M
UM129	HEAR	U of M
UM13	HEAR	U of M
UM130	HEAR	U of M
UM131	HEAR	U of M
UM132	HEAR	U of M
UM133	HEAR	U of M

Project Name	Quality[§]	Origin[¶]
UM134	HEAR	U of M
UM135	HEAR	U of M
UM136	HEAR	U of M
UM137	HEAR	U of M
UM138	HEAR	U of M
UM139	HEAR	U of M
UM14	HEAR	U of M
UM140	HEAR	U of M
UM141	HEAR	U of M
UM142	HEAR	U of M
UM143	HEAR	U of M
UM144	HEAR	U of M
UM145	HEAR	U of M
UM146	HEAR	U of M
UM147	HEAR	U of M
UM148	HEAR	U of M
UM149	HEAR	U of M
UM15	HEAR	U of M
UM150	HEAR	U of M
UM151	HEAR	U of M
UM152	HEAR	U of M
UM153	HEAR	U of M
UM154	HEAR	U of M
UM155	HEAR	U of M
UM156	HEAR	U of M
UM157	HEAR	U of M
UM158	HEAR	U of M
UM159	HEAR	U of M
UM16	HEAR	U of M
UM160	HEAR	U of M
UM161	HEAR	U of M
UM162	HEAR	U of M
UM163	HEAR	U of M
UM164	HEAR	U of M
UM165	HEAR	U of M
UM166	HEAR	U of M
UM167	HEAR	U of M
UM168	HEAR	U of M
UM169	HEAR	U of M
UM17	HEAR	U of M
UM170	HEAR	U of M
UM171	HEAR	U of M

Project Name	Quality[§]	Origin[¶]
UM172	HEAR	U of M
UM173	HEAR	U of M
UM177	HEAR	U of M
UM178	HEAR	U of M
UM179	HEAR	U of M
UM18	HEAR	U of M
UM180	HEAR	U of M
UM181	HEAR	U of M
UM182	HEAR	U of M
UM183	HEAR	U of M
UM184	HEAR	U of M
UM185	HEAR	U of M
UM186	HEAR	U of M
UM187	HEAR	U of M
UM188	HEAR	U of M
UM189	HEAR	U of M
UM19	HEAR	U of M
UM190	HEAR	U of M
UM191	HEAR	U of M
UM192	HEAR	U of M
UM193	HEAR	U of M
UM194	HEAR	U of M
UM195	HEAR	U of M
UM196	HEAR	U of M
UM197	HEAR	U of M
UM198	HEAR	U of M
UM199	HEAR	U of M
UM2	HEAR	U of M
UM20	HEAR	U of M
UM200	HEAR	U of M
UM201	HEAR	U of M
UM202	HEAR	U of M
UM203	HEAR	U of M
UM204	HEAR	U of M
UM205	HEAR	U of M
UM206	HEAR	U of M
UM207	HEAR	U of M
UM208	HEAR	U of M
UM209	HEAR	U of M
UM21	HEAR	U of M
UM210	HEAR	U of M
UM211	HEAR	U of M

Project Name	Quality[§]	Origin[¶]
UM212	HEAR	U of M
UM213	HEAR	U of M
UM214	HEAR	U of M
UM215	HEAR	U of M
UM216	HEAR	U of M
UM217	HEAR	U of M
UM218	HEAR	U of M
UM219	HEAR	U of M
UM22	HEAR	U of M
UM220	HEAR	U of M
UM221	HEAR	U of M
UM222	HEAR	U of M
UM223	HEAR	U of M
UM224	HEAR	U of M
UM225	HEAR	U of M
UM226	HEAR	U of M
UM227	HEAR	U of M
UM228	HEAR	U of M
UM229	HEAR	U of M
UM23	HEAR	U of M
UM230	HEAR	U of M
UM231	HEAR	U of M
UM232	HEAR	U of M
UM233	HEAR	U of M
UM234	HEAR	U of M
UM235	HEAR	U of M
UM236	HEAR	U of M
UM238	HEAR	U of M
UM239	HEAR	U of M
UM24	HEAR	U of M
UM240	HEAR	U of M
UM241	HEAR	U of M
UM242	HEAR	U of M
UM243	HEAR	U of M
UM244	HEAR	U of M
UM245	HEAR	U of M
UM246	HEAR	U of M
UM247	HEAR	U of M
UM248	HEAR	U of M
UM249	HEAR	U of M
UM25	HEAR	U of M
UM250	HEAR	U of M

