Phenotypic and Genotypic Investigation of the Seed Storage Protein Cruciferin in *Brassica napus* L.

By

Ashley Kendra Ammeter

A Thesis submitted to the Faculty of Graduate Studies of

The University of Manitoba

in partial fulfillment of the requirements of the degree of

MASTER OF SCIENCE

Department of Plant Science

University of Manitoba

Winnipeg

ACKNOWLEDGEMENTS

First, I would like to thank my advisor, Dr. Rob Duncan. I am grateful for this opportunity to gain valuable experience, and for the wealth of learning opportunities I received under your guidance. Your consistent support and confidence in my abilities has been invaluable.

To my committee members, Dr. Doug Cattani and Dr. Martin Nyachoti, I would like to extend my thanks for your time and guidance which has been so important in this research process.

Without the knowledge and expertise of the members of the *Brassica* breeding group, this project would not have been possible. I am particularly grateful to Ralph Kowatsch and Debbie Witko for their help and encouragement. I also wish to thank Sally Vail, Isobel Parkin, Steve Robinson, and Erin Higgins from Agriculture and Agri-Food Canada for their vital contributions of the *B. napus* population and genotypic data used for this research.

I would like to thank all of my fellow graduate students in the Department of Plant Science for their friendship and support. Especially to Kenny So and Jia Sun, I am so grateful for your help, guidance and friendship.

I would like to thank Research Manitoba, NSERC, DL Seeds and Bunge Canada for providing funding and financial assistance that made this research possible.

Finally, this would not have been possible without the unwavering love and encouragement of my family. You have shaped me into the person I am, giving me the opportunities that have allowed me to achieve my goals. From the very beginning of my educational journey, you have been unfailingly supportive, sharing in every success and frustration. I am so very grateful.

ABSTRACT

World-wide, there is an increasing demand for sources of high-quality protein, resulting in an interest in novel plant-based proteins. Canola (*Brassica napus* L.) is the second greatest produced oilseed crop worldwide, and a crop of high importance to the Canadian economy. *Brassica napus* meal is primarily used as a source of protein in livestock feed but may serve as a source of plant-based proteins for human consumption. The seed storage protein cruciferin makes up approximately 60% of the protein content of mature seeds and is of interest due to its functional properties. In order to optimize *B. napus* cruciferin protein profiles, plant breeding efforts require a thorough understanding of existing phenotypic variation in cruciferin content, as well as insight into the effect of genotype and environmental factors on this trait. This study used an enzyme-linked immunosorbent assay (ELISA)-based approach to determine cruciferin content in a diverse population of *B. napus* genotypes. Considerable variation in cruciferin content was observed, and the effects of genotype by site-year interactions were shown to significantly affect cruciferin content. Future breeding efforts will also require efficient methods to determine cruciferin content, and for this reason the suitability of near-infrared spectroscopy (NIRS) as a potential method was explored. Combining reference data provided from the ELISA-based quantification method with spectra produced by scanning whole seed samples of *B. napus* enabled the development of several NIRS calibration equations. Unfortunately, statistical analysis showed that these equations were poorly suited for the prediction of cruciferin content. Finally, a genome-wide association study (GWAS) was performed to provide a more thorough understanding of the genetic control of cruciferin content. A population of 51 *B. napus* genotypes was used for GWAS, which identified 144 SNP-trait associations across 47 loci significantly associated with cruciferin content. Based on these loci, 11 candidate genes were identified as having a potential role in the genetic control

of cruciferin content. This research provides a background on existing phenotypic variation and the genetic control of cruciferin content, building a valuable framework on which future efforts to develop *B. napus* with specialty protein profiles may be based.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1. General Introduction

Brassica napus L. (canola or rapeseed) is the second highest produced oilseed crop worldwide (FAO 2019), and is a highly important crop to Canadian agriculture. Annually, the canola sector contributes C$26.7 billion to the Canadian economy (LMC International 2016). In addition, Canada is the world's largest producer (Statistics Canada 2019a) and exporter (USDA 2019) of canola and rapeseed. It is clear, therefore, that the importance of this crop to the Canadian economy cannot be overstated. While *B. napus* is primarily grown for seed oil production, the secondary product, a protein-rich meal, presents an excellent opportunity to add value to an already major crop. Following oil extraction, this protein-rich meal has been shown to contain up to 40% crude protein on a dry matter basis (Wittkop et al. 2009; Wanasundara 2011; Rahman 2013). Currently, *B. napus* meal is used primarily as a source of protein for livestock feed. However, a variety of characteristics, including its well-balanced amino acid profile make it an excellent candidate for use as a source of protein for human consumption (Tan et al. 2011).

The proteins of *B. napus* seeds are found primarily in the form of the two seed storage proteins cruciferin and napin (Aider and Barbana 2011; Wanasundara 2011). Napin is a 2S albumin that makes up approximately 20% of the total protein content of mature seeds, while cruciferin is an 11S globulin that makes up approximately 60% of the protein content of mature seeds (Aider and Barbana 2011; Wanasundara 2011; Perera et al. 2016). The value of these *B. napus* proteins for human consumption lies in their functional properties, including foaming (Aider and Barbana 2011; Wanasundara et al. 2017) and solubility (Tan et al. 2011; Wanasundara and McIntosh 2013) properties for napin, and emulsifying (Cheung et al. 2015) and gelling properties (Tan et al. 2014b) for cruciferin.

While there is interest in the use of *B. napus* cruciferin protein for human consumption, the extent of variation in cruciferin content among genotypes has been largely unexplored. While a more thorough understanding of this variation would be valuable, phenotyping is often a costly and time consuming stage in the breeding process (Furbank and Tester 2011). To date, a number of methods have been used to quantify cruciferin proteins. These methods can most generally be divided into one of two categories: those that extract total soluble protein followed by separation and quantification (Kohno-Murase et al. 1995; Schatzki et al. 2014; Perera et al. 2016), or those that selectively extract the cruciferin protein fraction and then quantify (Wanasundara and McIntosh 2013; Cheung et al. 2014; Akbari and Wu 2015). These methods are effective, however, they also present a number of disadvantages. Each method results in the destruction of the test seed sample, may require relatively large seed amounts and they are also time consuming to varying degrees. In contrast, enzyme-linked immunosorbent assays (ELISAs) use an enzyme-labelled antibody to detect and quantify a protein of interest (Lin 2015), while requiring only small amounts of seed and maintaining the ability to handle many samples with relative efficiency (Crowther 2009).

While ELISAs are well-suited for use in *B. napus* breeding programs to quantify cruciferin, many programs already utilize near-infrared reflectance spectroscopy (NIRS) as a high-throughput , non-destructive method by which to analyze a number of seed quality traits (Font et al. 2006). The seed quality traits quantified by NIRS in *Brassica* includes chlorophyll content (Daun 1976; Tkachuk and Kuzina 1982), glucosinolate content (Daun et al. 1994; Velasco and Becker 1998a), oil content (Tkachuk 1981; Daun et al. 1994; Velasco et al. 1999b, 1999a), fatty acid composition (Sato et al. 1998; Velasco and Becker 1998b; Velasco et al. 1999b, 1999a; Kim et al. 2007), fiber content (Font et al. 2003, 2005; Dimov et al. 2012; Wittkop et al. 2012) and total protein content (Tkachuk 1981; Daun et al. 1994; Velasco and Möllers 2002). By correlating spectral data produced when a

seed sample is exposed to near-infrared radiation with a chemical or physical property as measured by an established reference method, a trait of interest may be efficiently evaluated (Shenk et al. 2008; Agelet and Hurburgh 2010). Given its widespread use in *Brassica* breeding programs, it would be favourable to also employ NIRS for the rapid determination of cruciferin content.

Beyond an understanding of the phenotypic variation present in cruciferin content, an understanding of the genetic control of cruciferin content would also be an important contribution towards improving the value of *B. napus* meal. Currently, research regarding the genetic control of cruciferin content in *B. napus* is limited. A quantitative trait loci (QTL) mapping study by Schatzki et al. (2014) analyzed the genetic control of cruciferin content in a doubled haploid population of winter oilseed rape, discovering two significantly associated QTL. Despite this, further validation and study is needed. Given the limited research in this area, further identification of the loci that are associated with cruciferin content in *B. napus* will be key in the advancement of breeding efforts.

In order for the potential value of *B. napus* meal to be fulfilled, it will be important for *B. napus* breeding efforts to focus on unique seed storage protein profiles. The research in this thesis addresses the need for greater knowledge surrounding *B. napus* cruciferin proteins, by evaluating the phenotypic variation in cruciferin content found in existing *B. napus* genotypes as well as the genetic control of cruciferin content. By making use of a diverse population of 52 *B. napus* genotypes, the following three projects were proposed:

1. Use ELISAs to evaluate cruciferin content in a population of *B. napus* genotypes across four site-years to explore the existing phenotypic variation in cruciferin content and shed light on the role of the environment in the expression of this trait.

2. Explore the potential application of NIRS to be used for the determination of cruciferin content in whole seeds of *B. napus* through the development of an NIRS calibration model.
3. Identify loci significantly associated with cruciferin content through an association mapping study, using genotype data generated using the *Brassica* 60K Illumina® Infinium SNP array.

2. LITERATURE REVIEW

2.1 Background

2.1.1 Introduction to Brassica Genus

The Brassicaceae family consists of 338 genera and 3709 species, the members of which are globally distributed (Al-Shehbaz et al. 2006; Warwick et al. 2006). A very diverse family, the Brassicaceae include many weedy species as well as economically important vegetable, condiment and oilseed crop species (Warwick 2011). Of particular interest are the model species *Arabidopsis thaliana* (L.) Heynh. and the economically important species within the genus *Brassica* (Schranz et al. 2006). Within the *Brassica* genus, six interrelated species are the most cultivated (Warwick 2011; Sharma et al. 2014). Three amphidiploid species, *Brassica carinata* A. Braun (2*n* =34, BBCC), *Brassica juncea* (L.) Czern. (2*n*=36 AABB), and *Brassica napus* L. (2*n*=38, AACC) resulted from hybridization and polyploidization events between three diploid species, *Brassica nigra* (L.) W.D.J. Koch (2*n*=16, BB), *Brassica oleracea* L. (2*n*=18, CC) and *Brassica rapa* L. (syn. *Brassica campestris*) (2*n*=20, AA) (Figure 2.1) (U 1935; Warwick 2011; Sharma et al. 2014). Among these cultivated *Brassica* species, remarkable morphological and genetic diversity exists. *Brassica napus* and *B. carinata* are cultivated primarily as oilseed crops, while *B. rapa* and *B. juncea* consist of both oilseed and leafy vegetable morphotypes (Warwick 2011; Cheng et al. 2014). Meanwhile, *B. oleracea* morphotypes include important crops such as kale, cabbages, Brussels sprouts, kohlrabi, cauliflower and broccoli (Warwick 2011; Cheng et al. 2014). In Canada, the species of primary importance are *B. napus, B. juncea* and *B. rapa* (Daun 2011).

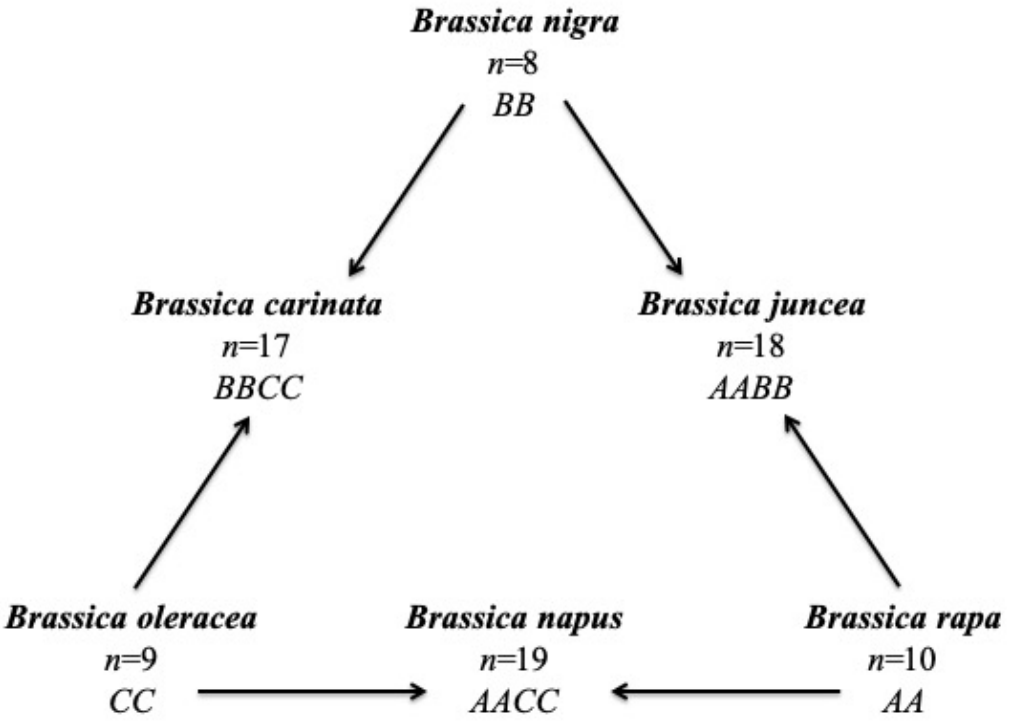


Figure 2.1 The "Triangle of U", a representation of the genetic relationships between the six cultivated species of *Brassica*. Adapted from U (1935).

2.1.2 History of Rapeseed and Canola (Brassica napus)

While *Brassica* oilseeds have been used by humans for thousands of years, rapeseed and canola were virtually unknown in Canada until just prior to World War II (Boulter 1983; Daun 2011). The first rapeseed grown in Canada was *B. rapa*, brought to Saskatchewan by Polish farmer Fred Salvoniuk in 1936 (Boulter 1983; Daun 2011). As a result, *B. rapa* is still known as Polish rapeseed today (Boulter 1983). During World War II, however, demand for rapeseed oil lubricant increased significantly (Downey 2007). In response, the Canadian government began planting experimental plots of the Polish *B. rapa* as well as *B. napus* seed imported from Argentina (Boulter 1983; Downey 2007; Daun 2011). Between the years 1943 and 1948 production of rapeseed in Western Canada increased by up to a factor of three times every year, and by the end of World War II (Boulter 1983), a significant amount of rapeseed was being produced (Daun 2011).

The rapeseed cultivars introduced during this time all had high levels of erucic acid (Daun 1983). However, in 1956 the Food and Drug Directorate ruled that due to the high levels of erucic acid, rapeseed oil would no longer be approved as an edible oil due to the results of animal feeding trials (Sauer and Kramer 1983). While the restrictions were later removed, significant effort from that point forward was made towards the development of rapeseed cultivars with low erucic acid content (Sauer and Kramer 1983). In 1968, the first *B. napus* cultivar with low erucic acid was developed, 'Oro' (Downey 2007), however it was grown on relatively few acres due to poor agronomic characteristics and low yield (Daun 1983). By 1971, breeding programs initiated at the University of Manitoba and the Agriculture Canada Research Station at Saskatoon had low erucic acid genotypes available, resulting in the release of 'Span' and 'Zephyr', to be replaced in 1973 by the agronomically improved and higher yielding varieties 'Midas' and 'Torch' (Daun 1983).

While a decrease in erucic acid content allowed for the use of rapeseed as a superior edible oil, the utilization of the meal fraction of the seed still remained a challenge (Downey 2007; Daun 2011). The presence of glucosinolates, a group of sulfur-containing secondary plant metabolites, limited the use of rapeseed meal as a feed for livestock due to their interference with thyroid gland function (Khajali and Slominski 2012). Ultimately, it was necessary to produce a cultivar with reduced glucosinolates in the seed and meal (Downey and Rakow 1987). In 1974, Dr. Baldur Stefansson from the University of Manitoba released the cultivar 'Tower', a *B. napus* cultivar with low erucic acid content and low glucosinolate content (Stefansson and Kondra 1975). This cultivar became known as the first "double low" variety (Downey and Rakow 1987). Today, these "double low" cultivars are commonly called canola (Stefansson 1983), and must meet the following official definition: "seeds of the genus *Brassica* (*Brassica napus*, *Brassica rapa,* or *Brassica juncea*) from which the oil shall contain less than 2% erucic acid in its fatty acid profile and the solid component shall contain less than 30 micromoles of any one or any mixture of 3-butenyl glucosinolate, 4-pentenyl glucosinolate, 2-hydroxy-3 butenyl glucosinolate and 2-hydroxy-4-pentenyl glucosinolate per gram of air-dry, oil-free solid" (Canola Council of Canada 2019a).

2.1.3 Production of Brassica napus

Since the introduction of the first "double low" cultivars, *B. napus* production has grown significantly. Contributing an average of $26.7 billion to the Canadian economy annually (LMC International 2016), canola is an important crop in Canadian agriculture. In the 2018-2019 crop year, over 70 million metric tonnes of canola and rapeseed were produced globally (USDA 2019). With over 19 million metric tonnes produced in Canada across over 8 million harvested hectares (Statistics Canada 2019b), Canada is the world's largest producer of canola and rapeseed, followed

by the European Union and China (USDA 2019). In addition, Canada is by far the world's largest exporter of canola seed, oil and meal, exporting 10.2 million metric tonnes, 3.3 million metric tonnes, and 4.7 million metric tonnes, respectively (USDA 2019).

2.2 *Brassica napus* End Use Products

Canola production results in three primary end-products: the seed, the oil, and the meal. While a large proportion of the seed produced in Canada is exported, Canada's 14 crushing facilities are also equipped to crush 10 million tonnes of seed annually, producing 4 million tonnes of meal and 3 million tonnes of oil (Canola Council of Canada 2019b). On average, *Brassica* seeds contain 40 to 45% oil (Wanasundara 2011). The removal of the oil fraction by the crushing process produces seed meal that is made up of 35 to 40% crude protein by weight (Wanasundara 2011).

2.2.1 Brassica napus Oil

In North America, the majority of edible canola oil comes from *B. napus* and *B. rapa* seeds (Przybylski and Eskin 2011). Primarily composed of triacylglycerols (94-99%), *Brassica* oil also contains small amounts of phospholipids and free fatty acids (Przybylski and Eskin 2011). Triacylglycerols are made up of three fatty acids esterified to the hydroxyl positions of glycerol, and thus the fatty acid makeup of canola oil represents the composition of the triacylglycerols present (Uppström 1995). The fatty acids found in *Brassica* oil typically include the following: palmitic acid (C16:0), palmitoleic acid (C16:1), stearic acid (C18:0), oleic acid (C18:1), linoleic acid (C18:2), alpha-linolenic acid (C18:3), arachidic acid (C20:0), eicosenoic acid (C20:1), behenic acid (C22:0), erucic acid (C22:1), lignoceric acid (C24:0) and nervonic acid (C24:1) (Uppström 1995; Przybylski and Eskin 2011). Modification of the fatty acid composition has

allowed for the development of edible canola oils with improved functional and nutritional properties (McVetty and Scarth 2002). As previously discussed, one of the initial modifications to the fatty acid composition of *B. napus* resulted in the development of *B. napus* genotypes with reduced erucic acid content (Daun 1983; Uppström 1995; Przybylski and Eskin 2011). Interest was also placed on the reduction of linolenic acid content to improve oil shelf life (Scarth and Tang 2006), resulting in the development of the canola cultivar "Stellar", a cultivar high in linoleic acid and low in linolenic acid (Scarth et al. 1988).

Along with specialty edible oils, efforts have also been directed towards the development of *B. napus* oils high in erucic acid for use in industrial purposes. While canola-quality varieties are bred to have very low erucic acid content (<2%), high erucic acid rapeseed (HEAR) is bred to have significantly higher erucic acid content (>50%), while maintaining low glucosinolates in the meal (McVetty et al. 2016). The primary use of erucic acid is in the production of erucamide, a derivative of erucic acid, which is used in the polymer industry as an anti-slip agent for plastic films. Erucic acid and HEAR oil may also be used in the production of lubricants, cosmetics, inks and textiles (Aukema and Campbell 2011).

The world's first HEAR cultivar, Reston, was developed at the University of Manitoba and released in 1982, followed shortly after by the cultivar Hero in 1989, which improved upon the erucic acid content of its predecessor (Zelmer and McVetty 2009). Since the development of the first HEAR cultivars, the University of Manitoba has continued to release cultivars, improving upon seed quality and agronomic performance (Zelmer and McVetty 2009). These cultivars include the non-herbicide tolerant open pollinated population (OPP) cultivars Mercury (Scarth et al. 1995), Neptune (McVetty et al. 1996a), Venus (McVetty et al. 1996b), Castor (McVetty et al.

1998), MillenniUM 01 (McVetty et al. 1999), MillenniUM 02 (McVetty et al. 2000a) and MillenniUM 03 (McVetty et al. 2000b). The "Red River" series of cultivars were the first successful glyphosate-tolerant (Roundup Ready™) HEAR cultivars available (McVetty et al. 2006a, 2006b, 2010, 2012). By 2014, HYHEAR 1, the first hybrid glyphosate-tolerant HEAR cultivar was released (McVetty et al. 2014), followed by HYHEAR 2 (Duncan et al. 2016) and HYHEAR 3 (Duncan et al. 2017).

2.2.2 Brassica napus Meal

Following the extraction of the oil from *B. napus* seeds, the remaining product is known as the meal. While the composition of canola meal may vary as a result of processing conditions, differences between cultivars, and the effect of environmental factors (Bell 1995), the basic composition includes small amounts of oil, protein, fibre, glucosinolates, sinapine, and phytic acid among other components (Table 2.1) (Canola Council of Canada 2015; Wanasundara et al. 2017). The presence of undesirable compounds such as glucosinolates does limit its functionality (Aider and Barbana 2011), however the presence of a good amino acid profile and a high protein efficiency ratio make canola meal a high quality protein source (Khattab and Arntfield 2009). For this reason, canola meal is widely used as a protein source in feed rations for the dairy cattle, poultry, swine and aquaculture industries (Bell 1995; Wanasundara 2011). *Brassica napus* meal functionality, end-uses and proteins will be discussed in greater detail in Section 2.3.

Table 2.1 Chemical composition of canola meal (12% moisture basis).[†]

Component	Average
Crude protein (N x 6.25, %)	36.7
Rumen escape protein (%)	43.5
Ether extract (%)	3.3
Linoleic acid (%)	0.67
Linolenic acid (%)	0.32
Ash (%)	6.7
Calcium (%)	0.65
Phosphorous (%)	0.99
Crude fibre (%)	11.2
Acid detergent fibre (%)	25.4
Neutral detergent fibre (%)	25.4
Total dietary fibre (%)	32.4
Sinapine (%)	1
Phytic acid (%)	2.3
Glucosinolates (µmol/g)	4.2

[†]Adapted from Canola Council of Canada (2015).

2.3 *Brassica napus* Meal Proteins

2.3.1 Brassica napus Meal Protein Composition

Seed storage proteins comprise up to 90% of the protein content in *B. napus,* with minor proteins such as oil body proteins making up the remainder (Campbell et al. 2016; Wanasundara et al. 2017). The storage proteins found in the seed are primarily in the form of cruciferin and napin (Campbell et al. 2016; Wanasundara et al. 2017). While the ratio of cruciferin to napin in *B. napus* can vary significantly (Wanasundara 2011), on average napin makes up approximately 20% of the protein content of mature seeds, and cruciferin approximately 60% of the total protein content of mature seeds (Aider and Barbana 2011; Wanasundara 2011; Perera et al. 2016).

2.3.1.1 Synthesis of Seed Storage Proteins During Embryo Development

Compounds such as oil, starch and seed storage proteins accumulate in the developing embryo of *B. napus* seeds and act as a source of energy to fuel seed germination and plant growth (Gacek et al. 2018). The temporal and spatial development of the seed storage proteins cruciferin and napin, as well as lipids, have been shown to be consistent with one another (Borisjuk et al. 2013). Genes associated with seed storage proteins are most highly expressed during the late stages of embryo development (Yu et al. 2010; Basnet et al. 2013; Zhang et al. 2014). Seed storage proteins are first detected during the cell expansion phase of development, at approximately 25 days (Crouch and Sussex 1981), and are found primarily in the radicle of the embryo during the early- and mid-storage stages (Borisjuk et al. 2013). By later stages of seed development, the accumulation of seed storage proteins ceases, with the accumulation of napin proteins ending when the water content of the embryo begins to decline, at day 42, while cruciferin accumulation continues until seed maturity is reached (Crouch and Sussex 1981).

2.3.1.2 Cruciferin

Cruciferin is an 11S-12S globulin heteromer (Schwenke et al. 1981; Sjödahl et al. 1991), with a molecular weight of 300-350 kDa (Campbell et al. 2016). *Brassica napus* seed protein is comprised of approximately 60% cruciferin at maturity (Crouch and Sussex 1981); however, cruciferin content has been shown to vary significantly (Malabat et al. 2003).

Like other similar proteins, cruciferin is organized into primary, secondary, tertiary and quaternary structures (Wanasundara et al. 2017). Five subunit isoforms exist, with polypeptide chains made up of 465 to 509 amino acid residues, the exact length of which depends on the gene involved in expression (Wanasundara 2011; Wanasundara et al. 2017). Each subunit consists of a heavy α-chain (30 kDa) and a light β-chain (20 kDa), linked to one another with a disulfide bond (Wanasundara 2011; Wanasundara et al. 2017).

Recent attention has been placed on canola proteins for use in human food products. With this in mind, it is important to consider the functional properties of cruciferin to determine its suitability for this use. Functional properties of importance include the solubility, foaming properties and emulsifying properties of the protein (Tan et al. 2011). Cruciferin-rich protein isolates exhibit a low solubility when compared to napin-rich isolates, as well as a lower ability to create and stabilize foams (Campbell et al. 2016; Wanasundara et al. 2017). However, cruciferin-rich isolates have been demonstrated to have superior emulsification properties when compared to napin-rich isolates (Wu and Muir 2008; Cheung et al. 2014, 2015), as well as superior emulsion capacity to soy protein isolates (Tan et al. 2014a). Cruciferin-rich protein isolates are able to form strong gels at pH 7 and 9 (Tan et al. 2014b). These functional properties indicate isolates of cruciferin protein would be best suited for the production of meat substitutes and baked goods (Akbari and Wu 2015).

2.3.1.3 Napin

Napin is a 2S albumin protein of the prolamin superfamily (Shewry et al. 1995). Multiple isoforms of napin have been identified, exhibiting a range of molecular masses from 12.5 to 16 kDa (Campbell et al. 2016; Perera et al. 2016). On average, napin makes up 20% of the total storage proteins of the mature seed (Crouch and Sussex 1981), but among *B. napus* rapeseed genotypes has been shown to have high variation in the range of 25-45% of the seed protein (Malabat et al. 2003).

The structure of napins have been found to display significant polymorphism (Wanasundara 2011), and are encoded by a multigene family with the number of genes involved in expression in *B. napus* ranging from 10 to 16 (Josefsson et al. 1987; Scofield and Crouch 1987). *Brassica napus* napin is composed of two subunits: a small (short, ~4 kDa) subunit and a large (long, ~10 kDa) subunit (Gehrig et al. 1996; Rico et al. 1996). These two polypeptides are linked to one another by two inter-chain disulfide bonds, while the large subunit is further stabilized by two intra-chain disulfide bonds (Rico et al. 1996).

As previously discussed, the functional properties of a protein play a crucial role in determining its suitability for use in food products. Napin exhibits a hydrophilic nature, and thus performs poorly in its emulsification capacity and stability compared to whey protein isolate and soy protein concentrate (Wanasundara et al. 2017). Napin-rich isolates have also been demonstrated to have poor gelling properties, forming weak gels that require high temperatures to form (Tan et al. 2014b). However, napin-rich protein extracts exhibit high solubility, particularly in the pH range of 2 to 10 (Wanasundara and McIntosh 2013), as well as the capacity to form highly stable foams, comparable in functionality to whey protein isolate (Wanasundara et al. 2017). As a result, napin

protein isolates are suited for used in beverages, baked and confectionary goods (Akbari and Wu 2015).

2.3.1.4 Minor Proteins

While the majority of protein found in *B. napus* seeds are in the form of the seed storage proteins cruciferin and napin, a variety of minor proteins are also present (Perera et al. 2016). For example, 2-8% of the seed protein is comprised of oleosins, low molecular weight (15-26 kDa) structural proteins that are associated with oil bodies (Huang 1992; Aider and Barbana 2011). Other minor proteins include lipid transfer proteins, which may be desirable for cosmetics and food additives, as well as thionins and trypsin inhibitors (Bérot et al. 2005).

2.3.2 Utilization of Brassica napus Meal and Meal Protein

2.3.2.1 Ruminant Feed

Ruminants are cud-chewing mammals that possess stomachs made up of four compartments (National Research Council 2007). The most common domesticated ruminants include sheep, cattle and goats (Hungate 1966). In North America and Europe, the most common end use of low glucosinolate *B. napus* (canola) meal is as a source of protein in dairy cow feed (Arntfield and Hickling 2011). The high palatability and economic competitiveness (Arntfield and Hickling 2011; Canola Council of Canada 2015) make canola meal well-suited for use in feed rations for dairy cows and other ruminants. Further, lysine and methionine are considered the most limiting amino acids for milk production in many feed rations (Schwab et al. 1976), thus the significant contribution of methionine by canola meal (Canola Council of Canada 2015) makes it a preferred choice.

Livestock diets are carefully considered and the protein and amino acid content must be balanced to ensure optimal growth, maintenance, and, in the case of dairy cattle, production of milk (Arntfield and Hickling 2011). Amino acids available to dairy cattle come from two primary sources: rumen-degradable protein (RDP), which is degraded into metabolizable amino acids by microbes in the rumen (Arntfield and Hickling 2011), and rumen-undegradable protein (RUP), which remains undigested after passing through the rumen (Canola Council of Canada 2016). In high-yielding cows, the amino acids supplied by rumen microbes must be supplemented by RUP (Paz et al. 2014). In canola meal, a number of studies have confirmed that a large fraction of soluble protein in canola is not degraded in the rumen, with less than half classifying as RDP (Hedqvist and Udén 2006; Stefański et al. 2013). While the RUP content in canola meal may vary significantly depending upon the method by which it is analyzed, a series of studies have concluded that the RUP of canola meal is on average higher than that present in soybean meal (Hedqvist and Udén 2006; Maxin et al. 2013). It has been found that the amino acid profile of RUP in canola meal is complementary to that supplied by the microbial protein fraction, and thus may improve milk production (Arntfield and Hickling 2011) and milk protein (Paz et al. 2014).

A great deal of research effort has been directed towards elucidating the value of canola meal towards milk production in dairy cattle. A recent meta-analysis comparing the feeding value of soybean and canola meal in dairy cattle demonstrated that rations including canola meal resulted in greater feed intake and improved milk yield by one kilogram per kilogram of protein supplied per day when compared to rations containing soybean meal (Huhtanen et al. 2011). A meta-analysis by Martineau et al. (2013) found a similar increase in milk yield and milk protein when canola meal was used as the protein source. These meta-analyses demonstrate that canola meal can have a positive effect when incorporated into dairy cow feed rations.

2.3.2.2 Swine Feed

Canola meal is an economical protein source that can support high levels of performance in swine production. However, a number of important factors must be considered to facilitate successful use. Primary among them are the antinutritional factors present in canola meal, including glucosinolates, fibre, phytic acid and sinapine (Mejicanos et al. 2016). The presence of glucosinolates can be a limiting factor in the inclusion of canola meal in swine diets, due to its association with a bitter taste that causes reduced palatability and therefore reduced feed intake (Tripathi and Mishra 2007). A study performed by Landero et al. (2012) showed that due to the presence of glucosinolates, weaned pigs showed an aversion to diets containing canola meal, when an option containing soybean meal was also provided. In addition, high levels of glucosinolates may also affect thyroid, kidney, and liver function (Mejicanos et al. 2016). Despite these concerns, in a series of studies examining the effects of glucosinolates on pig growth and health, Schöne et al. (1997a, 1997b) first recommended that glucosinolate levels up to 2.4 mmol/kg could be tolerated without an impact on growth or thyroid function, while in a subsequent study it was recommended that glucosinolate content in feed up to 2 mmol/kg is suitable. The efforts of breeders to reduce the glucosinolate content of the meal has improved the suitability of canola meal as an alternative to soybean meal in pig diets. The use of tail-end dehulling of canola meal has also been shown to effectively reduce dietary fibre and increase crude protein content in canola meal (Mejicanos et al. 2017), as well as result in increased apparent and standardized total tract digestibility of phosphorous in growing pigs when compared to non-dehulled canola meal (Mejicanos et al. 2018).

In addition to concerns due to anti-nutritional factors, the amino acid digestibility of canola meal

is less when compared to that of soybean meal (Canola Council of Canada 2015). As a result, it is critical to formulate pig diets based on digestible amino acid levels, rather than total amino acid levels to ensure growth is not affected (Canola Council of Canada 2015). This was demonstrated in feeding trials conducted in Western Canada in which diets were balanced based on digestible levels of the limiting amino acids lysine and threonine. Proper formulation of diets resulted in equivalent feed intake, pig performance and quality between diets supplemented with soybean meal and varying levels of canola meal (Hickling 1994).

2.3.2.3 Poultry Feed

Canola meal is suitable for use in a variety of poultry feeds, however its primary limitation is its low energy availability compared to soybean meal (Wanasundara 2011). As a result, canola meal is primarily used in egg layer and turkey feed rather than broiler feed where higher energy feed is required (Arntfield and Hickling 2011). Similar to swine diets, successful inclusion of canola meal in poultry diets relies on formulations based on available amino acids rather than total amino acids (Arntfield and Hickling 2011; Canola Council of Canada 2015). While the inclusion of canola meal in layer diets has been shown to result in a reduced egg size in some studies (e.g. Summers et al. 1988), other studies have disputed these results (Badshah et al. 2001; Perez-Maldonado 2003). It has been suggested that formulating diets based on the content of digestible amino acids will reduce the potential for negative outcomes (Arntfield and Hickling 2011). The development of yellow-seeded canola has potential to improve the suitability of canola meal for inclusion in poultry feed due to lower fibre and glucosinolate content as well as higher protein content (Khajali and Slominski 2012).

2.3.2.4 Aquaculture Feed

Canola meal is commonly used as a protein supplement in a variety of aquaculture diets, including tilapia, catfish, seabream and shrimp (Arntfield and Hickling 2011). Canola meal is best suited towards inclusion in the diets of herbivorous and omnivorous species that naturally have diets high in plant-based material and lower in protein (Canola Council of Canada 2015). Canola meal has been commonly used as an ingredient in salmon and trout feed (Enami 2011). Research has shown that the availability of amino acids and digestibility of canola meal is high for many species of fish, including chinook salmon (Hajen et al. 1993). One study has shown that canola meal crude protein has a digestibility of 84.5%, higher than that of soybean meal, although lower than other animal protein sources (Hajen et al. 1993). In contrast to swine and poultry, many fish species are highly tolerant to glucosinolates (Canola Council of Canada 2015), although the presence of glucosinolates may have an effect on weight gain (Hajen et al. 1993). Similar to other livestock, the continued increase in protein content and reduction of fibre and anti-nutritional compounds will be important steps towards improving the suitability of canola meal for the aquaculture feed market.

2.3.2.5 Human Consumption

Dietary proteins are essential to the maintenance of human growth, health and function (WHO 2007). Worldwide, approximately 65% of protein in human diets comes from plant-based sources, while 35% comes from animal sources (Wu et al. 2014). While demand is increasing for animal protein sources, factors including an increasing global population, environmental sustainability interests and production efficiency have resulted in greater interest in plant protein sources for human consumption (FAO 2011; Campbell et al. 2016). Human dietary protein and amino acid

requirements depend on a variety of factors including climate, sex and age. The World Health Organization states that the average protein requirement for healthy adults is 0.66 g of protein per kg body weight per day (WHO 2007). In humans, nine amino acids are considered indispensable: leucine, isoleucine, valine, lysine, threonine, tryptophan, methionine, phenylalanine and histidine (WHO 2007). In plant-based protein sources a number of these amino acids are generally considered limiting, including the sulfur-based amino acids (methionine and cystine), threonine, tryptophan and lysine (Leinonen et al. 2019). Research has shown that while rapeseed meal protein has a low digestibility compared to other plant-based proteins, it has very high levels of indispensable amino acids, particularly the often-limiting sulfur-containing amino acids (Bos et al. 2007).

The primary barrier towards using canola meal as a source of protein for human consumption is the presence of antinutritional factors including glucosinolates, fiber, and phenolics, much the same as livestock (Tan et al. 2011). These antinutritional compounds can result in poor colour of products, an unpleasant taste, and poor digestibility (Wu and Muir 2008). The presence of these compounds make it difficult to incorporate canola or rapeseed meal directly into human food, thus isolation and processing of meal proteins is required to produce products that are suitable for human use (Tan et al. 2011; Wanasundara et al. 2017).

In addition to the presence of antinutritional factors, the 2S albumin napin is a member of a group of allergenic proteins found in seeds of Brassicaceae (Wanasundara et al. 2017). While little information is available regarding the allergenicity of *B. napus* seed proteins, immune responses to napin proteins have been noted in the related species *B. juncea* and *Sinapis alba* (Monsalve et al. 2001; L'Hocine et al. 2019). In addition, one study found that napin and cruciferin proteins

from *B. napus* were recognized immunologically by patients who had previously displayed sensitivities or allergic responses to other *Brassica* seeds (Puumalainen et al. 2015). While the potential for allergic reactions from the ingestion of *B. napus* protein is largely unstudied, it has been recommended that foods containing *B. napus* protein or protein isolates are labelled to indicate the presence of potential allergens (GRAS Notice 327 2010). While canola proteins display great potential to be used in human food products, the risk of allergenicity must be taken into consideration.

Despite these challenges, a number of enterprises have been successful in developing canola protein products and introducing them to the commercial market. In 2010, Burcon NutraScience Corporation introduced Puratein® and Supertein® to the market, both of which have achieved generally recognized as safe (GRAS) status and have been approved by the US Food and Drug Administration for use in food (GRAS Notice 327 2010). With unique functional properties, each of these products are suited towards different end uses. Puratein® is made up of the globulin protein fraction, containing a minimum of 80% cruciferin protein (GRAS Notice 327 2010) and is ideal for use in baked goods, protein bars, dressings, sauces and meat substitutes due to its emulsification and gelling properties and ability to form heat-stable foams (Burcon NutraScience Corporation 2021a). In contrast, Supertein® consists primarily of the albumin protein fraction, containing a minimum of 80% napin (GRAS Notice 327 2010), and its solubility and foaming ability make it suited for beverages, confectionary, aerated desserts and protein bars (Burcon NutraScience Corporation 2021b). In 2019, Burcon NutraScience Corporation also announced its Nutratein® product family, which are made up of canola and pea protein blends (Burcon NutraScience Corporation 2019). The Puratein®, Supertein® and Nutratein® products are under licence by Merit Functional Foods (Burcon NutraScience Corporation 2021c). In February 2021,

Merit Functional Foods announced the opening of a facility in Winnipeg, Manitoba, for the production of food-grade canola and pea proteins (Merit Functional Foods 2021). The introduction and use of commercially available canola protein isolates demonstrates that in a market dominated by soybean protein, there is great potential for canola meal to be used as a source of protein for human consumption.

2.3.3 Alternatives to Brassica napus Meal

While canola meal proteins are an emerging source of protein for human consumption, other oil-producing crops including soybean, cottonseed, sunflower, peanut, corn, flax and hemp may also be used as a source of dietary protein (Arntfield 2017). With the exception of soybean, these crops are relatively minor in their usage (Arntfield 2017). Therefore, the focus of this section will be on soybean protein.

Soybean is generally considered to be canola's primary competitor in the plant protein market, for both livestock and human consumption. While global rapeseed production was 71.9 million metric tons in 2018-2019, global soybean production was 362.1 million metric tons over the same period (USDA 2019), making it the dominant oilseed crop worldwide. Aside from its abundance, soybean is also favoured due to its high protein content in addition to its well-balanced amino acid composition (Table 2.2) (Day 2013; Nishinari et al. 2017). Further, the digestibility and energy availability of soybean meal is superior to that of canola meal (Wanasundara 2011).

Table 2.2 Comparison of protein content and amino acid balance of canola and soybean meals.

Component	Canola Meal (Canada)†	Soybean Meal (US)‡
Crude Protein Content (%)	36.7	47.8
Amino Acids (%)		
Alanine	1.57	2.02
Arginine	2.38	3.43
Asparagine	2.61	5.42
Cystine	0.82	0.73
Glutamine	6.53	8.58
Glycine	1.77	1.99
Histidine	1.22	1.22
Isoleucine	1.25	2.10
Leucine	2.22	3.57
Lysine	2.13	2.99
Methionine	0.70	0.68
Phenylalanine	1.46	2.33
Proline	2.15	2.34
Serine	1.44	2.32
Threonine	1.54	1.85
Tryptophan	0.48	0.65
Tyrosine	0.90	0.40
Valine	1.78	2.26

†Based on a 12% moisture basis (Canola Council of Canada 2015).
‡ Dehulled, based on a 12% moisture basis (U.S. Soybean Export Council 2015).

As demand for plant proteins increase, the opportunity to use canola meal as a source of protein for humans and animals may allow for further value from this oilseed crop to be gained. However, the ability of canola meal proteins to compete with other oilseed proteins in the global market depends upon genetic improvements being made. While canola breeding efforts of the past have focused heavily on oil content and quality, there exists a great opportunity to add new breeding goals. One focus must be to increase protein content and alter amino acid and protein composition. In addition, improving digestibility by further reducing antinutritional factors such as fibre content and glucosinolates would improve the suitability of canola for use by livestock, poultry, and humans. Directing breeding efforts towards these goals will allow canola meal and protein products to compete economically and functionally with other oilseed proteins.

2.4 Extraction and Quantification of Seed Storage Proteins

2.4.1 Brassica napus Protein Extraction

As discussed in Section 2.3, the presence of antinutritional compounds in canola and rapeseed meal make it difficult to use directly in human food, necessitating the use of extraction and processing methods to produce protein isolates (Tan et al. 2011; Wanasundara et al. 2017). While many protocols have been studied and utilized in laboratories, on an industrial scale these methods generally recover protein from oil-extracted meal, through a variety of aqueous extraction processes (Figure 2.2).

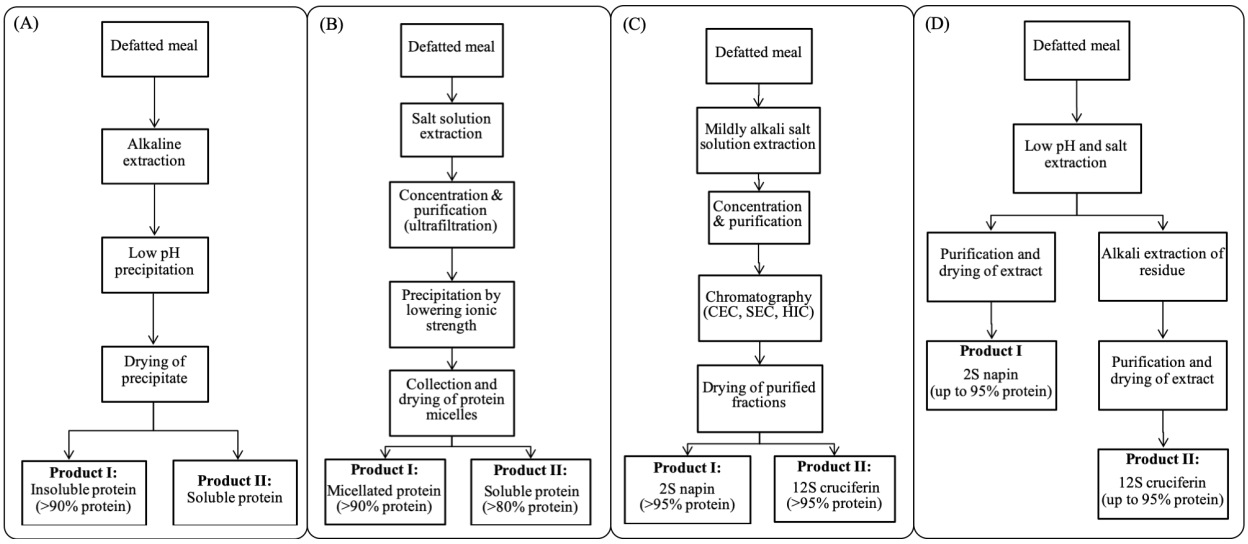


Figure 2.2 Summary of extraction and processing methods for production of canola protein products and isolates. (A) Aqueous alkaline solution method (Klockeman et al. 1997; Tan et al. 2011); (B) Protein micelle mass formation method (Ismond and Welsh 1992; Murray 1999); (C) Chromatographic separation method (Bérot et al. 2005); (D) Solubility-based separation method (Wanasundara and McIntosh 2013). Adapted from Wanasundara et al (2017).

