The Association of Seed Coat Color in Flax (*Linum usitatissimum* L.) Conditioned by the *b1* Locus with Flower Shape, Flower Color, Grain Yield, Fatty Acid Profile and its Linkage with Simple Sequence Repeat Markers (SSRs)

A Thesis Submitted to the College of
Graduate Studies and Research
In Partial Fulfillment of the Requirements
For the Degree of Masters of Science
In the Department of Plant Sciences
University of Saskatchewan
Saskatoon, Saskatchewan

By
Kayla Lindenback

PERMISSION TO USE

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

Requests for permission to copy or to make other use of material in this thesis in whole or part should be addressed to:

Head of the Department of Plant Sciences
51 Campus Drive
University of Saskatchewan
Saskatoon, Saskatchewan S7N 5A8

ABSTRACT

Previously seed coat color in flax has been used as a phenotypic marker for specialty quality traits and currently there is an increasing demand to use seed coat color in flax to market flax for human and animal nutrition uses. Seed coat color was studied to 1) understand the inheritance of seed coat color conditioned by the b1 locus, to 2) understand the relationship of other important flax traits with seed coat color as well as to 3) identify markers that are linked to seed coat color for future marker assisted selection of seed coat color.

Spearman's rank correlation and an allelism test was used to show the inheritance of the alleles at the *b1* locus. Bulked segregant analysis (BSA) was used to identify putatively linked markers with the *b1* locus, these were then screened on the CDC Bethune x M96006 recombinant inbred line population. Furthermore, the CDC Bethune x M96006 and CDC Bethune x USDA-ARS Crystal recombinant inbred line populations were used to identify any important flax traits that had a significant relationship with seed coat color.

It was shown that seed coat color conditioned by the *b1* locus was stably inherited and that b^{1vg} and b^{1} are allelic to one another. The results of the BSA showed that there were 17 candidates for linkage but when these markers were screened on the population only the Lu456 from linkage group (LG) six was identified to have linkage (χ^2=3.90; P<0.05) with the *b1* locus. Additionally, it was shown that the b^{1} seed coat color allele of the *b1* locus had a pleiotropic effect on flower color and flower shape and that seed coat color was associated with linolenic fatty acid content. None of the traits examined were found to be associated with the b^{1vg} allele of this locus.

These results show that the *b1* locus is likely present on linkage group six, more marker coverage on linkage group six of markers that are polymorphic between the two seed coat color parents would increase the accuracy of detection. Lastly, this study showed that plant breeders should consider using the b^{1vg} allele that conditions the variegated seed coat color to mark unique lines with important combinations of traits because it sorted independently for seed quality traits. Whereas, the yellow seed coat color conditioned by the b^{1} allele was found to be associated with higher linolenic fatty acid content and the semi-lethality of this allele would make it not suitable for use in parental lines.

ACKNOWLEDGEMENTS

First I would like to thank my supervisor Dr. Helen Booker for her support and guidance throughout this project. My gratitude goes out as well to the members of my committee Dr. Yuguang Bai, Dr. Bruce Coulman, Dr. Bunyamin Tar'an and Dr. Curtis Pozniak for their wealth of knowledge and helpful advice. I would also like to show appreciation to Dr. Bill McGregor for his mentorship on graduate studies as well as on the transition into plant sciences.

Thanks goes out to the collaborators of the flax seed coat color project at NRC-Saskatoon for their interest and constructive feedback on my work. I am grateful also to Lester Young the research associate of the Flax Breeding Program as well as the various technicians for all the time they spent helping me along the way. Additionally, I would like to give a special thanks to the graduate students Jenalee Mischkolz, Jia Sun, Kirby Nilsen, Courtney Thompson, Lisa Malo, Katherine Gui and Tao Zhang for their assistance at various stages of my research.

Financial support for this project from the Ministry of Agriculture through the Agriculture Development Fund and from the University of Saskatchewan is gratefully acknowledged.

Lastly, I would like to thank my parents and two sisters, Stephen, Sandra, Elyse and Tjyana. ☺

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AFLP	Amplified fragment length polymorphism
Bp	Base pair
BSA	Bulked segregant analysis
CDC	Crop Development Center
cM	Centimorgan
CTAB	Hexadecyltrimethylammonium bromide
Df	Degrees of freedom
DNA	Deoxyribonucleic acid
dNTP	Deoxyribonucleotide
EMS	Ethylmethane sulfonate
ESTs	Expressed sequence tags
FAD	Fatty acid desaturase
FAME	Fatty acid methyl ester
LG	Linkage group
LOD	Logarithm (base 10) of odds
NIR	Near infrared radiation
PCR	Polymerase chain reaction
PGRC	Plant Genome Resources Canada
RAPD	Random amplification of polymorphic DNA
Rcf	Relative centrifugal force
REP	Replication block
RIL	Recombinant inbred line
SNP	Single nucleotide polymorphism
SSR	Simple sequence repeat
Tufgen	Total Utilization Flax Genomics
USDA-ARS	United States Department of Agriculture - Agricultural Research Service

Chapter 1 Introduction

According to the Flax Council of Canada, flax (*Linum usitatissimum* L.) is one of the five major Canadian produced and exported crops. Canada is the number one producer of flax and according to statistics available from the Food and Agriculture Organization (FAO) Canadian farmers produced well over 700,000 tonnes as an annual average from the years 2000-2010.