Project Name	Quality[§]	Origin[¶]
UM251	HEAR	U of M
UM252	HEAR	U of M
UM253	HEAR	U of M
UM254	HEAR	U of M
UM255	HEAR	U of M
UM256	HEAR	U of M
UM257	HEAR	U of M
UM258	HEAR	U of M
UM259	HEAR	U of M
UM26	HEAR	U of M
UM260	HEAR	U of M
UM261	HEAR	U of M
UM263	HEAR	U of M
UM264	HEAR	U of M
UM265	HEAR	U of M
UM266	HEAR	U of M
UM267	HEAR	U of M
UM268	HEAR	U of M
UM269	HEAR	U of M
UM27	HEAR	U of M
UM270	HEAR	U of M
UM271	HEAR	U of M
UM272	HEAR	U of M
UM273	HEAR	U of M
UM274	HEAR	U of M
UM275	HEAR	U of M
UM276	HEAR	U of M
UM277	HEAR	U of M
UM278	HEAR	U of M
UM279	HEAR	U of M
UM28	HEAR	U of M
UM280	HEAR	U of M
UM281	HEAR	U of M
UM282	HEAR	U of M
UM283	HEAR	U of M
UM284	HEAR	U of M
UM285	HEAR	U of M
UM286	HEAR	U of M
UM287	HEAR	U of M
UM288	HEAR	U of M
UM289	HEAR	U of M
UM29	HEAR	U of M

Project Name	Quality[§]	Origin[¶]
UM290	HEAR	U of M
UM292	HEAR	U of M
UM293	HEAR	U of M
UM294	HEAR	U of M
UM295	HEAR	U of M
UM296	HEAR	U of M
UM297	HEAR	U of M
UM298	HEAR	U of M
UM299	HEAR	U of M
UM3	HEAR	U of M
UM30	HEAR	U of M
UM300	HEAR	U of M
UM301	HEAR	U of M
UM302	HEAR	U of M
UM303	HEAR	U of M
UM304	HEAR	U of M
UM305	HEAR	U of M
UM306	HEAR	U of M
UM307	HEAR	U of M
UM308	HEAR	U of M
UM309	HEAR	U of M
UM31	HEAR	U of M
UM310	HEAR	U of M
UM311	HEAR	U of M
UM312	HEAR	U of M
UM313	HEAR	U of M
UM314	HEAR	U of M
UM315	HEAR	U of M
UM316	HEAR	U of M
UM317	HEAR	U of M
UM318	HEAR	U of M
UM319	HEAR	U of M
UM32	HEAR	U of M
UM320	HEAR	U of M
UM321	HEAR	U of M
UM322	HEAR	U of M
UM323	HEAR	U of M
UM324	HEAR	U of M
UM325	HEAR	U of M
UM326	HEAR	U of M
UM327	HEAR	U of M
UM328	HEAR	U of M

Project Name	Quality[§]	Origin[¶]
UM329	HEAR	U of M
UM33	HEAR	U of M
UM330	HEAR	U of M
UM331	HEAR	U of M
UM332	HEAR	U of M
UM333	HEAR	U of M
UM334	HEAR	U of M
UM335	HEAR	U of M
UM337	HEAR	U of M
UM338	HEAR	U of M
UM339	HEAR	U of M
UM34	HEAR	U of M
UM340	HEAR	U of M
UM341	HEAR	U of M
UM342	HEAR	U of M
UM343	HEAR	U of M
UM344	HEAR	U of M
UM345	HEAR	U of M
UM346	HEAR	U of M
UM347	HEAR	U of M
UM348	HEAR	U of M
UM349	HEAR	U of M
UM35	HEAR	U of M
UM350	HEAR	U of M
UM351	HEAR	U of M
UM352	HEAR	U of M
UM353	HEAR	U of M
UM354	HEAR	U of M
UM355	HEAR	U of M
UM356	HEAR	U of M
UM357	HEAR	U of M
UM358	HEAR	U of M
UM359	HEAR	U of M
UM36	HEAR	U of M
UM360	HEAR	U of M
UM361	HEAR	U of M
UM362	HEAR	U of M
UM363	HEAR	U of M
UM364	HEAR	U of M
UM365	HEAR	U of M
UM367	HEAR	U of M
UM368	HEAR	U of M