2.4.1.1 Aqueous Alkaline Solution

The most widely reported method by which canola and rapeseed isolates are produced is alkaline extraction of proteins followed by dilute acid precipitation (Tan et al. 2011). Typically, an alkaline extraction buffer of NaOH (pH 11-12) and NaCl or CaO is used to extract protein from defatted meal (Wanasundara et al. 2017). The pH of the supernatant may then be adjusted to the isoelectric point, where protein solubility is minimized and the proteins precipitate (Wanasundara et al. 2017). The values of pH 3.5 (Klockeman et al. 1997; Diosady et al. 2005), pH 4.0 (Aluko and McIntosh 2001), pH 4.5-5.5 (Ghodsvali et al. 2005), or two pH levels in a step-wise manner (El Nockrashy et al. 1977; Pedroche et al. 2004) have all been reported as maximizing protein yield. The protein content of isolates produced in this manner are usually between 60-90% protein (Tan et al. 2011; Wanasundara et al. 2017).

2.4.1.2 Protein Micelle Mass Formation

Protein micelle mass (PMM) formation is a method of protein extraction that was developed to reduce the presence of antinutritional factors (Ismond and Welsh 1992). In this method, protein is extracted from defatted meal using a solution of NaCl. The solubilized protein is then concentrated and purified, often by ultrafiltration. Dilution of the protein concentrate reduces the ionic strength of the protein, resulting in precipitation in the form of a gelatinous protein micelle, which is collected and dried (Ismond and Welsh 1992; Murray 1999; Tan et al. 2011).

2.4.1.3 Chromatographic Separation

The production of canola protein isolates by chromatographic separation has also been reported. Bérot et al. (2005) report a method in which protein is extracted in a buffer containing 50 mM

Tris-HCl, 750 mM NaCl, 5 mM EDTA and 0.3% $NaHSO_3$. The concentrated proteins are then subjected to multiple stages of chromatography. First, cation exchange chromatography (CEC) is performed, which produces cruciferin in the unbound fraction and napin in the bound fraction. The cruciferin fraction is then further purified by size exclusion chromatography (SEC), and the napin fraction may be purified by hydrophobic interaction chromatography (HIC). These purified fractions may then be dried to produce protein isolates (Bérot et al. 2005).

2.4.1.4 Solubility-Based Separation

In 1897, Osborne suggested that it was possible to categorize maize proteins based on their solubilities, that is those soluble in water, salt solutions, alkaline solutions and alcohol (Osborne 1897). This knowledge allows us to isolate canola proteins based on their differing solubilities. Wanasundara and McIntosh (2013) describe a step-wise method in which napin is extracted under low pH and salt conditions. Cruciferin remains insoluble under these conditions and may then be extracted from the remaining residue under alkali conditions. The napin- and cruciferin- rich protein extracts may then be purified and dried to produce protein isolates.

2.4.2 Quantification of Brassica Proteins

Currently, a number of methods are commonly used for the quantification of proteins from seed extracts. One well-established method is the Kjeldahl method, in which acidic digestion of a sample converts protein to ammonia, which is then distilled into an acid and may be titrated to estimate the quantity of nitrogen in the original sample and therefore the amount of protein, based on a nitrogen conversion factor (Barthet and Daun 2011). This method has been used by researchers to quantify seed storage proteins in canola, following the preparation of cruciferin or

napin protein isolates (Cheung et al. 2014, 2015). While this method is well characterized and often automated, it is a tedious, multi-step process that also produces a number of dangerous waste products (Barthet and Daun 2011).

Similar to the Kjeldahl method, the combustion method of protein content determination has also been utilized. The high-temperature combustion of a sample releases nitrogen, which may then be measured by thermal conductivity and converted to a protein quantity by a nitrogen conversion factor (AOAC 2005). Similar to the Kjeldahl method, the combustion method may be used to estimate total protein, but has also been used to estimate protein content of isolated and purified canola seed storage proteins (Wanasundara et al. 2012; Marambe et al. 2014; Perera et al. 2016). One criticism of both the Kjeldahl method and the combustion methods, however, is that the nitrogen content conversion factor used (typically 6.25) assumes the nitrogen content of all proteins to be 16%. However, the actual nitrogen content of proteins differ depending on the amino acid composition (Mariotti et al. 2008). Therefore, the use of a conversion factor may provide an inaccurate estimate of the protein content of some samples.

Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) may also be used to quantify proteins in an aqueous sample. One of the most common techniques used to study protein biochemistry, electrophoresis separates charged protein molecules based on their movement through an electric field (Chiou and Wu 1999). SDS-PAGE separates proteins or polypeptides within a sample based on their varying molecular weights and surface charges, due to their varied ability to move through the pores of a polyacrylamide gel matrix (Srinivas 2012). The resulting separated bands of protein may then be visualized, often through staining with dyes such as Coomassie Blue (Srinivas 2012; Schatzki et al. 2014; Defaix et al. 2019). This allows for

comparison of the intensity of the resulting protein bands by densitometry, and the estimation of protein content (Defaix et al. 2019). By extracting total soluble protein from seed samples, Schatzki et al. (2014) utilized this method to quantify cruciferin and napin proteins in a winter oilseed rape population.

Most recently, Defaix et al. (2019) described a size-exclusion high performance liquid chromatography (SE-HPLC) method for quantifying the napin and cruciferin protein fractions in rapeseed. SE-HPLC separates proteins according to their size (Hong et al. 2012; Albe Slabi et al. 2019). By passing a sample through a column of porous particles, large proteins that are excluded from the pores elute first, and smaller proteins elute later (Hong et al. 2012). By calibration with markers of known molecular weight, unknown protein samples may be identified based on their retention time in the column (Schrag et al. 2014). The method outlined by Defaix et al. (2019) utilizes SE-HPLC to separate proteins, and detects and quantifies proteins based on UV signal (Defaix et al. 2019).

Enzyme-linked immunosorbent assays (ELISAs) may also be used to quantify proteins and will be discussed further in the following section.

2.4.3 Introduction to Enzyme-Linked Immunosorbent Assay (ELISA)

Many of the methods described in the previous section are time consuming, and if the quantification of a specific protein is desired, many require selective extraction and isolation of the protein of interest. In contrast to these methods, ELISAs are able to detect and quantify very small quantities of a protein of interest in a sample, with relative efficiency (Crowther 2009). First described by Engvall and Perlmann (1971) to quantify IgG in rabbit serum, the ELISA technique

is a variant of the previously described radioimmunoassay (Yalow and Berson 1960), and was developed by replacing the previously used radioisotope with an enzyme due to concerns of safety (Gan and Patel 2013). ELISAs use an enzyme-conjugated antibody to bind and detect a specific antigen, such as a protein, peptide or hormone (Gan and Patel 2013).

Four basic ELISA formats exist. The simplest, and upon which all others are based, is the direct ELISA. In this method, an antigen is bound to the solid phase (usually the well of a microtiter plate), and any un-bound antigen is washed away. Enzyme-conjugated antibodies targeting the antigen of interest are then added, and once again any un-bound antibody is washed away. Substrate solution is then added, and the enzyme catalyzes a reaction that produces a colour change which may be quantified with a spectrophotometer (Crowther 2009). A common modification to this protocol is known as the indirect ELISA, in which the primary antibody is unlabeled, and a secondary, enzyme-labelled anti-species antibody that targets the primary antibody is added, to enhance the signal (Figure 2.3) (Lin 2015). The sandwich ELISA format modifies the direct ELISA protocol by preparing the surface of the solid phase with an antibody designed to bind the antigen of interest. The antigen-containing sample is added, and the target antigens are captured by the bound antibody. An enzyme-conjugated primary antibody may then be added, "sandwiching" the antigen between two antibodies (Crowther 2009; Gan and Patel 2013). Finally, competitive ELISAs are based on a competition for the primary antibody between antigens bound to the wells of a microtiter plate and antigens in a sample. In this format, a primary antibody is incubated with a sample antigen, forming antibody-antigen complexes. This mixture is then added to the wells of a microtiter plate, to which the same antigen has been bound. After incubation and washing, substrate is added. In this format, the greater the quantity of antigen in the original sample, the less primary antibody is available to bind to the antigens coated to the well. Therefore,

an absence of colour change indicates a higher presence of antigen (Gan and Patel 2013).

While the earliest use of ELISA methods were in the fields of immunology, molecular biology and biotechnology, ELISAs are now also widely used in the fields of environmental sciences, veterinary sciences, medicine, pharmacology and toxicology as well as agriculture and biological sciences (Crowther 2009). A number of ELISA-based methods have been developed to detect seed proteins. In soybean seeds, sulfur-rich protein has been detected and quantified by ELISA (Monaghan et al. 2008). The abundant and potentially allergenic proteins beta-conglycinin (Moriyama et al. 2005; Hei et al. 2012) and glycinin (Chen et al. 2014) have also been detected and quantified in soybean seed by ELISA methods. Finally, Marambe et al. (2014) developed an indirect ELISA protocol to detect and quantify an allergenic 2S napin protein (Sin a 1) in yellow mustard (*Sinapis alba*). These studies demonstrate that ELISAs are a rapid and straight-forward method that may be successfully used for detection and screening of seed storage proteins.

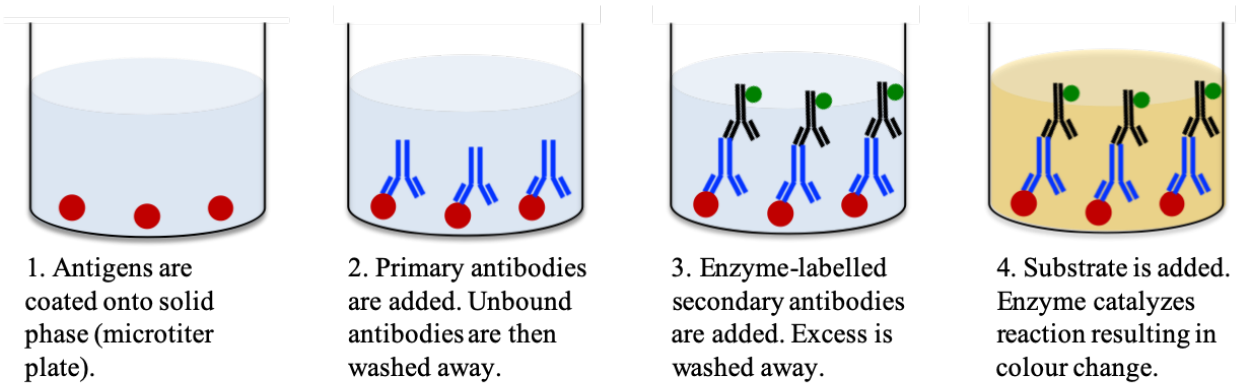


Figure 2.3 Indirect enzyme-linked immunosorbent assay (ELISA) technique. Antibodies bind an antigen of interest. Bound antibodies may then be detected by an anti-species enzyme-labelled antibody (Adapted from Crowther 2009; Gan and Patel 2013).

2.5 Introduction to Near-Infrared Spectroscopy

Electromagnetic energy in the form of light can be arranged in a spectrum based on wavelength and energy (Agelet and Hurburgh 2010). The near-infrared (NIR) region of the electromagnetic spectrum begins just above the visible light region, from 750 nm to 2500 nm, and is considered the most energetic infrared region (Agelet and Hurburgh 2010). When a material absorbs energy in the form of NIR radiation, information about the molecular bonds within the sample may be ascertained (Shenk et al. 2008). It is upon these principles that near-infrared spectroscopy (NIRS) is based.

When a material is exposed to light in the form of NIR radiation, interactions in the form of reflection, refraction, absorption, diffraction and transmission may occur. When the frequency at which the NIR radiation is vibrating matches the vibration frequency of the molecular bonds within a sample, the energy is absorbed, resulting in further bending, stretching and deformation of the molecular bonds (Shenk et al. 2008). When a spectrophotometer is used, the results of the absorption by molecular bonds may be visualised as a spectrum. In particular, NIR radiation is absorbed by the oxygen-hydrogen (O-H), carbon-hydrogen (C-H), nitrogen-hydrogen (N-H), sulfur-hydrogen (S-H) and carbon-oxygen (C=O) bonds in biological samples, and therefore the spectra produced yield information regarding the composition of the sample of interest (Chodak 2008; Agelet and Hurburgh 2010). Since NIR spectra contain information regarding the molecular bonds within a sample, this data may be correlated to chemical or physical properties of the samples through calibration with a reference method (Shenk et al. 2008). Samples of the material of interest are analyzed by an established laboratory method, to obtain a set of reference values. A calibration model may then be developed, correlating the spectral data with the chemical or

physical property of interest (Agelet and Hurburgh 2010).

Due to the efficiency and non-destructive nature of NIRS phenotyping, this method has been widely adopted into *B. napus* breeding programs to assess a variety of seed quality traits. These traits include chlorophyll content (International Organization for Standardization 2015), glucosinolate content (Daun et al. 1994; Velasco and Becker 1998a), oil content (Tkachuk 1981; Daun et al. 1994; Velasco et al. 1999b, 1999a), fatty acid composition (Sato et al. 1998; Velasco and Becker 1998b; Velasco et al. 1999b, 1999a; Kim et al. 2007), fiber content (Font et al. 2003, 2005; Dimov et al. 2012; Wittkop et al. 2012) and total protein content (Tkachuk 1981; Daun et al. 1994; Velasco and Möllers 2002).

2.6 Association Mapping

2.6.1 Introduction to Association Mapping

At its core, purpose of the study of genetics is to connect genotype with phenotype (Oraguzie and Wilcox 2007). In plant agriculture, the genetic inheritance of the majority of traits of interest is complex. These quantitative traits are influenced by the combined forces of many genes, the environment and interactions between the two (Stich and Melchinger 2010). One method by which the genetic basis of quantitative traits may be understood is through genetic mapping, in which inherited markers are identified that are physically close to the genes underlying quantitative traits of interest (Oraguzie and Wilcox 2007).

The two most commonly used methods by which the inheritance of quantitative traits are studied are quantitative trait loci (QTL) mapping and association mapping (Stich and Melchinger 2010). Association mapping is a method that reveals statistical associations between genetic

polymorphisms and phenotypic variation (Breseghello and Sorrells 2006). While QTL mapping uses populations of a known structure (often early generation crosses between two parents), association mapping uses populations of unrelated individuals (Oraguzie and Wilcox 2007). As such, while QTL mapping studies the variation and recombination events observed in the progeny of specific individuals, association mapping makes interpretations based on the natural variation and past recombination events of a species (Oraguzie et al. 2007).

The concept of linkage disequilibrium (LD) is central to association mapping. Linkage disequilibrium is "the nonrandom association of alleles at different loci" (Oraguzie et al. 2007). Loci are said to be in LD when two alleles at two separate loci are found together more often than if they were combining independently of one another (Oraguzie et al. 2007). Since association mapping studies use unstructured populations, recombination events over many years will have removed associations between markers and functional loci, unless the two are very closely linked (Stich and Melchinger 2010).

Association mapping studies begin with the assembly of germplasm to produce the mapping population. Populations may be developed from germplasm bank collections, elite breeding material, or synthetic populations (Breseghello and Sorrells 2006). The choice of germplasm is important, and affects the mapping resolution, statistical power and methods used (Stich and Melchinger 2010). Next, the population must be phenotyped to produce information about the trait(s) of interest. The population is then genotyped, producing molecular marker data (Hall et al. 2010; Stich and Melchinger 2010). Markers used may include random amplified polymorphic DNA (RAPD) markers, amplified fragment length polymorphism (AFLP) markers, and simple sequence repeats (SSRs) (Hall et al. 2010; Stich and Melchinger 2010). However, due to their

suitability for use in high-throughput detection systems as well as their high genome density and low mutation rate, single-nucleotide polymorphisms (SNPs) are often the preferred marker for association mapping studies (Stich and Melchinger 2010). Following the collection of phenotypic and genotypic data, association analysis may be carried out to statistically analyze marker-phenotype associations (Stich and Melchinger 2010). Since the first association mapping studies in plants were performed in maize (Bar-Hen et al. 1995) and rice (Virk et al. 1996) association mapping has been used in many crop plants, and is considered important in many breeding programs (Gupta et al. 2014)

2.6.2 Molecular Markers

Molecular markers arise from polymorphisms in the DNA sequences of individuals (Bernardo 2014). Once a linkage between a marker and a trait of interest has been discovered through mapping, the marker can be used for indirect selection for the trait of interest through marker assisted selection (MAS) (Singh and Singh 2015). As a result, molecular markers have become an invaluable tool to plant breeders through which to speed phenotyping and to facilitate mapping studies, the selection of parents, cultivar identification, and genetic diversity studies (Sleper and Poehlman 2006). A large number of marker types have been developed, a selection of which will be discussed in the following sections.

2.6.2.1 Restriction Fragment Length Polymorphisms (RFLPs)

The first DNA-based marker was restriction fragment length polymorphism (RFLP). First described by Botstein et al. (1980) for linkage mapping in humans, RFLPs are described as a co-dominant marker. By digesting the DNA of an individual with restriction enzymes, fragments are

produced that may be separated based on size by electrophoresis, transferred onto a membrane and hybridized with radioactive probes (Botstein et al. 1980). Genotypic differences between individuals are demonstrated by the differing fragment lengths produced (Botstein et al. 1980). While RFLPs have proven useful for the construction of the first whole-genome linkage maps in plants including tomato and maize (Bernatzky and Tanksley 1986; Helentjaris et al. 1986), a number of challenges exist, including the time-consuming nature of this technique, and the requirement for radioactive materials (Mohler and Schwarz 2005). Therefore, the use of RFLPs has largely fallen out of favour for use in breeding programs and has been replaced by other marker techniques.

2.6.2.2 Amplified Fragment Length Polymorphisms (AFLPs)

The development of polymerase chain reaction (PCR) by Mullis et al. (1986), lead to the discovery of PCR-based markers, including amplified fragment length polymorphisms (AFLPs). Developed by Vos et al. (1995), AFLPs are dominant markers (Sleper and Poehlman 2006) capable of revealing high levels of polymorphism (Lombard et al. 2000). This technique is based upon the ligation of oligonucleotides to restricted DNA and selective PCR amplification of the restriction fragments with adapter specific primers, which may then be visualized by gel electrophoresis (Vos et al. 1995; Gali and Sharpe 2011). Due to the reliability and reproducibility of AFLP markers, this technique has been used widely in plant research (Mohler and Schwarz 2005).

2.6.2.3 Simple Sequence Repeats (SSRs)

Similar to AFLP markers, simple sequence repeats (SSRs) are also PCR-based, and are considered a “second-generation” PCR-based marker (Gali and Sharpe 2011). These markers consist of 2-7

base pairs repeated many times in tandem (Gali and Sharpe 2011; Jonah et al. 2011), and are naturally distributed throughout the genome of eukaryotes (Hearne et al. 1992). SSR markers are codominant (Sleper and Poehlman 2006), highly polymorphic, and considered generally stable, and therefore are considered well-suited for constructing genetic maps (Hearne et al. 1992).

2.6.2.4 Single Nucleotide Polymorphisms (SNPs)

Considered the next generation of genetic markers, single nucleotide polymorphisms (SNPs) are individual nucleotide positions in the DNA sequence that differ between two individuals (Mohler and Schwarz 2005; Edwards et al. 2007). Thus, their identification relies upon the availability of sequence information (Gali and Sharpe 2011). Typically, SNPs are the most frequent form of genetic polymorphisms found in the genome (Mohler and Schwarz 2005; Edwards et al. 2007; Singh and Singh 2015). In addition, their nature as a codominant marker makes them capable of distinguishing between homozygous and heterozygous loci (Gali and Sharpe 2011). For these reasons, along with their ability to be utilized in automated, high-throughput genotyping, SNP markers are attractive for many uses including mapping studies and MAS (Mohler and Schwarz 2005).

As the availability of sequenced reference genomes for plant species has grown, so too has the usage of SNP markers in plants (Gali and Sharpe 2011). SNP discovery has been frequent among *Brassica* crop species, including *B. napus*. For example, Delourme et al. (2013) developed a high-density SNP-based map which allowed the study of linkage disequilibrium in *B. napus* collections with different seed quality types and growth habits. Raman et al (2014) used an Illumina Infinium™ array to develop a SNP-based genetic map to study flowering time and blackleg resistance in *B. napus*. Recently, a high-density Illumina Infinium™ array containing 52,157 SNP

markers was developed for *B. napus* (Clarke et al. 2016). This advancement provides a valuable resource for mapping and genome-wide association studies in *B. napus* (Clarke et al. 2016), and will be discussed further in a later section.

2.6.3 The Brassica napus Reference Genome

The availability of plant reference genomes is important to facilitate many breeding efforts. Since the model crucifer *Arabidopsis thaliana* was the first plant genome to be sequenced (The Arabidopsis Genome Initiative 2000), efforts to sequence the genome of other plant species have been undertaken. In 2014, a collaborative effort between many scientists around the world resulted in the assembly of the *B. napus* genome (Chalhoub et al. 2014). This undertaking used the European winter oilseed cultivar 'Darmor-*bzh*' as the genotype to be sequenced. A combination of 454 GS-FLX+ Titanium (Roche, Basel, Switzerland), Sanger sequencing and Illumina (San Diego, CA) HiSeq sequencing techniques were applied (Chalhoub et al. 2014). The previously assembled genomes for *B. rapa* (Wang et al. 2011) and *B. oleracea* (Liu et al. 2014) assisted in assigning scaffolds to the A and C sub-genomes (Chalhoub et al. 2014; Sun et al. 2018). A final assembly of 849.7 Mb was produced, covering approximately 79% of the genome. In addition, a SNP map was created, genetically anchoring 84% of the genome assembly (Chalhoub et al. 2014).

Since the sequencing of the winter-type 'Darmor-*bzh*', other types of *B. napus* have been sequenced, including the winter-type 'Tapidor' (Bayer et al. 2017) and the semi-winter type 'ZS11' (Sun et al. 2017). To more accurately represent the genetic diversity present among *B. napus,* a pan-genome has been constructed. The pan-genome was developed from eight newly assembled winter-, semi-winter- and spring-type accessions, in addition to the previously sequenced 'Darmour-*bhz*' (Song et al. 2020). The result is a set of genetic resources that play an

important role in efforts to exploit existing genetic variation for the improvement of *B. napus* cultivars and may be used for the identification of candidate genes for traits of interest and the development of genetic markers for MAS.

2.6.4 The Brassica 60K Illumina Infinium™ SNP Genotyping Array

The development and release of *Brassica* genome sequences has facilitated many research efforts, including the development of a high-density SNP Illumina Infinium™ array containing 52,157 markers for genotyping in *B. napus* (Clarke et al. 2016). The development of this array required significant consideration due to a number of challenges. The *B. napus* genome is highly complex, due to its allopolyploid nature and homologous sub-genomes (Sun et al. 2018). As a result, the development of this array included efforts to limit the impact of genome duplication. Reference mapping of sequences to the genomes of *B. rapa* and *B. oleraceae* (the diploid ancestors to *B. napus*) facilitated limiting the selected SNPs to those that were represented only a few times across the genome. The result is an array in which approximately 60% of the SNP assays selected display genome-specificity (Clarke et al. 2016).

Illumina Infinium™ assays are capable of genotyping up to 1 million polymorphic sites per sample, and multiple samples per experiment (Chagné et al. 2015). As described by Clarke et al. (2016), the development of the *Brassica* assay required that probes be designed to the sequences adjacent to the SNPs selected to be assayed. To use the assay, DNA is amplified by whole-genome amplification, is fragmented, and is then captured on the bead array by hybridization to the SNP primers. The probe is then extended by single labelled nucleotides, and florescence may be detected (Mason et al. 2017).

The development of the *Brassica* 60K Illumina Infinium™ SNP array provides a platform for genetic analysis of *B. napus*. Due to the ease of sample preparation, relatively simple data analysis, and high levels of reproducibility, this array has been widely adopted for use (Mason et al. 2017). In particular, it has been utilized in a number of genome-wide association mapping studies (GWAS) for the identification of loci associated with traits including *Sclerotinia* stem rot (Wei et al. 2016), plant architecture traits (Liu et al. 2016a), seed oil content (Liu et al. 2016b) as well as seed quality traits including erucic acid content, glucosinolate content, oil content and seed weight (Li et al. 2014)

2.6.5 Mapping Studies for Seed Quality Traits in Brassica napus

As genetic resources become more available, more and more efforts are being put forth to dissect *B. napus* seed quality traits of interest through quantitative trait loci (QTL) and association mapping studies. Since oil content is one of the major traits of economic value, many mapping studies in *B. napus* have been carried out with the goal of identifying markers for seed oil content (Gacek et al. 2018). A recent study by Liu et al. (2016b) genotyped 521 *B. napus* accessions with the *Brassica* 60K Illumina Infinium™ SNP array, and through a genome-wide association study (GWAS), identified 50 loci that are associated with seed oil content, and which explain approximately 80% of the total phenotypic variance. Li et al. (2014) also used GWAS to identify SNPs associated with seed quality traits. A panel of 457 *B. napus* accessions were genotyped with the *Brassica* 60K Illumina Infinium™ SNP array. Association mapping revealed two SNPs associated with erucic acid content on A08 and C03, four SNPs associated with glucosinolate content on A09, C02, C07 and C08, two SNPs associated with seed weight on A07 and A09, as well as one SNP associated with oil content on A08. Recently, Chao et al (2017) used QTL

mapping to identify QTL for seed oil and seed protein content. For seed oil content, 67 QTL were identified on A03, A08, A09, C03, C05 and C06 that explain 22% of the observed phenotypic variation. Additionally, 38 QTL for protein content on A03, A09, C03 and C05 were identified, and explain up to 27% of the phenotypic variation (Chao et al. 2017). Finally, one QTL mapping study identifying QTL for *B. napus* seed storage proteins has been performed. Schatzki et al. (2014) used a winter oilseed rape doubled haploid population and a set of 229 SSR and AFLP markers to detect QTL for cruciferin and napin proteins. A total of three QTL for napin were identified on A02, C06 and C09, explaining 47% of the phenotypic variation, while two QTL for cruciferin were identified on A02 and C09, which explain 35% of phenotypic variation (Schatzki et al. 2014). While many studies focusing on oil content have been performed, further study of the genetic control of protein content and specific seed storage proteins is required to facilitate efficient breeding efforts and the improvement of protein related traits.

2.8 Objectives

Through the use of a diverse panel of *B. napus* genotypes, the objective of this research was to explore existing variation in cruciferin content, investigate the genetic control of cruciferin content, and improve phenotyping efficiency to facilitate future breeding efforts. Three studies were developed toward this objective. First, diverse *B. napus* genotypes were grown across multiple site-years, and cruciferin content was analyzed to elucidate the extent of existing phenotypic variation for cruciferin content and the role of environment in its expression. Next, a study with the objective of exploring the potential for NIRS to be used for cruciferin content determination was performed. Finally, an association mapping study to identify significant SNP-trait associations was performed. As interest in alternative and plant-based protein grows, there

exists an opportunity to work towards the improvement of *B. napus* meal protein for existing livestock feed markets as well as the developing human food markets. The success of these efforts depends upon a greater understanding of protein-related traits in *B. napus*.

3. Assessment of Cruciferin Content in a Diverse Population of *Brassica napus* L.

3.1 Abstract

As the second greatest produced oilseed crop worldwide, and a crop of great economic importance to the Canadian economy, breeding efforts to improve the suitability of *Brassica napus* L. for various end-uses have been significant. While breeding interest from a seed quality perspective has often focused on improving oil content and quality and reducing the presence of anti-nutritional compounds, recent interest has been directed to increasing the value of *B. napus* meal proteins for a wider variety of end uses. The two major seed storage proteins found in the meal are napin (2S) and cruciferin (11S), which make up approximately 20% and 60% of the total protein content of mature seeds, respectively. This research aimed to explore phenotypic variation and the effect of genotype and environmental factors on cruciferin content and other seed quality traits in a diverse population of *B. napus*. After four years of field trials, an enzyme-linked immunosorbent assay (ELISA)-based approach was used to determine cruciferin content. Considerable variation in cruciferin content was observed, and several genotypes expressing consistently high or low cruciferin contents were identified. The effect of genotype by site-year interactions was shown to significantly affect cruciferin content. In addition, analysis of the correlation between cruciferin content and other seed quality traits revealed a negative correlation between cruciferin and glucosinolate content. As greater interest is directed towards plant-based proteins, this research will provide a foundation upon which breeding efforts to develop *B. napus* with specialty protein profiles may be based.

3.2 Introduction

Brassica napus L., commonly known as rapeseed or canola, is the second highest produced oilseed crop worldwide (FAO 2019), with an estimated global production of 69.54 million metric tons in 2019-2020 (USDA 2019). In Canada, the canola sector contributes an average of C$26.7 billion per year to the Canadian economy (LMC International 2016). Canola and rapeseed are grown primarily for seed oil production, while the protein-rich meal is often considered a by-product (Rahman 2013; Campbell et al. 2016).

Due to the primary use of canola as a source of oil, seed quality traits such as the improvement of oil content, modification of the fatty acid composition of the oil, and a decrease in glucosinolate content have historically been the primary focus of canola breeding efforts (McVetty and Scarth 2002; Malabat et al. 2003; Schatzki et al. 2014). While glucosinolate content can be as high as 100 µmol per gram of seed, modern canola cultivars have been developed to exhibit glucosinolate content well below 30 µmol per gram of seed (Wittkop et al. 2009; Barthet and Daun 2011; Rahman 2013). In addition, efforts to increase seed oil content have been successful, resulting in seed oil content up to 40-50% of the total seed biomass (Wittkop et al. 2009; Wanasundara 2011). While *B. napus* is primarily produced for oil, the protein-rich meal that remains following oil extraction can be up to 40% crude protein on a dry matter basis (Wittkop et al. 2009; Wanasundara 2011; Rahman 2013). The negative correlation between seed oil content and seed protein content has been well documented (Malabat et al. 2003; Zhao et al. 2006; Schatzki et al. 2014; Gu et al. 2017; Chao et al. 2017), presenting a challenge to efforts to widen the potential end uses of *B. napus* protein by increasing seed protein content without affecting seed oil content.

The seed protein of *B. napus* is found predominantly in the form of two storage proteins, cruciferin and napin (Aider and Barbana 2011; Wanasundara 2011). Napin is a 2S albumin that makes up approximately 20% of the total protein content of mature seeds, while cruciferin is a 11S globulin that makes up approximately 60% of the protein content of mature seeds (Aider and Barbana 2011; Wanasundara 2011; Perera et al. 2016). A negative correlation (-0.36) between cruciferin and napin content has been reported (Schatzki et al. 2014), and plants with anti-sense genes for napin accumulate 1.4 to 1.5 times more cruciferin (Kohno-Murase et al. 1994). The primary value of these *B. napus* proteins as a source of protein for human consumption lies in their functional properties in food products. Napin has been reported as providing good foaming properties (Aider and Barbana 2011; Wanasundara et al. 2017) and solubility (Tan et al. 2011; Wanasundara and McIntosh 2013). In contrast, cruciferin is characterized by its emulsifying (Cheung et al. 2015) and gelling properties (Tan et al. 2014b). Due to the potential value of cruciferin as a source of protein for human consumption, there is a need for research to improve our understanding of this protein. While it is hypothesized that phenotypic variability exists among *B. napus* genotypes, the extent of the variation in cruciferin content among genotypes is largely unknown. Thus, a more thorough understanding of this variation would assist in future efforts to breed *B. napus* with unique seed storage protein profiles.

In breeding programs, phenotyping is often costly and time consuming (Furbank and Tester 2011). Therefore, methods that are rapid and with a high capacity are important (Pieruschka and Schurr 2019). Currently, a number of methods have been utilized to quantify cruciferin protein. Such methods include selective extraction and quantification of cruciferin protein (Cheung et al. 2014) or the extraction of total soluble protein followed by SDS-PAGE (Schatzki et al. 2014). In contrast to these methods, enzyme-linked immunosorbent assays (ELISAs) are able to handle sizable

numbers of samples with relatively great efficiently, while requiring small seed sample sizes (Crowther 2009). First described by Engvall and Perlmann (1971) as a variant of the previously described radioimmunoassay (Yalow and Berson 1960), the ELISA technique allows for the detection and quantification of a protein of interest by the use of an enzyme-labelled antibody (Lin 2015). Widely used in such fields as medicine, immunology, microbiology and biological sciences (Crowther 2009), the ELISA technique is also well-suited for use in *B. napus* breeding programs.

The objective of this research was to utilize a diverse population of *B. napus* to determine the extent of existing variation in cruciferin content and determine the effect of genotype and environmental factors on cruciferin content in *B. napus.* This research hypothesizes that variation in cruciferin content will be discovered among genotypes, and that there will be a significant effect of the environment on cruciferin content in *B. napus*.

3.3 Materials and Methods

3.3.1 Plant Material

The *B. napus* genotypes utilized in this experiment represent the 51 parental genotypes (as well as an additional check genotype 'DH12075') of a nested association mapping population developed by Agriculture and Agri-Food Canada in Saskatoon, Saskatchewan (Parkin et al. 2017). The genotypes were selected to represent the phenotypic diversity across spring *B. napus* for a wide number of traits of interest.

3.3.2 Field Evaluation

In 2014 and 2015 phenotypic field evaluation was carried out near Saskatoon, Saskatchewan by Agriculture and Agri-Food Canada. The soil at both locations is classified as Sutherland orthic

dark brown soil (Acton and Ellis 1978). Field evaluation in both 2014 and 2015 was carried out in a randomized complete block design, with two replicates per genotype.

In 2014, the trial was located at Latitude 52.19, Longitude -106.50. Genotypes were planted in 3 m paired-row plots with a 60 cm row spacing. Planting took place on May 28. In the fall of 2013, 89.7 kg ha^{-1} nitrogen was applied in the form of anhydrous ammonia. In the spring, an additional 50.4 kg ha^{-1} nitrogen, 80.7 kg ha^{-1} phosphorous and 50.4 kg ha^{-1} sulfur were applied, along with 20.51 kg ha^{-1} of Edge™ Granular Herbicide (Gowan Canada, Winnipeg, MB). To allow sufficient dry down of plants prior to harvest, Reglone desiccant (Syngenta Canada, Guelph, ON) was applied. Plots were harvested on October 7 with a Wintersteiger small-plot combine (Nursery Master Classic, Wintersteiger, Salt Lake City, UT).

In 2015, the trial was located at Latitude 52.16, Longitude -106.57. Genotypes were planted in 6 m 4-row plots. Planting took place on May 13. In the fall of 2014, 84.1 kg ha^{-1} nitrogen was applied in the form of anhydrous ammonia. In the spring of 2015 prior to planting, 20.51 kg ha^{-1} of Edge™ Granular Herbicide was applied, as well as PrePass (Dow AgroSciences Canada Inc, Calgary, AB) at a rate of 98.8 mL ha^{-1} and Roundup (Monsanto Canada ULC, Winnipeg, MB) at a rate of 3.1 L ha^{-1}. Plots were differentially harvested with a Wintersteiger small-plot combine when direct-cut maturity was reached for each plot, between September 22 and October 2.

In 2016 and 2017, evaluation was carried out at the University of Manitoba Winnipeg research farm (Latitude 49.81, Longitude -97.12), where the soil is classified as Riverdale Silty Clay (Ehrlich 1953). Field evaluation was conducted in a randomized complete block design, with 3 replicates per genotype, planted in 3 m rows with a 40 cm row spacing. In 2016, planting took place on June 7, while in 2017 planting occurred on May 17. Agronomic management was the

same in 2016 and 2017. In the fall prior to planting, Edge™ Granular Herbicide was applied at a rate of 31.25 kg ha^{-1}, and fertilizer was applied at a rate of 130.0 kg ha^{-1} nitrogen, 56.3 kg ha^{-1} phosphorus and 27.0 kg ha^{-1} sulfur. Matador® (Syngenta Canada, Guelph, ON) insecticide was applied twice per season for control of insect pests including flea beetles at a rate of 0.84 L ha^{-1}. Once per season, at the four-leaf crop stage, a tank mixture of Poast® Ultra (BASF Canada Inc, Mississauga, ON) at a rate of 0.675 L ha^{-1}, Muster® (DuPont Canda, Mississauga, ON) at a rate of 30 g/ha, and Lontrel™ 360 (Dow AgroSciences Canada Inc., Calgary, AB) at a rate of 0.85 L ha^{-1} were applied along with Merge® adjuvant at a rate of 1 L per 109 L pesticide solution (BASF Canada Inc, Mississauga, ON).

In 2016, the trial was harvested on September 28 by cutting each row by hand and tying it into a bundle to dry in the field. When bundles were dry, they were threshed using a Wintersteiger small-plot combine and cleaned manually. Harvested seed samples (20-30 g) of each genotype were taken for seed quality analysis as well as cruciferin content analysis. In 2017, the trial was harvested between September 6 and September 8. A seed sample (20-30 g) from each row was harvested by hand, manually cleaned, and taken for seed quality and cruciferin content analysis.

3.3.3 Analysis of Cruciferin Protein by Enzyme-Linked Immunosorbent Assay

3.3.3.1 Seed Defatting

Prior to total soluble protein extraction and cruciferin quantification, seed samples were defatted. For each genotype, 1 g of whole seed was weighed and manually crushed using a mortar and pestle. The crushed seed was transferred to a glass test tube, and petroleum ether (1:10 w/v, meal: petroleum ether) was added. Sealed tubes were then allowed to shake gently overnight at room temperature at med-low speed on an orbital shaker (Fisher Scientific, Dubuque, IA). The following

day, the petroleum ether was changed, and the extraction continued for an additional 5 hours, at which point the petroleum ether was again decanted and the defatted meal was placed in a fume hood overnight to allow residual petroleum ether to evaporate.

3.3.3.2 Extraction of Total Soluble Protein

To prepare samples for the ELISA assay, total soluble protein (TSP) was extracted. For each genotype, 50 mg of defatted meal was weighed into a 2 mL microcentrifuge tube, and one copper bead was added. One mL of extraction buffer (50 mM Tris, 50 mM EDTA, pH 8.0) was added to each sample. Homogenization of samples was performed using a TissueLyser (Qiagen, Toronto, ON) at room temperature for five min at 20 Hz. Samples were then centrifuged at 17,000 g for five min at room temperature. The supernatant was then transferred to a 15 mL Falcon tube, to be held on ice between steps. One mL of extraction buffer was added to the meal residue from the previous extraction, samples were homogenized and centrifuged as previously described and the supernatant was pooled with the previous extract. The protocol above was repeated twice, for a total of four sequential extractions. From each sample, 1 mL of pooled supernatant was added to a 1.5 mL microcentrifuge tube, which was then centrifuged at room temperature at 17,000 g for five min to remove any remaining meal particles. The supernatant was then decanted and held on ice for analysis by Bradford assay and ELISA.

3.3.3.3 Bradford Assay

To normalize total soluble protein for dilution prior to the ELISA, the concentration of protein in each sample was determined by Bradford assay as described by Bradford (1976). Total soluble protein samples were diluted by combining 10 µL of sample with 290 µL Tris-buffered saline

(TBS) buffer. Bovine serum albumin (BSA) (Thermo Scientific, Rockford, IL) was diluted in TBS for use as a standard. In a 96-well high-binding microplate (Corning Incorporated, Kennebunk, ME), 10 µL of each diluted standard or sample was combined with 200 µL protein assay dye reagent (Bio Rad, Hercules, CA). The plate was then incubated at room temperature for 5 min, and the absorbance measured at 595 nm using a SpectraMax 340PC microplate reader (Molecular Devices, San Jose, CA).

3.3.3.4 Enzyme-Linked Immunosorbent Assay

Based on the results of the Bradford assay as outlined above, each sample was diluted to a concentration of 0.4 ng/µL in TBS buffer. In addition, a Westar standard and SK202 standard were diluted in TBS buffer. Diluted TSP was coated onto a high-binding 96-well microplate (Corning Incorporated, Kennebunk, ME) by loading 100 µL of each diluted standard or sample into the appropriate wells, which was then incubated for 2 hours at 37°C in a 1 cm deep water bath. The TSP was then removed, and the plate washed five times with 200 µL of phosphate buffered saline with 0.05% Tween 20 (PBST). Next, 200 µL of blocking buffer (PBST with 0.01% w/v BSA (Sigma-Aldrich, St. Louis, MO) was added to each well and the plate was blocked overnight at 4°C. Following the overnight incubation, the plate was allowed to incubate for an additional one hour at 37°C in a 1 cm deep water bath. The blocking buffer was then removed and the plate washed with PBST as described above. The primary rabbit polyclonal anti-cruciferin antibody (GenScript, Piscataway, NJ) was diluted 1:8000 in blocking buffer and added at a volume of 100 µL per well. The primary rabbit polyclonal anti-cruciferin antibody was left for one hour at 37°C in a 1 cm water bath. The plate was washed with PBST as described above. The mouse anti-rabbit horseradish peroxidase conjugated monoclonal antibody (GenScript, Piscataway, NJ) was diluted

to 1:2000 in blocking buffer and added at a volume of 100 µL per well, to incubate for one hour as described above. The plate was then washed as described above. To develop the plate, 100 µL per well of Ultra TMB ELISA substrate (Thermo Scientific, Rockford, IL) was added. The plate was incubated for seven min in the dark at 24°C in a 1 cm water bath. The development was stopped with 100 µL 1 M hydrochloric acid. The absorbance was then read at 450 nm using a SpectraMax 340PC microplate reader (Molecular Devices, San Jose, CA).

3.3.4 Analysis of Seed Quality Traits

Seed oil content, seed total protein content and glucosinolate content of whole seeds were determined using near infrared spectroscopy (NIR) with a FOSS XDS Rapid Content Analyzer (FOSS NIRSystems Inc., Maryland, USA) spectrometer. NIR analysis was conducted at the University of Manitoba seed quality lab (certified by the Canadian Grain Commission). Seed was measured into 5 g samples and placed into ring cups to allow for scanning in the spectrometer. Spectral data based on sample reflectance between 400 and 2500 nm was collected, and was statistically analysed using WinISI software (Version 4.9) (Tkachuk 1981; Daun et al. 1994).

3.3.5 Statistical Analysis

The Bradford assay and ELISA for each sample were performed with three technical replicates on each microplate. Cruciferin content for each sample is presented by expressing cruciferin content as a percentage of a standard Westar sample from the same microplate. The standard error of the difference of the laboratory method (SEL) was evaluated on a set of 20 samples. Each sample was analyzed in duplicate by the cruciferin quantification method on separate days. Error associated with the method was calculated by the equation $SEL = [\Sigma(y_1 - y_2)^2/n]^{0.5}$, where y_1 and y_2 are

ELISA values for two duplicates of the same sample and *n* is the number of samples set (FOSS NIRSystems Inc., Maryland, USA, personal communication).

Statistical analysis was performed using SAS version 9.4 (SAS Institute, Cary, NC). To determine the statistical significance of the effects of genotype and site-year on cruciferin content and seed quality traits, type 3 analysis of variance (ANOVA) was carried out using PROC MIXED. All variables were set as fixed variables. Analysis parameters were chosen based on the most favorable fit statistics (AIC). Pearson correlation coefficients between cruciferin content and seed quality traits were determined using PROC CORR.