For thousands of years flax has been grown for two separate utilities, its oil or its fibers. Fiber flax has long strong fibers within its stem that are increasingly important in the manufacturing of environmentally friendly products such as geotextiles, biofuels, biocomposites, and insulation. Oilseed flax, also known as linseed, has seeds that are high in fiber, contain 40-45% oil content, and 20-30% protein content. The oil is high in healthy polyunsaturated fatty acids, which are often also referred to as omega-3 fats. These qualities make flax oil and flax seed nutritionally beneficial to humans (Cunnane *et al.*, 1993). The oil has also proven to be industrially useful in oil-based paints and stains, lubricants and as well as in the curing of linoleum flooring. As well, the meal left behind after the seed is crushed is a healthy feed for cattle due its high protein content.

Canada primarily produces Linseed type flax, the majority of which is for export to other countries. The export of Canadian linseed has been valued at between 150-180 million dollars (SaskFlax 2013) for the 85% of flaxseed that is not consumed domestically (Ryan & Smyth, 2012). The province of Saskatchewan contributes up to 70% of Canadian production and export, and 25-30% of the world's total production (Food and Agriculture Organization of the United Nations, 2013). Most flaxseed exported is crushed in order to extract the oil for use in the industrial purposes listed in the previous paragraph; the left over meal is used in high quality animal feed. Historically, the European Union has imported the majority of the flax produced in Canada but since seeds of a genetically modified variety was detected in a shipment export to the EU has stopped. Now Canada mainly exports to the United States, China, the United Kingdom, Japan and Brazil (Ryan & Smyth, 2012). Of these markets Japan and Brazil are non-GM.

Until recently there have been two types of linseed flax, characterized by their seed coat color: brown and yellow. Regular flax has a brown seed coat color and contains greater

than 50 percent linolenic (omega-3) fatty acid in their oil. Yellow seeds were previously required to identify Solin, flax bred using mutation breeding from regular flax to contain less than five percent linolenic fatty acid, by the Canadian Grain Commissions standards (Mittapalli & Rowland, 2003). Solin was developed because it was more suitable for margarine and shortening than traditional flax oils (Saeidi & Rowland, 1997). However, there is no longer a market for Solin flax and there is shifting interest into developing a yellow seeded high linolenic linseed flax variety because the food market prefers the yellow seed coat as well as a healthier oil profile.

Due to this new demand for high linolenic yellow linseed, the Canadian Grain Commissions has removed the requirement for yellow seeds to be a phenotypic marker for the Solin oil profile. This requirement can be removed because the two traits are not linked and randomly assort (Saeidi & Rowland, 1999b). However, the requirement for farmers to have uniform seed at grading has not been removed. This is causing grade reductions in farmers' seed when there are amounts of yellow seed "contaminants" in their brown seed. This grade reduction may be unnecessary because it is likely that the brown and yellow seeds are not compositionally different from each other. Therefore, it is important for researchers to understand the genetic basis behind seed coat color development and its association with other important agronomic, quality, and morphological traits. This would allow for the mixing of seed coat colors by farmers without having grading penalties or the breeding of flax varieties with unique and useful combinations of quality traits and seed coat colors.

The following are the hypotheses tested by the research completed as partial fulfillment of my M.Sc.:

(1) The alleles b^1 and b^{1vg} will be found to be allelic to each other;
(2) Seed coat color will be found to have a significant relationship with flower color and shape, seed weight, oil content and all components of oil profile, but not grain yield for these populations;
(3) SSRs linked with seed coat color will be identified in the CDC Bethune x M96006 RIL population.

The research I carried out was with the following objectives:

(1) To confirm allelism of the two known alleles of the *b1* locus

(2) To identify relationships between seed coat color and other important flax quality traits;

(3) To locate the position of the *b1* seed coat color locus in the genome by identifying linked simple sequence repeat markers (SSRs).

Chapter 2 Literature Review

2.1 History of Cultivated Flax

It is believed that flax was first domesticated 8000 years ago in southeastern Turkey and southeastern Iran (Helbaek, 1959; van Zeist & Bakker-Heers, 1975). Molecular research of the *sad2* locus, which is known to influence unsaturated fatty acid content of the seed oil, suggests that flax was first cultivated for oil-use purposes (Allaby *et al.*, 2005). Further research using the *sad2* locus also has shown that fiber-types diverged from the oil-type 3,000 years ago.

"Pale flax" (*Linum bienne*), the wild progenitor of flax (*Linum usitatissimum* L.) (van Zeist & Bakker-Heers, 1975), is native to Turkey (Helbaek, 1959) and has smaller seeds than cultivated flax. Pale flax has a winter annual or perennial growth habit and is characterized by small seeds, narrow leaves, dehiscent capsules, highly variable vegetative plant parts, and large variation in growth habit (Diederichsen & Hammer, 1995). This is compared to cultivated flax that is characterized as being a summer annual that generally tends to have large seeds with variable dormancy, indehiscent capsule, fast growth, and highly variable vegetative plant parts (Diederichsen & Fu, 2006). Additionally, oil- and fiber-types of cultivated flax differ in their winter hardiness and capsular dehiscence (Diederichsen & Fu, 2006) because of divergent evolution due to selections for their different desired traits since cultivation.