Project Name	Quality[§]	Origin[¶]
UM369	HEAR	U of M
UM37	HEAR	U of M
UM370	HEAR	U of M
UM371	HEAR	U of M
UM372	HEAR	U of M
UM373	HEAR	U of M
UM374	HEAR	U of M
UM375	HEAR	U of M
UM376	HEAR	U of M
UM377	HEAR	U of M
UM378	HEAR	U of M
UM379	HEAR	U of M
UM38	HEAR	U of M
UM380	HEAR	U of M
UM381	HEAR	U of M
UM382	HEAR	U of M
UM383	HEAR	U of M
UM384	HEAR	U of M
UM385	HEAR	U of M
UM386	HEAR	U of M
UM387	HEAR	U of M
UM388	HEAR	U of M
UM389	HEAR	U of M
UM39	HEAR	U of M
UM390	HEAR	U of M
UM391	HEAR	U of M
UM392	HEAR	U of M
UM393	HEAR	U of M
UM394	HEAR	U of M
UM395	HEAR	U of M
UM396	HEAR	U of M
UM397	HEAR	U of M
UM398	HEAR	U of M
UM399	HEAR	U of M
UM4	HEAR	U of M
UM40	HEAR	U of M
UM400	HEAR	U of M
UM401	HEAR	U of M
UM402	HEAR	U of M
UM403	HEAR	U of M
UM404	HEAR	U of M
UM405	HEAR	U of M

Project Name	Quality[§]	Origin[¶]
UM406	HEAR	U of M
UM407	HEAR	U of M
UM408	HEAR	U of M
UM409	HEAR	U of M
UM41	HEAR	U of M
UM410	HEAR	U of M
UM411	HEAR	U of M
UM412	HEAR	U of M
UM413	HEAR	U of M
UM414	HEAR	U of M
UM415	HEAR	U of M
UM416	HEAR	U of M
UM417	HEAR	U of M
UM418	HEAR	U of M
UM419	HEAR	U of M
UM42	HEAR	U of M
UM420	HEAR	U of M
UM421	HEAR	U of M
UM422	Canola	U of M
UM423	Canola	U of M
UM424	Canola	U of M
UM425	HEAR	U of M
UM426	Canola	U of M
UM427	Low Linolenic	U of M
UM428	Low Linolenic	U of M
UM429	HOLLi	U of M
UM43	HEAR	U of M
UM430	HEAR	U of M
UM44	HEAR	U of M
UM440	HEAR	U of M
UM441	HEAR	U of M
UM45	HEAR	U of M
UM46	HEAR	U of M
UM47	HEAR	U of M
UM48	HEAR	U of M
UM49	HEAR	U of M
UM5	HEAR	U of M
UM50	HEAR	U of M
UM51	HEAR	U of M
UM52	HEAR	U of M
UM53	HEAR	U of M
UM54	HEAR	U of M

Project Name	Quality[§]	Origin[¶]
UM55	HEAR	U of M
UM56	HEAR	U of M
UM57	HEAR	U of M
UM58	HEAR	U of M
UM59	HEAR	U of M
UM6	HEAR	U of M
UM60	HEAR	U of M
UM61	HEAR	U of M
UM62	HEAR	U of M
UM63	HEAR	U of M
UM64	HEAR	U of M
UM65	HEAR	U of M
UM66	HEAR	U of M
UM67	HEAR	U of M
UM68	HEAR	U of M
UM69	HEAR	U of M
UM7	HEAR	U of M
UM70	HEAR	U of M
UM71	HEAR	U of M
UM72	HEAR	U of M
UM73	HEAR	U of M
UM75	HEAR	U of M
UM76	HEAR	U of M
UM77	HEAR	U of M
UM78	HEAR	U of M
UM79	HEAR	U of M
UM8	HEAR	U of M
UM80	HEAR	U of M
UM81	HEAR	U of M
UM82	HEAR	U of M
UM9	HEAR	U of M
UM97	HEAR	U of M
UM98	HEAR	U of M
UM99	HEAR	U of M

[†] Phenotypic groups established using morphological and quality traits by Ward's minimum variance and UPGMA methods.

[‡] Molecular groups established using marker data from genotyping-by-sequencing and sequence related amplified polymorphisms.

[§] HEAR, High Erucic Acid Rapeseed; HOLLi, high oleic, low linolenic acid.