3.4 Results

3.4.1 Statistical Analysis

Assessment of the error associated with the ELISA-based method of quantifying cruciferin protein revealed a SEL value of 8.46%. Cruciferin content data was analysed on an individual site-year basis to evaluate the effect of genotype and field replicate on the phenotypic expression of cruciferin content. In all four site-years, genotype by replicate interactions were found to significantly ($p<0.0001$) affect cruciferin content (Table 3.1). Combined analysis of all four site-years revealed that genotype by site-year interactions significantly ($p<0.0001$) affect cruciferin content in *B. napus* (Table 3.2, Figure 3.1).

Table 3.1 Analysis of variance for cruciferin content in *Brassica napus* L. NAM parental genotypes in 2014[†], 2015[‡], 2016[§] and 2017[¶].

	Source	Degrees of Freedom	Sum of Squares	Mean Square	F Value	Pr > F
2014	**Genotype**	50	137921.00	2758.42	12.52	<0.0001
	Rep[††]	1	178.96	178.96	0.81	0.3718
	Genotype*Rep	50	11019.00	220.37	6.58	<0.0001
	Residual	204	6829.77	33.48		
2015	**Genotype**	51	88513.00	1735.56	10.91	<0.0001
	Rep	1	1.68	1.68	0.01	0.9185
	Genotype*Rep	51	8116.21	159.14	7.13	<0.0001
	Residual	205	4574.32	22.31		
2016	**Genotype**	51	113635.00	2228.14	4.66	<0.0001
	Rep	2	36.14	18.07	0.04	0.9627
	Genotype*Rep	101	48511.00	480.31	14.42	<0.0001
	Residual	291	9692.09	33.31		
2017	**Genotype**	51	113016.00	2216.00	3.78	<0.0001
	Rep	2	237.18	118.59	0.20	0.8153
	Genotype*Rep	102	60032.00	588.54	16.22	<0.0001
	Residual	289	10489.00	36.30		

[†]2014, Saskatoon, Saskatchewan, Canada.
[‡]2015, Saskatoon, Saskatchewan, Canada.
[§]2016, Winnipeg, Manitoba, Canada.
[¶]2017, Winnipeg, Manitoba, Canada.
[††]Field replicate.

Table 3.2 Combined analysis of variance for cruciferin content in *Brassica napus* L. NAM parental genotypes in 2014[†], 2015[‡], 2016[§] and 2017[¶].

Source	Degrees of Freedom	Sum of Squares	Mean Square	F Value	Pr > F
Genotype	51	308281.00	6044.73	49.15	<0.0001
Site-Year	3	28118.00	9372.76	76.22	<0.0001
Genotype*Site-Year	152	147796.00	972.34	7.91	<0.0001
Residual	1299	159743.00	122.97		

[†]2014, Saskatoon, Saskatchewan, Canada.
[‡]2015, Saskatoon, Saskatchewan, Canada.
[§]2016, Winnipeg, Manitoba, Canada.
[¶]2017, Winnipeg, Manitoba, Canada.

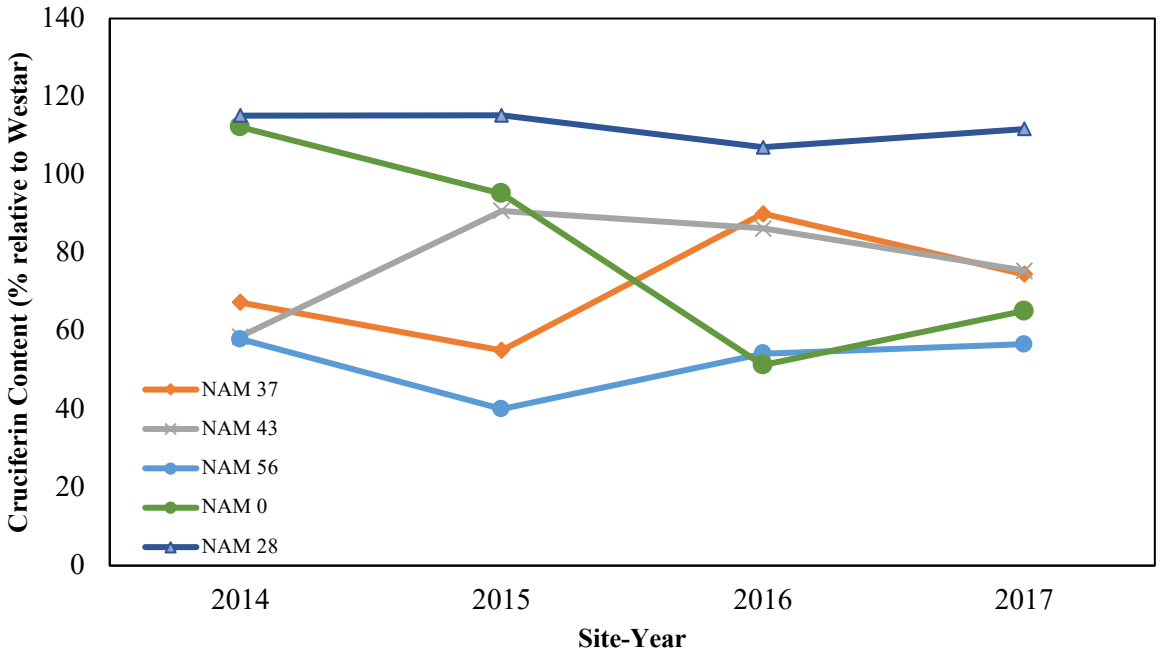


Figure 3.1 Cruciferin content in a selection of *Brassica napus* L. NAM parental genotypes in 2014 (Saskatoon, Saskatchewan), 2015 (Saskatoon, Saskatchewan), 2016 (Winnipeg, Manitoba) and 2017 (Winnipeg, Manitoba).

3.4.2 Phenotypic Variation

3.4.2.1 Seed Oil Content

Mean seed oil content was highest in 2017, where seed oil content ranged from 33.5% to 52.6% (Table 3.3, Table 3.4). The 2016 site-year had the lowest mean seed oil content, with a range from 28.3% to 44.8% (Table 3.3, Table 3.4). Analysis revealed a mean seed oil content of 41.7% across all genotypes and site-years (Table 3.4). Within the population, a number of genotypes expressed consistently high seed oil content levels across site-years. Genotype NAM 78 had 49.3% in 2015, 42.3% in 2016 and 52.4% in 2017 (data unavailable for 2014), with an overall mean of 48.2% across the site-years (Table 3.4). Genotype NAM 40 had 49.3% in 2014, 45.9% in 2015, 44.1% in 2016 and 50.4% in 2017, with an overall mean of 47.4% across all four site-years (Table 3.4).

3.4.2.2 Seed Protein Content

Analysis of total seed protein content revealed an overall mean of 27.1% across all genotypes and site-years (Table 3.5). With a mean seed protein content of 29.8% and a range of 23.5% to 35.9%, 2016 had the highest seed protein content (Table 3.3, Table 3.5). In contrast, 2017 had the lowest seed protein content, with a mean of 23.5% and a range of 17.5% to 32.2% (Table 3.3, Table 3.5). Genotypes NAM 36 and NAM 47 had consistently high seed protein contents across all four site-years. Genotype NAM 36 had a mean seed protein content of 31.1% in 2014, 33.0% in 2015, 34.7% in 2016 and 29.2% in 2017, with an overall mean of 32.0%. Similarly, genotype NAM 47 had a mean seed protein content of 30.7% in 2014, 32.6% in 2015, 31.5% in 2015 and 27.0% in 2017, with an overall mean of 30.5%.

Table 3.3 Phenotypic range and mean for oil content, protein content, glucosinolates content and cruciferin content for *Brassica napus* L. NAM parental genotypes in 2014[†], 2015[‡], 2016[§] and 2017[¶].

		2014	2015	2016	2017
Oil (%)	**Mean±SD**	42.05 ± 3.09	42.13 ± 4.15	37.35 ± 3.53	45.32 ± 4.01
	Range	34.76-50.22	29.82-50.42	28.26-44.76	33.48-52.59
Protein (%)	**Mean±SD**	27.25 ± 1.72	27.64 ± 2.95	29.83 ± 2.03	23.51 ± 2.78
	Range	22.12-31.25	21.99-34.26	23.45-35.89	17.50-32.24
Glucosinolates (µmol/g)	**Mean±SD**	40.80 ± 28.93	48.19 ± 33.73	46.59 ± 28.69	29.06 ± 22.96
	Range	5.93-88.34	10.92-107.84	10.86-102.34	3.27-86.93
Cruciferin (% of Westar)	**Mean±SD**	85.66 ± 22.18	77.01 ± 17.83	75.52 ± 19.12	74.09 ± 19.83
	Range	47.98-141.26	39.0-126.57	37.99-139.66	37.20-154.91

[†]2014, Saskatoon, Saskatchewan, Canada.
[‡]2015, Saskatoon, Saskatchewan, Canada.
[§]2016, Winnipeg, Manitoba, Canada.
[¶]2017, Winnipeg, Manitoba, Canada.

Table 3.4 Mean seed oil content (%) in *Brassica napus* L. NAM parental genotypes in 2014[†], 2015[‡], 2016[§] and 2017[¶].

Genotype	2014	2015	2016	2017	Mean
NAM 1	35.39	42.25	33.75	46.19	39.39
NAM 4	39.19	44.24	35.12	45.17	40.93
NAM 5	43.46	43.41	34.27	44.17	41.33
NAM 8	41.77	40.66	36.98	36.83	39.06
NAM 10	44.14	45.22	40.03	47.09	44.12
NAM 12	42.81	47.82	41.38	50.14	45.54
NAM 13	44.27	47.08	40.83	49.12	45.33
NAM 14	45.85	42.12	39.15	45.27	43.10
NAM 15	43.61	45.20	39.53	47.60	43.99
NAM 17	42.67	44.48	37.32	46.10	42.64
NAM 23	46.15	48.31	42.53	50.15	46.78
NAM 25	43.67	45.89	39.28	48.14	44.24
NAM 26	46.42	47.10	40.86	49.58	45.99
NAM 28	43.62	44.38	38.11	44.64	42.69
NAM 29	38.87	39.24	41.19	47.09	41.60
NAM 30	41.99	42.13	39.09	46.26	42.37
NAM 31	41.36	40.66	35.16	42.94	40.03
NAM 32	37.30	41.04	33.21	46.31	39.46
NAM 33	43.01	41.96	38.52	51.12	43.65
NAM 34	37.89	39.16	34.50	40.50	38.01
NAM 36	35.71	34.18	30.07	36.48	34.11
NAM 37	41.46	39.26	35.99	44.34	40.26
NAM 38	38.77	42.82	33.69	45.95	40.31
NAM 39	40.18	37.41	35.42	44.91	39.48
NAM 40	49.25	45.86	44.12	50.35	47.40
NAM 42	42.02	39.43	35.96	46.35	40.94
NAM 43	41.12	39.59	32.51	40.30	38.38
NAM 45	39.88	30.31	30.24	39.83	35.06
NAM 46	38.59	38.08	34.31	39.54	37.63
NAM 47	36.30	34.23	32.35	40.13	35.75
NAM 51	43.94	45.63	40.57	47.14	44.32
NAM 53	41.11	39.13	37.64	43.96	40.46
NAM 56	42.92	40.55	37.10	45.34	41.48
NAM 57	45.83	41.73	38.42	48.17	43.54
NAM 65	44.53	43.07	40.45	47.92	43.99

Table 3.4 continued Mean seed oil content (%) in *Brassica napus* L. NAM parental genotypes in 2014[†], 2015[‡], 2016[§] and 2017[¶].

Genotype	2014	2015	2016	2017	Mean
NAM 66	40.36	32.54	33.49	42.12	37.13
NAM 68	44.08	47.09	40.14	44.55	43.96
NAM 71	42.96	44.45	36.51	39.04	40.74
NAM 72	43.48	40.75	36.93	42.04	40.80
NAM 73	39.83	42.86	32.96	36.29	37.98
NAM 75	42.74	44.68	37.29	44.42	42.28
NAM 76	46.62	47.42	42.57	50.68	46.82
NAM 78	[††]	49.29	42.32	52.44	48.17
NAM 79	37.78	43.12	35.14	43.85	39.97
NAM 82	40.07	40.03	36.64	42.53	39.82
NAM 83	41.47	37.14	38.67	47.78	41.27
NAM 85	43.16	38.87	37.82	47.52	41.84
NAM 86	43.50	42.63	39.78	45.19	42.77
NAM 87	45.46	42.84	41.41	48.33	44.51
NAM 88	43.17	41.57	39.05	49.03	43.20
NAM 0	44.72	48.83	37.31	45.49	44.09
NAM 49	40.17	42.84	35.21	46.95	41.29
Mean	42.05	42.13	37.35	45.32	41.70
CV	2.60	1.97	3.69	3.54	5.05
LSD	2.20	1.67	2.24	2.61	2.95

[†]2014, Saskatoon, Saskatchewan, Canada.
[‡]2015, Saskatoon, Saskatchewan, Canada.
[§]2016, Winnipeg, Manitoba, Canada.
[¶]2017, Winnipeg, Manitoba, Canada.
[††]Genotype data unavailable.

Table 3.5 Mean seed protein content (%) in *Brassica napus* L. NAM parental genotypes in 2014[†], 2015[‡], 2016[§] and 2017[¶].

Genotype	2014	2015	2016	2017	Mean
NAM 1	28.86	25.39	31.12	22.71	27.02
NAM 4	26.83	23.30	29.07	21.80	25.25
NAM 5	25.73	25.30	32.17	22.84	26.51
NAM 8	26.32	28.49	28.56	27.80	27.80
NAM 10	24.64	26.20	28.44	22.09	25.34
NAM 12	29.50	24.28	29.40	21.04	26.06
NAM 13	24.40	22.98	27.11	19.63	23.53
NAM 14	27.44	29.73	29.82	25.00	28.00
NAM 15	25.86	25.05	28.62	21.89	25.35
NAM 17	28.00	25.75	30.73	23.07	26.89
NAM 23	25.35	24.67	28.38	20.85	24.81
NAM 25	26.10	25.55	29.53	21.77	25.74
NAM 26	22.61	22.65	26.18	18.61	22.51
NAM 28	28.04	27.79	31.48	25.94	28.31
NAM 29	29.15	30.63	28.72	22.86	27.84
NAM 30	26.13	26.60	28.23	22.85	25.95
NAM 31	26.98	27.16	29.36	21.75	26.31
NAM 32	29.41	27.65	31.17	22.68	27.73
NAM 33	24.40	25.80	27.28	17.76	23.81
NAM 34	27.23	28.60	30.30	27.86	28.50
NAM 36	31.13	32.99	34.71	29.23	32.01
NAM 37	27.04	29.83	30.52	23.45	27.71
NAM 38	26.42	24.19	29.19	19.37	24.79
NAM 39	27.84	29.85	29.71	23.50	27.73
NAM 40	26.10	29.08	28.03	24.00	26.80
NAM 42	28.14	30.11	31.53	23.61	28.35
NAM 43	26.78	29.51	33.18	27.73	29.30
NAM 45	27.99	33.55	32.58	24.97	29.77
NAM 46	30.54	30.97	30.72	25.43	29.41
NAM 47	30.71	32.59	31.48	27.03	30.45
NAM 51	26.89	25.05	29.41	22.95	26.07
NAM 53	26.85	28.55	28.89	23.59	26.97
NAM 56	26.73	29.54	29.82	24.45	27.64
NAM 57	27.95	30.36	31.27	22.92	28.13
NAM 65	29.08	30.83	30.91	24.91	28.93

Table 3.5 continued Mean seed protein content (%) in *Brassica napus* L. NAM parental genotypes in 2014[†], 2015[‡], 2016[§] and 2017[¶].

Genotype	2014	2015	2016	2017	Mean
NAM 66	27.29	33.32	30.69	25.01	29.08
NAM 68	26.24	24.53	29.65	24.05	26.12
NAM 71	26.64	27.14	31.80	29.74	28.83
NAM 72	27.34	28.39	29.22	24.22	27.30
NAM 73	26.64	25.00	31.27	27.44	27.59
NAM 75	28.76	27.97	32.53	25.55	28.71
NAM 76	25.90	25.12	28.41	21.45	25.22
NAM 78	[††]	23.14	27.96	18.61	23.28
NAM 79	29.61	27.42	31.32	24.93	28.32
NAM 82	27.17	26.74	29.28	24.49	26.92
NAM 83	29.57	31.72	28.81	23.65	28.44
NAM 85	27.02	31.14	28.07	22.73	27.24
NAM 86	25.45	25.26	24.32	22.97	24.50
NAM 87	28.08	29.95	29.07	24.08	27.80
NAM 88	26.45	28.76	28.58	21.13	26.23
NAM 0	26.44	24.09	29.97	23.94	26.11
NAM 49	28.11	26.99	31.15	22.18	27.11
Mean	27.25	27.64	29.83	23.51	27.06
CV	1.87	2.48	3.40	5.17	5.85
LSD	1.02	1.37	1.65	1.98	2.22

[†]2014, Saskatoon, Saskatchewan, Canada.
[‡]2015, Saskatoon, Saskatchewan, Canada.
[§]2016, Winnipeg, Manitoba, Canada.
[¶]2017, Winnipeg, Manitoba, Canada.
[††]Genotype data unavailable.

3.4.2.3 Glucosinolate Content

Glucosinolate content varied considerably across genotypes. Across all four site-years, the mean glucosinolate content was 41.2 µmol/g (Table 3.6). However, 2015 had the highest glucosinolate content with a mean of 48.2 µmol/g and a range of 10.9 µmol/g to 107.8 µmol/g, while 2017 had the lowest mean glucosinolate content of 29.1 µmol/g with a range of 3.3 µmol/g to 86.9 µmol/g (Table 3.3). Despite the range in glucosinolate content, several genotypes expressed consistently low values across site-years. Genotype NAM 30 had an overall mean glucosinolate content of 9.1 µmol/g, with 7.1 µmol/g in 2014, 14.0 µmol/g in 2015, 11.7 µmol/g in 2016 and 3.7 µmol/g in 2017 (Table 3.6). Similarly, NAM 23 had a mean glucosinolate content of 9.7 µmol/g, with 8.9 µmol/g in 2014, 12.6 µmol/g in 2016, 12.1 µmol/g in 2016 and 5.2 µmol/g in 2017 (Table 3.6).

3.4.2.4 Cruciferin Content

Cruciferin content varied greatly across genotypes and site-years, both of which significantly affected cruciferin content (Table 3.2). The 2014 site-year had the highest mean cruciferin content of 85.7% relative to a Westar standard and a range of 47.9 to 141.3% (Table 3.3), while 2017 had the lowest mean cruciferin content of 74.1% relative to the standard, and a range of 37.2 to 154.9%. While genotype by site-year interactions did significantly affect cruciferin content (Table 3.2), several genotypes had consistently low or high values. NAM 56 was consistently low, with an overall mean of 52.2%, and individual site-year means of 58.0% (2014), 40.1% (2015), 54.2% (2016) and 56.7% (2017) (Table 3.7). Similarly, NAM 86 had a mean cruciferin content of 59.9% in 2014, 43.2% in 2015, 51.4% in 2016 and 56.2% in 2017, (Table 3.7). In contrast, NAM 51 and NAM 28 were consistently high, with overall means of 105.3% and 112.3% relative to the standard, respectively (Table 3.7).

Table 3.6 Mean glucosinolate content (µmol/g) in *Brassica napus* L. NAM parental genotypes in 2014†, 2015‡, 2016§ and 2017¶.

Genotype	2014	2015	2016	2017	Mean
NAM 1	65.87	76.00	54.37	15.10	52.83
NAM 4	57.44	60.73	66.18	39.08	55.86
NAM 5	49.35	40.84	46.55	33.43	42.55
NAM 8	19.38	26.01	27.79	19.91	23.27
NAM 10	11.24	15.02	18.69	11.51	14.11
NAM 12	16.76	17.64	22.51	11.26	17.04
NAM 13	7.96	12.79	15.08	6.98	10.70
NAM 14	72.26	87.06	89.06	50.16	74.64
NAM 15	11.19	14.06	16.22	5.86	11.83
NAM 17	14.55	13.38	20.60	9.24	14.44
NAM 23	8.92	12.64	12.12	5.17	9.71
NAM 25	11.65	16.03	18.57	8.33	13.65
NAM 26	11.29	15.35	21.41	10.37	14.61
NAM 28	8.70	12.47	14.94	6.81	10.73
NAM 29	85.18	94.24	41.43	13.46	58.58
NAM 30	7.05	13.99	11.70	3.71	9.11
NAM 31	17.58	31.30	27.24	11.99	22.03
NAM 32	76.86	91.35	84.87	55.00	77.02
NAM 33	63.47	86.48	78.89	50.09	69.73
NAM 34	78.53	91.38	96.18	81.68	86.94
NAM 36	66.48	69.13	84.21	50.39	67.55
NAM 37	35.68	55.99	56.20	18.66	41.63
NAM 38	53.94	61.12	61.64	36.97	53.42
NAM 39	81.60	105.27	89.99	75.69	88.14
NAM 40	68.75	74.66	74.86	46.33	66.15
NAM 42	59.21	68.29	52.38	46.63	56.63
NAM 43	59.35	53.57	44.41	26.14	45.87
NAM 45	71.50	106.77	96.00	61.47	83.93
NAM 46	75.27	88.68	66.65	66.68	74.32
NAM 47	71.87	91.10	82.36	58.59	75.98
NAM 51	10.62	14.88	22.59	11.12	14.81
NAM 53	76.14	85.83	80.86	56.36	74.80
NAM 56	64.25	71.32	72.50	42.09	62.54
NAM 57	13.16	19.70	21.58	9.35	15.95
NAM 65	17.92	21.12	18.02	9.66	16.68

Table 3.6 continued Mean glucosinolate content (µmol/g) in *Brassica napus* L. NAM parental genotypes in 2014[†], 2015[‡], 2016[§] and 2017[¶].

Genotype	2014	2015	2016	2017	Mean
NAM 66	76.20	106.20	70.96	52.83	76.55
NAM 68	15.33	17.60	19.90	8.44	15.32
NAM 71	14.81	20.63	24.80	14.08	18.58
NAM 72	14.28	12.85	24.43	8.10	14.92
NAM 73	58.79	77.02	77.96	31.40	61.29
NAM 75	10.95	16.71	17.62	10.41	13.92
NAM 76	12.52	12.75	13.91	5.58	11.19
NAM 78	[††]	15.94	22.67	9.96	15.94
NAM 79	85.82	88.53	90.48	62.41	81.81
NAM 82	8.27	14.40	14.41	6.70	10.95
NAM 83	84.76	97.45	76.05	59.72	79.50
NAM 85	56.15	65.57	65.10	41.93	57.19
NAM 86	63.98	67.66	64.75	61.77	64.54
NAM 87	14.79	23.75	23.93	9.93	18.10
NAM 88	15.71	21.89	19.88	10.99	17.12
NAM 0	10.29	11.02	67.12	34.47	30.73
NAM 49	17.20	19.62	26.15	14.64	19.40
Mean	40.80	48.19	46.59	29.06	41.18
CV	5.36	4.55	10.96	15.97	23.12
LSD	4.39	4.41	8.31	7.55	13.28

[†]2014, Saskatoon, Saskatchewan, Canada.
[‡]2015, Saskatoon, Saskatchewan, Canada.
[§]2016, Winnipeg, Manitoba, Canada.
[¶]2017, Winnipeg, Manitoba, Canada.
[††]Genotype data unavailable.

Table 3.7 Mean cruciferin content (% relative to standard) in *Brassica napus* L. NAM parental genotypes in 2014†, 2015‡, 2016§ and 2017¶.

Genotype	2014	2015	2016	2017	Mean
NAM 1	58.11	67.53	66.05	66.43	64.53
NAM 4	70.00	64.23	44.74	71.99	62.74
NAM 5	71.59	71.83	52.25	46.21	60.47
NAM 8	93.46	90.64	76.99	53.77	78.72
NAM 10	82.25	67.52	61.18	68.70	69.91
NAM 12	124.87	79.37	76.41	71.77	88.11
NAM 13	92.25	73.39	80.85	75.90	80.60
NAM 14	70.37	64.34	61.55	48.01	61.07
NAM 15	100.28	93.41	96.74	83.51	93.49
NAM 17	113.26	91.68	84.50	100.25	97.42
NAM 23	89.81	90.10	77.04	55.73	78.17
NAM 25	109.61	106.66	91.51	84.35	98.03
NAM 26	77.32	97.90	72.29	69.48	79.25
NAM 28	115.22	115.24	107.07	111.72	112.31
NAM 29	78.72	89.16	77.32	78.25	80.86
NAM 30	88.85	74.16	74.71	85.82	80.89
NAM 31	112.87	104.82	96.13	96.11	102.48
NAM 32	63.48	54.63	76.69	61.18	64.00
NAM 33	60.34	55.82	54.27	59.17	57.40
NAM 34	72.72	87.41	61.73	77.23	74.77
NAM 36	61.00	67.20	59.62	47.52	58.84
NAM 37	67.33	55.07	90.08	74.53	71.75
NAM 38	63.82	70.99	80.07	62.87	69.44
NAM 39	54.08	54.12	67.95	63.07	59.81
NAM 40	79.01	85.09	89.22	83.14	84.12
NAM 42	72.34	75.18	110.89	65.00	80.85
NAM 43	58.63	90.75	86.29	75.46	77.78
NAM 45	55.69	57.98	59.93	50.48	56.02
NAM 46	69.25	88.15	77.77	53.05	72.06
NAM 47	72.98	78.65	79.43	67.66	74.68
NAM 51	95.35	117.01	103.90	104.87	105.28
NAM 53	55.95	54.08	69.02	62.95	60.50
NAM 56	57.99	40.08	54.22	56.65	52.24
NAM 57	116.76	70.97	98.06	72.39	89.55
NAM 65	99.47	81.47	104.64	94.27	94.96

Table 3.7 continued Mean cruciferin content (% relative to standard) in *Brassica napus* L. NAM parental genotypes in 2014[†], 2015[‡], 2016[§] and 2017[¶].

Genotype	2014	2015	2016	2017	Mean
NAM 66	72.31	61.52	53.19	65.92	63.24
NAM 68	108.61	74.32	78.46	70.71	83.03
NAM 71	95.88	81.41	68.16	94.79	85.06
NAM 72	100.28	73.11	76.64	77.96	82.00
NAM 73	86.61	64.04	54.69	70.78	69.03
NAM 75	127.72	78.35	85.93	85.67	94.42
NAM 76	91.02	74.73	78.46	83.68	81.97
NAM 78	[††]	79.25	85.30	88.68	86.98
NAM 79	68.63	56.49	55.08	56.50	59.18
NAM 82	113.31	84.46	82.24	83.49	90.88
NAM 83	84.38	67.10	81.15	76.71	77.34
NAM 85	111.14	77.92	84.66	89.80	90.88
NAM 86	59.93	43.16	51.38	56.21	52.67
NAM 87	126.28	89.81	88.49	107.00	102.90
NAM 88	97.58	72.86	72.37	92.36	83.79
NAM 0	112.21	95.17	51.33	65.18	80.97
NAM 49	87.49	104.18	61.84	87.74	85.31
Mean	85.66	77.01	75.52	74.09	78.05
CV	10.01	9.48	17.05	19.44	14.97
LSD	17.22	14.66	20.96	23.34	16.37

[†]2014, Saskatoon, Saskatchewan, Canada.
[‡]2015, Saskatoon, Saskatchewan, Canada.
[§]2016, Winnipeg, Manitoba, Canada.
[¶]2017, Winnipeg, Manitoba, Canada.
[††]Genotype data unavailable.

3.4.3 Correlation Between Cruciferin Content and Seed Quality Traits

A number of seed quality traits were found to be significantly correlated with one another, and with cruciferin content. As expected, seed oil content and seed protein content were negatively correlated (r=-0.86) (Table 3.8). Seed oil content was also negatively correlated with glucosinolate content (r=-0.52), while seed protein content was positively correlated with glucosinolate content (r=0.42) (Table 3.8). Interestingly, cruciferin content was found to be negatively correlated with glucosinolate content (r=-0.49), but not correlated with total seed protein content (Table 3.8).

Table 3.8 Pearson correlation coefficients for seed protein content, seed oil content, cruciferin content, and glucosinolate content in *Brassica napus* L. NAM parental genotypes in 2014[†], 2015[‡], 2016[§] and 2017[¶].

	Protein	Oil	Cruciferin
Oil	-0.86**	1.00	-
Cruciferin	-0.02	0.22**	1.00
Glucosinolates	0.42**	-0.52**	-0.49**

* Significant at $p < 0.05$.
** Significant at $p < 0.01$.
[†]2014, Saskatoon, Saskatchewan, Canada.
[‡]2015, Saskatoon, Saskatchewan, Canada.
[§]2016, Winnipeg, Manitoba, Canada.
[¶]2017, Winnipeg, Manitoba, Canada.

3.5 Discussion

As interest in developing *B. napus* cultivars with specialty protein compositions grows, successful efforts must not improve protein and protein related traits at the expense of traditionally important traits such as high oil content and low glucosinolates (Campbell et al. 2016). In commercially available cultivars tested by the Canadian Grain Commission's (CGC) harvest sample program of western Canadian canola, mean oil content of samples graded "Canola, No. 1 Canada" was 44.5% between 2013 and 2017 (Barthet 2018). While the mean seed oil content of NAM parental genotypes exceeded this value in the 2017 site-year (45.3%), the 2014, 2015 and 2016 site-years all fell below the current market mean (Table 3.3), and in only eight genotypes did the mean across all site-years meet or exceed 44.5% (Table 3.4). Based on samples collected between 2013 and 2017, the CGC found a mean total protein content of 20.2% in "Canola, No. 1 Canada" graded samples (Barthet 2018). Mean seed protein content in NAM parental genotypes well exceeded this value in all four site years (27.3% in 2014, 27.6% in 2015, 29.8% in 2016 and 23.5% in 2017). In fact, the mean seed protein content across all site-years exceeded 20.2% for every genotype examined in this study, with two genotypes (NAM 47 and NAM 36) exceeding the market average by greater than 10% (Table 3.5). As a result of extensive breeding efforts to maintain low glucosinolate contents in commercial canola varieties, glucosinolate content has been consistently low, with an average of 10 µmol/g between 2013 and 2017 among the CGC harvest samples (Barthet 2018). While glucosinolate content among the NAM parental genotypes varied significantly, the mean across all site-years was much higher than this market average, at 41.18 µmol/g (Table 3.6). Notably, when the mean glucosinolate content across all site-years is examined, only two genotypes (NAM 23 and NAM 30) were below 10 µmol/g. Overall, while the NAM parental genotypes are superior in terms of their total protein content, and a number of

genotypes are competitive in their of oil content, many of these genotypes have significantly higher glucosinolate content than would be considered acceptable in commercial canola cultivars.

With the future objective of breeding for specialty canola cultivars in mind, it important to consider that efforts to increase total protein content or alter seed storage protein profiles must not do so at the expense of other important seed quality traits. This study found a negative correlation (r= -0.86) between seed protein content and seed oil content (Table 3.8). This relationship has been extensively documented throughout the history of *B. napus* breeding efforts (Bhatty 1964; Grami et al. 1977; Daun and DeClercq 1995; Malabat et al. 2003; Zhao et al. 2006; Schatzki et al. 2014; Gu et al. 2017; Chao et al. 2017). While such a strong negative correlation between oil content and total seed protein content may prove challenging to efforts to breed for increased protein content without decreasing oil content, it has been suggested that *B. napus* genotypes with a yellow seed coat may be associated with increased oil and protein content when compared with black-seeded genotypes, due to a reduction in the meal fiber content (Liezhao et al. 2007). It may therefore be possible to select for increased levels of total seed protein while maintaining acceptable levels of seed oil, despite this negative correlation.

Interestingly, this research found no correlation between cruciferin content and total protein content (Table 3.8). These results are consistent with those found by Schatzki et al. (2014) in winter oilseed rape. This may be due to a compensatory relationship between differing types of seed storage proteins. The relationship between seed storage proteins has been demonstrated in *B. napus* in which antisense suppression of cruciferin expression reduced cruciferin levels, while levels of total protein remained unchanged (Kohno-Murase et al. 1995), and antisense suppression of napin expression resulted in a compensatory effect by cruciferin protein (Kohno-Murase et al. 1994).

Similarly, Withana-Gamage et al. (2013) showed that seed N content in a genotype in which no cruciferin genes were expressed remained comparable to that of the wild-type. In *Camelina sativa* L., the use of CRISPR/Cas9 gene editing demonstrated that that the knockout of genes encoding cruciferin protein did not result in an altered total seed protein content, due to a compensatory effect by other seed storage proteins (Lyzenga et al. 2019). This suggests that efforts to develop specialty cultivars with altered seed storage protein profiles may not have an effect on the total protein content of the seed. While there was no correlation between cruciferin content and total protein content, analysis did reveal a negative correlation (r= -0.49) between cruciferin content and glucosinolate content (Table 3.8). This is consistent with results from Schatzki et al. (2014) who found a similar negative correlation between cruciferin content and glucosinolate content in a doubled haploid population of winter oilseed rape, as well as the results reported by Malabat et al. (2003) that found that double low genotypes (those low in erucic acid and glucosinolates) were richer in cruciferin than genotypes rich in glucosinolates. The negative correlation between cruciferin content and glucosinolate content indicates that efforts to increase cruciferin content in *B.napus* would be possible, without sacrificing past improvements in the reduction of glucosinolate content.

To rapidly quantify cruciferin protein for this research, an ELISA-based method was developed and was found to have a SEL value of 8.46%. While sources of error within the ELISA-based method such as sensitivities to temperature and time (Crowther 2009; Gan and Patel 2013) may result in less precision and accuracy in the measurement of cruciferin protein content, this method is capable of utilizing relatively small seed amounts and assessing large numbers of samples, and thus fits the needs of this research.

While the population used for this research had not previously been examined for protein or protein related traits, when one considers that the NAM parental genotypes were selected to express variability in a variety of agronomically important traits (Parkin et al. 2017), it is not unexpected that these genotypes would also express considerable variation in seed quality traits, including cruciferin content (Table 3.7). Considerable variation in cruciferin content and the ratio between cruciferin and napin proteins has been previously demonstrated in other canola and winter oilseed rape populations (Malabat et al. 2003; Wanasundara 2011; Schatzki et al. 2014). Future breeding efforts may be applied to exploit this naturally existing variation.

When the overall mean cruciferin content from each of the four site-years is considered, one can see that that highest overall mean was observed in the Saskatoon location in 2014, while the lowest overall mean was observed in Winnipeg in 2017 (Table 3.7). In addition, both Saskatoon locations expressed higher overall mean cruciferin content values than each of the Winnipeg locations (Table 3.7). Based on data collected from weather stations near each of the field sites, annual precipitation at the Saskatoon locations in the years observed was considerably lower than that observed at the Winnipeg locations (Government of Canada 2019; University of Manitoba 2019). In addition, the highest daily maximum temperature in Saskatoon in 2014 was 31.5ºC (Government of Canada 2019), while in Winnipeg in 2017 the highest daily maximum temperature was considerably higher at 34.5ºC (University of Manitoba 2019). While this study observed a relatively small number of environments, and thus it is difficult to draw conclusions regarding the specific effects of climate and weather patterns on cruciferin content, associations between temperature and precipitation and other seed quality traits including oil and protein content have been widely observed (Si et al. 2003).

The analysis of the entire data set from all four site-years revealed that genotype by site-year interactions significantly affect cruciferin content (Table 3.2). In their analysis of a winter oilseed rape population, Schatzki et al. (2014) did not observe a significant effect of the interactions between genotype and year. This contrast may derive from differences between the populations observed, as well as the fact that while the above study had two locations in northwestern Germany, only two consecutive field seasons were carried out (Schatzki et al. 2014) compared to the four field seasons in this study. The significant effects of genotype by site-year interaction on cruciferin content in the NAM parental genotypes indicates that environment has a large impact on the expression of this trait. Many other seed quality traits in *B. napus* are also highly influenced by the environment, including oil content (Si et al. 2003; Javed et al. 2016; Fu et al. 2017; Guo et al. 2017; Chao et al. 2017), protein content (Si et al. 2003; Wu et al. 2005; Chao et al. 2017), fatty acid profile (Javed et al. 2016) and erucic acid content (Harvey and Downey 1964; Shi et al. 2003). Environmental and genotype by environmental effects can be a great challenge in breeding programs in the selection of genotypes that perform well across multiple environments (Annicchiarico 2002). Despite the significant effect of genotype by environmental interactions, a number of genotypes expressed consistently low (NAM 56 and NAM 86) or high cruciferin (NAM 51 and NAM 28) content across site-years (Table 3.7). Thus, the results of this study demonstrate that it is possible to identify genotypes stable in their cruciferin content across site-years. Genotypes that perform well for a given trait across environments are of particular value to breeding programs (Bernardo 2014), and therefore these genotypes may be of interest.

In summary, the objective of this research was to utilize a diverse population of *B. napus* to determine the extent of existing variation in cruciferin content and determine the effect of genotype and environmental factors on cruciferin content in *B. napus*. An ELISA-based method was

developed to allow for the assessment of cruciferin content using small seed samples and with relatively high efficiency. This research identified several genotypes expressing consistently high and low values for cruciferin content and added to the current understanding of effect of genotype by environment interactions on cruciferin content, as well as relationship between cruciferin content and other seed quality traits. The resulting data sheds further light onto a field of great interest and provides valuable information to future breeding efforts to enhance protein composition in canola.

4. Analysis of Cruciferin Content in Whole Seeds of *Brassica napus* L. by Near-Infrared Spectroscopy

4.1 Abstract

Globally, there is an increasing demand for sources of plant-based protein. While *Brassica napus* L. is an important oilseed crop worldwide, there is also interest in improving its ability to serve as a valuable source of plant-based proteins. Cruciferin, a seed storage protein that makes up 60% of the protein found in mature seeds of *B. napus*, is of interest as a source of protein for human consumption due to its functional properties. Existing methods for quantification of cruciferin protein are often time consuming and destructive of seed samples. This study explored the potential for the measurement of cruciferin protein content in whole seeds of *B. napus* by near-infrared spectroscopy (NIRS), to allow for efficient and non-destructive screening of breeding material. An enzyme-linked immunosorbent assay (ELISA)-based reference method was utilized to assess cruciferin content in a diverse population of *B. napus*. Scanning of whole seed samples produced spectra that were used to develop NIRS calibration equations. The statistical analysis of these calibration results indicates that these NIRS equations developed are poorly suited for prediction of cruciferin content.

4.2 Introduction

The combined effects of a growing worldwide population, rising incomes, and greater education on the importance of nutritious diets has resulted in an increased demand for protein (Henchion et al. 2017; Wanasundara and Hojilla-Evangelista 2018). *Brassica napus* L. is the second largest oilseed produced worldwide (USDA 2019). In Canada alone, 19 million metric tonnes of canola

and rapeseed are produced, making Canada the world's largest producer (Statistics Canada 2019a) and exporter (USDA 2019) of canola and rapeseed. While the primary use of *B. napus* is as an oilseed, its secondary product, a high-protein meal, has excellent potential to contribute to solving the challenge of the rising need for plant-based protein sources.

The proteins within *B. napus* meal are found primarily in the form of two seed storage proteins: cruciferin and napin (Wanasundara 2011; Campbell et al. 2016). Making up approximately 20% of the total protein content of mature seeds, napin is a 2S albumin (Aider and Barbana 2011; Wanasundara 2011). Of primary interest in this research, however, is the 11S globulin cruciferin, which makes up the majority of the total protein content of mature seeds (60%) (Aider and Barbana 2011; Wanasundara 2011). As the desire to add value and diversify the possible end-uses of *B. napus* products increases, it is therefore no surprise that breeding-related efforts have been increasingly turning towards specific seed storage proteins, including research on genetic variability (Malabat et al. 2003) and QTL mapping studies (Schatzki et al. 2014)

A number of methods have previously been used to quantify cruciferin content in *B. napus* meal. These methods typically employ a two-step process of either extracting total protein or selectively extracting cruciferin protein and then quantifying the protein. One such method of cruciferin quantification is the extraction of total soluble protein followed by separation and quantification by SDS-PAGE (Kohno-Murase et al. 1995; Schatzki et al. 2014). Other approaches require selective extraction of cruciferin protein by sequentially combining *B. napus* meal with acidic and then alkaline buffers (Wanasundara and McIntosh 2013; Akbari and Wu 2015). Cheung et al. (2014) employed a method in which a cruciferin rich protein isolate was prepared by dispersing ground meal in 0.2 M NaCl (pH 5.8-6.3) and collecting the resulting protein micelles, which were

then freeze-dried and the protein quantified by the Kjeldahal method of nitrogen content determination. Another method removed the phenolics using size exclusion chromatography, then cruciferin proteins were separated and further purified by cation exchange chromatography (Perera et al. 2016). This allowed for quantification by the combustion method of nitrogen-based protein content determination. In contrast to many of these other methods, enzyme-linked immunosorbent assays (ELISAs) are capable of detecting and quantifying proteins at low levels, with relative efficiency (Crowther 2009). Chapter 3 of this thesis discusses the analysis of cruciferin content in a diverse set of genotypes by an ELISA-based method. While each of these methods is effective, they do present a number of disadvantages. Primarily, each method results in the destruction of the test seed sample, and in some cases requires large seed amounts. These methods are also time consuming to varying degrees, resulting in significant limitations regarding the number of samples that may be feasibly analyzed.

Plant breeding programs often require evaluation of one or many traits of interest in many genotypes over multiple environments (Furbank and Tester 2011). Thus, phenotyping approaches that improve selection efficiency are essential (Chawade et al. 2019). For this reason, many *Brassica* breeding programs have adopted near-infrared reflectance spectroscopy (NIRS) as a method by which to analyze seed quality traits (Font et al. 2006), to allow for efficient phenotyping in a non-destructive manner.

By detecting molecular bonds within a biological sample, NIRS is able to yield a great deal of information regarding its composition (Shenk et al. 2008). Near-infrared (NIR) radiation produced by a spectrophotometer may be reflected, absorbed or transmitted by a sample of interest (Shenk et al. 2008). When NIR radiation vibrates at the same frequency as molecular bonds within the

sample, the energy is absorbed, and may be measured to produce a spectrum that is visualized (Shenk et al. 2008). Specifically, NIR radiation is absorbed by the oxygen-hydrogen (O-H), carbon-hydrogen (C-H), nitrogen-hydrogen (N-H), sulfur-hydrogen (S-H) and carbon-oxygen (C=O) bonds in biological samples, and therefore the differing spectra produced yield information regarding the composition of the sample of interest (Chodak 2008; Agelet and Hurburgh 2010). While the spectra contain valuable information regarding these functional groups within the sample, they must be correlated to a reference method to allow for measurement of a trait of interest (Shenk et al. 2008). A calibration model or equation facilitates the correlation of the spectral data with a chemical or physical property as measured by the reference method. Following the development of a calibration equation, future samples may be simply scanned by a spectrophotometer to analyze the trait of interest (Agelet and Hurburgh 2010).

The use of NIRS has been widely adopted across *B. napus* breeding programs for the determination of a variety of seed quality traits including chlorophyll content (Daun 1976; Tkachuk and Kuzina 1982), glucosinolate content (Daun et al. 1994; Velasco and Becker 1998a), oil content (Tkachuk 1981; Daun et al. 1994; Velasco et al. 1999b, 1999a), fatty acid composition (Sato et al. 1998; Velasco and Becker 1998b; Velasco et al. 1999b, 1999a; Kim et al. 2007), fiber content (Font et al. 2003, 2005; Dimov et al. 2012; Wittkop et al. 2012) and total protein content (Tkachuk 1981; Daun et al. 1994; Velasco and Möllers 2002). Since NIRS is already so widely used in *Brassica* breeding, it would be opportune to also employ it for the rapid determination of cruciferin content, resulting in a reduction in the need for potentially hazardous chemicals, improving efficiency, and allowing for analysis to be carried out in a non-destructive manner (Font et al. 2006).