Other molecular studies of homologous segments of several plant genomes including flax elucidated that flax, now a self-pollinating diploid with 15 sets of chromosomes, was once an ancient polyploid (Lin & Paterson, 2011). Further evidence for this ancestral duplication event was discovered when in some cases two paralogous loci were amplified using a subset of recently developed molecular markers (Cloutier *et al.*, 2009). This study showed the importance of understanding the history and evolution of flax to help interpret results of genetic studies.

2.2 Seed Coat Color Inheritance Studies

In flax, seed coat color was initially considered to be governed by 3 different factors, G, D and B', where D and B' were modifying factors of the primary factor G. The G factor was thought to be responsible for the brown seed coat, while D had an inhibitory

effect and when present with G would result in a grayish phenotype. The presence of G with D and B' would result in a brown seed coat because the B' factor would cancel out the inhibitory action of D. When G was not present, the result would be no seed coat color and the yellow cotyledons would be visible through the seed coat, resulting in the yellow type (Tammes, 1922).

Further studies identified a fourth factor, which was called M by some groups and Sc by another (Afzal Naz, 1976; Beard & Comstock, 1965; Shaw *et al.*, 1931). It was concluded that yellow was the basal color in flax seed, and when any of the four factors were present it would result in a colored phenotype. For example if factors M and D were present together it would result in a fawn color, while if G, D and M were all present it would be brown (Shaw *et al.*, 1931).

The further understanding of DNA and genetics, has led to better clarification of seed coat color inheritance. It became known that these G, D and B' factors were in fact different loci that controlled seed coat color in the genome. Furthermore, it was discovered that brown was the dominant phenotype and yellow seed coat was a result of one or more of these loci being in the homozygous recessive condition (Barnes *et al.*, 1960). A more recent study by Mittapalli and Rowland (2003) found that seed coat color is in fact conditioned by four independently inherited loci, which they called *Y1, g, d,* and *b1*. At the *Y1* locus there is a dominant yellow allele while the *g, d* and *b1* all have recessive yellow alleles. In all these cases the alternative phenotype is always brown. They also found what they suggest to be a second recessive allele for the *b1* locus that they called b^{1vg}. This allele conferred a variegated or mottled seed coat color (Saeidi & Rowland, 1997). They concluded that yellow was epistatic to variegated at the loci where yellow was a recessive trait. The variegated trait is always recessive. Additionally, when all other loci were homozygous for the recessive allele they were epistatic to the one carrying a dominant allele and therefore would result in a yellow phenotype.

With advancements in molecular genetic research, flax molecular marker development and flax genomic sequencing it is now possible to map these loci as well as identify putative genes present that might be responsible for controlling seed coat color.

2.3 Flavonoid Biosynthesis

Flavonoids are a class of plant secondary metabolites known to have a role in red, blue and purple pigments in seeds and flowers (Winkel-Shirley, 2001). These molecules have important roles in embryo protection by filtering UV radiation or for pollinator attraction through coloration of the flowers. They are also thought to have roles in seed dormancy, seed viability, male fertility, regulation of auxin transport, antimicrobial activity, signaling between plants and microbes, and their bitter taste decreases its palatability to animals trying to eat it (Lepeniec *et al.*, 2006; Winkel-Shirley, 2001). Additionally, flavonoids are known to affect the densities of some of the different tissues of the seed (Buer *et al.*, 2010), they can influence plant architecture (Taylor & Grotewold, 2005) and they have antioxidant activity which plays a role in human health when ground seeds are consumed (Ross & Kasum, 2002).

Flavonoids are aromatic molecules are derived from phenylalanine and malonyl-coenzyme A and are represented by the chalcones, flavones, flavonols, flavandiols, anthocyanins and condensed tannins (Winkel-Shirley, 2001) found in all higher plants including flax. There is not much known about these compounds in flax including their biosynthetic pathway and their associated genes. But there is a lot known in *Arabidopsis*.

Arabidopsis has been used as a model to study the flavonoid biosynthetic pathway (Fig. 2.1) because it is a well-known model and all the genes of this pathway are present in single copy (Winkel-Shirley, 2001). It is known that condensed tannins are responsible for the wild-type brown seed coat colors in *Arabidopsis.* The accumulation and oxidation of the condensed tannins, which function as an embryotic protection in the seed coat during desiccation, leads to this formation of brown pigments in mature *Arabidopsis* seeds (Debeaujon *et al.*, 2003; Lepeniec *et al.*, 2006). These molecules when present in differing levels are responsible for seed coat colors ranging from tan to yellow (Debeaujon *et al.*, 2000). Several genetic loci known to have a role in the biosynthetic pathway have been identified in *Arabidopsis*. These genes cause the formation of flavonoids and when mutagenized result in absence or reduced pigmentation of the seed coat (Table 2.1) (Winkel-Shirley, 2001).