Appendix B. List of 231 *Brassica napus* L. genotypes that were hierarchically clustered using the UPGMA method, by evaluating presence/absence of 291,782 SNPs gathered using Genotyping-by-Sequencing.

Identity	GBS Group	Intra-Group Order
UM389	A	1
UM211	A	2
UM196	A	3
UM4	A	4
UM400	A	5
UM1	A	6
UM2	A	7
UM9	A	8
UM29	A	9
UM241	A	10
UM27	A	11
UM393	A	12
UM401	A	13
UM3	A	14
UM3	A	15
UM10	A	16
UM13	A	17
UM14	A	18
UM11	A	19
UM12	A	20
UM17	A	21
UM28	A	22
UM30	A	23
UM45	A	24
UM98	A	25
UM129	A	26
UM101	A	27
UM135	A	28
UM102	A	29
UM390	A	30
UM405	A	31
UM406	A	32
UM407	A	33
UM410	A	34
UM409	A	35

Identity	GBS Group	Intra-Group Order
UM387	A	36
UM413	A	37
UM414	A	38
UM415	A	39
UM416	A	40
UM31	A	41
UM394	A	42
UM399	A	43
UM408	A	44
UM397	A	45
UM396	A	46
UM411	A	47
UM430	A	48
UM202	A	49
UM412	A	50
UM426	A	51
UM427	A	52
UM428	A	53
UM429	A	54
UM423	A	55
UM422	A	56
UM424	A	57
UM425	A	58
UM391	B	1
UM5	B	2
UM33	B	3
UM40	B	4
UM34	B	5
UM35	B	6
UM37	B	7
UM38	B	8
UM39	B	9
UM81	B	10
UM82	B	11
UM42	B	12
UM43	B	13
UM47	B	14
UM48	B	15
UM49	B	16
UM76	B	17
UM80	B	18
UM32	B	19

Identity	GBS Group	Intra-Group Order
UM50	B	20
UM61	B	21
UM58	B	22
UM60	B	23
UM36	B	24
UM70	B	25
UM54	B	26
UM57	B	27
UM68	B	28
UM72	B	29
UM52	B	30
UM53	B	31
UM41	B	32
UM395	B	33
UM46	B	34
UM21	B	35
UM62	B	36
UM16	B	37
UM16	B	38
UM23	B	39
UM23	B	40
UM24	B	41
UM22	B	42
UM242	B	43
UM19	B	44
UM19	B	45
UM20	B	46
UM15	B	47
UM15	B	48
UM18	B	49
UM232	B	50
UM26	B	51
UM239	B	52
UM25	B	53
UM51	C	1
UM55	C	2
UM56	C	3
UM67	C	4
UM69	C	5
UM392	D	1
UM44	D	2
UM73	D	3

Identity	GBS Group	Intra-Group Order
UM73	D	4
UM6	D	5
UM78	D	6
UM63	D	7
UM65	D	8
UM64	D	9
UM66	D	10
UM75	D	11
UM75	D	12
UM77	D	13
UM79	D	14
UM71	D	15
UM7	E	1
UM8	E	2
UM398	E	3
UM97	E	4
UM97	E	5
UM100	F	1
UM127	F	2
UM99	F	3
UM136	F	4
UM130	F	5
UM103	F	6
UM402	F	7
UM404	F	8
UM403	F	9
UM386	F	10
UM388	F	11
UM131	G	4
UM138	G	1
UM172	G	2
UM139	G	3
UM132	G	5
UM106	G	5
UM133	G	7
UM137	G	8
UM123	H	1
UM115	H	2
UM111	H	3
UM108	H	4
UM119	H	6
UM124	H	7

Identity	GBS Group	Intra-Group Order
UM109	I	1
UM110	I	2
UM117	I	3
UM118	I	4
UM107	I	5
UM126	I	6
UM128	I	7
UM128	I	8
UM112	J	1
UM113	J	2
UM113	J	3
UM125	J	4
UM105	J	5
UM121	J	6
UM122	J	7
UM150	J	8
UM251	K	1
UM168	K	2
UM164	K	3
UM165	K	4
UM161	K	5
UM163	K	6
UM162	K	7
UM166	K	8
UM167	K	9
UM158	K	10
UM160	K	11
UM159	K	12
UM169	K	13
UM170	K	14
UM171	K	15
UM140	K	16
UM141	K	17
UM142	K	18
UM144	K	19
UM145	K	20
UM143	K	21
UM147	K	22
UM151	K	23
UM154	K	24
UM152	K	25
UM148	K	26