By utilizing the diverse genotypes of a nested association mapping population and an enzyme-linked immunosorbent assay (ELISA)-based reference method, the objective of this research was to investigate the potential for improved phenotyping efficiency by developing a NIRS calibration equation for the determination of cruciferin content in *B. napus*. The ability to use NIRS to quantify cruciferin rather than the current, less efficient methods, would aid future breeding efforts to develop *B. napus* cultivars with specialty protein profiles.

4.3 Materials and Methods

4.3.1 Plant Material, Field Evaluation and Reference Analysis of Cruciferin Content

Details regarding the *B. napus* genotypes (the parental genotypes of a nested association mapping (NAM) population) used in this study may be found in Chapter 3 (Section 3.3.1 Plant Material). Field evaluation conditions and agronomic management are described in Chapter 3 (Section 3.3.2 Field Evaluation). Briefly, plots were grown in four site-years: Saskatoon in 2014, Saskatoon in 2015, Winnipeg in 2016 and Winnipeg in 2017. Field evaluation in 2014 and 2015 (Saskatoon) was carried out in a randomized complete block design, with two replicates per genotype, while in 2016 and 2017 (Winnipeg), evaluation was carried out in a randomized complete block design with three replicates per genotype. Reference values for cruciferin content used for the development of near-infrared spectroscopy (NIRS) calibrations were determined through an enzyme-linked immunosorbent assay (ELISA)-based approach, as detailed in Chapter 3 (Section 3.3.3 Analysis of Cruciferin Protein by Enzyme-Linked Immunosorbent Assay).

4.3.4 Statistical Analysis of Cruciferin Content Reference Data

The Bradford assay and ELISA for each sample were performed with three technical replicates on each microplate. Cruciferin content for each sample is presented by expressing cruciferin content as a percentage of a standard Westar sample from the same microplate.

The standard error of the difference of the laboratory method (SEL) was evaluated on a set of 20 samples. Each sample was analyzed in duplicate by the reference method, on separate days. Error associated with the reference method was calculated by the equation $SEL = [\Sigma(y_1 - y_2)^2/n]^{0.5}$, where y_1 and y_2 are laboratory reference (ELISA) values for two replicates of the same sample and n is the number of samples set (FOSS NIRSystems Inc., Maryland, USA, personal communication).

4.3.5 Development of Near Infrared Spectroscopy Calibration

4.3.5.1 Acquiring Sample Spectra and Population Structuring

Seed samples from each of the NAM parental genotypes at each location were cleaned, measured into 5 g samples, and placed into ring cups. Samples were then scanned with a Foss XDS Rapid Content Analyzer (FOSS NIRSystems Inc., Maryland, USA). Spectra were acquired at 0.5 nm intervals over the entire wavelength range (400 – 2500 nm), to include both visible and NIR regions of the spectrum.

To remove spectral outliers, population structuring was carried out. Outliers were identified with principal component analyses (PCA) through the WinISI software (version 4.9.0, FOSS Analytical

A/S, Hillerod, Denmark, and Infrasoft International LLC, State College, PA). Spectral outliers were identified based on the Global H value, with a cut-off H-value of 3.0 (Larkin 2012).

4.3.5.2 Calibration Development

The cruciferin content values based on the ELISA reference method were imported using the WinISI 4 software, producing the calibration file. The modified partial least squares method was selected to correlate the cruciferin reference values with the NIR spectra utilizing the full wavelength spectrum (400 – 2500 nm). Commonly used for seed quality analysis, the modified partial least squares regression technique utilizes the spectra and the reference data, maximizing the variance in the spectral data, as well as maximizing correlation with the reference data (Font et al. 2006). Cross-validation was used to identify outliers and to confirm the calibration equation (Font et al. 2005).

For the removal of outliers, a critical "T" value of 2.5 was used, eliminating samples in which there is a large difference between the reference value and the NIR prediction. A critical "X" outlier value of 10 was used, through which the software can identify unusual spectra. A critical global H (GH) outlier value of 10 was selected. Two outlier elimination passes were used.

For the calibrations developed in this study, the scatter correction setting "SNV (standard normal variate) and detrend" was utilized. Scatter corrections are used to minimize spectral distortion, due to the potential effect of differences in particle size among samples (Font et al. 2006; Agelet and Hurburgh 2010) and to reduce the effect of the signal-to-noise ratio (Rinnan et al. 2009).

Three calibration equations were developed. The first two used a total of 429 samples. Calibration A used a math treatment of 1, 16, 16, 1 (derivative, gap, first smooth, second smooth). A second

calibration equation (Calibration B) used a math treatment of 2, 16, 16, 1 (derivative, gap, first smooth, second smooth). Each of these equations produced poor calibration and cross-validation statistics, making them unsuitable for further use. To determine if more favorable statistics may be produced, an additional 131 samples were manually removed, based on their residuals that exceeded a value of 15.

With the remaining sample set, a third calibration was developed. Calibration C used a total of 298 samples, and a math treatment of 2, 16, 16, 1 (derivative, gap, first smooth, second smooth) was used.

The performance of the equations was assessed by examining the statistics, as calculated by the WinISI software: a low standard error of calibration (SEC), and coefficient of determination (R^2) close to one were used as measures of quality of the equations. Internal cross-validation also allowed for assessment of prediction ability through the use of the following statistics: a low standard error of cross validation (SECV), and a value close to one for one minus the variance ratio (1-VR), which is an estimation of the coefficient of determination for the cross-validation.

4.3.5.3 External Equation Validation

External validation of Calibrations A, B and C was carried out to assess their performance. Prior to calibration development, a subset of 74 samples (every 7^{th} sample of the entire data set) were selected and removed to be used as an external validation set. These samples were analyzed using the ELISA method previously described but were not utilized for the development of the calibration equations.

The "Monitor" program within the WinISI software was used for statistical comparison of the laboratory reference values and predicted values for the external validation set samples, based on the calibration equations developed. External validation produced a number of statistics through which the prediction ability of the calibration equations was assessed. A coefficient of determination (R^2) and slope close to one were considered optimal. In addition, a low standard error of prediction (SEP) was superior. Equations with high ratios of the standard deviation (SD) of the reference data to the SEP (RPD) and ratios of the range in the reference data to the SEP (RER) were also preferred.

4.4 Results

4.4.1 Variation in Cruciferin Content as Determined by the Reference Method

Assessment of the error associated with the reference method revealed a SEL value of 8.46%. Samples selected for use in the development of the NIRS calibration equations displayed large variation in cruciferin contents. For Calibration A, the mean cruciferin content was 76.17%, with a range of 37.96 to 128.18%, as measured by the ELISA-based reference method. In Calibration B, cruciferin content in the reference samples ranged from 37.96 to 128.18%, with a mean cruciferin content of 76.35%, while Calibration C had a slightly lower mean (75.31%) and narrower range (43.20 to 123.26%) (Table 4.1). A detailed assessment of phenotypic variation in cruciferin content observed among a diverse set of genotypes of *Brassica napus*, its correlation with other seed quality traits, as well as the effects of genotype and environment on cruciferin content may be found in Chapter 3 of this thesis.

Table 4.1 Calibration and cross-validation statistics of three NIRS calibrations for cruciferin content (% relative to Westar) in *Brassica napus* L.

				Calibration			Cross-Validation	
Calibration[†]	**N**[‡]	**Mean**	**Range**	**SD**[§]	**SEC**[¶]	**R^2**[††]	**SECV**[‡‡]	**1-VR**[§§]
A	417	76.17	37.96-128.18	18.51	14.12	0.42	14.78	0.36
B	416	76.35	37.96-128.18	18.69	13.64	0.47	14.31	0.41
C	298	75.31	43.20-123.26	15.09	7.91	0.73	8.48	0.68

[†]Calibration A math treatment: 1,16,16,1 (derivative, gap, first smooth, second smooth), Calibration B math treatment 2,16,16,1, Calibration C math treatment: 2,16,16,1.
[‡]Number of samples.
[§]Standard deviation of calibration set samples.
[¶]Standard error of the calibration.
[††]Coefficient of determination of the calibration.
[‡‡]Standard error of the cross-validation.
[§§]One minus the variance ratio (1-VR), an estimation of the coefficient of determination for the cross-validation.

4.4.2 Spectral Analysis

Spectra were collected between 400-2500 nm (Figure 4.1). Absorption bands were viewed in the raw (log 1/R) spectra. Functional groups correlated with peak wavelengths were identified through WinISI. The peaks were identified at 1200 nm (C-H bend second overtone), 1720 nm (C-H stretch first overtone), 2300 nm (CH bend second overtone), 1930 nm (O-H bend second overtone), 1500 nm (N-H stretch first overtone) and 2060 nm (N-H bend second overtone).

4.4.3 NIRS Calibration Equations for Cruciferin Content and Cross Validation

Calibrations A and B resulted in poor calibration and cross-validation statistics (Table 4.1, Figure 4.2). After removal of samples with high residuals, the additional Calibration C was produced.

The SEC, a description of how well the calibration set samples were fit in the equation, was 14.12, 13.64 and 7.91 for Calibrations A, B, and C, respectively. Calibration C had the highest coefficient of determination (R^2), a measure of the correlation between NIRS predicted values and the reference values in the calibration set (Table 4.1). The relationship between the NIRS values and the reference values for samples in the calibration data is found in Figure 4.2.

During equation development in WinISI software, internal cross-validation of the calibrations was carried out. In this process, groups of samples are removed from the calibration set, and are predicted based on the samples remaining, providing an estimate of the performance of the calibration equation (Shenk and Westerhaus 1996). The lowest SECV, a measure of the standard error between the reference values and NIRS predicted values in the cross validation, as well as the highest 1-VR, an estimate of the coefficient of determination for the cross validation, were obtained in Calibration C, with values of 8.48 and 0.68, respectively (Table 4.1).

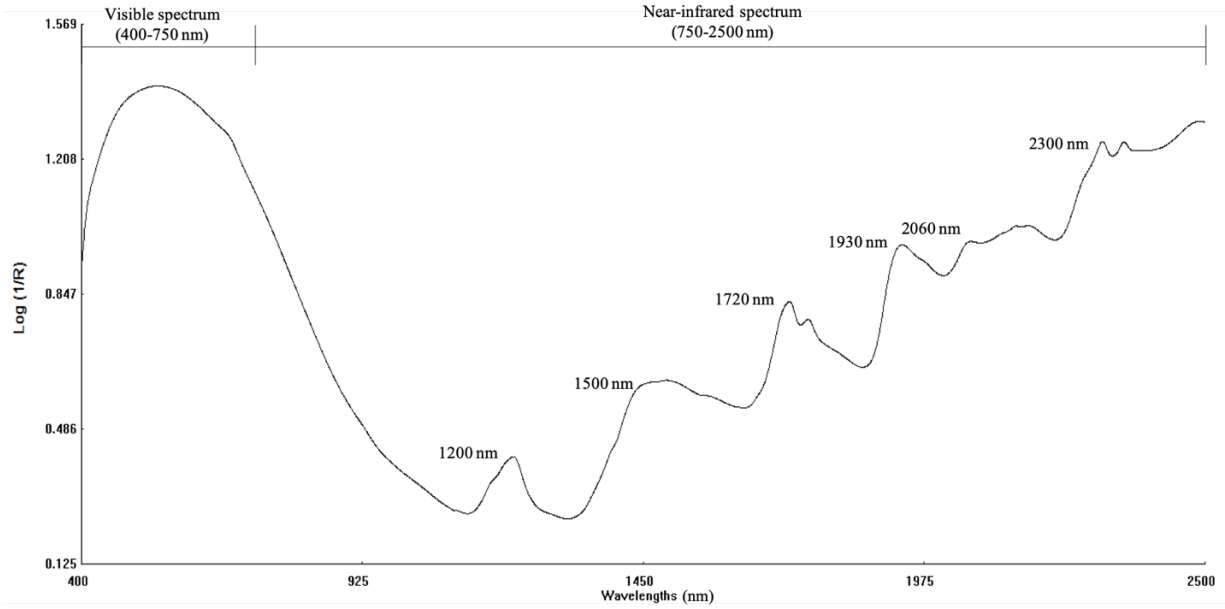


Figure 4.1 Average raw near-infrared reflectance (NIR) spectra of whole seeds of *Brassica napus* L.

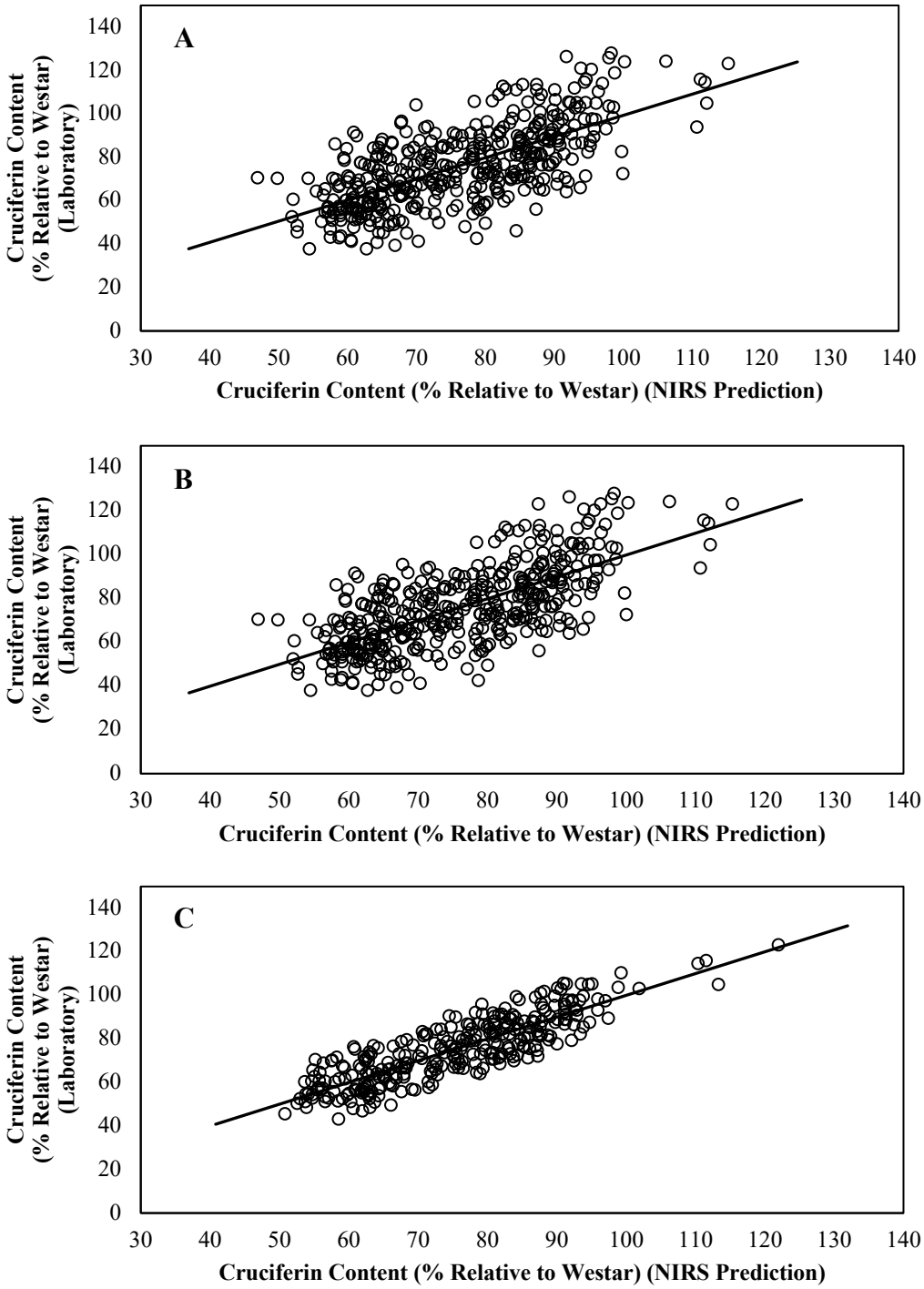


Figure 4.2 Calibration scatter plot of laboratory (ELISA) vs NIRS prediction values for cruciferin content in *Brassica napus* L. whole seed samples as determined by: (A) Calibration A: math treatment 1,16,16,1 (derivative, gap, first smooth, second smooth); (B) Calibration B: math treatment 2,16,16,1; (C) Calibration C: math treatment 2,16,16,1.

4.2.3 External Validation of NIRS Calibration Equations

Following the development of the NIRS calibrations, external validation was carried out using a set of 74 samples that had not been used in the development of said equations. The validation set samples represent similar variation in cruciferin content as those samples found in the calibration set (Table 4.2).

The results of the statistical analysis of the external validation may be found in Table 4.2. While the R^2 for Calibrations A and B were similar between the external validation (Table 4.2) and the calibration (Table 4.1), the R^2 for the external validation (Table 4.2) for Calibration C is significantly lower than that of the calibration (Table 4.1). Of the three calibration equations, the R^2 of the external validation for Calibration B was slightly higher, with a value of 0.47, when compared to the other two equations (Table 4.2, Figure 4.3). In addition to a higher coefficient variation of the calibration, Calibration B also had a slightly lower SEP (a measure of the standard error between the reference values and NIRs predicted values for the validation set) with a value of 15.12, as well as a slightly higher RPD and RER (measures that relate the SEP to the variation in the reference data (Williams and Sobering 1996)), with values of 1.36 and 6.78, respectively. In contrast, it can be noted that of the three calibration equations, Calibration C has a slope closer to one, with a value of 1.11 (Figure 4.3, Table 4.2).

Table 4.2 External validation statistics of two NIRS calibration equations for cruciferin content (% relative to Westar) in *Brassica napus* L.

Calibration[†]	N[‡]	Mean	Range	SD[§]	$R^{2¶}$	Slope	SEP[††]	RPD[‡‡]	RER[§§]
A	74	77.44	37.20-139.66	20.54	0.43	1.23	15.65	1.31	6.55
B	74	77.44	37.20-139.66	20.54	0.47	1.16	15.12	1.36	6.78
C	74	77.44	37.20-139.66	20.54	0.43	1.11	15.60	1.32	6.57

[†]Calibration A math treatment: 1,16,16,1, (derivative, gap, first smooth, second smooth), Calibration B math treatment: 2,16,16,1, Calibration C math treatment: 2,16,16,1.
[‡]Number of samples.
[§]Standard deviation of validation set samples.
[¶]Coefficient of determination for external validation.
[††]Standard error of prediction.
[‡‡]Ratio of the standard deviation of the reference data for the validation set samples to the SEP.
[§§]Ratio of the range in reference data for the validation set samples to the SEP.

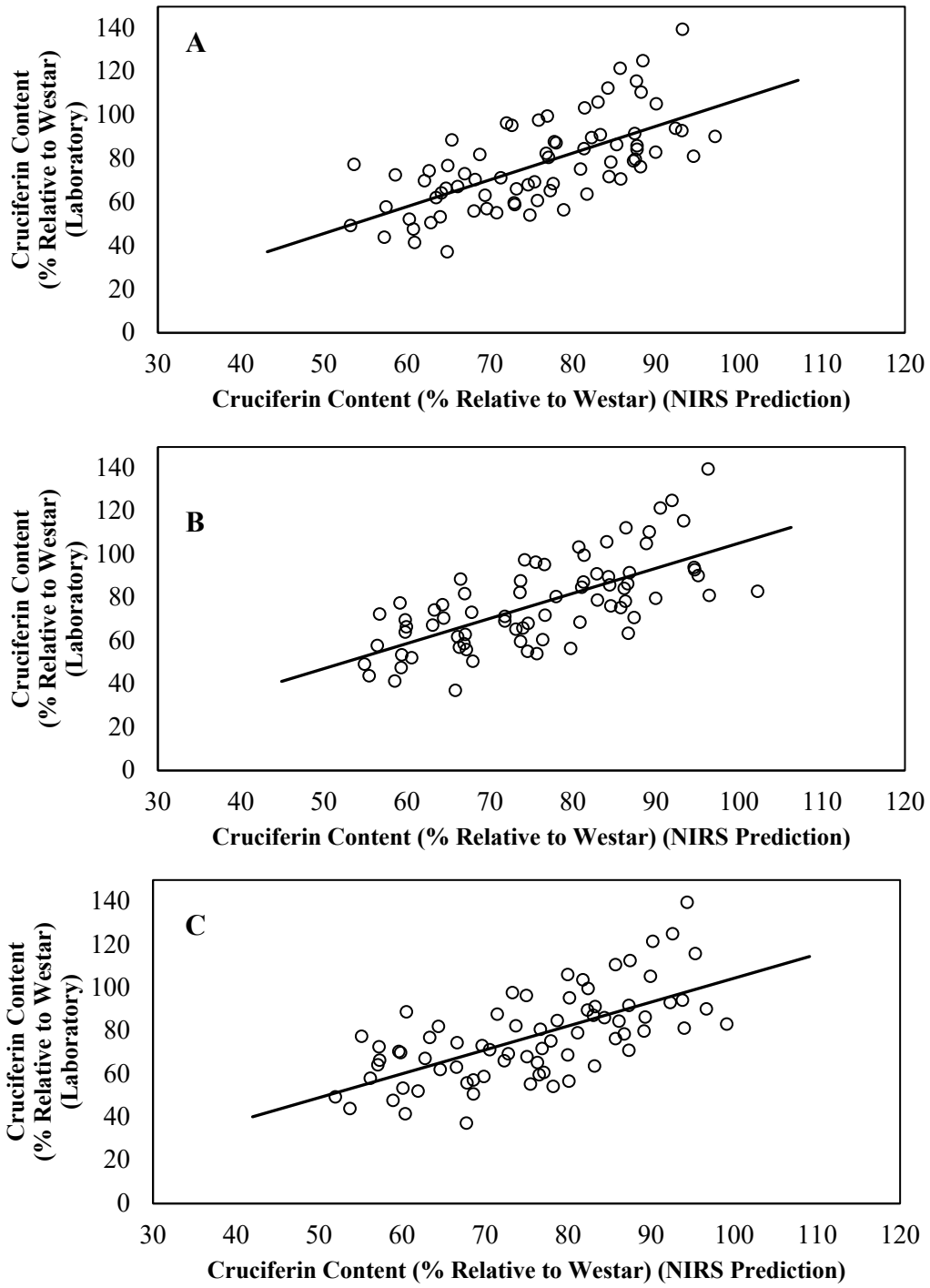


Figure 4.3 External validation scatter plot of laboratory (ELISA) vs NIRS prediction values for cruciferin content in 74 *Brassica napus* L. whole seed samples as determined by: (A) Calibration A: math treatment 1,16,16,1 (derivative, gap, first smooth, second smooth); (B) Calibration B: math treatment 2,16,16,1; and (C) Calibration C: math treatment 2,16,16,1

4.5 Discussion

Plant breeding programs depend heavily on efficient methods through which plants may be phenotyped (Chawade et al. 2019). As such, many *B. napus* breeding programs have adopted near-infrared reflectance spectroscopy (NIRS) as a method by which to assess seed quality traits (Font et al. 2006). Due to the interest in expanding the use of canola seed storage proteins such as cruciferin, breeding efforts would therefore benefit from the ability to utilize NIRS to assess cruciferin content with greater efficiency.

In this study, samples of whole *B. napus* seed were used to develop NIRS calibration equations for cruciferin content. The spectra produced peaks associated with oil (1200, 1720 and 2300 nm), proteins (1500 and 2060 nm) and water (1930 nm) (Workman and Weyer 2012) (Figure 4.1). The spectra collected in this research are consistent with those observed in *B. napus* by other researchers (Tkachuk 1981; Sato et al. 1998; Williams and Norris 2001; Golebiowski et al. 2004; Kaur et al. 2016), and are also similar to those of other common oilseed crops, including flaxseed, palm kernel, safflower, peanut, cottonseed, sesame and soybean (Williams and Norris 2001).

Using whole seeds of *B. napus,* three calibrations were developed, utilizing a number of different math treatments. Math treatments may be used to reduce background noise and improve the ability to differentiate between chemical compounds in the sample (Agelet and Hurburgh 2010). Derivatives are able to minimize the effect of overlapping peaks, remove the effect of the drift of spectra base line intensity over different wavelengths, and remove baseline slope effects (Font et al. 2006; Agelet and Hurburgh 2010). The wavelengths over which the derivative is calculated is known as the gap. Averaging spectral data over several wavelength points, known as smoothing, can also improve calibrations.

The development of Calibration C required the removal of reference samples with high residual values from the original reference sample set. High residual values are an indication that the sample was poorly predicted by the model, and thus removing them may improve the calibration (Agelet and Hurburgh 2010). This does risk, however, the removal of important variation present within the sample set, and a reduction in the ability to predict unknown samples (Shenk and Westerhaus 1996). The analysis of the calibration and cross-validation statistics associated with Calibrations A and B showed that they were unsuitable for estimating cruciferin content (Table 4.1). Therefore, samples were removed in an effort to improve the calibration equation.

In order to evaluate the suitability of the calibration equations for their use as a tool to measure cruciferin content, the statistics generated by the calibration, cross-validation, and external validation must be taken into consideration. When the calibration statistics are considered, a low SEC and an R^2 value close to one are considered favorable (Windham et al. 1989). Calibration B, with values of 13.64 and 0.47 for the SEC and R^2, respectively, may be considered the best performing of the first two calibration equations, while Calibration C, with a SEC of 7.91 and a R^2 of 0.73 would be the best performing overall (Table 4.1). However, the SEC is often considered a poor indication of the accuracy of a prediction, due to its nature as simply an indicator of how well the reference values fit the calibration developed using those same values (Shenk and Westerhaus 1996). Therefore, the more accurate predictors of the accuracy of a calibration equation are the cross-validation statistics, SECV and 1-VR. Similar to the SEC and R^2 of the calibration, a low SECV and a 1-VR close to 1 are considered ideal. When these statistics are taken into consideration, the conclusions based on the SEC and R^2 are further supported, showing that Calibration C has the lowest SECV and highest 1-VR of the three calibration equations (Table 4.1).

Following internal cross-validation, external validation is necessary to assess the performance of the calibration equation on samples not included in the original sample set used for calibration development (Windham et al. 1989). Based on an evaluation of the external validation statistics of the equations reported here, none of the three equations would be considered suitable for the measurement of cruciferin content by NIRS. As shown by Figure 4.2, as well as the R^2 value of the external validation (Table 4.4), it can be seen that there is a poor correlation between the values predicted by the NIRS calibrations and the reference values measured by ELISA. It may be noted that while the R^2 of the external validation for Calibrations A and B (Table 4.2) are similar to their respective R^2 of the calibration (Table 4.1), the R^2 of the external validation for Calibration C is significantly lower than the R^2 observed for calibration development, illustrating the risk that removal of samples from a calibration set can result in a reduced ability to predict unknown samples (Shenk and Westerhaus 1996).

In addition to the R^2, the slope of the external validation is also an important statistic to consider. A slope near one is preferred, and should not be greater than 1.05 or less than 0.95 (Windham et al. 1989). A slope that exceeds this range is an indicator that the calibration equation may underestimate high values and overestimate low values (Westerhaus 1989). All three calibration equations exceed the range recommended by Windham et al (1989) (Table 4.2, Figure 4.3). The final external validation statistics to consider are the RPD and the RER. These statistics put the SEP into the context of the standard deviation and the range of the reference data, and thus are also important indicators of the ability of the equations to predict unknown samples. Williams and Sobering (1996) recommend that the RPD should be at least three and the RER should be at least 10. Thus, the values obtained indicate a low predictive performance of the equations developed in this research (Table 4.2).

NIRS has been applied for use in the assessment of *B. napus* seed characteristics such as total protein content (Tkachuk 1981; Velasco and Möllers 2002), oil content (Tkachuk 1981; Velasco et al. 1999b; Golebiowski et al. 2004), fatty acid composition (Sato et al. 1998; Velasco and Becker 1998b; Velasco et al. 1999b; Kim et al. 2007; Kaur et al. 2016), chlorophyll content (Daun 1976; Tkachuk and Kuzina 1982), fibre content (Font et al. 2003, 2005; Dimov et al. 2012; Wittkop et al. 2012) and glucosinolate content (Velasco and Becker 1998a). In addition, equations have been developed for several amino acids, including aspartic acid, glutamic acid, glycine, alanine, valine, leucine, lysine, histidine and arginine (Chen et al. 2011).

However, NIRS has not been applied to screen for specific seed storage protein fractions, such as cruciferin in *B. napus*. However, some effort has been made towards the development of NIRS calibrations for seed storage protein fractions in soybean. Pazdernik et al. (1996) developed a NIRS calibration equation for the 11S (glycinin) protein fraction of soybean that had an SEC of 7.7 g/kg and an R^2 of the calibration of 0.84. However, the reported SECV was high, and 1-VR statistic was low. A more recent study by Delwiche et al. (2007) was unable to develop successful calibrations for the assessment of the 11S (glycinin) and 7S (β-conglycinin) fractions, with R^2 values of the calibration ranging from 0.198 to 0.218 for the 7S fraction and 0.296 to 0.305 for the 11S fraction, as well as high SEC, SECV and SEP values.

Delwiche et al. (2007) and Pazdernik et al. (1996) each propose that a reason for the poor ability of NIRS to determine the content of specific seed storage proteins is a lack of capacity to detect the spectra specific to individual protein fractions, a suggestion that may extend to the lack of success in developing calibrations for the cruciferin fraction in this study. These studies also propose that error associated with the reference method may also have contributed to their apparent

lack of success (Pazdernik et al. 1996; Delwiche et al. 2007). It is widely known that the ability of an NIRS calibration equation to predict a constituent of interest is dependent upon the reference method error (Windham et al. 1989; Shenk and Westerhaus 1996). In this study, the reference method was found to have a SEL of 8.46%. Thus, along with a lack of sensitivity of NIRS to individual protein fractions, sources of error within the ELISA reference method, such as sensitivity to temperature and time (Crowther 2009; Gan and Patel 2013), may also have contributed to the performance of the equations developed in this study.

While the lack of success by other researchers suggests that NIRS may not be sensitive enough to detect individual protein fractions, other reference methods may be considered for future research. While a number of quantification methods exist, recently Defaix et al. (2019) described a method of quantifying *B. napus* cruciferin and napin proteins through size-exclusion high performance liquid chromatography (SE-HPLC) that allows for minimal sample preparation and relatively swift analysis, which may be a good alternative.

Finally, the use of ground rather than whole seeds for NIRS calibration equation development may be of value. Some studies have found that calibration equations for intact seeds are less precise than those for ground seeds, due to variation in large variations seed size among intact seed samples (Font et al. 2003; Wang et al. 2014). Thus, reducing variation due to seed size by developing a calibration equation based on ground seed samples may be useful.

In order to improve the ability of breeding programs to efficiently screen for seed storage protein content in whole seeds of *B. napus*, this research aimed to investigate the potential for NIRS as a method to make the phenotyping process more efficient, non-destructive, and reduce the need for laboratory chemical methods. The calibration equations developed in this study have proven to be

incompatible with accurately determining cruciferin content. If future efforts are undertaken to explore the potential of this method for the analysis of cruciferin content, points of interest may include increasing the size of the calibration set and improvements to the reference method.

5. GENOME-WIDE ASSOCIATION MAPPING FOR CRUCIFERIN CONTENT IN *BRASSICA NAPUS* L.

5.1 Abstract

Canada is the world's largest producer and exporter of canola (*Brassica napus* L.), making the canola industry a key contributor to the Canadian economy. As such, breeding efforts that will add value and diversify canola end-use products by enhancing *B. napus* meal protein suitability for human consumption are important to accelerate. The seed storage protein cruciferin makes up approximately 60% of the protein content of mature seeds and is of interest for use as a source of plant-based proteins for human consumption due to its functional properties. Despite this, research investigating the underlying genetic control of cruciferin content in *B. napus* is limited. This study aimed to broaden the understanding of the genetic control of cruciferin content by identifying loci significantly associated with cruciferin content. To this end, the *Brassica* 60K Illumina® Infinium SNP array was used to genotype a population of 51 diverse *B. napus* genotypes. STRUCTURE and principle coordinates analyses indicated the existence of two sub-populations separated largely on the basis of geographical origin. A genome-wide association study was performed using the GLM+Q model and identified 144 SNP-trait associations across 47 loci significantly associated with cruciferin content. Subsequently, a total of 11 genes were identified as candidate genes, including *BnaA02g28280D*, an ortholog of the *Arabidopsis thaliana* transcription factor *FUS3*. As interest turns towards alternative and plant-based sources of protein, this research will facilitate breeding efforts that will position *B. napus* as a competitive source of plant-based proteins for human consumption.

5.2 Introduction

Behind soybeans, canola (*Brassica napus* L.) is the second most widely produced oilseed crop worldwide (FAO 2019). The canola industry contributes an average of C$26.7 billion per year to the Canadian economy (LMC International 2016), making Canada the world's largest producer (Statistics Canada 2019a) and exporter (USDA 2019) of canola and rapeseed. In order to ensure the Canadian canola industry continues to grow and develop, adding value and exploring end-use diversification will be an important mechanism to contribute to the Canadian economy.

Brassica napus is primarily used as an oilseed (Arntfield 2011). The high-protein meal that remains following oil extraction is considered a secondary product (Rahman 2013; Campbell et al. 2016), and is currently used as a protein source for livestock (Tan et al. 2011). The proteins within the meal of *B. napus* seeds are found predominantly in the form of two seed storage proteins, cruciferin and napin (Aider and Barbana 2011; Wanasundara 2011). The 11S globulin cruciferin makes up the largest proportion of the seed storage proteins, at approximately 60% of the protein content of mature seeds (Aider and Barbana 2011; Wanasundara 2011; Perera et al. 2016). Cruciferin in the form of protein isolates has potential value as a source of protein for human consumption due to its functional properties for food products (Tan et al. 2011), including its emulsifying (Cheung et al. 2015) and gelling properties (Tan et al. 2014b). As interest in alternative forms of proteins for human consumption increases, breeding efforts to develop *B. napus* genotypes with specialty seed storage protein profiles will be essential to ensure *B. napus* is competitive in the plant-based protein industry.

To facilitate breeding efforts to improve seed storage protein profiles, an understanding of the genetic control of cruciferin content will be critical. By identifying genes that are associated with

cruciferin content, their associated markers may be used in breeding programs through marker-assisted and genomic selection (Gacek et al. 2018). By allowing breeders to select for a phenotype of interest based upon a known marker, marker-assisted selection can improve plant breeding efficiency and reduce the cost of phenotyping (Sleper and Poehlman 2006).

The two most common methods by which the genetic control of traits are studied in order to facilitate the discovery of marker-trait associations are quantitative trait loci (QTL) mapping and association mapping (Stich and Melchinger 2010). QTL mapping uses populations of known structure (often through early generation crosses between two parents) to study recombination events in the progeny of specific individuals (Oraguzie and Wilcox 2007). In contrast, association mapping (genome-wide association study, GWAS) uses populations of diverse, unrelated individuals to study the historical recombination and natural variation within a species (Oraguzie and Wilcox 2007).

To date, the study of the genetic control of cruciferin content has been limited. A single QTL mapping study identifying QTL for cruciferin content has been conducted, using a doubled haploid population of winter oilseed rape (Schatzki et al. 2014). Schatzki et al. (2014) identified two QTL for cruciferin content, located on chromosomes A02 and C09. The study by Schatzki et al. (2014) used a marker map consisting of 229 SSR and AFLP markers. In contrast, the *Brassica* 60K Illumina® Infinium SNP array contains 52,157 markers for genotyping in *B. napus* (Clarke et al. 2016). While Schatzki et al. (2014) have been successful in identifying QTL for cruciferin content, further study and the use of the *Brassica* 60K Illumina® Infinium SNP array will further elucidate the genetic control of cruciferin content.

The objective of this research is to provide information regarding the genetic control of cruciferin content in *B. napus* seeds by identifying significant SNP-trait associations and candidate genes for the control of cruciferin content. A genome-wide association mapping study was conducted using genotypic data generated using the *Brassica* 60K Illumina® Infinium SNP array to identify loci significantly associated with cruciferin content. This research will provide a framework for future association mapping studies and facilitate breeding efforts to develop *B. napus* genotypes with improved seed storage protein profiles.

5.3 Materials and Methods

5.3.1 Plant Material, Field Evaluation and Phenotyping

Details regarding the 51 diverse *B. napus* genotypes (the parental genotypes of a nested association mapping (NAM) population) used in this study may be found in Chapter 3 (Section 3.3.1 Plant Material). For association analysis, the check genotype DH12075 was not used. Field evaluation and agronomic management are described in Chapter 3 (Section 3.3.2 Field Evaluation). Briefly, plots were grown in four site-years: Saskatoon in 2014, Saskatoon in 2015, Winnipeg in 2016 and Winnipeg in 2017. Field evaluation in 2014 and 2015 (Saskatoon) was carried out in a randomized complete block design, with two replicates per genotype, while in 2016 and 2017 (Winnipeg), evaluation was carried out in a randomized complete block design with three replicates per genotype. Cruciferin content was determined through an enzyme-linked immunosorbent assay (ELISA)-based approach, as detailed in Chapter 3 (Section 3.3.3 Analysis of Cruciferin Protein by Enzyme-Linked Immunosorbent Assay).

5.3.2 Statistical Analysis

Statistical analysis was performed using SAS version 9.4 (SAS Institute, Cary, NC). Broad-sense heritability was estimated for each of the two environments (Saskatoon 2014 and 2015, Winnipeg 2016 and 2017) separately. The estimates of the variance components were calculated using PROC MIXED, as described by Holland et al (2003). Broad-sense heritability was calculated as $h^2 = \sigma_g^2/(\sigma_g^2 + \sigma_{gy}^2/y + \sigma_e^2/yr)$, where σ_g^2 is the genetic variance, σ_{gy}^2 is the interaction variance between the genotype and year, σ_e^2 is the error variance, *y* is the number of years, and *r* is the number of replications. Best linear unbiased predictions (BLUPs) (Piepho et al. 2008) for each genotype in each site-year, as well as across the four site-years were calculated using PROC MIXED, for use as the trait values for association analysis.

5.3.3 SNP Genotyping and Filtration

SNP genotyping of the 51 NAM parental genotypes was performed with the *Brassica* 60K Illumina® Infinium SNP array by Agriculture and Agri-Food Canada in Saskatoon, Saskatchewan, according to the manufacturer's protocol (Illumina Inc, San Diego, CA). Filtration of SNP marker data was performed using the Illumina GenomeStudio Genotyping Module v2.0 software. From an initial 52,157 SNPs, 15,569 SNPs with a call frequency of < 0.7, 3,719 SNPs with a minor allele frequency of < 0.05, 858 SNPs with an AB frequency of > 0.1 and 3,370 SNPs with a number of no calls > 5 were excluded from the final set of SNPs. In addition, a further 2,908 markers that were mapped to unknown chromosomes were excluded. A final set of 30,796 SNPs were used for further analyses.

5.3.4 Population Structure and Kinship Analysis

Using PLINK (Purcell et al. 2007) -indep-pairwise 50 5 0.2, a subset of 2,700 markers evenly distributed across the genome was selected for use in population structure analysis. The population structure within 51 *B. napus* genotypes was analysed using STRUCTURE 2.3.4 software (Pritchard et al. 2000). Analysis was performed using the admixture model, with a burn-in length and MCMC (Markov Chain Monte Carlo) reps of 100,000. The number of sub-populations (*K*) was set from 1 to 10, with 5 iterations for each *K*. To determine the most likely number of sub-populations, the output from STRUCTURE was visualized using STRUCTURE HARVESTER (Earl and VonHoldt 2012) to analyze for the delta K value (Δ*K*), as described by Evanno et al. (2005). Each genotype was assigned to a sub-population based on their membership probability (*Q*). Genotypes with a probability > 0.7 were assigned to respective sub-populations, while those with < 0.7 were assigned to a mixed group. The population structure (*Q*) matrix was utilized for genome-wide association analyses. Principle coordinates analysis (PCoA) was performed using GenAlEx v6.5 (Peakall and Smouse 2012), and the pruned set of 2,700 SNPs. PCoA was performed using the codominant genotypic distance and covariance-standardized methodology options. The relative kinship matrix (*K)* was calculated using TASSEL 5.0 (Bradbury et al. 2007), using the entire 30,796 SNP set.

5.3.5 Linkage Disequilibrium

Linkage disequilibrium (LD) analysis was performed between all pairs of the 30,796 SNP set using PLINK (Purcell et al. 2007), with the parameters -r^2 -ld-window-r^2 0 -ld-window 999999. To obtain a clear visualization of LD decay and reduce the influence of outliers, SNP pairs were

combined into intervals of 50 kb (Liu et al. 2016b), and the mean the r^2 was plotted against the physical distance using the R package ggplot2 (Wickham 2016).

5.3.6 Genome-Wide Association Analysis

Trait-SNP association analysis was performed using TASSEL 5.0 (Bradbury et al. 2007), with three different models: (1) the general linear model (GLM), controlling for population structure (*Q*) (GLM+Q), (2) the mixed linear model (MLM), controlling for relative kinship *(K*) (MLM+K), and (3) the MLM, controlling for *Q* and K (MLM+Q+K). The *Q* matrix was obtained from STRUCTURE analysis, and the *K* matrix was obtained with TASSEL. Association analysis was performed for cruciferin content in each of the four site-years individually, as well as the site-years combined. To determine the significance of trait-SNP associations, the threshold $P < 3.25 \times 10^{-5}$ ($P = 1/n$, where n = the number of SNPs used; $-\log_{10}(P) = -\log_{10}(1/30796) = 4.5$) was used. Quantile-quantile (QQ) plots, displaying the $-\log_{10}(P)$ of each SNP compared to the expected were generated using the R package ggplot2 (Wickham 2016), while Manhattan plots were generated using the R package qqman (Turner 2014).

5.3.7 Candidate Gene Identification

Haplotype block estimation was carried out using PLINK (Purcell et al. 2007), with a window size of 10,000 kb, using the definition of a haplotype block suggested by Gabriel et al. (2002), in which pairs are defined as being in "strong LD" with an upper confidence interval limit of D' of >0.98, and a lower limit of 0.7, and a block is created if 95% of comparisons are in "strong LD". To establish regions of interest for candidate gene identification, all genes present between the flanking markers of a significantly associated marker were taken into consideration (Hatzig et al.

2015). When the significantly associated marker was a member of a haplotype block, as defined above, all genes within the entire haplotype region were taken into consideration (Yao et al. 2020). Genes in regions of interest were determined using the *B. napus* reference genome sequence (http://www.genoscope.cns.fr/brassicanapus) (Chalhoub et al. 2014), and were annotated using Blast2Go (Conesa et al. 2005), using the default settings. Candidate genes were selected based on gene ontology (GO) terms, with genes with GO terms for positive regulation of abscisic acid biosynthetic process, response to abscisic acid, seed maturation, protein targeting to vacuole and amino acid transport highlighted as candidate genes. Orthologous genes were determined by BlastP search against the *Arabidopsis thaliana* database (http://www.arabidopsis.org/Blast/index.jsp).

5.4 Results

5.4.1 Phenotypic Variation in Cruciferin Content

An overview of the summary statistics of the BLUP values calculated for cruciferin content for the genotypes evaluated in this study can be found in Table 5.1. In Saskatoon, the mean cruciferin content relative to Westar was 87.45% and 103.24% for 2014 and 2015, respectively. In Winnipeg, the mean cruciferin content relative to Westar was 62.70% and 86.72% for 2016 and 2017, respectively. Broad-sense heritability was calculated for each of the two field locations (Saskatoon in 2014 and 2015, and Winnipeg 2016 and 2017) separately. Broad-sense heritability was calculated as 70.34% in Saskatoon, and 68.94% in Winnipeg.

Table 5.1 Phenotypic range, mean, and broad sense heritability for cruciferin content (% relative to standard) in *Brassica napus* L. NAM parental genotypes.

Location	Year	Minimum	Maximum	Mean±SD	Broad-sense heritability (%)
Saskatoon	**2014**	56.67	128.44	87.45 ± 20.90	70.34
	2015	67.48	142.26	103.42 ± 16.69	
Winnipeg	**2016**	34.21	95.88	62.70 ± 15.07	68.94
	2017	60.98	123.75	86.72 ± 15.12	

5.4.2 SNP Filtration

The 52,157 SNPs of the *Brassica* 60K SNP array were used to genotype 51 parental genotypes of the NAM population. After exclusion of low-quality SNPs, a total of 30,796 SNPs were used for association analysis. These SNPs were distributed across all 19 chromosomes. Chromosome C04 had the most, with a total of 3,083 SNPs and a SNP density of 15.86 kb/SNP, while chromosome C05 had the fewest SNPs, with a total of 910 and a SNP density of 47.11 kb/SNP (Table 5.2).