There are 26 seed coat color associated genes in *Arabidopsis*. These genes were found through mutational screens. All genes were found to be recessive and were expressed

in the maternal tissues of the seed coat (Lepeniec *et al.*, 2006). The mutants were all named as *transparent testa (tt)* mutants because the seed coat of the mutant phenotype is transparent which leads to the yellow cotyledons showing through the coat (Diederichsen & Richards, 2003; Tammes, 1922). In table 2.1 it is evident that these genes fall into three main categories, enzyme coding genes, transcriptional factor genes and then others of unknown function. Seventeen of the genes of known function all have important roles in condensed tannin synthesis, structure, regulation and compartmentalization. Conversely, in legumes and in *Brassica carinata* it is known that chalcone synthase is the primary gene product involved in seed coat pigmentation, as well as flavonone 3'-hydroxylase has also been found to have a role (Moise *et al.*, 2005).

The genes found in *Arabidopsis* can be used in addition to *Arabidopsis* flavonoid biochemical pathways (Fig. 2.1) for homology and candidate gene determination in flax (Lepeniec *et al.*, 2006) to help elucidate flax flavonoid genes and its biochemical pathway.

Table 2.1. The seventeen seed coat color associated loci in *Arabidopsis* identified through screening for mutants with impaired seed pigmentation. (Used with permission from Copyright Clearing Center, © 2006, Annual Review of Plant Biology, Palo Alto, California.)

	Locus	Seed Coat Color	Gene Product
Protein-coding	*tt3*	Yellow	Dihydroflavonol reductase (DFR)
	tt4	Yellow	Chalcone synthase (CHS)
	tt5	Yellow	Chalcone isomerase (CHI)
	tt6	Pale brown spotted	Flavanone-3-hydroxylase (F3H)
	tt7	Pale brown spotted	Flavanone-3'-hydroxylase (F3'H)
	tt10	Dark yellow/brown C*	Polyphenol oxydase (PPO)
	tt12	Pale brown	MATE secondary transporter
	tt15	Pale brown/brown CM*	Glycosyltransferase (GT)
	tt18/tds4/tt11	Yellow	Leucoanthocyanidin dioxygenase (LDOX)
	tt19/tt14	Dark yellow	Glutathione S-transferase (GST)
	ban	Pale grey/grey CM*	Anthocyanidin reductase (ANR)
	aha10	Pale brown	Autoinhibited H^+-ATPase isoform 10
Transcription factors	*tt1*	Yellow/brown CM*	Transcription factor WIP-type Zn-Finger
	tt2	Yellow/brown CM*	Transcription factor AtMYB123
	tt8	Yellow	Transcription factor AtbHLH042
	tt16/abs	Yellow	Transcription factor MADS AtAGL32
	ttg1	Yellow	Regulatory Protein ("WD40" or "WDR")
	ttg2	Yellow	Transcription factor AtWRKY44
Unknown	*tt9*	Pale grey/dark CM*	Unknown
	tt13	Pale brown	Unknown
	tt17	Pale brown	Unknown
	tds1,3,5,6	Pale brown	Unknown
	tds2	Pale brown	Unknown