Identity	GBS Group	Intra-Group Order
UM149	K	27
UM153	K	28
UM155	K	29
UM157	K	30
UM156	K	31
UM114	L	1
UM104	L	2
UM178	L	3
UM183	L	4
UM120	L	5
UM146	M	1
UM190	M	2
UM190	M	3
UM180	M	4
UM182	M	5
UM181	M	6
UM185	M	7
UM189	M	8
UM177	M	9
UM184	M	10
UM186	M	11
UM188	M	12
UM187	M	13
UM417	N	1
UM420	N	2
UM419	N	3
UM418	N	4
UM421	N	5

Appendix C. Name, group membership and intra-group order of 160 Brassica napus genotypes hierarchically clustered based using Ward's method with presence/absence variation in 230 sequence related amplified polymorphism markers as criteria.

Identity	SRAP Group	Intra-Group Order
UM336	A	1
UM376	A	2
UM383	A	3
UM374	A	4
UM382	A	5
UM233	A	6
UM348	A	7
UM337	A	8
UM359	A	9
UM360	A	10
UM372	A	11
UM360	A	12
UM373	A	13
UM61	B	1
UM106	B	2
UM306	B	3
UM289	B	4
UM173	B	5
UM309	B	6
UM310	B	7
UM178	B	8
UM189	B	9
UM387	B	10
UM388	B	11
UM103	B	12
UM384	B	13
UM244	C	1
UM313	C	2
UM42	C	3
UM251	C	4
UM105	C	5
UM65	C	6
UM76	C	7
UM177	C	8
UM113	C	9
UM365	C	10
UM79	C	11

Identity	SRAP Group	Intra-Group Order
UM80	C	12
UM364	C	13
UM114	C	14
UM320	C	15
UM64	D	1
UM323	D	2
UM316	D	3
UM66	D	4
UM322	D	5
UM301	D	6
UM324	D	7
UM67	D	8
UM127	D	9
UM110	D	10
UM82	D	11
UM380	D	12
UM63	D	13
UM81	D	14
UM62	D	15
UM335	D	16
UM371	D	17
UM128	D	18
UM304	D	19
UM346	D	20
UM58	D	21
UM59	D	22
UM57	D	23
UM305	D	24
UM60	D	25
UM325	D	26
UM301	D	27
UM277	D	28
UM214	D	29
UM130	D	30
UM73	E	1
UM370	E	2
UM284	E	3
UM299	E	4
UM206	E	5
UM338	E	6
UM344	E	7
UM124	E	8

Identity	SRAP Group	Intra-Group Order
UM333	E	9
UM355	E	10
UM302	E	11
UM356	E	12
UM334	E	13
UM56	E	14
UM121	E	15
UM357	E	16
UM358	E	17
UM314	E	18
UM117	E	19
UM186	E	20
UM312	E	21
UM332	E	22
UM312	E	23
UM353	E	24
UM328	E	25
UM334	E	26
UM342	E	27
UM386	E	28
UM315	E	29
UM343	E	30
UM311	E	31
UM119	E	32
UM321	E	33
UM184	E	34
UM366	E	35
UM354	E	36
UM330	E	37
UM331	E	38
UM367	E	39
UM329	E	40
UM71	E	41
UM379	E	42
UM120	E	43
UM72	E	44
UM347	E	45
UM23	F	1
UM8	F	2
UM22	F	3
UM265	F	4
UM327	F	5

Identity	SRAP Group	Intra-Group Order
UM361	F	6
UM317	F	7
UM318	F	8
UM188	F	9
UM43	F	10
UM224	F	11
UM6	F	12
UM68	F	13
UM69	F	14
UM325	F	15
UM212	F	16
UM183	F	17
UM286	G	1
UM112	G	2
UM341	G	3
UM75	G	4
UM203	G	5
UM363	G	5
UM196	G	6
UM340	G	7
UM352	G	8
UM125	G	9
UM326	G	10
UM70	G	11
UM351	G	12
UM349	H	1
UM350	H	2
UM385	H	3
UM377	H	4
UM378	H	5
UM339	H	6
UM362	H	7
UM235	H	8
UM77	H	9
UM18	H	10
UM288	H	11
UM111	H	12
UM308	H	13
UM319	H	14