5.4.3 Population Structure

A subset of 2,700 SNPs from the 30,796 SNP data set was used for the analysis of population structure. STRUCTURE analysis showed that ΔK showed the highest peak at $K = 2$ (Figure 5.1), indicating the presence of two sub-populations within this *B. napus* population (referred to as S1 and S2) (Figure 5.2). Genotypes were assigned to sub-populations based on *Q* matrix values, with values ≥ 0.7 resulting in genotypes being assigned to one of the two sub-populations, while those with a value < 0.7 were assigned to the admixed group (AD). Six genotypes (11.76%) were assigned to S1, 35 (68.63%) were assigned to S2, while the remaining 10 genotypes (19.61%) were assigned to the admixed group (Table 5.3, Figure 5.2). The distribution of genotypes into sub-populations was largely consistent with their geographic origin. S1 consisted exclusively of genotypes originating from Asia, while S2 consisted of genotypes from North America (8), Europe (13), Australia (4), Asia/South Asia (8), and South America (2) (Table 5.3).

Results from PCoA based on the 2,700 SNP subset could explain 10.04% of the genetic variance in the first principal component, and 5.86% in the second. The results of PCoA analysis are consistent with those from the STRUCTURE analysis (Figure 5.3).

Table 5.2 Summary of single nucleotide polymorphism density for each chromosome across 51 parental genotypes of a *Brassica napus* L. NAM population.

Chromosome	Length (kb)	Number of SNPs	SNP density (kb/SNP)
A01	22793.57	1309	17.41
A02	24751.61	949	26.08
A03	29734.32	1924	15.45
A04	19103.33	1373	13.91
A05	23023.76	1471	15.65
A06	24387.67	1413	17.26
A07	23956.72	1706	14.04
A08	18934.51	1069	17.71
A09	33569.22	1385	24.24
A10	17271.59	1479	11.68
Sub-genome A mean	237526.30	14078	16.87
C01	38550.28	2464	15.65
C02	46139.69	2230	20.69
C03	60556.90	2609	23.21
C04	48892.70	3083	15.86
C05	42867.05	910	47.11
C06	36940.77	1351	27.34
C07	44631.33	1682	26.53
C08	38386.40	1434	26.77
C09	48126.39	955	50.39
Sub-genome C mean	405091.51	16718	24.23
A + C	642617.81	30796	20.87

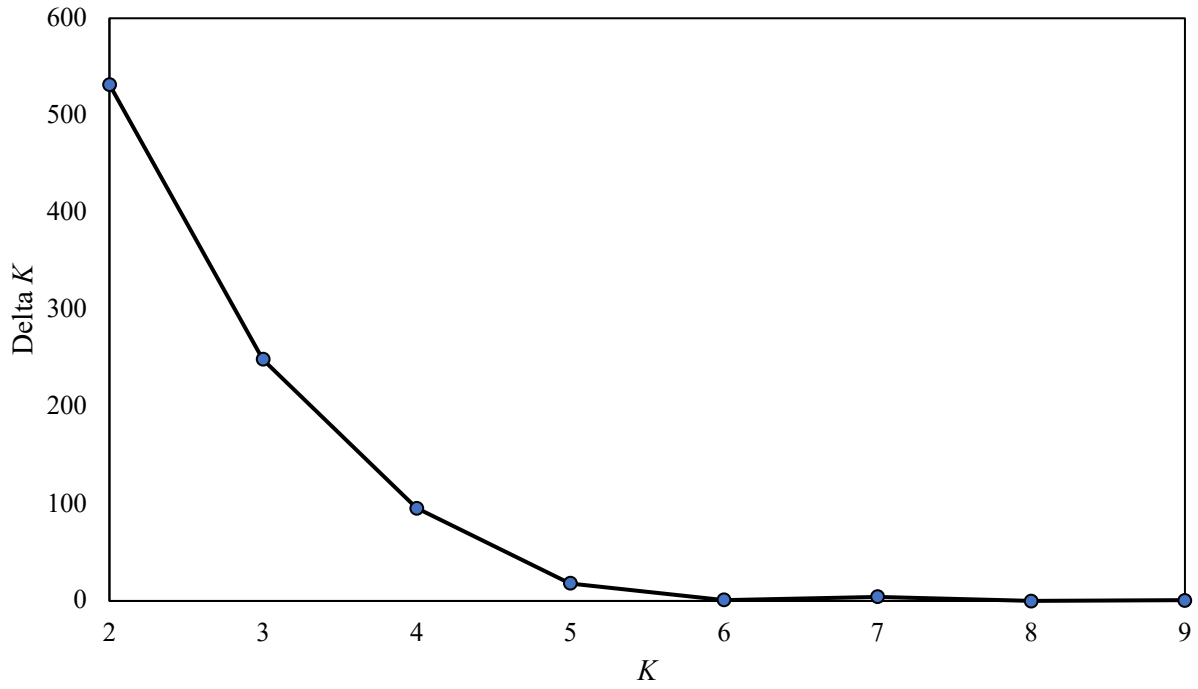


Figure 5.1 Estimation of Δ*K* based on STRUCTURE analysis shows the most likely number of subpopulations (K) in 51 parental genotypes of a *Brassica napus* L. NAM population.

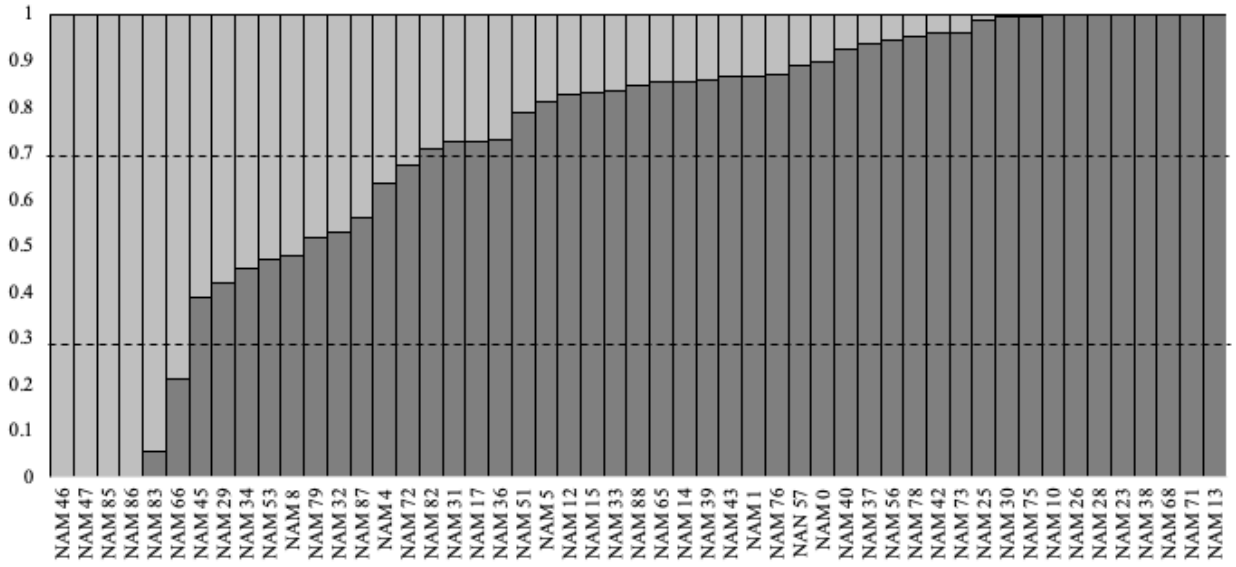


Figure 5.2 Population structure analysis performed by STRUCTURE for K = 2 sub-populations. Each of the 51 parental genotypes of the *Brassica napus* L. NAM population are represented by a single column and categorized as belonging to one of two sub-populations or the admixed group as indicated based on the relative proportions of the light (S1) and dark (S2) grey bars.

Table 5.3 *Brassica napus* L. NAM parental genotypes, their geographical origins, and their sub-population membership based on STRUCTURE analysis.

Genotype	Country of Origin	Continent of Origin	Sub-population[†]
NAM 46	South Korea	Asia	S1
NAM 47	South Korea	Asia	S1
NAM 85	South Korea	Asia	S1
NAM 86	South Korea	Asia	S1
NAM 83	China	Asia	S1
NAM 66	Unknown	Asia	S1
NAM 45	South Korea	Asia	AD
NAM 34	South Korea	Asia	AD
NAM 53	South Korea	Asia	AD
NAM 8	Unknown	Asia	AD
NAM 32	South Korea	Asia	AD
NAM 4	Australia	Australia	AD
NAM 87	France	Europe	AD
NAM 72	Canada	North America	AD
NAM 29	RSYN[‡]	Other	AD
NAM 79	Pakistan	South Asia	AD
NAM 31	Korea	Asia	S2
NAM 36	Unknown	Asia	S2
NAM 33	South Korea	Asia	S2
NAM 39	Korea	Asia	S2
NAM 23	North Korea	Asia	S2
NAM 82	Australia	Australia	S2
NAM 15	Australia	Australia	S2
NAM 37	Australia	Australia	S2
NAM 38	Australia	Australia	S2
NAM 88	Poland	Europe	S2
NAM 65	Sweden or Denmark	Europe	S2
NAM 14	Sweden	Europe	S2
NAM 1	Poland	Europe	S2
NAN 57	Denmark	Europe	S2
NAM 40	Poland	Europe	S2
NAM 56	Poland	Europe	S2
NAM 73	Denmark	Europe	S2
NAM 30	Unknown	Europe	S2

Table 5.3 continued *Brassica napus* L. NAM parental genotypes, their geographical origins, and their sub-population membership based on STRUCTURE analysis.

Genotype	Country of Origin	Continent of Origin	Sub-population[†]
NAM 10	Sweden or Denmark	Europe	S2
NAM 28	Sweden	Europe	S2
NAM 68	Germany	Europe	S2
NAM 13	Germany	Europe	S2
NAM 17	Canada	North America	S2
NAM 51	Canada	North America	S2
NAM 12	Canada	North America	S2
NAM 76	Canada	North America	S2
NAM 0	Canada	North America	S2
NAM 78	Canada	North America	S2
NAM 75	Canada	North America	S2
NAM 71	Canada	North America	S2
NAM 25	Argentina	South America	S2
NAM 26	Argentina	South America	S2
NAM 5	India	South Asia	S2
NAM 43	Bangladesh	South Asia	S2
NAM 42	Bangladesh	South Asia	S2

[†]S1 is sub-population 1, S2 is sub-population 2, AD is the admixed sub-population.
[‡] Resynthesized *Brassica napus* genotype. This genotype was derived from a cross between a *Brassica oleracea* line and a *Brassica rapa* line.

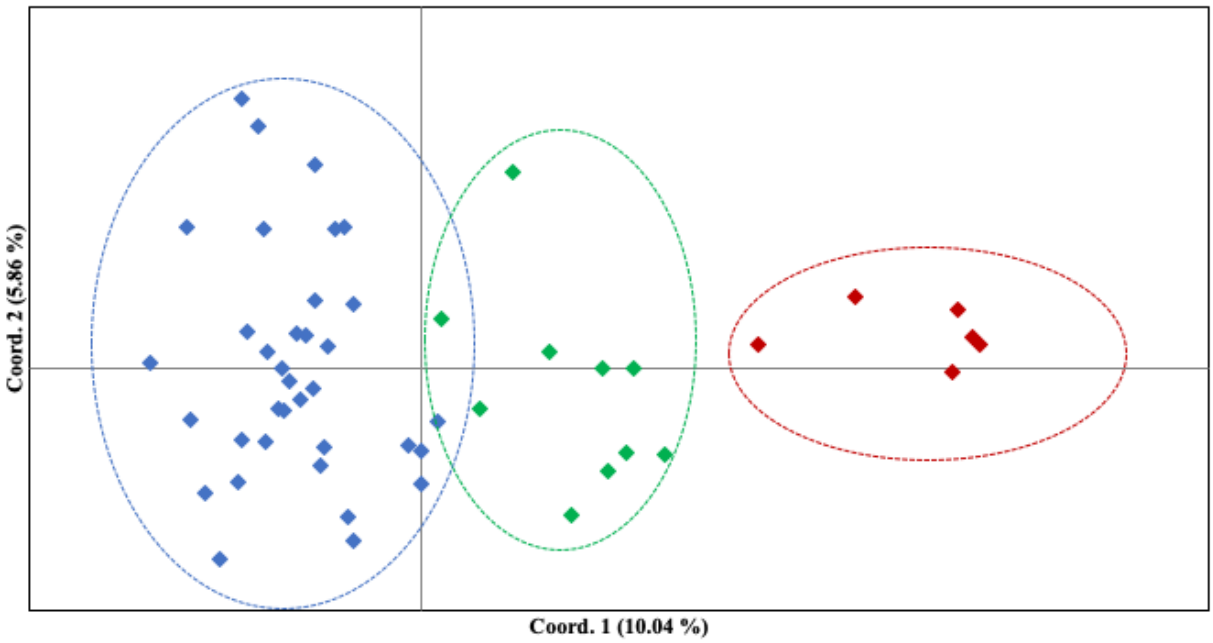


Figure 5.3 Principal coordinates analysis (PCoA) for the *Brassica napus* L. NAM parental genotypes based on a subset of 2,700 SNPs. Genotypes were coloured based on their sub-population as defined by STRUCTURE analysis (red is sub-population 1 (S1), blue is sub-population 2 (S2) and green is the admixed group (AD)).

5.4.4 Linkage Disequilibrium

Linkage disequilibrium (LD) analysis was done on the set of 30,796 SNPs, using r^2 to calculate LD. Genome-wide LD decay for the A and C sub-genomes as well as the entire genome is shown in Figure 5.4. For the entire genome, r^2 dropped to 0.2 within 2.10 Mb and to 0.1 within 6.20 Mb. For sub-genome A, r^2 dropped to 0.2 within 0.15 Mb and to 0.1 within 0.9 Mb, while a significantly slower LD decay rate was observed in sub-genome C, in which r^2 dropped to 0.2 within 4.60 Mb and to 0.1 within 7.85 Mb. In addition, a great deal of variation in the rate of LD decay was observed between each of the 19 chromosomes of sub-genome A and sub-genome C (Figure 5.4).

5.4.5 Genome-Wide Association Analysis

Using a total of 30,796 SNPs, association analysis was carried out using three GWAS models. As can be seen in the QQ plots, the GLM+Q model likely resulted in a number of false positives, in particular for the 2014-2017 combined analysis and the individual 2014 analysis, while the incorporation of kinship and population structure in the MLM+K and MLM+Q+K models resulted in p-values closer to the expected (Figure 5.5). Of the three GWAS models utilized, the GLM+Q model was the sole model to result in significant SNP-trait associations in any of the site-years in this study (Figure 5.6, Appendix 5.1, Appendix 5.2). A total of 144 SNP-trait associations were detected for cruciferin content across each of the individual site-years and in the combined analysis at a significance threshold of $P < 3.25 \times 10^{-5}$ ($P = 1/n$, where n = the number of SNPs used; $-\log_{10}(P) = -\log_{10}(1/30796) = 4.5$) (Figure 5.6). Of these SNP-trait associations, none were identified in 2016, and only 20 were detected in more than one year. The largest number of SNP-trait associations was found on chromosome C02 (72), with the second greatest number found on C09 (28).

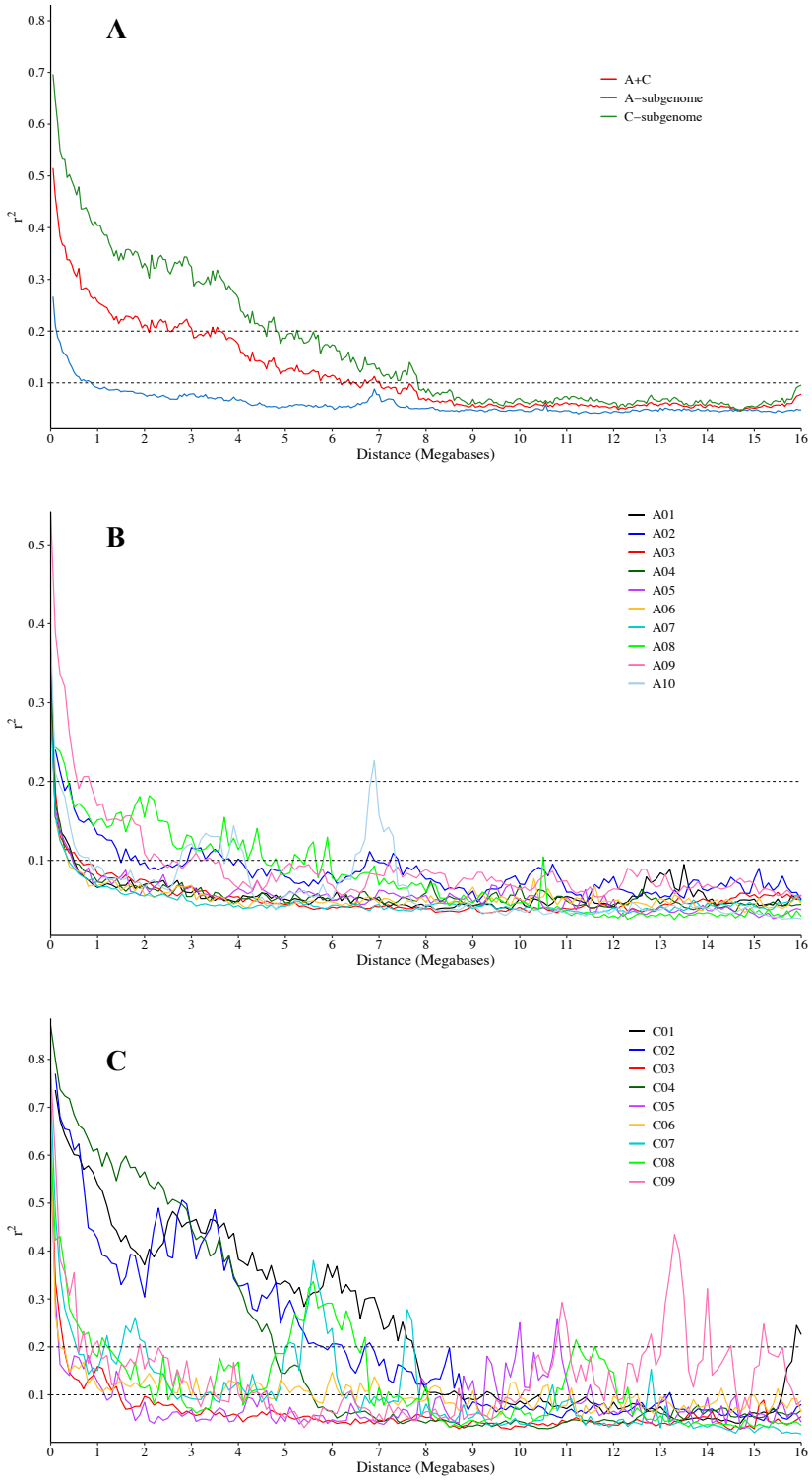


Figure 5.4 Linkage disequilibrium (LD) decay on the A and C sub-genomes of *Brassica napus* L. (A) LD decay in the A and C sub-genomes. (B) LD decay for each of the chromosomes in the A sub-genome. (C) LD decay for each of the chromosomes in the C sub-genome.

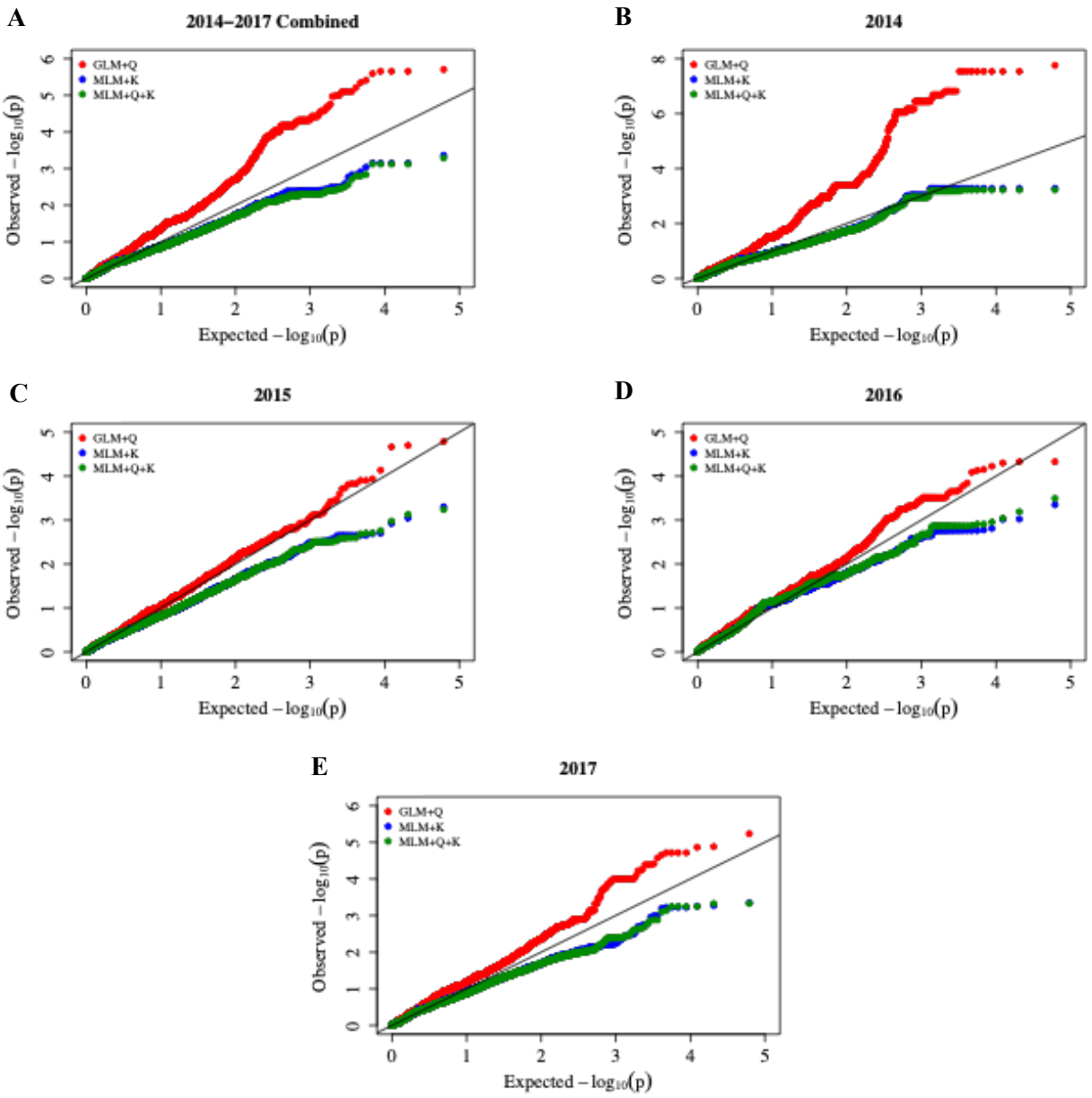


Figure 5.5 Quantile-quantile (QQ) plots of the observed p values plotted against the expected p values from association analysis for three different genome-wide association models. QQ plots for 2014-2017 combined (A), 2014 (B), 2015 (C), 2016 (D), and 2017 (E) are shown.

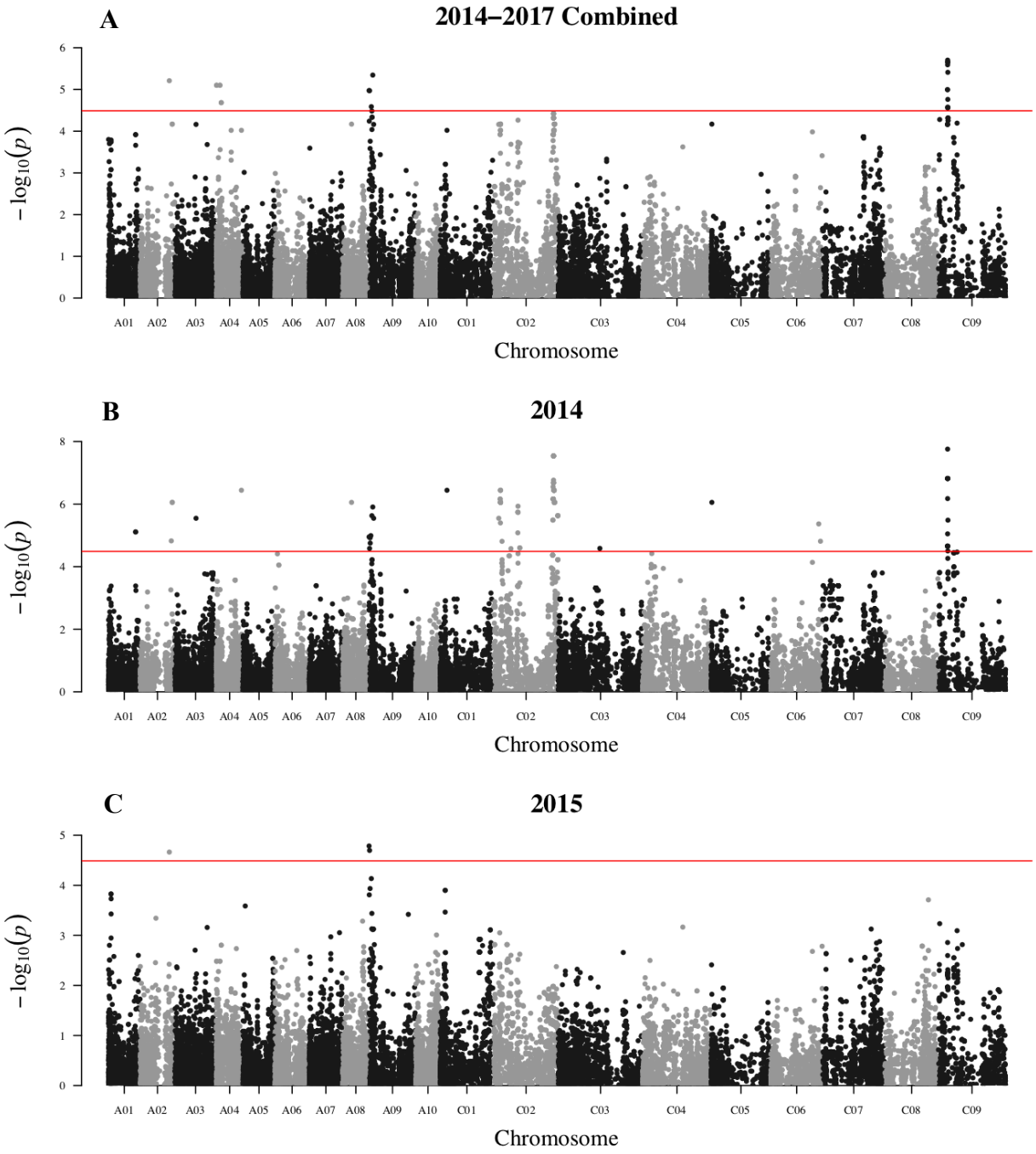


Figure 5.6 Manhattan plots showing single nucleotide polymorphisms with significant associations to cruciferin content in the GLM+Q model in: (A) 2014-2017 combined, (B) 2014, (C) 2015, (D) 2016, and (E) 2017. The horizontal red line represents the significance threshold of $-\log_{10}(P) = 4.5$.

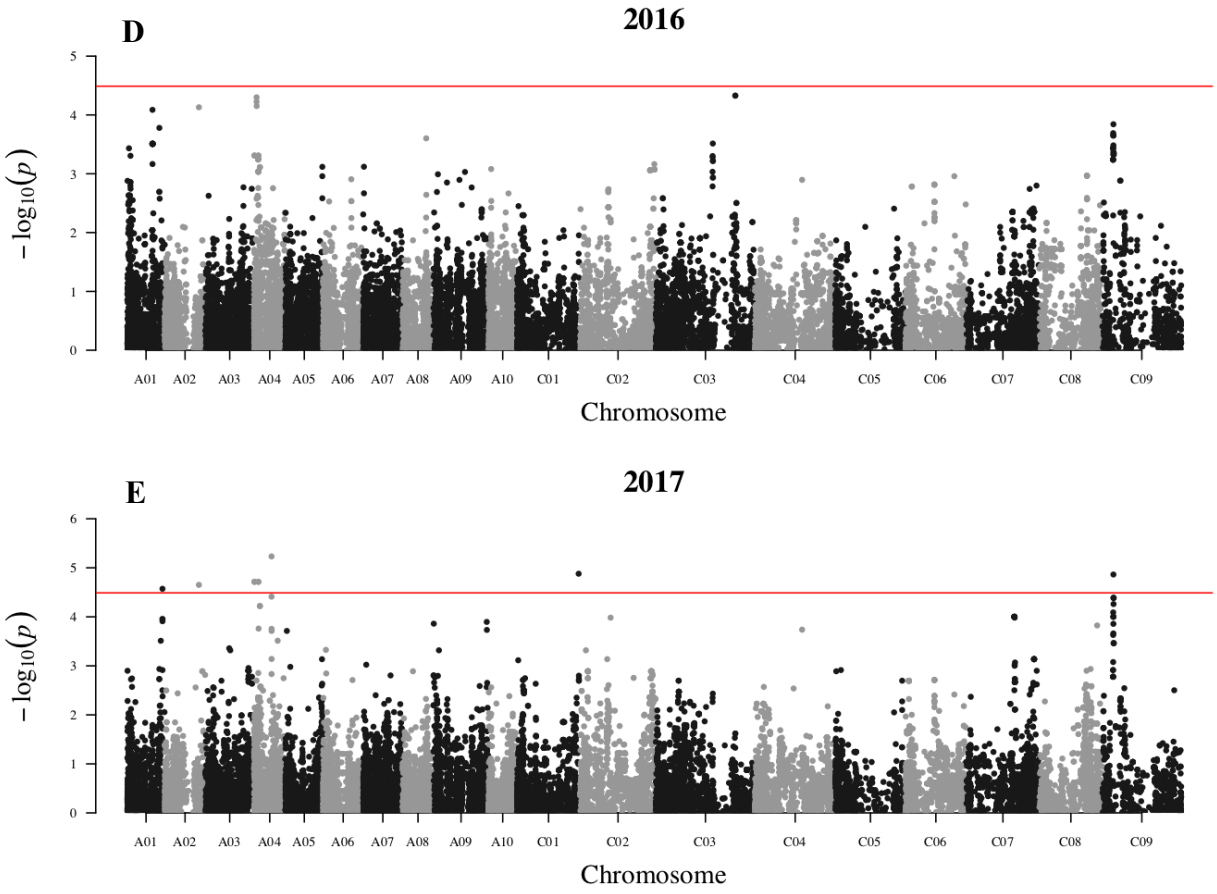


Figure 5.6 continued Manhattan plots showing single nucleotide polymorphisms with significant associations to cruciferin content in the GLM+Q model in: (A) 2014-2017 combined, (B) 2014, (C) 2015, (D) 2016, and (E) 2017. The horizontal red line represents the significance threshold of $-\log_{10}(P) = 4.5$.

5.4.6 Candidate Gene Identification

To explore potential candidate genes, the 144 significant SNPs were blocked according to haplotype blocks, resulting in 47 loci for cruciferin content, distributed across 12 chromosomes (Table 5.4, Appendix 5.3). A total of 936 genes were discovered within the 47 SNP regions significantly associated with cruciferin content. Based on GO terms identified to have potential to be related to the control of cruciferin content (positive regulation of abscisic acid biosynthetic process, response to abscisic acid, seed maturation, protein targeting to vacuole, amino acid transport), 11 candidate genes were identified (Table 5.5). On chromosome A02, a candidate gene was identified which encodes an ortholog of the *Arabidopsis thaliana FUS3* transcription factor (*BnaA02g28280D*). Significant associations in four separate haplotype blocks on chromosome A09 contained genes orthologous to *A. thaliana EZA1* (*BnaA09g00500D*), *SUCROSE SYNTHASE 3* (*SUS3*) (*BnaA09g00710D*), *SUGAR TRANSPORT PROTEIN 13 (STP13) (BnaA09g04260D)* and *VESICLE TRANSPORT V-SNARE 13* (*VTI13*) (*BnaA09g04260D*). Candidate genes within a haplotype block on chromosome C02 were found to be orthologous to the *A. thaliana* genes *ARABIDOPSIS THALIANA RING AND DOMAIN OF UNKNOWN FUNCTION 1117 2 (ATRDUF2)* (*BnaC02g0991D*) and two candidate genes orthologous to *LIPID TRANSFER PROTEIN 4 (LTP4)* (*BnaC0210050D* and *BnaC02g10060D*). A significant association within another haplotype block of chromosome C02 was orthologous to the *A. thaliana* gene *ARABIDOPSIS THALIANA MYB DOMAIN PROTEIN 44 (AtMYB44)* (*BnaC02g16640D*), while an additional haplotype block contained a candidate gene orthologous to *LYSINE-HISTIDINE LIKE TRANSPORTER 5* (*LHT5)* (*BnaC02g20660*). A final haplotype block on chromosome C02 contained a gene orthologous to *ARABIDOPSIS TOXICOS EN LEVADURA 31 (ATL31)* (*BnaC02g39700D*).

Table 5.4 Single nucleotide polymorphisms (SNPs) significantly associated with cruciferin content based on the genome-wide association model GLM+Q. For each haplotype block, SNPs with the lowest p-value in the given block are shown.

Haplotype block	SNP	Chr[†]	Position (bp)	-log$_{10}$ (p)	R^2 [‡] (%)	Location[§]
1.1	Bn-A01-p22983721	A01	19474151	5.11	0.38	2014
1.2	Bn-A01-p22999151	A01	19494342	5.11	0.38	2014
1.3	Bn-scaff_15712_3-p818174	A01	21408820	4.57	0.35	2017
2.1	Bn-scaff_21884_1-p819724	A02	20834407	4.65-5.21	0.30-0.33	2015, 2017, ALL
2.2	Bn-scaff_15918_1-p77494	A02	22242156	4.82	0.32	2014
2.3	Bn-scaff_17623_1-p32036	A02	22864476	6.05	0.39	2014
2.4	Bn-scaff_17721_1-p758760	A02	22987017	6.05	0.39	2014
3.1	Bn-scaff_16139_1-p1103077	A03	15222567	5.55	0.41	2014
4.1	Bn-A04-p2963387	A04	199198	4.72-5.10	0.31-0.32	2017, ALL
4.2	Bn-A04-p2999394	A04	212801	4.72-5.10	0.31-0.33	2017, ALL
4.3	Bn-A04-p2944723	A04	2605833	4.72-5.10	0.31-0.34	2017, ALL
4.4	Bn-A04-p2954824	A04	2614835	4.72-5.10	0.31-0.35	2017, ALL
4.5	Bn-A04-p3933714	A04	3540219	4.68	0.30	ALL
4.6	Bn-A04-p9511386	A04	10644573	5.23	0.34	2017
4.7	Bn-scaff_17623_1-p234259	A04	17992643	6.44	0.41	2014
8.1	Bn-scaff_17721_1-p841884	A08	6402565	6.05	0.39	2014
9.1	Bn-A09-p596014	A09	22672	4.79-4.97	0.36-0.37	2014, 2015, ALL
9.2	Bn-scaff_19783_1-p1060183	A09	376961	4.58-4.97	0.30-0.32	2014, 2015, ALL
9.3	Bn-A01-p27004740	A09	745841	4.92	0.32	2014
9.4	Bn-scaff_17526_1-p2206429	A09	836994	4.75	0.31	2014
9.5	Bn-A09-p2116947	A09	1287776	4.99	0.38	2014
9.6	Bn-A09-p616919	A09	1637130	4.59	0.34	ALL
9.7	Bn-A09-p1929245	A09	2026507	4.48-5.63	0.34-0.41	2014, ALL
9.8	Bn-A09-p2733282	A09	2677575	5.34-5.91	0.34-0.38	2014, ALL
9.9	Bn-scaff_16139_1-p1102847	A09	3308335	5.55	0.41	2014
11.1	Bn-scaff_17623_1-p252454	C01	4983595	6.44	0.41	2014
11.2	Bn-scaff_15712_3-p639153	C01	37651052	4.88	0.37	2017
12.1	Bn-scaff_16139_1-p1103215	C02	3529557	5.55	0.41	2014
12.2	Bn-scaff_17623_1-p797758	C02	4482108	6.16	0.40	2014
12.3	Bn-scaff_17623_1-p837616	C02	4555662	6.16	0.40	2014
12.4	Bn-scaff_17623_1-p141721	C02	4645680	6.44	0.41	2014

Table 5.4 continued Single nucleotide polymorphisms (SNPs) significantly associated with cruciferin content based on the genome wide association model GLM+Q. For each haplotype block, SNPs with the lowest p-value in the given block are shown.

Haplotype block	Lead SNP	Chr[†]	Position (bp)	-log$_{10}$ (p)	R^2 [‡] (%)	Location[§]
12.5	Bn-scaff_17721_1-p906353	C02	4808394	6.44	0.41	2014
12.6	Bn-scaff_16934_1-p98107	C02	5861950	4.81	0.32	2014
12.7	Bn-scaff_16139_1-p380924	C02	12253220	4.58	0.30	2014
12.8	Bn-scaff_17079_1-p145741	C02	17316002	5.93	0.38	2014
12.9	Bn-scaff_16369_1-p182886	C02	18593794	4.60	0.35	2014
12.10	Bn-scaff_16162_1-p79842	C02	42315091	5.49	0.40	2014
12.11	Bn-scaff_17623_1-p715915	C02	42720839	7.54	0.46	2014
12.12	Bn-scaff_16139_1-p670451	C02	45841870	5.63	0.41	2014
12.13	Bn-scaff_16139_1-p662939	C02	45848494	5.63	0.41	2014
13.1	Bn-scaff_18917_1-p430795	C03	29800829	4.58	0.30	2014
15.1	Bn-scaff_17721_1-p811886	C05	585524	6.05	0.39	2014
16.1	Bn-scaff_17799_1-p2273825	C06	34271977	5.37	0.40	2014
16.2	Bn-scaff_17799_1-p1070542	C06	35525817	4.81	0.32	2014
19.1	Bn-scaff_17487_1-p421663	C09	6621292	4.57-7.76	0.35-0.47	2014, ALL
19.2	Bn-scaff_17487_1-p441639	C09	6640709	5.70-6.18	0.40-0.44	2014, ALL
19.3	Bn-scaff_17487_1-p464536	C09	6674714	4.50-6.82	0.31-0.47	2014, 2017, ALL

[†]Chromosome.
[‡]Percentage of phenotypic variance explained by the lead SNP.
[§] 2014 in Saskatoon, Saskatchewan, Canada. 2015 Saskatoon, Saskatchewan, Canada. 2016 in Winnipeg, Manitoba, Canada. 2017 in Winnipeg, Manitoba, Canada. ALL is the combined analysis for 2014-2017.

Table 5.5 Candidate genes for cruciferin content in *Brassica napus* L. based on the genome-wide association model GLM+Q in a population of 51 NAM parental genotypes.

Chr†	Haplotype block	Lead SNP	Candidate gene	Arabidopsis ortholog	Description
A02	2.1	Bn-scaff_21884_1-p819724	BnaA02g28280D	At3g26790	B3 domain-containing transcription factor FUS3
A09	9.2	Bn-scaff_19783_1-p1060183	BnaA09g00500D	At4g02020	Histone-lysine N-methyltransferase EZA1-like
	9.2	Bn-scaff_19783_1-p1060183	BnaA09g00710D	At4g02280	Sucrose synthase 3 (SUS3)
	9.5	Bn-A09-p2116947	BnaA09g02560D	At3g29001	Vesicle transport v-SNARE 13 (VTI13)
	9.7	Bn-A09-p1929245	BnaA09g04260D	At5g26340	Sugar transport protein 13 (STP13)
C02	12.6	Bn-scaff_16934_1-p98107	BnaC02g09910D	At5g59550	Arabidopsis thaliana ring and domain of unknown function 1117 2 (AtARDUF2)
	12.6	Bn-scaff_16934_1-p98107	BnaC02g10050D	At5g59310	Lipid transfer protein 4 (LTP4)
	12.6	Bn-scaff_16934_1-p98107	BnaC02g10060D	At5g59310	Lipid transfer protein 4 (LTP4)
	12.7	Bn-scaff_16139_1-p380924	BnaC02g16640D	At5g67300	Arabidopsis thaliana myb domain protein 44 (AtMYB44)
	12.8	Bn-scaff_17079_1-p145741	BnaC02g20660D	At1g71680	Lysine histidine transporter-like 5
	12.11	Bn-scaff_17623_1-p715915	BnaC02g39700D	At5g27420	Arabidopsis toxicos en levadura 31 (ATL31)

†Chromosome.

4.5 Discussion

The improvement of *B. napus* cultivars for traits such as altered seed storage protein profiles depends upon an understanding of the genetic control of these traits. By carrying out a GWAS on a diverse population of 51 *B. napus* genotypes, this study broadens our knowledge surrounding the genetic control of cruciferin content to further these breeding efforts.

An understanding of broad-sense heritability is an important tool in optimizing breeding strategies (Holland et al. 2003). Broad-sense heritability is a measure of the proportion of phenotypic variation that is due to genotypic effects, and thus is an important factor by which expected gain from selection may be predicted (Holland et al. 2003; Schmidt et al. 2019). Heritability estimates are specific to a certain population within certain environments, and may vary significantly amongst populations and environments (Bernardo 2014). For this reason, comparisons of broad-sense heritability estimates between studies are difficult. In this study, broad-sense heritability was 70.34% in Saskatoon, and 68.94% in Winnipeg (Table 5.1). Thus, in this population of genotypes in these two environments, a significant proportion of the observed phenotypic variation was due to genotypic effects.

The presence of population structure is an important factor to consider when carrying out an association mapping study. Often arising due to natural selection or selection by breeders (Luo et al. 2019), population structure refers to the presence of sub-populations within a larger population in which those individuals are more closely related to one another than the average random pair of individuals within the larger population (Breseghello and Sorrells 2006). When unaccounted for, the presence of population structure in an association mapping study can lead to an increase in the frequency of false positive associations (Sneller et al. 2009; Korte and Farlow 2013; Gupta et al.

2014). In the current study, the software STRUCTURE was used to determine population structure among the panel of 51 genotypes used for the association mapping study. This analysis revealed the panel consisted of two sub-populations (Figure 5.2), and principle coordinates analysis (PCoA) confirmed the clustering of these genotypes (Figure 5.3). These sub-populations are separated primarily based on their geographic origins (Table 5.3). Previous studies have found that *B. napus* population structure can be explained largely by morphology and growth habit (Gyawali et al. 2013, 2016; Gazave et al. 2016; Raman et al. 2016; Schiessl et al. 2017). However, in a study of 782 *B. napus* accessions with both winter and spring growth habits, Gazave et al. (2016) demonstrated that while the accessions separated primarily by growth habit, their secondary grouping was by geographical origin. Xiao et al. (2012) demonstrated similar separation based on geographical origin. While sub-population clustering observed in the current study was largely consistent with the geographical origins of the genotypes, there are some exceptions. One reason for this is that geographical origin for each genotype is assigned based on breeding records. When one considers the extensive germplasm exchange between *B. napus* breeding programs, it would not be unexpected for a genotype to be incorrectly assigned (Gazave et al. 2016). Additionally, clustering has been observed based on agronomically important traits (Hu et al. 2007), and thus factors other than growth habit, morphology and geographical origin may contribute to sub-population separation.