*C, Chalaze; M, Micropyle

Source: Lepeniec et al., 2006 Annual Review of Plant Biology 57:405-430

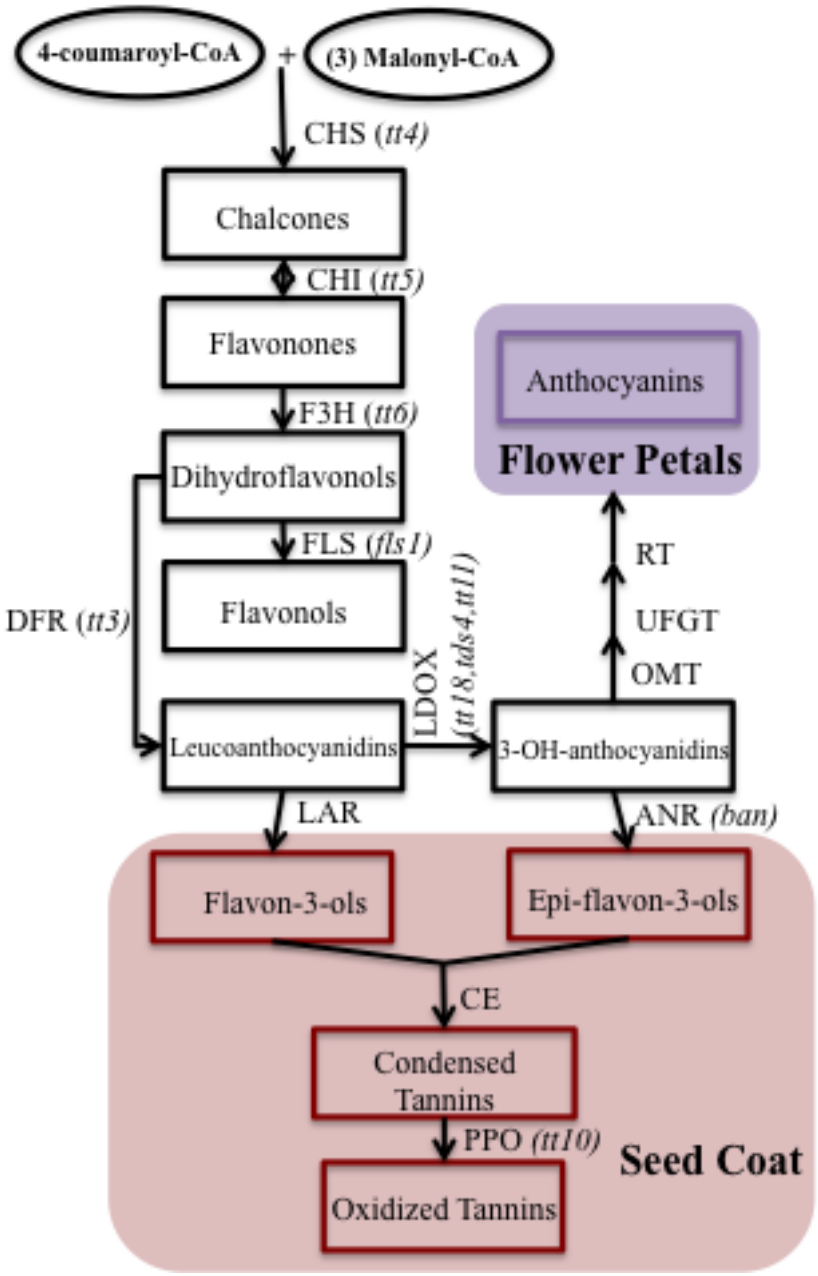


Figure 2.1. *Arabidopsis* flavonoid biosynthetic pathway.

Enzyme names are abbreviated as follows: chalcone synthase (CHS), chalcone isomerase (CHI), flavanone-3-hydroxylase (F3H), flavonols synthase (FLS), dihydroflavonol reductase (DFR), leucoanthocyanidin dioxygenase (LDOX), O-methyltransferase (OMT), UDPG-flavonoid glucosyl transferase (UFGT) rhamnosyl transferase (RT), leucoanthocyanidin reductase (LAR), anthocyanidin reductase (ANR), condensation enzyme (CE) and polyphenol oxidase (PPO). (Used with permission from the Copyright Clearance Center, © 2006, Annual Review of Plant Biology, Palo Alto, California and D. McCauley, © 2001, Plant Physiology, Rockville Maryland.)

2.4 Seed Coat Structure

When studying seed coat color it is not only important to know the biochemistry behind the production of the pigment molecule and the genes involved, it is also important to understand the tissues in which those genes are expressed. Again, not much literature is available on flax seed coat structure but there is information available in *Arabidopsis*, legumes and the *Brassicaceae* family.

Seed coats of all plants develop from the maternal integuments and the chalazal tissue of the ovule after fertilization (Moise *et al.*, 2005). This means that seed coat color is maternally derived (Lepeniec *et al.*, 2006) and thus for example, the genotype and therefore the color of the seed coat surrounding the F1 embryo is the maternal genotype. As the seed develops the seed coat and endosperm develops first, followed by the development of the embryo after which, maturity of the seed coat and of the embryo occur (Weber *et al.*, 2005).

The seed coats of *Arabidopsis* have 4 distinct layers. Working from the outside of the seed under the epidermal layer is the outermost layer of the outer integument. This layer is often only one cell layer thick and contains the mucilage in flax and *Arabidopsis*, for this reason it is called the mucilaginous layer. The second layer is innermost layer of this outer integument is also usually one cell thick; it is a sclerotic and often called the palisade layer. This layer plays a role in the permeability of the seed to water (Ma *et al.*, 2004). These two layers were both of the outer integuments of the ovule; the next two layers are of the inner integuments. The third layer is called the parenchymatic and the fourth is the endothelium. These layers are made up of parenchyma cells that are compressed together at maturity. These two layers are often collectively called the pigment layer because this is where the flavonoids accumulate (Bouman, 1975; Lepeniec *et al.*, 2006; Vaughan & Whitehouse, 1971).

The information available about the seed coats of *Arabidopsis,* legumes and the *Brassicaceae* family is helpful and useful in understanding generalities amongst different species seed coat structure, but it is not sufficient to provide specific information about each crop. By further genetic studies looking at seed coat color through genes involved in pigmentation of the inner integument of flax, this process will be better understood.

2.5 Grading Regulation of Seed Coat Color in Flax

The Canadian Grain Commission (CGC) is the scientific research organization responsible for monitoring the quality of flaxseed that is produced an exported. Additionally, it is also the official certifier of all Canadian grains. Canadian grain is known worldwide for its reputation for consistent and reliable quality. To assure the quality of flaxseed with regards to seed coat color, the CGC had set limits to the amount of off-type contaminants for the different grade classes of flax (Table 2.2). The grade that the grain receives determines the amount that a producer will be paid for their harvest. As well it also determines the end uses of grains where higher grades are used for human consumption and lower grades are used for animal feeds.

Table 2.2. Canadian Grain Commission frequency limit of off-type seed coat color for the grading of brown and yellow flaxseed.