In GWAS, the rate of linkage disequilibrium (LD) decay over physical distance is a significant factor that restricts mapping resolution (Liu et al. 2016b; Sun et al. 2016), and the decay of LD may vary considerably between different populations of the same species (Qian et al. 2014). In the population used in the present study, LD for the entire genome was observed to decay within 6.20 Mb (r^2=0.1) (Figure 5.4). Further, it was observed that LD decay was significantly different

between the A (0.9 Mb) and C (7.85 Mb) sub-genomes. The marked difference in LD decay between the two sub-genomes has been observed in other studies (Qian et al. 2014; Li et al. 2016; Liu et al. 2016b). This may be due to a great number of crosses between *B. rapa* and *B. napus* in historical breeding programs (Qian et al. 2006; Liu et al. 2016b), resulting in greater allelic diversity in the A sub-genome than the C sub-genome (Sun et al. 2016). In addition to the significant difference of LD decay between sub-genomes, it was also observed that LD decay varied extensively across chromosomes. In the A sub-genome, longer LD decays were observed on chromosomes A08 and A09, while in the C sub-genome, longer LD decays were observed on chromosomes C01 and C02, as has also been observed by Liu et al. (2016) and Li et al. (2016), suggesting that there have been fewer recombination events on these chromosomes compared to others.

In association mapping studies, the central purpose is to associate marker loci to phenotypes of interest (Kaler et al. 2020). To this end, several models may be used to identify these associations, each with their own strengths and weaknesses. In this study, three GWAS models were tested to identify the model that best located the SNP-trait associations in this population. The most common way in which models are evaluated in their ability to control false positives and false negatives is through the evaluation of QQ plots. In a QQ plot, a straight line close to the 1:1 line with a upward tail indicates a model that has controlled for false positives and false negatives, and is showing true SNP-trait associations (Price et al. 2010; Kaler et al. 2020). Examination of the QQ plots for the GLM+Q model showed that the observed p-values deviated from expected, indicating the possible presence of false positives (Figure 5.5). This has been an observed weakness of the GLM model for association mapping studies, and thus must be taken into consideration (Zhang et al. 2015; Liu et al. 2016b; Fredua-Agyeman et al. 2020). In contrast,

however, the MLM model (MLM+K, MLM+Q+K) is more stringent in its control of false positives and thus may result in false negative associations due to overfitting for population structure and kinship (Zhao et al. 2011; Gupta et al. 2014; Chen et al. 2018) (Figure 5.5). Despite the risk of false negative associations, many studies find that MLM models that control for population structure and kinship are more accurate and reliable (Wang et al. 2015; Lu et al. 2017; Zheng et al. 2017). Of the three models used, the GLM+Q model was the sole model to detect significant SNP-trait associations at a significance threshold of $P < 3.25 \times 10^{-5}$ [5] ($P = 1/n$, where n = the number of SNPs used; $-\log_{10}(P) = -\log_{10}(1/30796) = 4.5$) in any of the site-years in this study (Figure 5.6, Appendix 5.1, Appendix 5.2).

The genome-wide association study using the GLM+Q model led to the identification of 144 SNP-trait associations for cruciferin content. Significant SNPs were identified on both the A (A01, A02, A03, A04, A08, A09) and C (C01, C02, C03, C05, C06, C09) sub-genomes (Figure 5.6). To date, only one other study has been conducted to identify loci associated with cruciferin content in *B. napus*. Schatzki et al. (2014) performed a quantitative trait loci (QTL) mapping study on a population of 224 winter oilseed rape doubled haploid (DH) lines using a marker map consisting of 80 simple sequence repeat (SSR) markers and 149 amplified fragment length polymorphism (AFLP) markers, leading to the discovery of two QTL for cruciferin content (Schatzki et al. 2014). The first QTL was located in the 35.5 – 79.9 cM range of chromosome A02. In *B. napus*, the average relationship between physical and genetic distance is that 1 cM is equal to approximately 500 kb (Suwabe et al. 2006; Ecke et al. 2010). Therefore, the confidence interval of this QTL corresponds to the physical range of approximately 17750 – 39950 kb, which includes all SNPs significantly associated with cruciferin content found on chromosome A02 in the current study. The second QTL discovered by Schatzki et al. (2014) was located in the 30.0 – 38.5 cM range of

chromosome C09, corresponding to the physical range of approximately 15000 – 19250 kb. While the current study did identify SNPs associated with cruciferin content on chromosome C09, they did not coincide with the QTL discovered by Schatzki et al. (2014). Due to the limited extent of research focused on the discovery of loci associated with cruciferin content, these significant SNP-trait associations provide a valuable foundation for improving the understanding of the genetic control of cruciferin content.

Following the identification of SNPs significantly associated with cruciferin content, a total of 11 candidate genes were identified. Candidate genes identified did not correspond with known cruciferin encoding genes (Pang et al. 1988; Breen and Crouch 1992), but were recognized based on their potential connections to the control of cruciferin content (Table 5.5). On chromosome A02, the gene *BnaA02g28280D* was identified as a candidate gene. *BnaA02g28280D* encodes an ortholog of the *Arabidopsis thaliana FUS3* transcription factor, which plays an important role in the regulation of seed storage protein accumulation during seed maturation in *A. thaliana* and *B. napus* (Baumienin et al. 1994; Gacek et al. 2018). The *fus3* mutation of *A. thaliana* has been shown to affect seed development and result in the absence of cruciferin and napin proteins (Baumienin et al. 1994). Several candidate genes have also been identified on chromosome A09. The gene *BnaA09g00500D* encodes an ortholog of the *A. thaliana* gene *EZA1*, which plays a role in the initiation of endosperm development and aids in the control of seed initiation (Wang et al. 2006). During seed development, the transport of sugars and amino acids from the leaves is important to facilitate the synthesis of seed storage proteins (Fujiwara et al. 2002). For this reason, *BnaA09g04260D* was identified as a candidate gene. *BnaA09g04260D* is orthologous to *A. thaliana SUGAR TRANSPORT PROTEIN* (*STP13*), a gene that is induced by abiotic and biotic stressors, but also plays a role in transporting sucrose from source to sink tissues (Lee and Seo

2021). Similarly, *BnaA09g00710D* was found to be orthologous to the *A. thaliana* gene *SUCROSE SYNTHASE 3* (*SUS3*). Sucrose synthase plays a role in seed development by supplying the precursors for seed storage protein biosynthesis (Fallahi et al. 2008), with its strongest expression taking place in *A. thaliana* between 15 and 18 days after flowering, the time at which seed storage proteins accumulate in the seed (Angeles-Núñez and Tiessen 2012). The gene *BnaA09g02560D* was also on chromosome A09 and is orthologous to *A. thaliana* gene *VESICLE TRANSPORT V-SNARE 13* (*VTI13*). While little information is available regarding *VTI13*, it likely plays a role in transporting proteins to the vacuole (Larson et al. 2014). Since seed storage proteins are produced in the endoplasmic reticulum and are then transported to protein storage vacuoles (Gacek et al. 2018), this is an important step in the synthesis of seed storage proteins. *VTI13* is a member of the VTI-type SNAREs family, of which *VTI11* is also a member. A more extensively studied member of this family, *VTI11* is part of a complex that facilitates the fusion of the prevacuolar compartment and the vacuole, and has been shown to be involved in seed storage protein synthesis in *A. thaliana* (Ebine et al. 2008).

Candidate genes were also identified among haplotype blocks located on chromosome C02. The gene *BnaC02g39700D* was found to be orthologous to the gene *ARABIDOPSIS TOXICOS EN LEVADURA 31* (*ATL31*), which is a regulator of plant response to carbon and nitrogen nutrient availability (Li et al. 2017; Fredes et al. 2019). Li et al. (2017) suggested that *ATL31* plays a role in tomato fruit development, and when one considers that nutrient availability is a crucial factor in seed protein synthesis (Tabe et al. 2002), it may also play a role in the control of cruciferin content in *B. napus*. The gene *BnaC02g16640D* is orthologous to *ARABIDOPSIS THALIANA MYB DOMAIN PROTEIN 44* (*AtMYB44*), which has been shown to regulate glucosinolate biosynthesis by affecting the expression of 11 genes that function at 5 of the 6 steps in glucosinolate

biosynthesis (Lü et al. 2013). Given the established negative correlation between cruciferin content and glucosinolate content (Malabat et al. 2003; Schatzki et al. 2014), it may be that this gene also indirectly affects cruciferin content. The gene *BnaC02g20660* is orthologous to the *A. thaliana* gene *LYSINE-HISTIDINE LIKE TRANSPORTER 5* (*LHT5). LHT5* is a member of the amino acid/auxin permease family, and is involved in transporting amino acids into the cell (Tegeder and Ward 2012). This gene is expressed in male and female floral tissues, and is involved in sexual reproduction (Foster et al. 2008). However, little is known about this gene, and it may be expressed elsewhere and have other functions (Tegeder and Ward 2012). The presence of ABA has been shown to be an important contributor to seed storage protein gene expression during seed development (DeLisle and Crouch 1989). To this end, the candidate gene *BnaC02g0991D* was found to be orthologous to the *ARABIDOPSIS THALIANA RING AND DOMAIN OF UNKNOWN FUNCTION 1117 2* (*AtRDUF2*) gene. Kim et al. (2012) found that *AtRDUF2* plays a role in ABA-dependant responses to drought stress, by acting as a positive regulator of ABA response. Similarly, *BnaC0210050D* and *BnaC02g10060D* are orthologous to the *A. thaliana* gene *LIPID TRANSFER PROTEIN 4* (*LTP4*), which is strongly upregulated by ABA (Arondel et al. 2000). In sesame, *LTP4* is expressed in the seed-pod walls, and it has been hypothesized that it contributes to the plant's ability to adapt to drought conditions during seed development (Choi et al. 2008).

Along with the study discussed above by Schatzki et al. (2014), the current study is one of few that investigate the control of cruciferin content and aim to identify loci associated with this trait in *B. napus.* While this study identified significant SNP-trait associations and candidate genes with functions relating to the control of cruciferin content, it is important to note that the size of the population used for association mapping in this study was small. Genome-wide association studies often use population sizes of over 300 individuals (Soumya et al. 2021). In *B. napus* studies,

population size has ranged from 66 (Zhang et al. 2015) up to 950 individuals (Jan et al. 2019). By using a small sample size in GWAS, there is a decrease in statistical power to find meaningful SNP-trait associations (Korte and Farlow 2013), and an increased risk in the presence of false-positive associations (Finno et al. 2014). In future studies, it will be valuable to use a larger mapping population to improve the ability to discover significant SNP-trait associations. In addition to performing studies with larger mapping populations, the validation of candidate genes identified in this study is also essential. In the identification of SNP-trait associations, GWAS only gives indirect, statistical evidence for the function of a genome region (Stich and Melchinger 2010). For this reason, validation of candidate genes through GWAS in a different population or by performing candidate gene knockout or overexpression experiments is a necessary follow-up to any association study (Liu and Yan 2019).

By conducting a genome-wide association study, this research was able to identify 47 loci that were significantly associated with cruciferin content, and 11 candidate genes that may contribute to the control of cruciferin content in *B. napus*. As interest in plant-based proteins for human consumption continues to rise, alternative plant-based proteins such as cruciferin will become increasingly important. This study provides a framework for further studies that will broaden the understanding of the genetic control of cruciferin content, and aid in the improvement of *B. napus* cultivars with altered seed storage protein profiles.

6. General Discussion and Conclusions

Canola (*Brassica napus* L.), is the second highest produced oilseed crop worldwide (FAO 2019), and is considered Canada's most valuable crop (Wells and Slade 2021). To ensure the Canadian canola industry continues to remain competitive, adding value to this important crop and exploring end-use diversification will be critical. A well-balanced amino acid profile and desirable functional properties make *B. napus* meal protein a suitable candidate for plant-based protein for human consumption (Chmielewska et al. 2020). Due to its relative abundance, the seed storage protein cruciferin is of particular interest. One of the challenges facing efforts to enhance *B. napus* meal suitability for human consumption is a lack of information regarding existing phenotypic variation and the genetic control of cruciferin content in *B. napus*. The research described in this thesis provides valuable insight into this topic.

In Chapter 3, the need for a deeper understanding of the existing phenotypic variation in cruciferin content and the role of the environment on cruciferin content was addressed. To facilitate the quantification of cruciferin protein, an enzyme-linked immunosorbent assay (ELISA)-based approach was developed. Considerable variation in cruciferin content was observed amongst the experimental genotypes. The variation observed in this research is supported by previous studies demonstrating significant variation in cruciferin content in both canola and winter oilseed rape populations (Malabat et al. 2003; Wanasundara 2011; Schatzki et al. 2014). In addition, this research demonstrated that the effects of genotype, site-year and genotype by site-year interactions all significantly affect cruciferin content. However, several genotypes were identified that were consistently low or high in cruciferin content across site years, and thus may be candidates for inclusion in future breeding research. In accordance with previous studies that demonstrate a

negative correlation between cruciferin content and glucosinolate content (Malabat et al. 2003; Schatzki et al. 2014), this research demonstrated that it will be possible to increase cruciferin content while not sacrificing the improvements made in *B. napus* glucosinolate levels in the meal. It is important to consider that the primary use of *B. napus* is as an oilseed (Rahman 2013; Campbell et al. 2016), and thus any efforts to alter seed storage protein profiles must not do so at the expense of oil content or quality.

While Chapter 3 used an ELISA-based approach to cruciferin content determination, Chapter 4 focused on the need for an easily adopted, non-destructive method to allow for more efficient quantification of cruciferin in *B. napus* breeding programs. While the ELISA-based method used in Chapter 3 has strengths when compared to the cruciferin quantification methods described in other studies (Wanasundara and McIntosh 2013; Cheung et al. 2014; Schatzki et al. 2014; Akbari and Wu 2015; Perera et al. 2016), it still results in the destruction of the seed sample, and is more time consuming when compared to near-infrared spectroscopy (NIRS). Using the ELISA-based reference method, this study developed three NIRS calibration equations for cruciferin content in whole seeds of *B. napus*. Statistical analysis of these calibration equations showed that they were unable to accurately predict cruciferin content. Efforts to develop NIRS calibration equations for specific seed storage protein fractions in soybean have had limited (Pazdernik et al. 1996) or no success (Delwiche et al. 2007). While this may indicate that the NIRS method lacks the capacity to detect spectra specific to individual seed storage protein fractions, as has been suggested (Pazdernik et al. 1996; Delwiche et al. 2007), improvements to the reference method and increasing the size of the population used for calibration may be points of interest for future studies. This study was unsuccessful in developing an efficient, non-destructive quantification method for cruciferin content. However, large-scale breeding efforts to develop genotypes with altered seed

storage protein profiles will require further exploration of methods to determine cruciferin content that fit these criteria.

To complement the understanding of phenotypic variation in cruciferin content established in Chapter 3, Chapter 5 used a genome-wide association study (GWAS) to investigate the genetic control of cruciferin content within the same population of 51 *B. napus* genotypes. STRUCTURE and principle coordinates analysis (PCoA) indicated the existence of two sub-populations among the 51 genotypes, which were separated primarily on the basis of geographical origin. Along with morphology and growth habit (Gyawali et al. 2013, 2016; Gazave et al. 2016; Raman et al. 2016; Schiessl et al. 2017), geographical origin is a known contributor to sub-population separation in *B. napus* (Xiao et al. 2012; Gazave et al. 2016). While a single previous quantitative trait loci (QTL) mapping study was successful in identifying QTL for cruciferin content in a population of winter oilseed rape (Schatzki et al. 2014), the research in Chapter 5 was the first to explore the genetic control of cruciferin content in spring-type *B. napus*. The population used in this study is much smaller than is typical for genome-wide association studies (Soumya et al. 2021), and this likely resulted in the presence of false-positives and a decrease in statistical power. Despite this, a total of 47 loci significantly associated with cruciferin content were identified, located in both sub-genome A (chromosomes A01, A02, A03, A04, A08 and A09) and sub-genome C (chromosomes C01, C02, C03, C05, C06 and C09). In their study, Schatzki et al. (2014) discovered two QTL for cruciferin content, located on chromosomes A02 and C09. While the confidence interval region for the QTL on A02 includes all of the significant SNPs on chromosome A02 in the current study, the QTL on C09 found by Schatzki et al. (2014) is not consistent with the results of this research. In addition, 11 genes were identified as potential candidates responsible for the control of cruciferin content, including a gene orthologous to the *Arabidopsis thaliana FUS3* transcription

factor, which plays an important role in regulating seed storage protein accumulation in *A. thaliana* and *B. napus* (Baumienin et al. 1994; Gacek et al. 2018). This study was able to provide valuable preliminary information regarding the genetic control of cruciferin content, which will be important to future studies.

This research addressed the need for greater knowledge surrounding *B. napus* cruciferin proteins from a plant breeding point of view. While efforts to develop an efficient NIRS method to quantify cruciferin protein were unsuccessful, studies evaluating phenotypic variation and the genetic control of cruciferin content were informative. These findings will contribute to future studies as well as breeding efforts that will ensure *B. napus* continues to be a valuable and important crop in Canada.

7. FUTURE RESEARCH RECOMMENDATIONS

This study provided a framework of information to support innovative efforts to breed canola (*Brassica napus* L.) genotypes with altered seed storage protein profiles. Therefore, further research will be valuable to build upon these results and validate the study findings.

Current canola end use targets rely on maximizing oil content and quality and minimizing undesirable compounds in meal. As such, these traits have largely been the focus of breeding efforts to date (McVetty and Scarth 2002; Malabat et al. 2003; Schatzki et al. 2014). While efforts to alter seed storage protein profiles have the potential to add value, they must not come at the cost of established traits of importance. While the current study provided insight into the correlation of cruciferin content with a number of seed quality traits, further correlation analysis will be valuable to assist in directing breeding efforts.

This study developed an enzyme linked immunosorbent assay (ELISA)-based method by which to quantify cruciferin protein that was able to use small seed amounts to assess samples with relative efficiency. However, sources of error within the ELISA-based method such as sensitivities to temperature and time (Crowther 2009; Gan and Patel 2013) may result in some loss of precision and accuracy. Thus, enquiry into alternate methods to quantify cruciferin protein, such as the size-exclusion high performance liquid chromatography (SE-HPLC) method described by Defaix et al. (2019), may be useful.

Efforts to develop a near-infrared reflectance spectroscopy (NIRS) calibration model for cruciferin content were unsuccessful in this study. While Delwiche et al. (2007) and Pazdernik et al. (1996) propose that NIRS may lack the capacity to detect spectra specific to individual seed storage

protein fractions, it may be valuable to attempt calibration development with an alternate reference method and a larger calibration set, as well as ground seed samples rather than whole seeds.

The genome-wide association study (GWAS) in this research was carried out on a very small population. By using a small sample size in GWAS, there is a decrease in statistical power to find meaningful SNP-trait associations (Korte and Farlow 2013), and an increased risk in the presence of false-positive associations (Finno et al. 2014). Thus, to further elucidate the genetic control of cruciferin content and validate the results of this study, a genome-wide association study using a larger population should be conducted. In addition, validation of the candidate genes proposed in this study should be conducted by performing candidate gene knockout or overexpression experiments.

To further assist efforts to introduce *B. napus* as a competitive source of plant-based protein, similar research needs to be conducted on the other *B. napus* seed storage protein of importance, napin. The results from this research and supplementation and validation by future research will lay the foundation for breeding efforts to develop *B. napus* genotypes with alternate seed storage protein profiles.

8. REFERENCE MATTER

8.1 Literature Cited

Acton, D.F., and Ellis, J.G. 1978. The soils of the Saskatoon map area 73-B Saskatchewan. Saskatoon, SK.

Agelet, L.E., and Hurburgh, C.R. 2010. A tutorial on near infrared spectroscopy and its calibration. Crit. Rev. Anal. Chem. **40**: 246–260. doi:10.1080/10408347.2010.515468.

Aider, M., and Barbana, C. 2011. Canola proteins: composition, extraction, functional properties, bioactivity, applications as a food ingredient and allergenicity - a practical and critical review. Trends Food Sci. Technol. **22**: 21–39. doi:10.1016/j.tifs.2010.11.002.

Akbari, A., and Wu, J. 2015. An integrated method of isolating napin and cruciferin from defatted canola meal. LWT - Food Sci. Technol. **64**: 308–315. doi:10.1016/j.lwt.2015.05.046.

Al-Shehbaz, I.A., Beilstein, M.A., and Kellogg, E.A. 2006. Systematics and phylogeny of the Brassicaceae (Cruciferae): an overview. Plant Syst. Evol. **259**: 89–120. doi:10.1007/s00606-006-0415-z.

Albe Slabi, S., Mathé, C., Framboisier, X., Defaix, C., Mesieres, O., Galet, O., and Kapel, R. 2019. A new SE-HPLC method for simultaneous quantification of proteins and main phenolic compounds from sunflower meal aqueous extracts. Anal. Bioanal. Chem. **411**: 2089–2099. doi:10.1007/s00216-019-01635-2.

Aluko, R.E., and McIntosh, T. 2001. Polypeptide profile and functional properties of defatted meals and protein isolates of canola seeds. J. Sci. Food Agric. **81**: 391–396. doi:10.1002/1097-0010(200103)81:4<391::AID-JSFA823>3.0.CO;2-S.

Angeles-Núñez, J.G., and Tiessen, A. 2012. Regulation of *AtSUS2* and *AtSUS3* by glucose and

the transcription factor LEC2 in different tissues and at different stages of Arabidopsis seed development. Plant Mol. Biol. **78**: 377–392. doi:10.1007/s11103-011-9871-0.

Annicchiarico, P. 2002. Genotype x environment interactions: Challenges and opportunities for plant breeding and cultivar recommendations. Food and Agriculture Organization of the United Nations. [Online] Available: http://www.fao.org/3/y4391e/y4391e00.htm [2019 Oct. 24].

AOAC 2005. AOAC Official Method 990.03: Protein (crude) in animal feed, combustion method. Page *in* D.G.. Latimer and D.W. Horwitz, eds. Official methods of analysis of AOAC International, 18th edition. AOAC International, Rockville, MD.

Arntfield, S., and Hickling, D. 2011. Meal Nutrition and Utilization. Pages 281–312 *in* J.K. Daun, M.N.A. Eskin, and D. Hickling, eds. Canola: chemistry, production, processing, and utilization. AOCS Press, Urbana, IL.

Arntfield, S.D. 2011. Canola and other oilseed proteins. Pages 289–315 *in* G.O. Phillips and P.A. Williams, eds. Handbook of food proteins. Woodhead Publishing Limited, Cambridge, UK.

Arntfield, S.D. 2017. Proteins from oil-producing plants. Pages 187–221 *in* R.Y. Yada, ed. Proteins in food processing: second edition. Elsevier, Duxford, UK.

Arondel, V., Vergnolle, C., Cantrel, C., and Kader, J.C. 2000. Lipid transfer proteins are encoded by a small multigene family in *Arabidopsis thaliana*. Plant Sci. **157**: 1–12. doi:10.1016/S0168-9452(00)00232-6.

Aukema, H., and Campbell, L. 2011. Oil nutrition and utilization. Pages 245–280 *in* J.K. Daun, M.N.A. Eskin, and D. Hickling, eds. Canola: chemistry, production, processing, and utilization. AOCS Press, Urbana, IL.

Badshah, A., Zeb, A., Bibi, N., Ali, S., Chaudry, M.A., and Sattar, A. 2001. Utilization of

rapeseed meal/cake in poultry feed. Part II. Effect of incorporating higher levels of rapeseed cake in poultry diet on laying performance of brown-egg layer (Tetra SL). Pak. J. Sci. Ind. Res. **44**: 171–174.

Bar-Hen, A., Charcosset, A., Bourgoin, M., and Guiard, J. 1995. Relationship between genetic markers and morphological traits in a maize inbred lines collection. Euphytica, **84**: 145–154. doi:10.1007/BF01677953.

Barthet, V.J. 2018. Quality of western Canadian canola 2018. Winnipeg, MB.

Barthet, V.J., and Daun, J.K. 2011. Seed morphology, composition, and quality. Pages 119–162 *in* J.K. Daun, N.A.M. Eskin, and D. Hickling, eds. Canola: chemistry, production, processing and utilization. AOSC Press, Urbana, IL.

Basnet, R.K., Moreno-Pachon, N., Lin, K., Bucher, J., Visser, R.G.F., Maliepaard, C., and Bonnema, G. 2013. Genome-wide analysis of coordinated transcript abundance during seed development in different *Brassica rapa* morphotypes. BMC Genomics, **14**: 840. doi:10.1186/1471-2164-14-840.

Baumienin, H., Misera, S., LeurBen, H., Kolle, K., Horstmann, C., Worbus, U., and Muller, A.J. 1994. The *FUS3* gene of *Arabidopsis thaliana* is a regulator of gene expression during late embryogenesis. Plant J. **6**: 379–387.

Bayer, P.E., Hurgobin, B., Golicz, A.A., Chan, C.K.K., Yuan, Y., Lee, H.T., Renton, M., Meng, J., Li, R., Long, Y., Zou, J., Bancroft, I., Chalhoub, B., King, G.J., Batley, J., and Edwards, D. 2017. Assembly and comparison of two closely related *Brassica napus* genomes. Plant Biotechnol. J. **15**: 1602–1610. doi:10.1111/pbi.12742.

Bell, J.M. 1995. Meal and by-product utilization in animal nutrition. Pages 301–337 *in* D.S. Kimber and D.I. McGregor, eds. *Brassica* oilseeds production and utilization. CAB

International, Oxon, UK.

Bernardo, R. 2014. Essentials of plant breeding. Stemma Press, Woodbury (MN).

Bernatzky, R., and Tanksley, S.D. 1986. Toward a saturated linkage map in tomato based on isozymes and random cDNA sequences. Genetics, **112**: 887–98. doi:10.1007/978-1-4684-4556-5_4.

Bérot, S., Compoint, J.P., Larré, C., Malabat, C., and Guéguen, J. 2005. Large scale purification of rapeseed proteins (*Brassica napus* L.). J. Chromatogr. B **818**: 35–42. doi:10.1016/j.jchromb.2004.08.001.

Bhatty, R.S. 1964. Influence of nitrogen fertilization on the yield, protein, and oil content of two varieties of rape. Can. J. Plant Sci. **44**: 215–217.

Borisjuk, L., Neuberger, T., Schwender, J., Heinzel, N., Sunderhaus, S., Fuchs, J., Hay, J.O., Tschiersch, H., Braun, H.P., Denolf, P., Lambert, B., Jakob, P.M., and Rolletschek, H. 2013. Seed architecture shapes embryo metabolism in oilseed rape. Plant Cell, **25**: 1625–1640. doi:10.1105/tpc.113.111740.

Bos, C., Airinei, G., Mariotti, F., Benamouzig, R., Bérot, S., Evrard, J., Fénart, E., Tomé, D., and Gaudichon, C. 2007. The poor digestibility of rapeseed protein is balanced by its very high metabolic utilization in humans. J. Nutr. **137**: 594–600. doi:10.1093/jn/137.3.594.

Botstein, D., White, R.L., Skolnick, M., and Davis, R.W. 1980. Construction of a genetic linkage map in Man using restriction fragment length polymorphisms. Am. J. Hum. Genet. **32**: 314–331.

Boulter, G.S. 1983. The history and marketing of rapeseed oil in Canada. Pages 61–83 *in* K. J.K.G, F.D. Sauer, and W.J. Pigden, eds. High and low erucic acid rapeseed oils. Academic Press Canada, Don Mills, ON.

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. doi:10.1093/bioinformatics/btm308.

Bradford, M.M. 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal. Biochem. **72**: 248–254.

Breen, J.P., and Crouch, M.L. 1992. Molecular analysis of a cruciferin storage protein gene family of *Brassica napus*. Plant Mol. Biol. **19**: 1049–1055. doi:10.1007/BF00040536.

Breseghello, F., and Sorrells, M.E. 2006. Association analysis as a strategy for improvement of quantitative traits in plants. Crop Sci. **46**: 1323–1330. doi:10.2135/cropsci2005.09-0305.

Burcon NutraScience Corporation 2019. Burcon introduces Nutratein® plant protein ingredients – new pea + canola protein blends provide protein quality equivalent to or exceeding dairy and meat. [Online] Available: https://www.burcon.ca/2019/05/burcon-introduces-nutratein-plant-protein-ingredients-new-pea-canola-protein-blends-provide-protein-quality-equivalent-to-or-exceeding-dairy-and-meat/ [2019 Nov. 27].

Burcon NutraScience Corporation 2021a. Puratein. [Online] Available: https://www.burcon.ca/products/canola-proteins/puratein/ [2021 Jun. 24].

Burcon NutraScience Corporation 2021b. Supertein. [Online] Available: https://www.burcon.ca/products/canola-proteins/supertein/ [2021 Jun. 24].

Burcon NutraScience Corporation 2021c. Canola proteins. [Online] Available: https://burcon.ca/products/canola-proteins/ [2021 Jun. 24].

Campbell, L., Rempel, C., and Wanasundara, J. 2016. Canola/rapeseed protein: Future opportunities and directions—Workshop proceedings of IRC 2015. Plants **5**: 1–7. doi:10.3390/plants5020017.

Canola Council of Canada 2015. Canola meal feeding guide, 5th edition. [Online] Available: www.canolacouncil.org, [2019 Nov. 5].

Canola Council of Canada 2016. What Is RUP and what does it mean for my dairy ration? [Online] Available: http://canolamazing.com/what-is-rup-and-what-does-it-mean-for-my-ration/ [2019 Nov. 26].

Canola Council of Canada 2019a. What is canola? [Online] Available: https://www.canolacouncil.org/oil-and-meal/what-is-canola/ [2019 Nov. 4].

Canola Council of Canada 2019b. Industry overview. [Online] Available: https://www.canolacouncil.org/markets-stats/industry-overview/ [2019 Nov. 3].

Chagné, D., Bianco, L., Lawley, C., Micheletti, D., and Jacobs, J.M.E. 2015. Methods for the design, implementation, and analysis of Illumina Infinium SNP assays in plants. Pages 281–298 *in* J. Batley, ed. Plant genotyping: methods and protocols. Springer, New York, NY.

Chalhoub, B., Denoeud, F., Liu, S., Parkin, I.A.P., Tang, H., Wang, X., Chiquet, J., Belcram, H., Tong, C., Samans, B., Corréa, 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.D., Chalabi, S., Hu, Q., Fan, C., Tollenaere, R., Lu, Y., Battail, C., Shen, J., Sidebottom, C.H.D., Wang, X., Canaguier, A., Chauveau, A., Bérard, 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., Adams, K.L., Batley, J., Snowdon, R.J., Tost, J., Edwards, D., Zhou, Y., Hua, W., Sharpe, A.G., Paterson, A.H., Guan, C., and Wincker, P.

2014. Early allopolyploid evolution in the post-neolithic *Brassica napus* oilseed genome. Science, **345**: 950–953. doi:10.1126/science.1253435.

Chao, H., Wang, H., Wang, X., Guo, L., Gu, J., Zhao, W., Li, B., Chen, D., Raboanatahiry, N., and Li, M. 2017. Genetic dissection of seed oil and protein content and identification of networks associated with oil content in *Brassica napus*. Sci. Rep. **7**: 46295. doi:10.1038/srep46295.

Chawade, A., Van Ham, J., Blomquist, H., Bagge, O., Alexandersson, E., and Ortiz, R. 2019. High-throughput field-phenotyping tools for plant breeding and precision agriculture. Agronomy, **9**: 1–18. doi:10.3390/agronomy9050258.

Chen, G.L., Zhang, B., Wu, J.G., and Shi, C.H. 2011. Nondestructive assessment of amino acid composition in rapeseed meal based on intact seeds by near-infrared reflectance spectroscopy. Anim. Feed Sci. Technol. **165**: 111–119. doi:10.1016/j.anifeedsci.2011.02.004.

Chen, J., Wang, J., Song, P., and Ma, X. 2014. Determination of glycinin in soybean and soybean products using a sandwich enzyme-linked immunosorbent assay. Food Chem. **162**: 27–33. doi:10.1016/j.foodchem.2014.04.065.

Chen, L., Wan, H., Qian, J., Guo, J., Sun, C., Wen, J., Yi, B., Ma, C., Tu, J., Song, L., Fu, T., and Shen, J. 2018. Genome-wide association study of cadmium accumulation at the seedling stage in rapeseed (*Brassica napus* L.). Front. Plant Sci. **9**: 1–15. doi:10.3389/fpls.2018.00375.

Cheng, F., Wu, J., and Wang, X. 2014. Genome triplication drove the diversification of *Brassica* plants. Hortic. Res. **1**: 14024. doi:10.1007/978-3-662-47901-8_10.

Cheung, L., Wanasundara, J., and Nickerson, M.T. 2014. The effect of pH and NaCl levels on

the physicochemical and emulsifying properties of a cruciferin protein isolate. Food Biophys. **9**: 105–113. doi:10.1007/s11483-013-9323-2.

Cheung, L., Wanasundara, J., and Nickerson, M.T. 2015. Effect of pH and NaCl on the emulsifying properties of a napin protein isolate. Food Biophys. **10**: 30–38. doi:10.1007/s11483-014-9350-7.

Chiou, S.H., and Wu, S.H. 1999. Evaluation of commonly used electrophoretic methods for the analysis of proteins and peptides and their application to biotechnology. Anal. Chim. Acta **383**: 47–60. doi:10.1016/S0003-2670(98)00487-5.

Chmielewska, A., Kozłowska, M., Rachwał, D., Wnukowski, P., Amarowicz, R., Nebesny, E., and Rosicka-Kaczmarek, J. 2020. Canola/rapeseed protein–nutritional value, functionality and food application: a review. Crit. Rev. Food Sci. Nutr. **0**: 1–21. doi:10.1080/10408398.2020.1809342.

Chodak, M. 2008. Application of near infrared spectroscopy for analysis of soils, litter and plant materials. Polish J. Environ. Stud. **17**: 631–642.

Choi, A.M., Lee, S.B., Cho, S.H., Hwang, I., Hur, C.G., and Suh, M.C. 2008. Isolation and characterization of multiple abundant lipid transfer protein isoforms in developing sesame (*Sesamum indicum* L.) seeds. Plant Physiol. Biochem. **46**: 127–139. doi:10.1016/j.plaphy.2007.10.003.

Clarke, W.E., Higgins, E.E., Plieske, J., Wieseke, R., Sidebottom, C., Khedikar, Y., Batley, J., Edwards, D., Meng, J., Li, R., Lawley, C.T., Pauquet, J., Laga, B., Cheung, W., Iniguez-Luy, F., Dyrszka, E., Rae, S., Stich, B., Snowdon, R.J., Sharpe, A.G., Ganal, M.W., and Parkin, I.A.P. 2016. A high-density SNP genotyping array for *Brassica napus* and its ancestral diploid species based on optimised selection of single-locus markers in the

allotetraploid genome. Theor. Appl. Genet. **129**: 1887–1899. doi:10.1007/s00122-016-2746-7.

Conesa, A., Götz, S., García-Gómez, J.M., Terol, J., Talón, M., and Robles, M. 2005. Blast2GO: A universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics, **21**: 3674–3676. doi:10.1093/bioinformatics/bti610.

Crouch, M.L., and Sussex, I.M. 1981. Development and storage-protein synthesis in *Brassica napus* L. embryos in vivo and in vitro. Planta, **153**: 64–74. doi:10.1007/BF00385319.

Crowther, J.R. 2009. The ELISA guidebook. 2nd edition. Humana Press, New York, NY.

Daun, J.K. 1976. A rapid procedure for the determination of chlorophyll in rapeseed by reflectance spectroscopy. J. Am. Oil Chem. Soc. **53**: 767–770. doi:10.1007/BF02635481.

Daun, J.K. 1983. The introduction of low erucic acid rapeseed varieties into Canadian production. Pages 162–180 *in* J.K.. Kramer, F.D. Sauer, and W.J. Pigden, eds. High and low erucic acid rapeseed oils. Academic Press Canada, Don Mills, ON.

Daun, J.K. 2011. Origin, distribution, and production. Pages 1–27 *in* J.K. Daun, M.N.A. Eskin, and D. Hickling, eds. Canola: chemistry, production, processing and utilization. AOCS Press, Urbana, IL.

Daun, J.K., Clear, K.M., and Williams, P. 1994. Comparison of three whole seed near-infrared analyzers for measuring quality components of canola seed. J. Am. Oil Chem. Soc. **71**: 1063–1068. doi:10.1007/BF02675897.

Daun, J.K., and DeClercq, D.R. 1995. Interrelationships between quality factors and yield in Canadian canola from harvest surveys, 1956 to 1994. Pages 336–338 *in* Proceedings of the 9th International Rapessed Congress. Cambridge, UK.

Day, L. 2013. Proteins from land plants – Potential resources for human nutrition and food

security. Trends Food Sci. Technol. **32**: 25–42. doi:10.1016/j.tifs.2013.05.005.

Defaix, C., Aymes, A., Albe Slabi, S., Basselin, M., Mathé, C., Galet, O., and Kapel, R. 2019. A new size-exclusion chromatography method for fast rapeseed albumin and globulin quantification. Food Chem. **287**: 151–159. doi:10.1016/j.foodchem.2019.01.209.

DeLisle, A.J., and Crouch, M.L. 1989. Seed storage protein transcription and mRNA levels in *Brassica napus* during development and in response to exogenous abscisic acid. Plant Physiol. **91**: 617–623. doi:10.1104/pp.91.2.617.

Delourme, R., Falentin, C., Fomeju, B.F., Boillot, M., Lassalle, G., André, I., Duarte, J., Gauthier, V., Lucante, N., Marty, A., Pauchon, M., Pichon, J.P., Ribière, N., Trotoux, G., Blanchard, P., Rivière, N., Martinant, J.P., and Pauquet, J. 2013. High-density SNP-based genetic map development and linkage disequilibrium assessment in *Brassica napus* L. BMC Genomics, **14**: 1–18. doi:10.1186/1471-2164-14-120.

Delwiche, S.R., Pordesimo, L.O., Panthee, D.R., and Pantalone, V.R. 2007. Assessing glycinin (11S) and β-conglycinin (7S) fractions of soybean storage protein by near-infrared spectroscopy. J. Am. Oil Chem. Soc. **84**: 1107–1115. doi:10.1007/s11746-007-1144-3.

Dimov, Z., Suprianto, E., Hermann, F., and Mollers, C. 2012. Genetic variation for seed hull and fibre content in a collection of European winter oilseed rape material (*Brassica napus* L.) and development of NIRS calibrations. Plant Breed. **131**: 361–368. doi:10.1111/j.1439-0523.2012.01951.x.

Diosady, L.L., Xu, L., and Chen, B.-K. 2005. Production of high-quality protein isolates from defatted meals of *Brassica* seeds. US Patent 6,905,713 B2.

Downey, R.K. 2007. Rapeseed to canola: rags to riches. Pages 67–76 *in* A. Eaglehan and R.W.F. Hardy, eds. Agricultural biotechnology: economic growth through new products,

partnerships and workforce development. National Agricultural Biotechnology Council, Ithica, NY.

Downey, R.K., and Rakow, G. 1987. Rapeseed and mustard. Page *in* W.R. Fehr, E.L. Fehr, and H.J. Jessen, eds. Principles of cultivar development, Vol 2. Macmillan, New York, NY.

Duncan, R.W., McVetty, P.B.E., Duguid, J.L., Fox, S.L., Paulmann, W., Sass, O., Fernando, W.G.D., and Li, G. 2016. HYHEAR 2 Roundup Ready® high erucic acid, low glucosinolate hybrid summer rape. Can. J. Plant Sci. **96**: 711–713. doi:10.1139/cjps-2015-0329.

Duncan, R.W., McVetty, P.B.E., Duguid, J.L., Fox, S.L., Paulmann, W., Sass, O., Fernando, W.G.D., and Li, G. 2017. HYHEAR 3 Roundup Ready®, high erucic acid, low glucosinolate hybrid summer rape. Can. J. Plant Sci. **97**: 1185–1187. doi:10.1139/cjps-2016-0229.

Earl, D.A., and VonHoldt, B.M. 2012. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv. Genet. Resour. **4**: 359–361.

Ebine, K., Okatani, Y., Uemura, T., Goh, T., KeikoShoda, Niihama, M., Morita, M.T., Spitzer, C., Otegui, M.S., Nakano, A., and Ueda, T. 2008. A SNARE complex unique to seed plants is required for protein storage vacuole biogenesis and seed development of *Arabidopsis thaliana*. Plant Cell, **20**: 3006–3021. doi:10.1105/tpc.107.057711.

Ecke, W., Clemens, R., Honsdorf, N., and Becker, H.C. 2010. Extent and structure of linkage disequilibrium in canola quality winter rapeseed (*Brassica napus* L.). Theor. Appl. Genet. **120**: 921–931. doi:10.1007/s00122-009-1221-0.

Edwards, D., Forster, J.W., Chagné, D., and Batley, J. 2007. What are SNPs? Pages 41–52 *in*

N.C. Oraguzie, E.H.A. Rikkerink, S.E. Gardiner, and H.N. De Silva, eds. Association mapping in plants. Springer, New York, NY.

Ehrlich, W.A. 1953. Report of reconnaissance soil survey of Winnipeg and Morris map sheet areas. Winnipeg, MB.

Enami, H.R. 2011. A review of using canola/rapeseed meal in aquaculture feeding. J. Fish. Aquat. Sci. **6**: 22–36. doi:10.3923/jfas.2011.22.36.

Engvall, E., and Perlmann, P. 1971. Enzyme-linked immunosorbent assay (ELISA) quantitative assay of immunoglobulin G. Immunochemistry, **8**: 871–874.

Evanno, G., Regnaut, S., and Goudet, J. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. Mol. Ecol. **14**: 2611–2620. doi:10.1111/j.1365-294X.2005.02553.x.

Fallahi, H., Scofield, G.N., Badger, M.R., Chow, W.S., Furbank, R.T., and Ruan, Y.L. 2008. Localization of sucrose synthase in developing seed and siliques of *Arabidopsis thaliana* reveals diverse roles for SUS during development. J. Exp. Bot. **59**: 3283–3295. doi:10.1093/jxb/ern180.

FAO 2011. Dietary protein quality evaluation in human nutrition. FAO Food and Nutrition Paper 92. Rome, IT.

FAO 2019. Food outlook - biannual report on global food markets. Food and Agriculture Organization of the United Nations, Rome, IT.

Finno, C.J., Aleman, M., Higgins, R.J., Madigan, J.E., and Bannasch, D.L. 2014. Risk of false positive genetic associations in complex traits with underlying population structure: A case study. Vet. J. **202**: 543–549. doi:10.1016/j.tvjl.2014.09.013.

Font, R., del Río-Celestino, M., and de Haro-Bailón, A. 2006. The use of near-infrared

spectroscopy (NIRS) in the study of seed quality components in plant breeding programs. Ind. Crops Prod. **24**: 307–313. doi:10.1016/j.indcrop.2006.06.012.

Font, R., Del Río, M., Fernández, J.M., and De Haro, A. 2003. Acid detergent fiber analysis in oilseed *Brassicas* by near-infrared spectroscopy. J. Agric. Food Chem. **51**: 2917–2922. doi:10.1021/jf021011+.

Font, R., Wittkop, B., Badani, A.G., Del Río-Celestino, M., Friedt, W., Lühs, W., and De Haro-Bailón, A. 2005. The measurements of acid detergent fibre in rapeseed by visible and near-infrared spectroscopy. Plant Breed. **124**: 410–412. doi:10.1111/j.1439-0523.2005.01138.x.

Foster, J., Lee, Y.H., and Tegeder, M. 2008. Distinct expression of members of the LHT amino acid transporter family in flowers indicates specific roles in plant reproduction. Sex. Plant Reprod. **21**: 143–152. doi:10.1007/s00497-008-0074-z.

Fredes, I., Moreno, S., Díaz, F.P., and Gutiérrez, R.A. 2019. Nitrate signaling and the control of Arabidopsis growth and development. Curr. Opin. Plant Biol. **47**: 112–118. doi:10.1016/j.pbi.2018.10.004.

Fredua-Agyeman, R., Yu, Z., Hwang, S.F., and Strelkov, S.E. 2020. Genome-wide mapping of loci associated with resistance to clubroot in *Brassica napus* ssp. *napobrassica* (rutabaga) accessions from Nordic countries. Front. Plant Sci. **11**: 1–17. doi:10.3389/fpls.2020.00742.

Fu, Y., Zhang, D., Gleeson, M., Zhang, Y., Lin, B., Hua, S., Ding, H., Frauen, M., Li, J., Qian, W., and Yu, H. 2017. Analysis of QTL for seed oil content in *Brassica napus* by association mapping and QTL mapping. Euphytica, **213**: 1–15. doi:10.1007/s10681-016-1817-9.

Fujiwara, T., Nambara, E., Yamagishi, K., Goto, D.B., and Naito, S. 2002. Storage proteins. Arab. B. **2002**: e0020. doi:10.1199/tab.0020.

Furbank, R.T., and Tester, M. 2011. Phenomics: technologies to relieve the phenotyping

bottleneck. Trends Plant Sci. **16**: 635–644. doi:10.1016/j.tplants.2011.09.005.

Gabriel, S.B., Schaffner, S.F., Nguyen, H., Moore, J.M., Blumenstiel, B., Higgins, J., Defelice, M., Lochner, A., Faggart, M., Liu-cordero, S.N., Rotimi, C., Adeyemo, A., Cooper, R., Ward, R., Lander, E.S., Daly, M.J., Altshuler, D., Gabriel, S.B., Schaffner, S.F., Nguyen, H., Moore, J.M., Roy, J., Blumenstiel, B., Higgins, J., Defelice, M., Lochner, A., Faggart, M., Liu-cordero, S.N., Rotimi, C., Adeyemo, A., Cooper, R., Ward, R., Lander, E.S., Daly, M.J., and Altshulerl-, D. 2002. The structure of haplotype blocks in the human genome. Science, **296**: 2225–2229.

Gacek, K., Bartkowiak-Broda, I., and Batley, J. 2018. Genetic and molecular regulation of seed storage proteins (SSPs) to improve protein nutritional value of oilseed rape (*Brassica napus* L.) seeds. Front. Plant Sci. **9**: 1–9. doi:10.3389/fpls.2018.00890.

Gali, K.K., and Sharpe, A.G. 2011. Molecular linkage maps: strategies, resources and achievements. Pages 85–129 *in* D. Edwards, J. Bately, I. Parkin, and C. Kole, eds. Genetics, genomics and breeding of oilseed *Brassicas*. CRC Press Taylor & Francis Group, Boca Raton, FL.

Gan, S.D., and Patel, K.R. 2013. Enzyme immunoassay and enzyme-linked immunosorbent assay. J. Invest. Dermatol. **133**: 1–3. doi:10.1038/jid.2013.287.

Gazave, E., Tassone, E.E., Ilut, D.C., Wingerson, M., Datema, E., Witsenboer, H.M.A., Davis, J.B., Grant, D., Dyer, J.M., Jenks, M.A., Brown, J., and Gore, M.A. 2016. Population genomic analysis reveals differential evolutionary histories and patterns of diversity across subgenomes and subpopulations of *Brassica napus* L. Front. Plant Sci. **7**: 1–16. doi:10.3389/fpls.2016.00525.

Gehrig, P.M., Krzyzaniak, A., Barciszewski, J., and Biemann, K. 1996. Mass spectrometric

amino acid sequencing of a mixture of seed storage proteins (napin) from *Brassica napus*, products of a multigene family. Proc. Natl. Acad. Sci. U. S. A. **93**: 3647–3652.

Ghodsvali, A., Khodaparast, M.H.H., Vosoughi, M., and Diosady, L.L. 2005. Preparation of canola protein materials using membrane technology and evaluation of meals functional properties. Food Res. Int. **38**: 223–231. doi:10.1016/j.foodres.2004.10.007.

Golebiowski, T., Leong, A.S., and Panozzo, J.F. 2004. Near infrared reflectance spectroscopy of oil in intact canola seed (*Brassica napus* L.). I. Attributes of the intact seed spectrum. J. Near Infrared Spectrosc. **12**: 325–330. doi:10.1255/jnirs.556.

Government of Canada 2019. Saskatoon RCS Saskatchewan historical data. [Online] Available: https://climate.weather.gc.ca/climate_data/daily_data_e.html?hlyRange=2008-12-02%7C2019-12-15&dlyRange=2008-12-02%7C2019-12-15&mlyRange=%7C&StationID=47707&Prov=SK&urlExtension=_e.html&searchType=stnName&optLimit=specDate&StartYear=2013&EndYear=2015&selRo.

Grami, B., Baker, R.J., and Stefanssont, B.R. 1977. Genetics of protein and oil content in summer rape: heritability, number of effective factors and correlations. Can. J. Plant Sci. **57**: 937–943.

GRAS Notice 327 2010. GRAS notification for cruciferin-rich and napin-rich protein isolates derived from canola/rapeseed (Puratein and Supertein). [Online] Available: http://www.fda.gov/Food/FoodIngredientsPackaging/GenerallyRecognizedasSafeGRAS/GRASListings/default.htm [2019 Nov. 6].

Gu, J., Chao, H., Wang, H., Li, Y., Li, D., Xiang, J., Gan, J., Lu, G., Zhang, X., Long, Y., and Li, M. 2017. Identification of the relationship between oil body morphology and oil content by microstructure comparison combining with QTL analysis in *Brassica napus*. Front. Plant

Sci. **7**: 1989. doi:10.3389/fpls.2016.01989.

Guo, Y., Si, P., Wang, N., Wen, J., Yi, B., Ma, C., Tu, J., Zou, J., Fu, T., and Shen, J. 2017. Genetic effects and genotype×environment interactions govern seed oil content in *Brassica napus* L. BMC Genet. **18**: 1–11. doi:10.1186/s12863-016-0468-0.

Gupta, P.K., Kulwal, P.L., and Jaiswal, V. 2014. Association mapping in crop plants: opportunities and challenges. Pages 109–147 *in* T. Friedmann, J.C. Dunlap, and S.F. Goodwin, eds. Advances in genetics. Elsevier, Waltham, MA.

Gyawali, S., Harrington, M., Durkin, J., Horner, K., Parkin, I.A.P., Hegedus, D.D., Bekkaoui, D., and Buchwaldt, L. 2016. Microsatellite markers used for genome-wide association mapping of partial resistance to *Sclerotinia sclerotiorum* in a world collection of *Brassica napus*. Mol. Breed. **36**: 1–13. doi:10.1007/s11032-016-0496-5.

Gyawali, S., Hegedus, D.D., Parkin, I.A.P., Poon, J., Higgins, E., Horner, K., Bekkaoui, D., Coutu, C., and Buchwaldt, L. 2013. Genetic diversity and population structure in a world collection of *Brassica napus* accessions with emphasis on South Korea, Japan, and Pakistan. Crop Sci. **53**: 1537–1545. doi:10.2135/cropsci2012.10.0614.

Hajen, W.E., Beames, R.M., Higgs, D.A., and Dosanjh, B.S. 1993. Digestibility of various feedstuffs by post-juvenile chinook salmon (*Oncorhynchus tshawytscha*) in sea water. 1. Validation of technique. Aquaculture, **112**: 321–332. doi:10.1016/0044-8486(93)90393-D.

Hall, D., Tegstrom, C., and Ingvarsson, P.K. 2010. Using association mapping to dissect the genetic basis of complex traits in plants. Brief. Funct. Genomics **9**: 157–165. doi:10.1093/bfgp/elp048.

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

Hatzig, S. V., Frisch, M., Breuer, F., Nesi, N., Ducournau, S., Wagner, M.H., Leckband, G., Abbadi, A., and Snowdon, R.J. 2015. Genome-wide association mapping unravels the genetic control of seed germination and vigor in *Brassica napus*. Front. Plant Sci. **6**: 1–13. doi:10.3389/fpls.2015.00221.

Hearne, C.M., Ghosh, S., and Todd, J.A. 1992. Microsatellites for linkage analysis of genetic traits. Trends Genet. **8**: 288–294. doi:10.1016/0168-9525(92)90256-4.

Hedqvist, H., and Udén, P. 2006. Measurement of soluble protein degradation in the rumen. Anim. Feed Sci. Technol. **126**: 1–21. doi:10.1016/j.anifeedsci.2005.05.011.

Hei, W., Li, Z., Ma, X., and He, P. 2012. Determination of beta-conglycinin in soybean and soybean products using a sandwich enzyme-linked immunosorbent assay. Anal. Chim. Acta **734**: 62–68. doi:10.1016/j.aca.2012.05.009.

Helentjaris, T., Slocum, M., Wright, S., Schaefer, A., and Nienhuis, J. 1986. Construction of genetic linkage maps in maize and tomato using restriction fragment length polymorphisms. Theor. Appl. Genet. **72**: 761–769. doi:10.1007/BF00266542.

Henchion, M., Hayes, M., Mullen, A.M., Fenelon, M., and Tiwari, B. 2017. Future protein supply and demand: strategies and factors influencing a sustainable equilibrium. Foods, **6**: 1–21. doi:10.3390/foods6070053.

Hickling, D. 1994. Canola meal hog-feeding trials in Western Canada. Winnipeg, MB.

Holland, J.B., Nyquist, W.E., and Cervantes-Martínez, C.T. 2003. Estimating and interpreting heritability for plant breeding: an update. Pages 9–112 *in* J. Janick, ed. Plant breeding reviews. John Wiley & Sons, Inc., Hoboken, NJ.

Hong, P., Koza, S., and Bouvier, E.S.P. 2012. A review size-exclusion chromatography for the analysis of protein biotherapeutics and their aggregates. J. Liq. Chromatogr. Relat. Technol.

35: 2923–2950. doi:10.1080/10826076.2012.743724.

Hu, S., Yu, C., Zhao, H., Sun, G., Zhao, S., Vyvadilova, M., and Kucera, V. 2007. Genetic diversity of *Brassica napus* L. germplasm from China and Europe assessed by some agronomically important characters. Euphytica, **154**: 9–16. doi:10.1007/s10681-006-9263-8.

Huang, A.H.C. 1992. Oil bodies and oleosins in seeds. Annu. Rev. Plant Physiol. Plant Mol. Biol. **43**: 177–200. doi:10.1146/annurev.pp.43.060192.001141.

Huhtanen, P., Hetta, M., and Swensson, C. 2011. Evaluation of canola meal as a protein supplement for dairy cows: a review and a meta-analysis. Can. J. Anim. Sci. **91**: 529–543. doi:10.4141/cjas2011-029.

Hungate, R.E. 1966. The rumen and its microbes. Academic Press Inc., New York, NY.

International Organization for Standardization 2015. Rapeseed — Determination of chlorophyll content — Spectrometric method (ISO Standard No. 10519:2015). [Online] Available: https://www.iso.org/standard/66390.html [2019 Oct. 30].

Ismond, M.A.H., and Welsh, W.D. 1992. Application of new methodology to canola protein isolation. Food Chem. **45**: 125–127. doi:10.1016/0308-8146(92)90022-T.

Jan, H.U., Guan, M., Yao, M., Liu, W., Wei, D., Abbadi, A., Zheng, M., He, X., Chen, H., Guan, C., Nichols, R.A., Snowdon, R.J., Hua, W., and Qian, L. 2019. Genome-wide haplotype analysis improves trait predictions in *Brassica napus* hybrids. Plant Sci. **283**: 157–164. doi:10.1016/j.plantsci.2019.02.007.

Javed, N., Geng, J., Tahir, M., McVetty, P.B.E., Li, G., and Duncan, R.W. 2016. Identification of QTL influencing seed oil content, fatty acid profile and days to flowering in *Brassica napus* L. Euphytica, **207**: 191–211. doi:10.1007/s10681-015-1565-2.

Jonah, P.M., Bello, L.L., Lucky, O., Midau, A., and Moruppa, S.M. 2011. Review: The importance of molecular markers in plant breeding programs. Glob. J. Sci. Front. Res. **11**: 5–12.

Josefsson, L.G., Lenman, M., Ericson, M.L., and Rask, L. 1987. Structure of a gene encoding the 1.7 S storage protein, napin, from *Brassica napus*. J. Biol. Chem. **262**: 12196–12201.

Kaler, A.S., Gillman, J.D., Beissinger, T., and Purcell, L.C. 2020. Comparing different statistical models and multiple testing corrections for association mapping in soybean and maize. Front. Plant Sci. **10**: 1–13. doi:10.3389/fpls.2019.01794.

Kaur, B., Sangha, M.K., and Kaur, G. 2016. Calibration of NIRS for the estimation of fatty acids in *Brassica juncea*. J. Am. Oil Chem. Soc. **93**: 673–680. doi:10.1007/s11746-016-2802-0.

Khajali, F., and Slominski, B.A. 2012. Factors that affect the nutritive value of canola meal for poultry. Poult. Sci. **91**: 2564–2575.

Khattab, R.Y., and Arntfield, S.D. 2009. Functional properties of raw and processed canola meal. LWT - Food Sci. Technol. **42**: 1119–1124. doi:10.1016/j.lwt.2009.02.009.

Kim, K.S., Park, S.H., Choung, M.G., and Jang, Y.S. 2007. Use of near-infrared spectroscopy for estimating fatty acid composition in intact seeds of rapeseed. J. Crop Sci. Biotechnol. **10**: 15–20.

Kim, S.J., Ryu, M.Y., and Kim, W.T. 2012. Suppression of *Arabidopsis* RING-DUF1117 E3 ubiquitin ligases, *AtRDUF1* and *AtRDUF2*, reduces tolerance to ABA-mediated drought stress. Biochem. Biophys. Res. Commun. **420**: 141–147. doi:10.1016/j.bbrc.2012.02.131.

Klockeman, D.M., Toledo, R., and Sims, K.A. 1997. Isolation and characterization of defatted canola meal protein. J. Agric. Food Chem. **45**: 3867–3870. doi:10.1021/jf970026i.

Kohno-Murase, J., Murase, M., Ichikawa, H., and Imamura, J. 1994. Effects of an antisense

napin gene on seed storage compounds in transgenic *Brassica napus* seeds. Plant Mol. Biol. **26**: 1115–1124. doi:10.1007/BF00040693.

Kohno-Murase, J., Murase, M., Ichikawa, H., and Imamura, J. 1995. Improvement in the quality of seed storage protein by transformation of *Brassica napus* with an antisense gene for cruciferin. Theor. Appl. Genet. **91**: 627–631. doi:10.1007/BF00223289.

Korte, A., and Farlow, A. 2013. The advantages and limitations of trait analysis with GWAS: a review. Plant Methods, **9**: 1–9.

L'Hocine, L., Pitre, M., and Achouri, A. 2019. Detection and identification of allergens from Canadian mustard varieties of *Sinapis alba* and *Brassica juncea*. Biomolecules, **9**: 489. doi:10.3390/biom9090489.

Landero, J.L., Beltranena, E., and Zijlstra, R.T. 2012. Growth performance and preference studies to evaluate solvent-extracted *Brassica napus* or *Brassica juncea* canola meal fed to weaned pigs. J. Anim. Sci. **90**: 406–408. doi:10.2527/jas.53955.

Larkin, J.L. 2012. Innovative approaches to assessing seed quality in *Brassicas* . University of Manitoba.

Larson, E.R., Domozych, D.S., and Tierney, M.L. 2014. SNARE VTI13 plays a unique role in endosomal trafficking pathways associated with the vacuole and is essential for cell wall organization and root hair growth in arabidopsis. Ann. Bot. **114**: 1147–1159. doi:10.1093/aob/mcu041.

Lee, H.G., and Seo, P.J. 2021. Transcriptional activation of *SUGAR TRANSPORT PROTEIN 13* mediates biotic and abiotic stress signaling. Plant Signal. Behav. **2021**: e1920759. doi:10.1080/15592324.2021.1920759.

Leinonen, I., Iannetta, P.P.M., Rees, R.M., Russell, W., Watson, C., and Barnes, A.P. 2019.

Lysine supply is a critical factor in achieving sustainable global protein economy. Front. Sustain. Food Syst. **3**: 1–11. doi:10.3389/fsufs.2019.00027.

Li, F., Chen, B., Xu, K., Gao, G., Yan, G., Qiao, J., Li, J., Li, H., Li, L., Xiao, X., Zhang, T., Nishio, T., and Wu, X. 2016. A genome-wide association study of plant height and primary branch number in rapeseed (*Brassica napus*). Plant Sci. **242**: 169–177. doi:10.1016/j.plantsci.2015.05.012. \.

Li, F., Chen, B., Xu, K., Wu, J., Song, W., Bancroft, I., Harper, A.L., Trick, M., Liu, S., Gao, G., Wang, N., Yan, G., Qiao, J., Li, J., Li, H., Xiao, X., Zhang, T., and Wu, X. 2014. Genome-wide association study dissects the genetic architecture of seed weight and seed quality in rapeseed (*Brassica napus* L.). DNA Res. **21**: 355–367. doi:10.1093/dnares/dsu002.

Li, X., Hasegawa, Y., Lu, Y., and Sato, T. 2017. Ubiquitin related enzymes and plant-specific ubiquitin ligase ATL family in tomato plants. Plant Biotechnol. **34**: 71–78. doi:10.5511/plantbiotechnology.17.0306a.

Liezhao, L., Jinling, M., Na, L., Li, C., Tang, Z., Zhang, X., and Li, J. 2007. QTL mapping of seed coat color of yellow-seeded *Brassica napus*. Pages 377–381 *in* Proceedings of the 12th International Rapeseed Congress. Wuhan, CN.

Lin, A. V. 2015. Indirect ELISA. Pages 51–59 *in* R. Hnasko, ed. ELISA: methods and protocols. Springer Science, New York, NY.

Liu, H.J., and Yan, J. 2019. Crop genome-wide association study: a harvest of biological relevance. Plant J. **97**: 8–18. doi:10.1111/tpj.14139.

Liu, J., Wang, W., Mei, D., Wang, H., Fu, L., Liu, D., Li, Y., and Hu, Q. 2016a. Characterizing variation of branch angle and genome-wide association mapping in rapeseed (*Brassica napus* L.). Front. Plant Sci. **7**: 1–10. doi:10.3389/fpls.2016.00021.

Liu, S., Fan, C., Li, J., Cai, G., Yang, Q., Wu, J., Yi, X., Zhang, C., and Zhou, Y. 2016b. A genome-wide association study reveals novel elite allelic variations in seed oil content of *Brassica napus*. Theor. Appl. Genet. **129**: 1203–1215. doi:10.1007/s00122-016-2697-z.

Liu, S., Liu, Y., Yang, X., Tong, C., Edwards, D., Parkin, I. a P., Zhao, M., Ma, J., Yu, J., Huang, S., Wang, X.X.X., Wang, J.J.J., Lu, K., Fang, Z., Bancroft, I., Yang, T., Hu, Q., Yue, Z., Li, H., Yang, L.L., Wu, J., Zhou, Q., Wang, W., King, G.J., Pires, J.C., Lu, C., Wu, Z., Sampath, P., Wang, Z., Guo, H., Pan, S., Min, J., Zhang, D., Jin, D., Li, W., Belcram, H., Tu, J., Guan, M., Qi, C., Du, D., Li, J.J., Jiang, L., Batley, J., Sharpe, A.G., Park, B.-S., Ruperao, P., Cheng, F., Waminal, N.E., Huang, Y., Dong, C., Wang, L., Hu, Z., Zhuang, M., Huang, J., Shi, J., Mei, D., Liu, J., Lee, T.-H., Jin, H., Li, Z., Li, X., Zhang, J., Xiao, L., Zhou, Y., Liu, Z., Liu, X., Qin, R., Tang, X., Liu, W., Wang, Y., Zhang, Y., Lee, J., Kim, H.H., Denoeud, F., Xu, X., Liang, X., Hua, W., Chalhoub, B., and Paterson, A.H. 2014. The *Brassica oleracea* genome reveals the asymmetrical evolution of polyploid genomes. Nat. Commun. **5**: 3930. doi:10.1038/ncomms4930.

LMC International 2016. The economic impact of canola on the Canadian economy. Winnipeg, MB.

Lombard, V., Baril, C.P., Dubreuil, P., Blouet, F., and Zhang, D. 2000. Genetic relationships and fingerprinting of rapeseed cultivars by AFLP: consequences for varietal registration. Crop Sci. **40**: 1417–1425. doi:10.2135/cropsci2000.4051417x.

Lü, B.B., Li, X.J., Sun, W.W., Li, L., Gao, R., Zhu, Q., Tian, S.M., Fu, M.Q., Yu, H.L., Tang, X.M., Zhang, C.L., and Dong, H.S. 2013. AtMYB44 regulates resistance to the green peach aphid and diamondback moth by activating EIN2-affected defences in *Arabidopsis*. Plant Biol. **15**: 841–850. doi:10.1111/j.1438-8677.2012.00675.x.

Lu, K., Peng, L., Zhang, C., Lu, J., Yang, B., Xiao, Z., Liang, Y., Xu, X., Qu, C., Zhang, K., Liu, L., Zhu, Q., Fu, M., Yuan, X., and Li, J. 2017. Genome-wide association and transcriptome analyses reveal candidate genes underlying yield-determining traits in *Brassica napus*. Front. Plant Sci. **8**: 1–15. doi:10.3389/fpls.2017.00206.

Luo, Z., Brock, J., Dyer, J.M., Kutchan, T., Schachtman, D., Augustin, M., Ge, Y., Fahlgren, N., and Abdel-Haleem, H. 2019. Genetic diversity and population structure of a *Camelina sativa* spring panel. Front. Plant Sci. **10**: 1–12. doi:10.3389/fpls.2019.00184.

Lyzenga, W.J., Harrington, M., Bekkaoui, D., Wigness, M., Hegedus, D.D., and Rozwadowski, K.L. 2019. CRISPR/Cas9 editing of three CRUCIFERIN C homoeologues alters the seed protein profile in *Camelina sativa*. BMC Plant Biol. **19**: 292. doi:10.1186/s12870-019-1873-0.

Malabat, C., Atterby, H., Chaudhry, Q., Renard, M., and Guéguen, J. 2003. Genetic variability of rapeseed protein composition. Pages 1–3 *in* 11th International Rapeseed Congress. Copenhagen, DK.

Marambe, H.K., McIntosh, T.C., Cheng, B., and Wanasundara, J.P.D. 2014. Quantification of major 2S allergen protein of yellow mustard using anti-Sin a 1 epitope antibody. Food Control **44**: 233–241. doi:10.1016/j.foodcont.2014.03.053.

Mariotti, F., Tomé, D., and Mirand, P. 2008. Converting nitrogen into protein – beyond 6.25 and Jones ' factors. Crit. Rev. Food Sci. Nutr. **48**: 177–184.

Martineau, R., Ouellet, D.R., and Lapierre, H. 2013. Feeding canola meal to dairy cows: A meta-analysis on lactational responses. J. Dairy Sci. **96**: 1701–1714. doi:10.3168/jds.2012-5740. Elsevier.

Mason, A.S., Higgins, E.E., Snowdon, R.J., Batley, J., Stein, A., Werner, C., and Parkin, I.A.P.

2017. A user guide to the *Brassica* 60K Illumina Infinium™ SNP genotyping array. Theor. Appl. Genet. **130**: 621–633. doi:10.1007/s00122-016-2849-1.

Maxin, G., Ouellet, D.R., and Lapierre, H. 2013. Effect of substitution of soybean meal by canola meal or distillers grains in dairy rations on amino acid and glucose availability. J. Dairy Sci. **96**: 7806–7817. doi:10.3168/jds.2013-6976.

McVetty, P.B.E.-169. pd., Duncan, R.W., Fernando, W.G.D., Li, G., and Zelmer, C.D. 2012. Red River 1861 Roundup Ready™ high erucic acid, low glucosinolate summer rape. Can. J. Plant Sci. **92**: 1407–1409. doi:10.4141/CJPS2012-126.

McVetty, P.B.E., Cuthbert, J.L., Marwede, V., Paulmann, W., Sass, O., Duncan, R.W., Fernando, W.G.D., Li, G., and Zelmer, C.D. 2014. HYHEAR 1 hybrid Roundup Ready™ high erucic acid, low glucosinolate summer rape. Can. J. Plant Sci. **94**: 453–455. doi:10.4141/CJPS2013-169.

McVetty, P.B.E., Fernando, W.G.D., Li, G., Tahir, M., and Zelmer, C.D. 2010. Red River 1997 Roundup Ready™ high erucic acid, low glucosinolate summer rape. Can. J. Plant Sci. **90**: 711–713. doi:10.4141/CJPS10068.

McVetty, P.B.E., Fernando, W.G.D., Scarth, R., and Li, G. 2006a. Red River 1852 Roundup Ready™ high erucic acid, low glucosinolate summer rape. Can. J. Plant Sci. **86**: 1181–1182. doi:10.4141/CJPS2012-126.

McVetty, P.B.E., Fernando, W.G.D., Scarth, R., and Li, G. 2006b. Red River 1826 Roundup Ready™ high erucic acid, low glucosinolate summer rape. Can. J. Plant Sci. **86**: 1179–1180. doi:10.4141/CJPS2012-126.

McVetty, P.B.E., Mietkiewska, E., Omonov, T., Curtis, J., Taylor, D.C., and Weselake, R.J. 2016. *Brassica* spp. Oils. Pages 113–156 *in* T.A. McKeon, D.G. Hayes, D.F. Hildebrand,

and R.J. Weselake, eds. Industrial oil crops. Elsevier, London, UK.

McVetty, P.B.E., Rimmer, S.R., Scarth, R., and van den Berg, C.G.J. 1996a. Neptune high erucic acid, low glucosinolate summer rape. Can. J. Plant Sci. **76**: 343–344. doi:10.4141/P97-083.

McVetty, P.B.E., Rimmer, S.R., Scarth, R., and van den Berg, C.G.J. 1996b. Venus high erucic acid, low glucosinolate summer rape. Can. J. Plant Sci. **76**: 341–342. doi:10.4141/P97-083.

McVetty, P.B.E., and Scarth, R. 2002. Breeding for improved oil quality in *Brassica* oilseed species. J. Crop Prod. **5**: 345–369. doi:10.1300/J144v05n01.

McVetty, P.B.E., Scarth, R., and Rimmer, S.R. 1998. Castor high erucic acid, low glucosinolate summer rape. Can. J. Plant Sci. **78**: 305–306. doi:10.4141/P98-093.

McVetty, P.B.E., Scarth, R., and Rimmer, S.R. 1999. MillenniUM 01 high erucic acid, low glucosinolate summer rape. Can. J. Plant Sci. **79**: 251–252. doi:10.4141/P98-093.

McVetty, P.B.E., Scarth, R., and Rimmer, S.R. 2000a. MilleniUM 02 summer rape. Can. J. Plant Sci. **80**: 609–610.

McVetty, P.B.E., Scarth, R., and Rimmer, S.R. 2000b. MillenniUM 03 summer rape. Can. J. Plant Sci. **80**: 611–612. doi:10.4141/P00-010.

Mejicanos, G., Sanjayan, N., Kim, I.H., and Nyachoti, C.M. 2016. Recent advances in canola meal utilization in swine nutrition. J. Anim. Sci. Technol. **58**: 1–13. doi:10.1186/s40781-016-0085-5.

Mejicanos, G.A., Kim, J.W., and Nyachoti, C.M. 2018. Tail-end dehulling of canola meal improves apparent and standardized total tract digestibility of phosphorus when fed to growing pigs. J. Anim. Sci. **96**: 1430–1440. doi:10.1093/jas/sky040.

Mejicanos, G.A., Rogiewicz, A., Nyachoti, C.M., and Slominski, B.A. 2017. Fractionation of canola meal using sieving technology. Can. J. Anim. Sci. **97**: 613–621. doi:10.1139/cjas-

2016-0229.

Merit Functional Foods 2021. Merit Functional Foods opens world-class plant-based protein facility. [Online] Available: https://meritfoods.com/merit-functional-foods-opens-world-class-plant-based-protein-facility/ [2021 Jun. 24].

Mohler, V., and Schwarz, G. 2005. Genotyping tools in plant breeding: From restriction fragment length polymorphisms to single nucleotide polymorphisms. Pages 23–38 *in* H. Lorz and G. Wenzel, eds. Molecular marker systems in plant breeding and crop improvement. Springer-Verlag, Berlin, Germany.

Monaghan, E.K., Venkatachalam, M., Seavy, M., Beyer, K., Sampson, H.A., Roux, K.H., and Sathe, S.K. 2008. Enzyme-linked immunosorbent assay (ELISA) for detection of sulfur-rich protein (SRP) in soybeans (*Glycine max* L.) and certain other edible plant seeds. J. Agric. Food Chem. **56**: 765–777. doi:10.1021/jf072079y.

Monsalve, R.I., Villalba, M., and Rodríguez, R. 2001. Allergy to mustard seeds : the importance of 2S abumins as food allergens. Internet Symp. Food Allergens **3**: 57–69.

Moriyama, T., Machidori, M., Ozasa, S., Maebuchi, M., Urade, R., Takahashi, K., Ogawa, T., and Maruyama, N. 2005. A novel enzyme-linked immunosorbent assay for quantification of soybean β-conglycinin, a major soybean storage protein, in soybean and soybean food products. J. Nutr. Sci. Vitaminol. (Tokyo). **51**: 34–39. doi:10.3177/jnsv.51.34.

Mullis, K., Faloona, F., Scharf, S., Saiki, R., Horn, G., and Erlich, H. 1986. Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. Cold Spring Harb. Symp. Quant. Biol. **51**: 263–273.

Murray, E.D. 1999. Oil seed protein extraction. US Patent 6,005,076.

National Research Council 2007. Nutrient requirements of small ruminants: sheep, goats,

cervids, and new world camelids. Washington, DC.

Nishinari, K., Fang, Y., Nagano, T., Guo, S., and Wang, R. 2017. Soy as a food ingredient. Pages 149–186 *in* R.Y. Yada, ed. Proteins in food processing: second edition. Elsevier, Duxford, UK.

El Nockrashy, A.S., Mukherjee, K.D., and Mangold, H.K. 1977. Rapeseed protein isolates by countercurrent extraction and isoelectric precipitation. J. Agric. Food Chem. **25**: 193–197. doi:10.1021/jf60209a022.

Oraguzie, N.C., and Wilcox, P.L. 2007. An overview of association mapping. Pages 1–9 *in* N.C. Oraguzie, E.H.A. Rikkerink, S.E. Gardiner, and H.N. De Silva, eds. Association mapping in plants. Springer, New York, NY.

Oraguzie, N.C., Wilcox, P.L., Rikkerink, E.H.A., and de Silva, H.N. 2007. Linkage disequilibrium. Pages 11–39 *in* N.C. Oraguzie, E.H.A. Rikkerink, S.E. Gardiner, and H.N. De Silva, eds. Association mapping in plants. Springer, New York, NY.

Osborne, T.B. 1897. The amount and properties of the proteids of the maize kernel. J. Am. Chem. Soc. **19**: 525–532. doi:10.1021/ja02081a002.

Pang, P.P., Pruitt, R.E., and Meyerowitz, E.M. 1988. Molecular cloning, genomic organization, expression and evolution of 12S seed storage protein genes of *Arabidopsis thaliana*. Plant Mol. Biol. **11**: 805–820. doi:10.1007/BF00019521.

Parkin, I., Vail, S., and Robinson, S. 2017. Development of a germplasm resource to dissect complex traits in *Brassica napus*, ADF Project #20110155 Final Report.

Paz, H.A., Klopfenstein, T.J., Hostetler, D., Fernando, S.C., Castillo-Lopez, E., and Kononoff, P.J. 2014. Ruminal degradation and intestinal digestibility of protein and amino acids in high-protein feedstuffs commonly used in dairy diets. J. Dairy Sci. **97**: 6485–6498.

doi:10.3168/jds.2014-8108.

Pazdernik, D.L., Plehn, S.J., Halgerson, J.L., and Orf, J.H. 1996. Effect of temperature and genotype on the crude glycinin fraction (11S) of soybean and its analysis by near-infrared reflectance spectroscopy (near-IRS). J. Agric. Food Chem. **44**: 2278–2281. doi:10.1021/jf960225m.

Peakall, R., and Smouse, P.E. 2012. GenALEx 6.5: Genetic analysis in Excel. Population genetic software for teaching and research-an update. Bioinformatics, **28**: 2537–2539. doi:10.1093/bioinformatics/bts460.

Pedroche, J., Yust, M.M., Lqari, H., Girón-Calle, J., Alaiz, M., Vioque, J., and Millán, F. 2004. *Brassica carinata* protein isolates: chemical composition, protein characterization and improvement of functional properties by protein hydrolysis. Food Chem. **88**: 337–346. doi:10.1016/j.foodchem.2004.01.045.

Perera, S., McIntosh, T., and Wanasundara, J. 2016. Structural properties of cruciferin and napin of *Brassica napus* (canola) show distinct responses to changes in pH and temperature. Plants, **5**: 1–24. doi:10.3390/plants5030036.

Perez-Maldonado, R.A. 2003. Canola meal and cottonseed meal in broiler and layer diets. Hurstville, AU.

Piepho, H.P., Möhring, J., Melchinger, A.E., and Büchse, A. 2008. BLUP for phenotypic selection in plant breeding and variety testing. Euphytica, **161**: 209–228. doi:10.1007/s10681-007-9449-8.

Pieruschka, R., and Schurr, U. 2019. Plant phenotyping: past, present, and future. Plant Phenomics, **2019**: 1–6. doi:10.34133/2019/7507131.

Price, A.L., Zaitlen, N.A., Reich, D., and Patterson, N. 2010. New approaches to population

stratification in genome-wide association studies. Nat. Rev. Genet. **11**: 459–463. doi:10.1038/nrg2813.

Pritchard, J.K., Stephens, M., and Donnelly, P. 2000. Inference of population structure using multilocus genotype data. Genetics, **155**: 945–959. doi:10.1007/s10681-008-9788-0.

Przybylski, R., and Eskin, M.N.A. 2011. Oil composition and properties. Pages 189–227 *in* J.K. Daun, M.N.A. Eskin, and D. Hickling, eds. Canola: chemistry, production, processing and utilization. AOCS Press, Urbana, IL.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J., Sklar, P., De Bakker, P.I.W., Daly, M.J., and Sham, P.C. 2007. PLINK: A tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. **81**: 559–575. doi:10.1086/519795.

Puumalainen, T.J., Puustinen, A., Poikonen, S., Turjanmaa, K., Palosuo, T., and Vaali, K. 2015. Proteomic identification of allergenic seed proteins, napin and cruciferin, from cold-pressed rapeseed oils. Food Chem. **175**: 381–385. doi:10.1016/j.foodchem.2014.11.084.

Qian, L., Qian, W., and Snowdon, R.J. 2014. Sub-genomic selection patterns as a signature of breeding in the allopolyploid *Brassica napus* genome. BMC Genomics, **15**: 1–17. doi:10.1186/1471-2164-15-1170.

Qian, W., Meng, J., Li, M., Frauen, M., Sass, O., Noack, J., and Jung, C. 2006. Introgression of genomic components from Chinese *Brassica rapa* contributes to widening the genetic diversity in rapeseed (*B. napus* L.), with emphasis on the evolution of Chinese rapeseed. Theor. Appl. Genet. **113**: 49–54. doi:10.1007/s00122-006-0269-3.

Rahman, H. 2013. Review: Breeding spring canola (*Brassica napus* L.) by the use of exotic germplasm. Can. J. Plant Sci. **93**: 363–373. doi:10.4141/CJPS2012-074.

Raman, H., Dalton-Morgan, J., Diffey, S., Raman, R., Alamery, S., Edwards, D., and Batley, J. 2014. SNP markers-based map construction and genome-wide linkage analysis in *Brassica napus*. Plant Biotechnol. J. **12**: 851–860. doi:10.1111/pbi.12186.

Raman, H., Raman, R., Coombes, N., Song, J., Prangnell, R., Bandaranayake, C., Tahira, R., Sundaramoorthi, V., Killian, A., Meng, J., Dennis, E.S., and Balasubramanian, S. 2016. Genome-wide association analyses reveal complex genetic architecture underlying natural variation for flowering time in canola. Plant Cell Environ. **39**: 1228–1239. doi:10.1111/pce.12644.

Rico, M., Bruix, M., González, C., Monsalve, R.I., and Rodríguez, R. 1996. 1H NMR assignment and global fold of napin BnIb, a representative 2S albumin seed protein. Biochemistry, **35**: 15672–15682. doi:10.1021/bi961748q.

Rinnan, Å., Berg, F. van den, and Engelsen, S.B. 2009. Review of the most common pre-processing techniques for near-infrared spectra. Trends Anal. Chem. **28**: 1201–1222. doi:10.1016/j.trac.2009.07.007.

Sato, T., Uezono, I., Morishita, T., and Tetsuka, T. 1998. Nondestructive estimation of fatty acid composition in seeds of *Brassica napus* L. by near-infrared spectroscopy. J. Am. Oil Chem. Soc. **75**: 1877–1881. doi:10.1007/s11746-998-0344-9.

Sauer, F.D., and Kramer, J.K.. 1983. The problems associated with the feeding of high erucic acid rapeseed oils and some fish oils to experimental animals. Pages 253–288 *in* J.K.. Kramer, F.D. Sauer, and W.J. Pigden, eds. High and low erucic acid rapeseed oils. Academic Press Canada, Don Mills, ON.

Scarth, R., McVetty, P.B.E., and Rimmer, S.R. 1995. Mercury high erucic low glucosinolate summer rape. Can. J. Plant Sci. **75**: 205–206. doi:10.4141/cjps95-038.

Scarth, R., McVetty, P.B.E., Rimmer, S.R., and Stefansson, B.R. 1988. Stellar low linolenic-high linoleic acid summer rape. Can. J. Plant Sci. **68**: 509–511.

Scarth, R., and Tang, J. 2006. Modification of *Brassica* oil using conventional and transgenic approaches. Crop Sci. **46**: 1225–1236. doi:10.2135/cropsci2005.08-0245.

Schatzki, J., Ecke, W., Becker, H.C., and Möllers, C. 2014. Mapping of QTL for the seed storage proteins cruciferin and napin in a winter oilseed rape doubled haploid population and their inheritance in relation to other seed traits. Theor. Appl. Genet. **127**: 1213–1222. doi:10.1007/s00122-014-2292-0.

Schiessl, S., Huettel, B., Kuehn, D., Reinhardt, R., and Snowdon, R. 2017. Post-polyploidisation morphotype diversification associates with gene copy number variation. Sci. Rep. **7**: 1–18. doi:10.1038/srep41845.

Schmidt, P., Hartung, J., Rath, J., and Piepho, H.P. 2019. Estimating broad-sense heritability with unbalanced data from agricultural cultivar trials. Crop Sci. **59**: 525–536. doi:10.2135/cropsci2018.06.0376.

Schöne, F., Groppel, B., Hennig, A., Jahreis, G., and Lange, R. 1997a. Rapeseed meals, methimazole, thiocyanate and iodine affect growth and thyroid. Investigations into glucosinolate tolerance in the pig. J. Sci. Food Agric. **74**: 69–80. doi:10.1002/(SICI)1097-0010(199705)74:1<69::AID-JSFA771>3.0.CO;2-0.

Schöne, F., Rudolph, B., Kirchheim, U., and Knapp, G. 1997b. Counteracting the negative effects of rapeseed and rapeseed press cake in pig diets. Br. J. Nutr. **78**: 947–962. doi:10.1079/bjn19970211.

Schrag, D., Corbier, M., and Raimondi, S. 2014. Size exclusion-high-performance liquid chromatography (SEC-HPLC). Pages 151–181 *in* V. Ossipow and N. Fischer, eds.

Monoclonal antibodies: methods and protocols. Humana Press, Totowa, NJ.

Schranz, M.E., Lysak, M.A., and Mitchell-Olds, T. 2006. The ABC's of comparative genomics in the Brassicaceae: building blocks of crucifer genomes. Trends Plant Sci. **11**: 535–542. doi:10.1016/j.tplants.2006.09.002.

Schwab, C.G., Satter, L.D., and Clay, A.B. 1976. Response of lactating dairy cows to abomasal infusion of amino acids. J. Dairy Sci. **59**: 1254–1270. doi:10.3168/jds.S0022-0302(76)84354-8.

Schwenke, K.D., Raab, B., Linow, K.J., Platz, W., and Uhlin, J. 1981. Isolation of the 12S globulin from rapeseed (*Brassica napus* L.) and characterization as a neutral protein. Nahrung, **25**: 271–280.

Scofield, S.R., and Crouch, M.L. 1987. Nucleotide sequence of a member of the napin storage protein family from *Brassica napus*. J. Biol. Chem. **262**: 12202–12208.

Sharma, A., Li, X., and Lim, Y.P. 2014. Comparative genomics of Brassicaceae crops. Breed. Sci. **64**: 3–13. doi:10.1270/jsbbs.64.3.

Shenk, J.S., and Westerhaus, M.O. 1996. Calibration the ISI way. Pages 198–202 *in* Proceedings of the 7th International Conference on Near Infrared Spectroscopy. Montreal, CA.

Shenk, J.S., Workman, J.J., and Westerhaus, M.O. 2008. Application of NIR spectroscopy to agricultural products. Pages 348–386 *in* D.A. Burns and E.W. Ciurczak, eds. Handbook of near-infrared analysis, 3rd edition. CRC Press Taylor & Francis Group, Boca Raton, FL.

Shewry, P.R., Napier, J.A., and Tatham, A.S. 1995. Seed storage proteins: structures and biosynthesis. Plant Cell, **7**: 945–956. doi:10.1105/tpc.7.7.945.

Shi, C., Zhang, H., Wu, J., Li, C., and Ren, Y. 2003. Genetic and genotype x environment interaction effects analysis for erucic acid content in rapeseed (*Brassica napus* L.).

Euphytica, **130**: 249–254. doi:10.1023/A:1022867100199.

Si, P., Mailer, R.J., Galwey, N., and Turner, D.W. 2003. Influence of genotype and environment on oil and protein concentrations of canola (*Brassica napus* L.) grown across southern Australia. Crop Pasture Sci **54**: 397–407.

Singh, B.D., and Singh, A.K. 2015. Introduction to marker-assisted crop improvement. Pages 3–16 *in* Marker-assisted plant breeding: principles and practices. Springer, New Delhi, India.

Sjödahl, S., Rödin, J., and Rask, L. 1991. Characterization of the 12S globulin complex of *Brassica napus*: Evolutionary relationship to other 11–12S storage globulins. Eur. J. Biochem. **196**: 617–621. doi:10.1111/j.1432-1033.1991.tb15857.x.

Sleper, D.A., and Poehlman, J.M. 2006. Breeding field crops. 5th edition. Blackwell Publishing Ltd, Ames, Iowa.

Sneller, C.H., Mather, D.E., and Crepieux, S. 2009. Analytical approaches and population types for finding and utilizing QTL in complex plant populations. Crop Sci. **49**: 363–380. doi:10.2135/cropsci2008.07.0420.

Song, J.M., Guan, Z., Hu, J., Guo, C., Yang, Z., Wang, S., Liu, D., Wang, B., Lu, S., Zhou, R., Xie, W.Z., Cheng, Y., Zhang, Y., Liu, K., Yang, Q.Y., Chen, L.L., and Guo, L. 2020. Eight high-quality genomes reveal pan-genome architecture and ecotype differentiation of *Brassica napus*. Nat. Plants, **6**: 34–45. doi:10.1038/s41477-019-0577-7. Springer US.

Soumya, P.R., Burridge, A.J., Singh, N., Batra, R., Pandey, R., Kalia, S., Rai, V., and Edwards, K.J. 2021. Population structure and genome-wide association studies in bread wheat for phosphorus efficiency traits using 35 K Wheat Breeder's Affymetrix array. Sci. Rep. **11**: 1–17. doi:10.1038/s41598-021-87182-2.

Srinivas, P.R. 2012. Introduction to protein electrophoresis. Pages 23–28 *in* B.T. Kurien and

R.H. Scofield, eds. Protein electrophoresis: methods and protocols. Humana Press, Totowa, NJ.

Statistics Canada 2019a. Table 32-10-0359-01 Estimated areas, yield, production, average farm price and total farm value of principal field crops, in metric and imperial units. [Online] Available: https://www150.statcan.gc.ca/t1/tbl1/en/cv.action?pid=3210035901 [2019 Nov. 3].