Grade	**Brown Seed**	**Yellow Seed**
	% Yellow	% Brown
No. 1	2	2
No. 2	3	2
No. 3	4	2
Grade if No. 3 not met	50% or less-Flaxseed 50% or over-Solin	

Source: Official Grain Grading Guide, 2012. Canadian Grain Commission

2.6 Methods for Measuring Flax Seed Coat Color

There are several methods used for measuring flax seed coat color. Most commonly measurements with three parameters are taken to differentiate seed coat colors from one another (Wiesnerova & Wiesner, 2008). Traditionally, a qualitative approach where seeds are scored visually and assigned a descriptor based on a predefined scale was used (Wiesnerova & Wiesner, 2008). However, there are numerous and complex factors influencing seed coat color (Diederichsen & Richards, 2003). In order to unambiguously describe minute differences caused by these numerous factors, quantitative measurements can be used (Wiesnerova & Wiesner, 2008).

A system for qualitative measurement of seed coat color in flax has been established. It involves using a seven-point descriptor list based on Munsell color standards

(Diederichsen & Richards, 2003). The Munsell color system is over 100 years old and it is based on a standard sequence of *hue*, *value* and *chroma* measurement (Kuehni, 2002). There are five principle *hues* that are arranged into horizontal rings: red, yellow, green, blue and purple. Intermediates can also be measured that are combinations of these principle *hues*. Within each *hue* you can measure the *value* and *chroma*. *Value* is measured vertically within the *hue* circles and ranges from 0 for black and 10 for white. *Chroma* is related to saturation and is measured radially within the *hue* circles. The seven-point descriptor list uses these Munsell measurements to categorize seed coat color on a scale of (1) light brown, (2) brown, (3) dark brown, (4) yellow, (5) olive, and (6) variegated (Diederichsen & Richards, 2003). In this scale brown is separated into three categories because it is the most common flax seed coat color.

When qualitative color measurement is not sufficient a quantitative measurement can be taken to distinguish the continuous transitions between seed coat colors. Recently the L*, a* and b* gamets of the Commission Internationale de l'Eclairage (CIE) color coordinates are used to measure color opponent dimensions quantitatively. On this scale L* measures lightness from 0-100 where 0 is white and 100 is black, a* measures the green to magenta color opposition and b* measures yellow to blue. Using these measurement parameters the L* and a* gamets alone have been found to describe three quarters of the total variability of seed coat color (Wiesnerova & Wiesner, 2008).

When measuring seed coat color, the method used will depend on the samples to be measured and the type of analyses that will be performed. As shown above, qualitative measurements are sufficient but when it is important to distinguish between minute differences in seed coat color, then a more mechanized form of quantitative color measurement is best for unambiguous results.

2.7 Correlation of Flax Seed Coat Color with Other Traits

Seed coat color in flax has been found by several different researchers to be associated with important breeding traits. It is important to understand these associations because these traits may be genetically linked. This means that when you select for seed coat color, you may be selecting for another trait at the same time. The genes involved in seed coat color may even be pleiotropic, meaning that a single gene may play a role in the expression of multiple phenotypic traits. By understanding the relationships between seed

coat color and other traits, breeders can use seed coat color as a phenotypic marker, or they can avoid making selections for deleterious traits when selecting for color. It can also show that seed coat color is not linked with many quality traits and therefore grade reductions due to mixing of seeds with different coat colors need not occur.

Yield is a complicated trait as genes involved in yield are pleiotropic to many other traits such as, height, germination, stand and maturity. Studies in *Brassicaceae* showed that the absence of flavonoids in the yellow seed coat lines is correlated with higher grain yield (Simbaya *et al.*, 1995). However, in flax the opposite was found where lines with yellow seed coats tended to have lower yields (Culbertson & Kommendahl, 1956), which was attributed to decreased seed vigor, poorer stands (Saeidi & Rowland, 1999b) and lower germination rate (Saeidi & Rowland, 1999a). No published studies have been done that compare the yield of variegated seed coat color to that of brown or yellow.

Oil profile is another trait that has associations to seed coat color. In the case of Solin, yellow seed coat color was used as a phenotypic marker for a low linolenic fatty acid profile. In an EMS mutation study a line with high palmitic fatty acid content was found with a variegated seed coat color. However, further investigation into this line showed that that high palmitic fatty acid trait segregated independently from the variegated seed coat color trait (Saeidi & Rowland, 1997). It was originally thought that variegated seed coat color could be a useful marker to distinguish high palmitic and low linolenic oil profiles from that of Solin or traditional brown seeds (Saeidi & Rowland, 1997), but a market for this type of oil never developed. It is possible, however, that variegated seed could be a marker for a high linolenic oil profile that is in demand for the current health market.

Additionally, another trait known to have associations with seed coat color is oil content. Oil content is the amount of oil within the cotyledons of the embryo, and is a very important trait to linseed type flax. In the *Brassicaceae* family yellow seed coat lines have been found to have higher oil content (Simbaya *et al.*, 1995). This is also the case in flax but only in seeds with a high linolenic fatty acid profile (Culbertson & Kommendahl, 1956; Saeidi & Rowland, 1999b). This higher oil content was found not to be due to larger seed size because seed size, unlike oil content, was found not to be associated with seed coat color in most populations (Saeidi & Rowland, 1999b). It is generally accepted that this

higher oil content is due to thinner seed coats resulting from the absence of flavonoids in the inner seed coat layers.