Statistics Canada 2019b. Table 32-10-0359-01 Estimated areas, yield, production, average farm price and total farm value of principal field crops, in metric and imperial units. [Online] Available: https://www150.statcan.gc.ca/t1/tbl1/en/cv.action?pid=3210035901 [2019 Sep. 24].

Stefański, T., Ahvenjärvi, S., Huhtanen, P., and Shingfield, K.J. 2013. Metabolism of soluble rapeseed meal (*Brassica rapa* L.) protein during incubations with buffered bovine rumen contents in vitro. J. Dairy Sci. **96**: 440–450. doi:10.3168/jds.2012-5642.

Stefansson, B.R. 1983. The development of improved rapeseed cultivars. Pages 144–163 *in* J.K.. Kramer, F.D. Sauer, and W.J. Pigden, eds. High and low erucic acid rapeseed oils. Academic Press Canada, Don Mills, ON.

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

Stich, B., and Melchinger, A.E. 2010. An introduction to association mapping in plants. CAB Rev. Perspect. Agric. Vet. Sci. Nutr. Nat. Resour. **5**: 1–9. doi:10.1079/PAVSNNR20105039.

Summers, J.D., Leeson, S., and Spratt, D. 1988. Canola meal and egg size. Can. J. Anim. Sci. **68**: 907–913. doi:10.4141/cjas88-100.

Sun, C., Wang, B., Wang, X., Hu, K., Li, K., Li, Z., Li, S., Yan, L., Guan, C., Zhang, J., Zhang,

Z., Chen, S., Wen, J., Tu, J., Shen, J., Fu, T., and Yi, B. 2016. Genome-Wide Association Study Dissecting the Genetic Architecture Underlying the Branch Angle Trait in Rapeseed (*Brassica napus* L.). Sci. Rep. **6**: 1–11. doi:10.1038/srep33673.

Sun, F., Chalhoub, B., Liu, S., and Hua, W. 2018. Deciphering genome organization of the polyploid *Brassica napus*. Pages 87–97 *in* S. Liu, R. Snowdon, and B. Chalhoub, eds. The *Brassica napus* genome. Springer Nature Switzerland, Cham, Switzerland.

Sun, F., Fan, G., Hu, Q., Zhou, Y., Guan, M., Tong, C., Li, J., Du, D., Qi, C., Jiang, L., Liu, W., Huang, S., Chen, W., Yu, J., Mei, D., Meng, J., Zeng, P., Shi, J., Liu, K., Wang, X., Wang, X., Long, Y., Liang, X., Hu, Z., Huang, G., Dong, C., Zhang, H., Li, J., Zhang, Y., Li, L., Shi, C., Wang, J., Lee, S.M.Y., Guan, C., Xu, X., Liu, S., Liu, X., Chalhoub, B., Hua, W., and Wang, H. 2017. The high-quality genome of *Brassica napus* cultivar 'ZS11' reveals the introgression history in semi-winter morphotype. Plant J. **92**: 452–468. doi:10.1111/tpj.13669.

Suwabe, K., Tsukazaki, H., Iketani, H., Hatakeyama, K., Kondo, M., Fujimura, M., Nunome, T., Fukuoka, H., Hirai, M., and Matsumoto, S. 2006. Simple sequence repeat-based comparative genomics between *Brassica rapa* and *Arabidopsis thaliana*: The genetic origin of clubroot resistance. Genetics, **173**: 309–319. doi:10.1534/genetics.104.038968.

Tabe, L., Hagan, N., and Higgins, T.J.V. 2002. Plasticity of seed protein composition in response to nitrogen and sulfur availability. Curr. Opin. Plant Biol. **5**: 212–217. doi:10.1016/S1369-5266(02)00252-2.

Tan, S.H., Mailer, R.J., Blanchard, C.L., and Agboola, S.O. 2011. Canola proteins for human consumption: extraction, profile, and functional properties. J. Food Sci. **76**: R16–R28. doi:10.1111/j.1750-3841.2010.01930.x.

Tan, S.H., Mailer, R.J., Blanchard, C.L., and Agboola, S.O. 2014a. Emulsifying properties of proteins extracted from Australian canola meal. LWT - Food Sci. Technol. **57**: 376–382. doi:10.1016/j.lwt.2013.12.040. Elsevier Ltd.

Tan, S.H., Mailer, R.J., Blanchard, C.L., Agboola, S.O., and Day, L. 2014b. Gelling properties of protein fractions and protein isolate extracted from Australian canola meal. Food Res. Int. **62**: 819–828. doi:10.1016/j.foodres.2014.04.055.

Tegeder, M., and Ward, J.M. 2012. Molecular evolution of plant AAP and LHT amino acid transporters. Front. Plant Sci. **3**: 1–11. doi:10.3389/fpls.2012.00021.

The Arabidopsis Genome Initiative. 2000. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature, **408**: 796–815. doi:10.1134/S1022795411020074.

Tkachuk, R. 1981. Oil and protein analysis of whole rapeseed kernels by near infrared reflectance spectroscopy. J. Am. Oil Chem. Soc. **58**: 819–822.

Tkachuk, R., and Kuzina, F.D. 1982. Chlorophyll analysis of whole rapeseed kernels by near infrared reflectance. Can. J. Plant Sci. **62**: 875–884. doi:10.4141/cjps82-131.

Tripathi, M.K., and Mishra, A.S. 2007. Glucosinolates in animal nutrition: a review. Anim. Feed Sci. Technol. **132**: 1–27. doi:10.1016/j.anifeedsci.2006.03.003.

Turner, S.D. 2014. qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. bioRxiv: 1–2. doi:10.1086/519795.

U.S. Soybean Export Council 2015. U.S. Soybean Meal.

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

University of Manitoba 2019. The Point archive data. [Online] Available: http://home.cc.umanitoba.ca/~grieger/archive.html [2019 Dec. 16].

Uppström, B. 1995. Seed chemistry. Pages 217–242 *in* D.S. Kimber and D.I. McGregor, eds. *Brassica* oilseeds production and utilization. CAB International, Oxon, UK.

USDA 2019. Oilseeds: World markets and trade.

Velasco, L., and Becker, H.C. 1998a. Analysis of total glucosinolate content and individual glucosinolates in *Brassica* spp. by near-infrared reflectance spectroscopy. Plant Breed. **117**: 97–102. doi:10.1111/j.1439-0523.1998.tb01459.x.

Velasco, L., and Becker, H.C. 1998b. Estimating the fatty acid composition of the oil in intact-seed rapeseed (*Brassica napus* L.) by near-infrared reflectance spectroscopy. Euphytica, **101**: 221–230. doi:10.1023/A:1018358707847.

Velasco, L., Goffman, F.D., and Becker, H. 1999a. Development of calibration equations to predict oil content and fatty acid composition in Brassicaceae germplasm by near-infrared reflectance spectroscopy. J. Am. Oil Chem. Soc. **76**: 25–30.

Velasco, L., and Möllers, C. 2002. Nondestructive assessment of protein content in single seeds of rapeseed (*Brassica napus* L.) by near-infrared reflectance spectroscopy. Euphytica, **123**: 89–93. doi:10.1023/A:1014452700465.

Velasco, L., Möllers, C., and Becker, H.C. 1999b. Estimation of seed weight, oil content and fatty acid composition in intact single seeds of rapeseed (*Brassica napus* L.) by near-infrared reflectance spectroscopy. Euphytica, **106**: 79–85.

Virk, P.S., Ford-Lloyd, B. V., Jackson, M.T., Pooni, H.S., Clemeno, T.P., and Newbury, H.J. 1996. Predicting quantitative variation within rice germplasm using molecular markers. Heredity, **76**: 296–304. doi:10.1038/hdy.1996.43.

Vos, P., Hogers, R., Bleeker, M., Reijans, M., van de Lee, T., Hornes, M., Frijters, A., Pot, J., Paleman, J., Kuiper, M., and Zabeau, M. 1995. AFLP: A new technique for DNA

fingerprinting. Nucleic Acids Res. **23**: 4407–4414. doi:10.1093/nar/23.21.4407.

Wanasundara, J.P.D. 2011. Proteins of Brassicaceae oilseeds and their potential as a plant protein source. Crit. Rev. Food Sci. Nutr. **51**: 635–677. doi:10.1080/10408391003749942.

Wanasundara, J.P.D., Abeysekara, S.J., McIntosh, T.C., and Falk, K.C. 2012. Solubility differences of major storage proteins of Brassicaceae oilseeds. J. Am. Oil Chem. Soc. **89**: 869–881. doi:10.1007/s11746-011-1975-9.

Wanasundara, J.P.D., and Hojilla-Evangelista, M.P. 2018. Satisfying protein demand with plant and alternative proteins. J. Am. Oil Chem. Soc. **95**: 879–881. doi:10.1002/aocs.12138.

Wanasundara, J.P.D., and McIntosh, T.C. 2013. Process of aqueous protein extraction from Brassicaceae oilseeds. US Patent 8,557,963.

Wanasundara, J.P.D., Tan, S., Alashi, A.M., Pudel, F., and Blanchard, C. 2017. Proteins from canola/rapeseed: current status. Pages 285–304 *in* S.R. Nadathur, J.P.D. Wanasundara, and L. Scanlin, eds. Sustainable protein sources. Elsevier, London, UK.

Wang, D., Tyson, M.D., Jackson, S.S., and Yadegari, R. 2006. Partially redundant functions of two SET-domain polycomb-group proteins in controlling initiation of seed development in *Arabidopsis*. Proc. Natl. Acad. Sci. U. S. A. **103**: 13244–13249. doi:10.1073/pnas.0605551103.

Wang, J., Jian, H., Wei, L., Qu, C., Xu, X., Lu, K., Qian, W., Li, J., Li, M., and Liu, L. 2015. Genome-wide analysis of seed acid detergent lignin (ADL) and hull content in rapeseed (*Brassica napus* L.). PLoS ONE, **10**: 1–18. doi:10.1371/journal.pone.0145045.

Wang, J., Liu, H., and Ren, G. 2014. Near-infrared spectroscopy (NIRS) evaluation and regional analysis of Chinese faba bean (*Vicia faba* L.). Crop J. **2**: 28–37. doi:10.1016/j.cj.2013.10.001. Crop Science Society of China and Institute of Crop Sciences,

CAAS.

Wang, X.X., Wang, H.H.H., Wang, J.J.J., Sun, R., Wu, J., Liu, S., Bai, Y., Mun, J.-H., Bancroft, I., Cheng, F., Huang, S.S., Li, X., Hua, W., Wang, J.J.J., Wang, X.X., Freeling, M., Pires, J.C., Paterson, A.H., Chalhoub, B., Wang, B., Hayward, A., Sharpe, A.G., Park, B.-S., Weisshaar, B., Liu, B.B., Li, B., Liu, B.B., Tong, C., Song, C., Duran, C., Peng, C., Geng, C., Koh, C., Lin, C., Edwards, D., Mu, D., Shen, D., Soumpourou, E., Li, F., Fraser, F., Conant, G., Lassalle, G., King, G.J., Bonnema, G., Tang, H., Wang, H.H.H., Belcram, H., Zhou, H., Hirakawa, H., Abe, H., Guo, H., Wang, H.H.H., Jin, H., Parkin, I. a P., Batley, J., Kim, J. a J.-S., Just, J., Li, J.J., Xu, J., Deng, J., Kim, J. a J.-S., Li, J.J., Yu, J., Meng, J., Wang, J.J.J., Min, J., Poulain, J., Wang, J.J.J., Hatakeyama, K., Wu, K., Wang, L., Fang, L., Trick, M., Links, M.G., Zhao, M., Jin, M., Ramchiary, N., Drou, N., Berkman, P.J., Cai, Q., Huang, Q., Li, R., Tabata, S., Cheng, S., Zhang, S., Zhang, S., Huang, S.S., Sato, S., Sun, S., Kwon, S.-J., Choi, S.-R., Lee, T.-H., Fan, W., Zhao, X., Tan, X., Xu, X., Wang, Y.Y., Qiu, Y., Yin, Y., Li, Y., Du, Y., Liao, Y., Lim, Y., Narusaka, Y., Wang, Y.Y., Wang, Z.Z., Li, Z., Wang, Z.Z., Xiong, Z., and Zhang, Z. 2011. The genome of the mesopolyploid crop species *Brassica rapa*. Nat. Genet. **43**: 1035–1039. doi:10.1038/ng.919.

Warwick, S.I. 2011. Brassicaceae in agriculture. Pages 33–65 *in* R. Schmidt and I. Bancroft, eds. Genetics and genomics of the Brassicaceae. Springer New York, New York, NY.

Warwick, S.I., Francis, A., and Al-Shehbaz, I.A. 2006. Brassicaceae: Species checklist and database on CD-Rom. Plant Syst. Evol. **259**: 249–258. doi:10.1007/s00606-006-0422-0.

Wei, L., Jian, H., Lu, K., Filardo, F., Yin, N., Liu, L., Qu, C., Li, W., Du, H., and Li, J. 2016. Genome-wide association analysis and differential expression analysis of resistance to *Sclerotinia* stem rot in *Brassica napus*. Plant Biotechnol. J. **14**: 1368–1380.

doi:10.1111/pbi.12501.

Wells, J., and Slade, P. 2021. The effect of the Canada–China canola trade dispute on canola prices. Can. J. Agric. Econ. **69**: 141–149. doi:10.1111/cjag.12258.

Westerhaus, M.O. 1989. Validation. Page 40 *in* G.. Marten, J.S. Shenk, and F.E. Barton, eds. Near infrared reflectance spectroscopy (NIRS): analysis of forage quality. U.S. Department of Agriculture, Washington, DC.

WHO 2007. Protein and amino acid requirements in human nutrition.

Wickham, H. 2016. ggplot2: elegant graphics for data analysis. Springer-Verlag New York.

Williams, P., and Norris, K. (eds.) 2001. Near-infrared technology in the agricultural and food industries. 2nd edition. American Association of Cereal Chemists, St. Paul, MN.

Williams, P.C., and Sobering, D.C. 1996. How do we do it: a brief summary of the methods we use in developing near infrared calibrations. Pages 185–188 *in* Proceedings of the 7th International Conference on Near Infrared Spectroscopy. Montreal, CA.

Windham, W.R., Mertens, D.R., and Barton, F.E. 1989. Protocol for NIRS calibration: sample selection and equation development and validation. Pages 69–103 *in* G.. Marten, J.S. Shenk, and F.E. Barton, eds. Near infrared reflectance spectroscopy (NIRS): analysis of forage quality. U.S. Department of Agriculture, Washington, DC.

Withana-Gamage, T.S., Hegedus, D.D., Qiu, X., Yu, P., May, T., Lydiate, D., and Wanasundara, J.P.D. 2013. Characterization of *Arabidopsis thaliana* lines with altered seed storage protein profiles using synchrotron-powered FT-IR spectromicroscopy. J. Agric. Food Chem. **61**: 901–912. doi:10.1021/jf304328n.

Wittkop, B., Snowdon, A.R.J., and Friedt, A.W. 2009. Status and perspectives of breeding for enhanced yield and quality of oilseed crops for Europe. Euphytica, **170**: 131–140.

doi:10.1007/s10681-009-9940-5.

Wittkop, B., Snowdon, R.J., and Friedt, W. 2012. New NIRS calibrations for fiber fractions reveal broad genetic variation in *Brassica napus* seed quality. J. Agric. Food Chem. **60**: 2248–2256. doi:10.1021/jf204936f.

Workman, J., and Weyer, L. 2012. Practical guide and spectral atlas for interpretive near-infrared spectroscopy. 2nd edition. CRC Press Taylor & Francis Group, Boca Raton, FL.

Wu, G., Fanzo, J., Miller, D.D., Pingali, P., Post, M., Steiner, J.L., and Thalacker-Mercer, A.E. 2014. Production and supply of high-quality food protein for human consumption: Sustainability, challenges, and innovations. Ann. N. Y. Acad. Sci. **1321**: 1–19. doi:10.1111/nyas.12500.

Wu, J., and Muir, A.D. 2008. Comparative structural, emulsifying, and biological properties of 2 major canola proteins, cruciferin and napin. J. Food Sci. **73**: C200–C216. doi:10.1111/j.1750-3841.2008.00675.x.

Wu, J.G., Shi, C.H., and Zhang, H.Z. 2005. Genetic analysis of embryo, cytoplasmic, and maternal effects and their environment interactions for protein content in *Brassica napus* L. Aust. J. Agric. Res. **56**: 69–73. doi:10.1071/AR04089.

Xiao, Y., Cai, D., Yang, W., Ye, W., Younas, M., Wu, J., and Liu, K. 2012. Genetic structure and linkage disequilibrium pattern of a rapeseed (*Brassica napus* L.) association mapping panel revealed by microsatellites. Theor. Appl. Genet. **125**: 437–447. doi:10.1007/s00122-012-1843-5.

Yalow, B.R.S., and Berson, S.A. 1960. Immunoassay of endogenous plasma insulin in man. J. Clin. Invest. **39**: 1157–1175.

Yao, M., Guan, M., Zhang, Z., Zhang, Q., Cui, Y., Chen, H., Liu, W., Jan, H.U., Voss-Fels, K.P.,

Werner, C.R., He, X., Liu, Z., Guan, C., Snowdon, R.J., Hua, W., and Qian, L. 2020. GWAS and co-expression network combination uncovers multigenes with close linkage effects on the oleic acid content accumulation in *Brassica napus*. BMC Genomics, **21**: 1–12. doi:10.1186/s12864-020-6711-0.

Yu, B., Gruber, M., Khachatourians, G.G., Hegedus, D.D., and Hannoufa, A. 2010. Gene expression profiling of developing *Brassica napus* seed in relation to changes in major storage compounds. Plant Sci. **178**: 381–389. doi:10.1016/j.plantsci.2010.02.007.

Zelmer, C.D., and McVetty, P.B.E. 2009. Industrial products. Pages 343–359 *in* S.K. Gupta, ed. Biology and breeding of crucifers. CRC Press Taylor & Francis Group, Boca Raton, FL.

Zhang, J., Mason, A.S., Wu, J., Liu, S., Zhang, X., Luo, T., Redden, R., Batley, J., Hu, L., and Yan, G. 2015. Identification of putative candidate genes for water stress tolerance in canola (*Brassica napus)*. Front. Plant Sci. **6**: 1–13. doi:10.3389/fpls.2015.01058.

Zhang, Y., Peng, L., Wu, Y., Shen, Y., Wu, X., and Wang, J. 2014. Analysis of global gene expression profiles to identify differentially expressed genes critical for embryo development in *Brassica rapa*. Plant Mol. Biol. **86**: 425–442. doi:10.1007/s11103-014-0238-1.

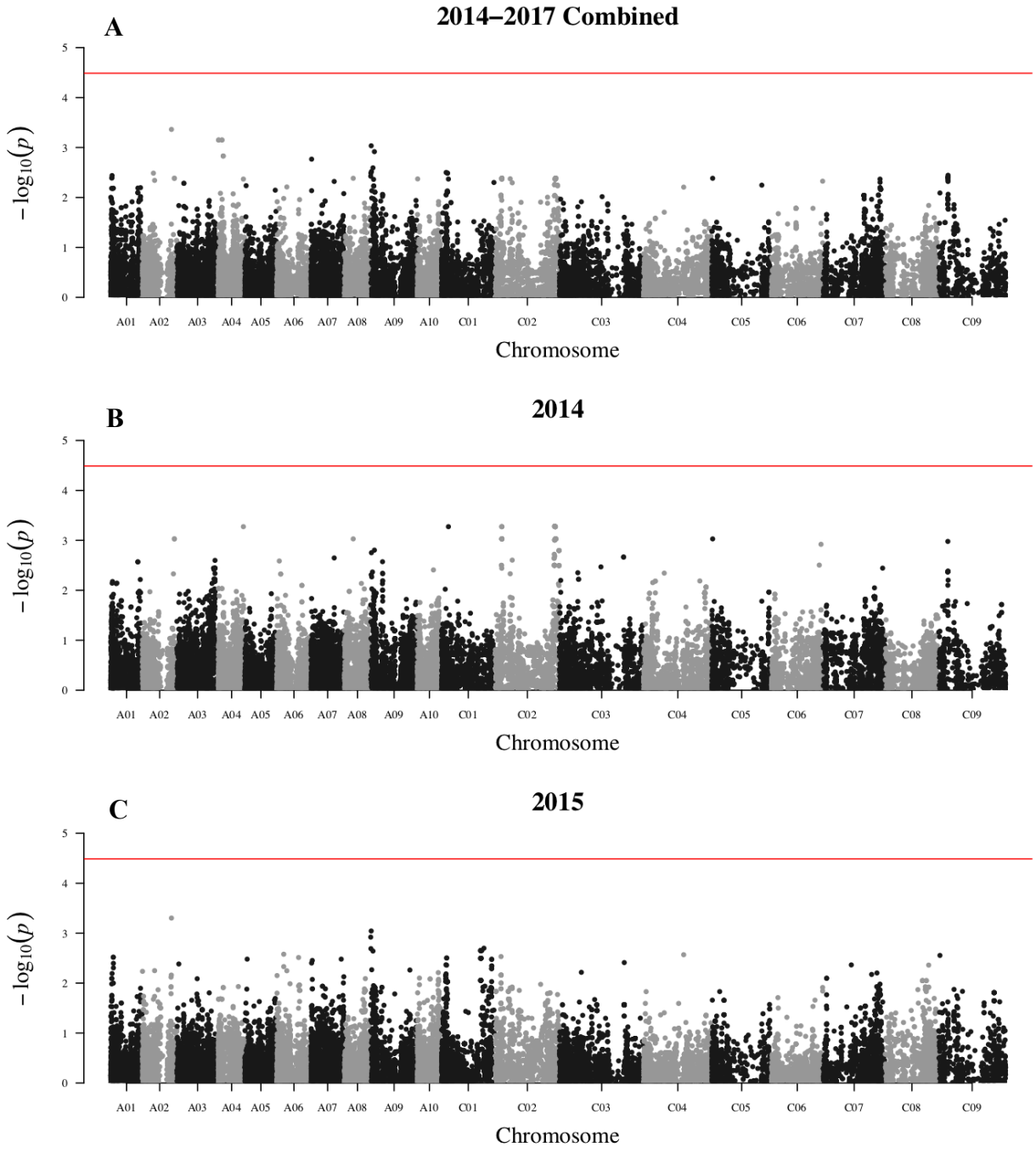
Zhao, J., Becker, H.C., Zhang, D., Zhang, Y., and Ecke, W. 2006. Conditional QTL mapping of oil content in rapeseed with respect to protein content and traits related to plant development and grain yield. Theor. Appl. Genet. **113**: 33–38. doi:10.1007/s00122-006-0267-5.

Zhao, K., Tung, C.W., Eizenga, G.C., Wright, M.H., Ali, M.L., Price, A.H., Norton, G.J., Islam, M.R., Reynolds, A., Mezey, J., McClung, A.M., Bustamante, C.D., and McCouch, S.R. 2011. Genome-wide association mapping reveals a rich genetic architecture of complex

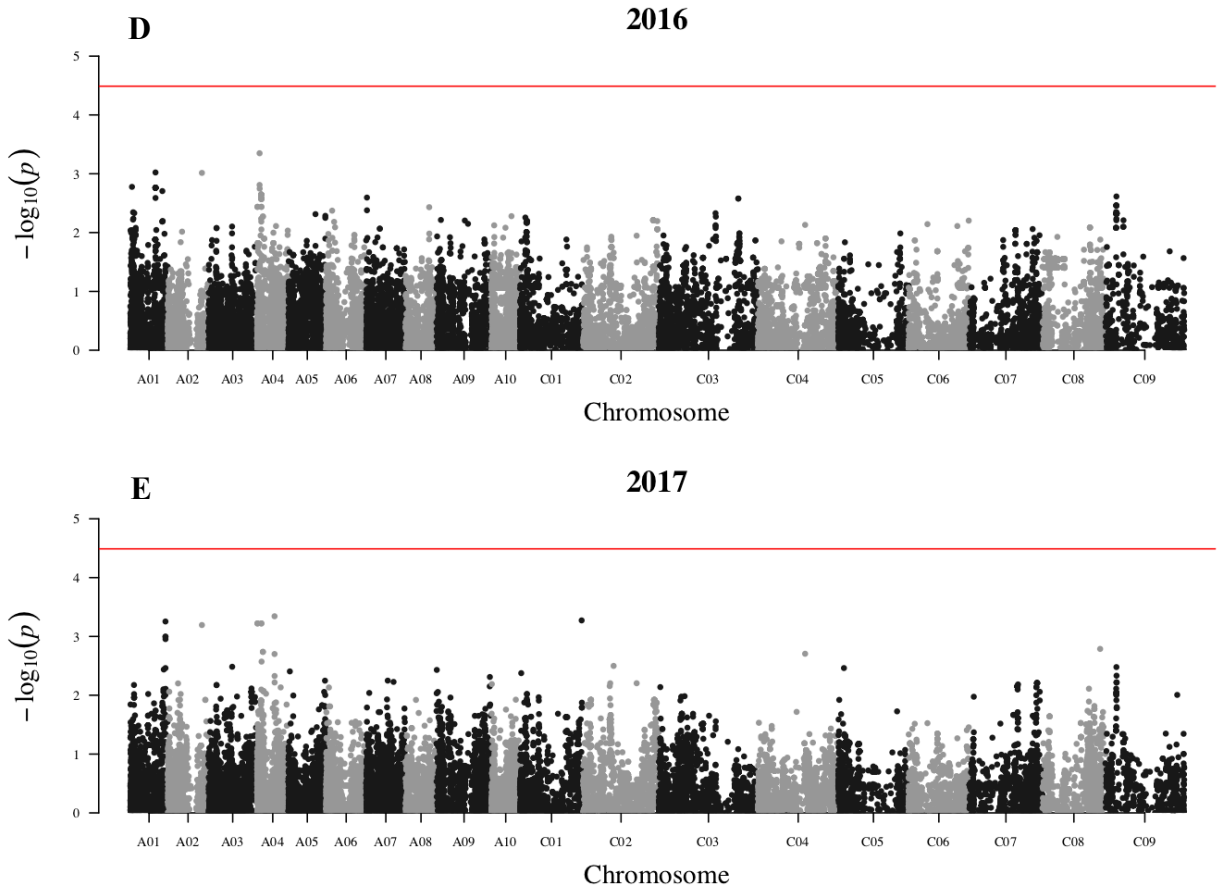
traits in *Oryza sativa*. Nat. Commun. **2**: 1–10. doi:10.1038/ncomms1467.

Zheng, M., Peng, C., Liu, H., Tang, M., Yang, H., Li, X., Liu, J., Sun, X., Wang, X., Xu, J., Hua, W., and Wang, H. 2017. Genome-wide association study reveals candidate genes for control of plant height, branch initiation height and branch number in rapeseed (*Brassica napus* L.). Front. Plant Sci. **8**: 1–13. doi:10.3389/fpls.2017.01246.

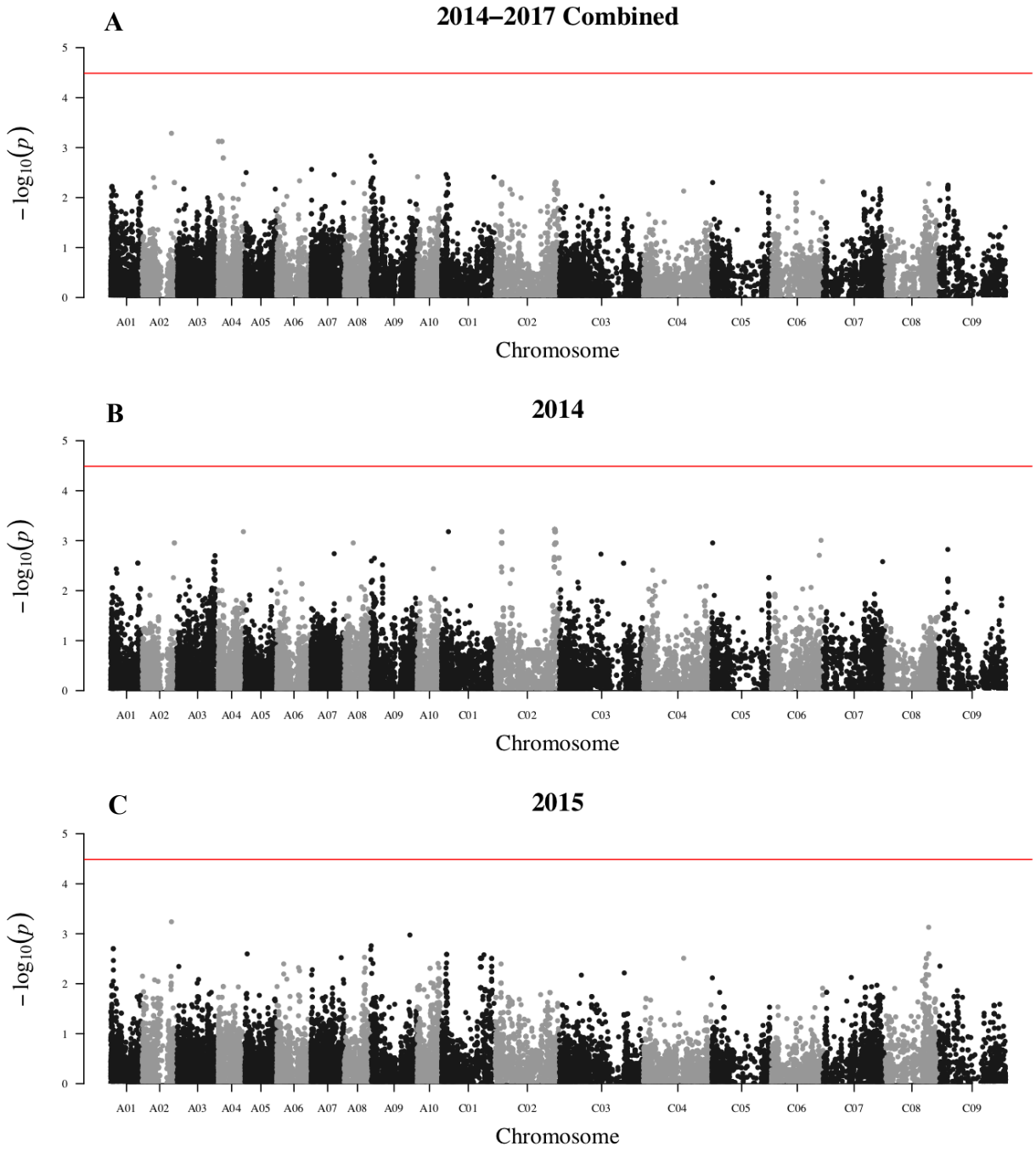
8.2 Appendices



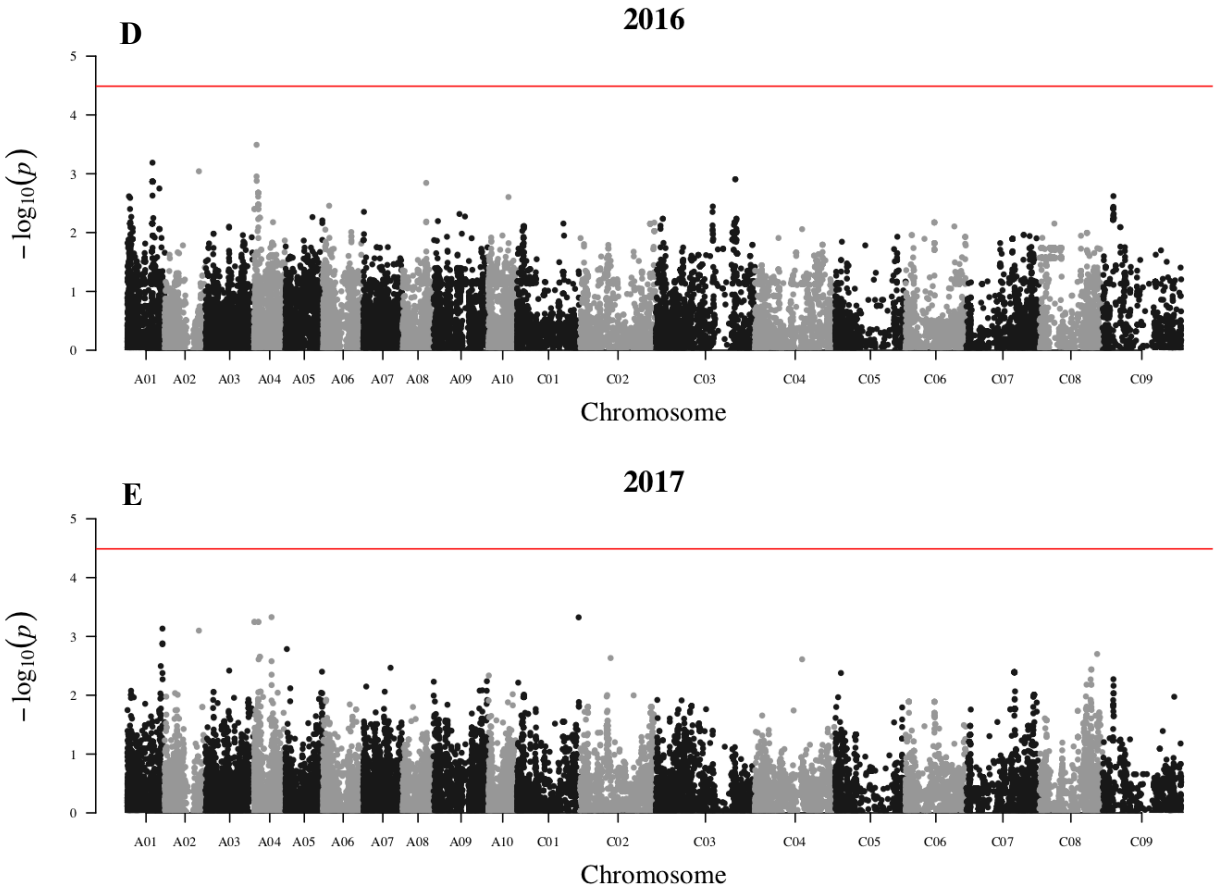
Appendix 5.1 Manhattan plots showing single nucleotide polymorphisms with significant associations to cruciferin content in the MLM+K model in: (A) 2014-2017 combined, (B) 2014, (C) 2015, (D) 2016, and (E) 2017. The horizontal red line represents the significance threshold of $-\log_{10}(P) = 4.5$.



Appendix 5.1 continued Manhattan plots showing single nucleotide polymorphisms with significant associations to cruciferin content in the MLM+K model in: (A) 2014-2017 combined, (B) 2014, (C) 2015, (D) 2016, and (E) 2017. The horizontal red line represents the significance threshold of $-\log_{10}(P) = 4.5$.



Appendix 5.2 Manhattan plots showing single nucleotide polymorphisms with significant associations to cruciferin content in the MLM+Q+K model in: (A) 2014-2017 combined, (B) 2014, (C) 2015, (D) 2016, and (E) 2017. The horizontal red line represents the significance threshold of $-\log_{10}(P) = 4.5$.



Appendix 5.2 continued Manhattan plots showing single nucleotide polymorphisms with significant associations to cruciferin content in the MLM+Q+K model in: (A) 2014-2017 combined, (B) 2014, (C) 2015, (D) 2016, and (E) 2017. The horizontal red line represents the significance threshold of $-\log_{10}(P) = 4.5$.

Appendix 5.3 Single nucleotide polymorphisms significantly associated with cruciferin content based on the genome-wide association model GLM+Q.

Chr†	Haplotype block	Start (bp)	End (bp)	Significant SNPs	Location(s) ‡
A01	1.1	19474151	19474151	Bn-A01-p22983721	2014
	1.2	19494342	19513183	Bn-A01-p22999151	2014
	1.3	21366875	21417205	Bn-scaff_15712_3-p818174	2017
A02	2.1	20834407	20834407	Bn-scaff_21884_1-p819724	ALL, 2015, 2017
	2.2	22242156	22242156	Bn-scaff_15918_1-p77494	2014
	2.3	22864476	22864476	Bn-scaff_17623_1-p32036	2014
	2.4	22987017	22987017	Bn-scaff_17721_1-p758760	2014
A03	3.1	15222567	15222567	Bn-scaff_16139_1-p1103077	2014
A04	4.1	171605	212801	Bn-A04-p2963387	ALL, 2017
				Bn-A04-p2999394	ALL, 2017
	4.2	2572063	2623118	Bn-A04-p2944723	ALL, 2017
				Bn-A04-p2954824	ALL, 2017
	4.3	3540219	3541177	Bn-A04-p3933714	ALL
				Bn-A04-p3934671	ALL
	4.4	10644573	10673100	Bn-A04-p9511386	2017
	4.5	17992643	17992643	Bn-scaff_17623_1-p234259	2014
A08	8.1	6402565	6402565	Bn-scaff_17721_1-p841884	2014
A09	9.1	22672	22672	Bn-A09-p596014	ALL, 2014, 2015
	9.2	376961	376961	Bn-scaff_19783_1-p1060183	ALL, 2014, 2016
	9.3	745841	745841	Bn-A01-p27004740	2014
	9.4	836994	836994	Bn-scaff_17526_1-p2206429	2014
	9.5	1287776	1300927	Bn-A09-p2116947	2014
	9.6	1637130	1637130	Bn-A09-p616919	ALL
	9.7	2026507	2079503	Bn-A09-p1929245	ALL, 2014
	9.8	2677575	2677575	Bn-A09-p2733282	ALL, 2014
	9.9	3308335	3308335	Bn-scaff_16139_1-p1102847	2014
C01	11.1	4983595	4983595	Bn-scaff_17623_1-p252454	2014
	11.2	37651052	37651052	Bn-scaff_15712_3-p639153	2017
C02	12.1	3529557	3529557	Bn-scaff_16139_1-p1103215	2014
	12.2	4482108	4482108	Bn-scaff_17623_1-p797758	2014
	12.3	4555662	4555662	Bn-scaff_17623_1-p837616	2014
	12.4	4645680	4645680	Bn-scaff_17623_1-p141721	2014
	12.5	4662183	4895386	Bn-scaff_15712_5-p1084403	2014
				Bn-scaff_15712_5-p1078471	2014

Appendix 5.3 continued Single nucleotide polymorphisms significantly associated with cruciferin content based on the genome-wide association model GLM+Q.

Chr†	Haplotype block	Start (bp)	End (bp)	Significant SNPs	Location(s) ‡
C02	12.5	4662183	4895386	Bn-scaff_15712_5-p1078277	2014
				Bn-scaff_15712_5-p1078140	2014
				Bn-scaff_17721_1-p831928	2014
				Bn-scaff_17721_1-p832008	2014
				Bn-scaff_17721_1-p906353	2014
				Bn-scaff_17721_1-p909770	2014
				Bn-scaff_17721_1-p909842	2014
				Bn-scaff_17721_1-p913943	2014
	12.6	5564924	6064087	Bn-scaff_16934_1-p98107	2014
	12.7	12058888	12253220	Bn-scaff_16139_1-p380924	2014
	12.8	17158582	17329880	Bn-scaff_16130_1-p696599	2014
				Bn-scaff_17079_1-p21658	2014
				Bn-scaff_17079_1-p28593	2014
				Bn-scaff_17079_1-p145741	2014
				Bn-scaff_21188_1-p9203	2014
	12.9	18593794	18594998	Bn-scaff_16369_1-p182886	2014
				Bn-scaff_16369_1-p181890	2014
	12.10	42315091	42316910	Bn-scaff_16162_1-p79842	2014
				Bn-scaff_16162_1-p78008	2014
	12.11	42420861	43658302	Bn-scaff_27946_1-p1047	2014
				Bn-scaff_17623_1-p1013272	2014
				Bn-scaff_17623_1-p1006646	2014
				Bn-scaff_17623_1-p966870	2014
				Bn-scaff_17623_1-p966804	2014
				Bn-scaff_17623_1-p956003	2014
				Bn-scaff_17623_1-p895939	2014
				Bn-scaff_17623_1-p894398	2014
				Bn-scaff_17623_1-p738869	2014
				Bn-scaff_17623_1-p715915	2014
				Bn-scaff_17623_1-p714325	2014
				Bn-scaff_17623_1-p706105	2014
				Bn-scaff_17623_1-p701672	2014
				Bn-scaff_17623_1-p701501	2014
				Bn-scaff_17623_1-p697243	2014
				Bn-scaff_17623_1-p697128	2014

Appendix 5.3 continued Single nucleotide polymorphisms significantly associated with cruciferin content based on the genome-wide association model GLM+Q.

Chr†	Haplotype block	Start (bp)	End (bp)	Significant SNPs	Location(s) ‡
C02	12.11	42420861	43658302	Bn-scaff_17623_1-p696488	2014
				Bn-scaff_17623_1-p688369	2014
				Bn-scaff_17623_1-p667042	2014
				Bn-scaff_17623_1-p664390	2014
				Bn-scaff_17623_1-p664320	2014
				Bn-scaff_17623_1-p663777	2014
				Bn-scaff_17623_1-p663609	2014
				Bn-scaff_17623_1-p662677	2014
				Bn-scaff_17623_1-p662478	2014
				Bn-scaff_17623_1-p256991	2014
				Bn-scaff_17623_1-p255719	2014
				Bn-scaff_17623_1-p255145	2014
				Bn-scaff_17623_1-p254765	2014
				Bn-scaff_17623_1-p254694	2014
				Bn-scaff_17623_1-p254624	2014
				Bn-scaff_17623_1-p254113	2014
				Bn-scaff_17623_1-p243181	2014
				Bn-scaff_17623_1-p213755	2014
				Bn-scaff_17623_1-p165270	2014
				Bn-scaff_17623_1-p73661	2014
				Bn-scaff_17623_1-p71246	2014
				Bn-scaff_17623_1-p71045	2014
				Bn-scaff_17721_1-p795449	2014
				Bn-scaff_17721_1-p794270	2014
				Bn-scaff_17721_1-p794032	2014
				Bn-scaff_17721_1-p792414	2014
				Bn-scaff_17721_1-p767812	2014
	12.12	45841870	45841870	Bn-scaff_16139_1-p670451	2014
	12.13	45848494	46044731	Bn-scaff_16139_1-p662939	2014
				Bn-scaff_16139_1-p660500	2014
				Bn-scaff_16139_1-p655249	2014
C03	13.1	29800829	29808168	Bn-scaff_18917_1-p430795	2014
C05	15.1	585524	585524	Bn-scaff_17721_1-p811886	2014

Appendix 5.3 continued Single nucleotide polymorphisms significantly associated with cruciferin content based on the genome-wide association model GLM+Q.

Chr†	Haplotype block	Start (bp)	End (bp)	Significant SNPs	Location(s) ‡
C06	16.1	34271977	34271977	Bn-scaff_17799_1-p2273825	2014
	16.2	35525817	35525817	Bn-scaff_17799_1-p1070542	2014
C09	19.1	6552767	6624169	Bn-scaff_17487_1-p381689	ALL, 2014
				Bn-scaff_17487_1-p382382	ALL, 2014
				Bn-scaff_17487_1-p383153	2014
				Bn-scaff_17487_1-p384829	2014
				Bn-scaff_17487_1-p415797	2014
				Bn-scaff_17487_1-p416361	2014
				Bn-scaff_17487_1-p421663	ALL, 2014
				Bn-scaff_17487_1-p424557	ALL, 2014
	19.2	6640709	6640709	Bn-scaff_17487_1-p441639	ALL, 2014
	19.3	6674714	6899036	Bn-scaff_17487_1-p464536	ALL, 2014
				Bn-scaff_17487_1-p488896	ALL
				Bn-scaff_17487_1-p491031	ALL, 2014
				Bn-scaff_17487_1-p511597	ALL, 2014
				Bn-scaff_17487_1-p512535	ALL, 2014, 2017
				Bn-scaff_17487_1-p527387	ALL, 2014
				Bn-scaff_17487_1-p561594	ALL, 2014

†Chromosome.
‡2014 in Saskatoon, Saskatchewan, Canada. 2015 Saskatoon, Saskatchewan, Canada. 2016 in Winnipeg, Manitoba, Canada. 2017 in Winnipeg, Manitoba, Canada. ALL is the combined analysis for 2014-2017.