As already discussed, yellow seed coats are a result of loss of pigment in the layers of the seed coat structure. These pigments are a product of the flavonoid biosynthesis biochemical pathway that is also responsible for the pigmentation of the flower (Lepeniec *et al.*, 2006; Taylor & Grotewold, 2005; Winkel-Shirley, 2001). Therefore, it is sensible that the *b1* locus has been found to be pleiotropic to seed coat color, flower color and flower shape (Barnes *et al.*, 1960; Culbertson & Kommendahl, 1956), where the USDA-ARS Crystal flax variety (Arny, 1944) has yellow seeds and crimped white flowers and carries the recessive b^1 allele. This same pleiotropic effect was not mentioned in a study done with another flax line that is suggested to carry the b^{1vg} recessive allele of the *b1* locus (Mittapalli & Rowland, 2003; Saeidi & Rowland, 1997).

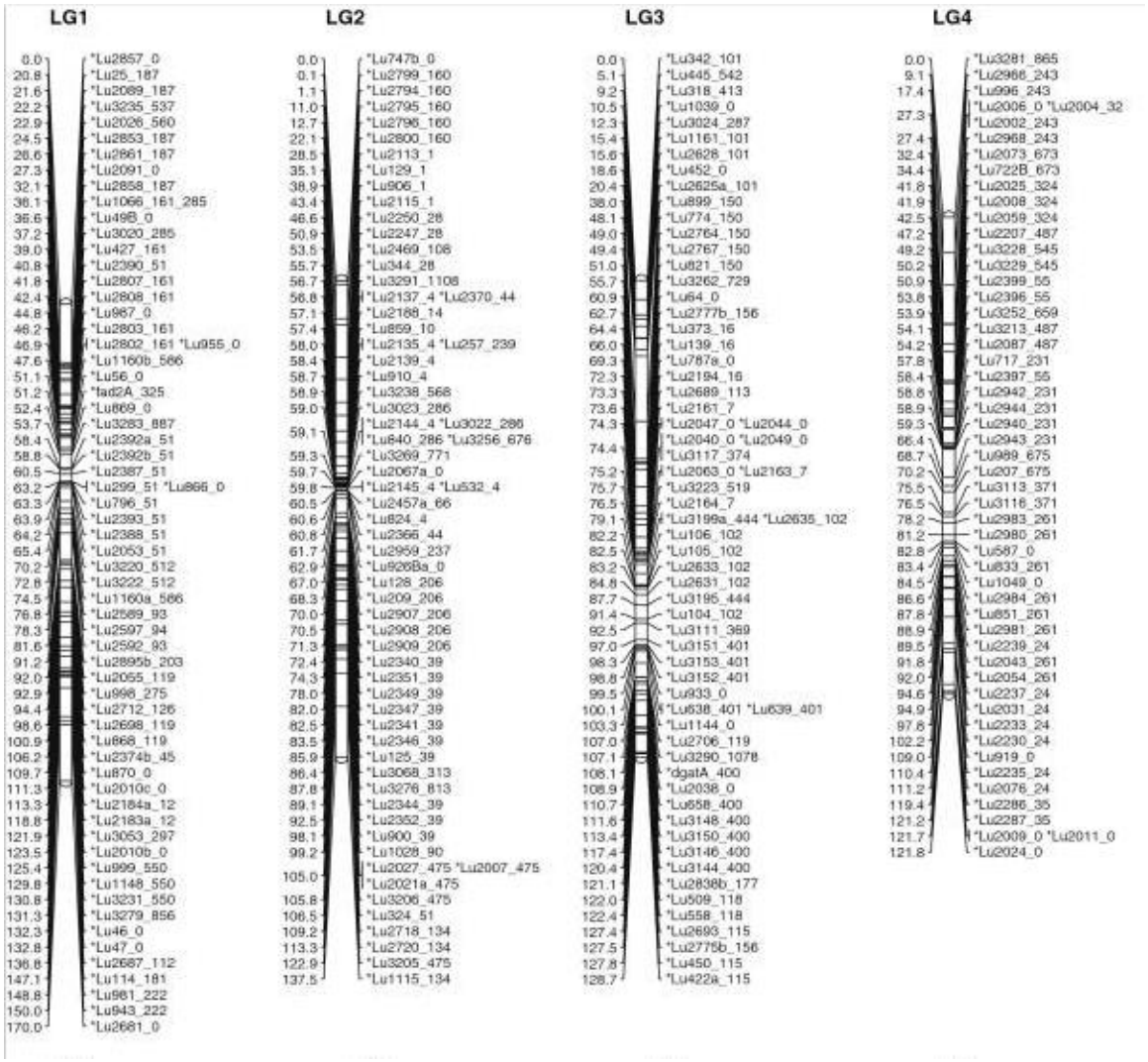
2.8 Molecular Markers

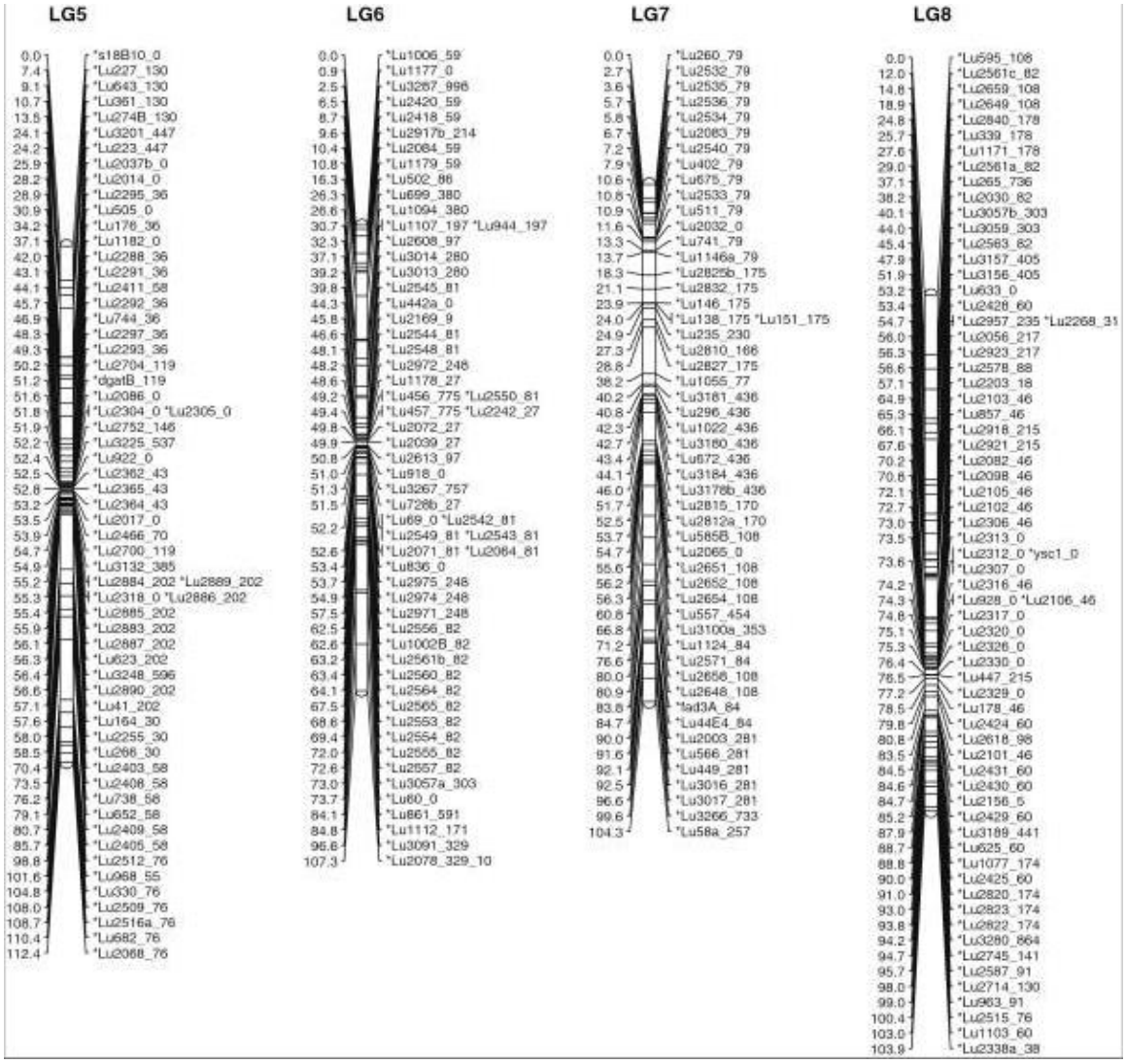
Molecular markers are useful for studying variation in crops because they are highly abundant throughout the genome; they are present in all cells of all tissues and they are not influenced by environment. Until recently, few molecular markers had been developed for use in flax. The previous markers available had been randomly amplified polymorphic DNA (RAPD), restriction fragment length polymorphism (RFLP), very few simple sequence repeat (SSR), and no single nucleotide polymorphism (SNP) type markers (Cloutier *et al.*, 2009). The markers that were available were limited in number and were not overly informative (Cloutier *et al.*, 2012b). Recently, work to further develop flax genetic resources was a focus of an international project called Total Utilization Flax Genomics (Tufgen). This project brought the efforts of several research groups together and over the decade of its existence, has led to developments in flax research including molecular marker development.

SSRs are the primary marker type that has been developed through the Tufgen project. SSRs are repeated motifs of 2-6 nucleotides that vary in the number of tandem repeats. SSRs are useful as molecular markers because they are abundant, evenly distributed within the genome in flax, ubiquitous, codominant, multiallelic and highly polymorphic (Powell *et al.*, 1996). Novel and polymorphic SSRs were developed using a CDC Bethune bacterial artificial chromosome (BAC) library, where the BAC end

sequences were mined with expressed sequence tags (ESTs) that were available from GenBank (Cloutier *et al.*, 2009). Work has also been done in developing SNP markers in flax, which would offer better genomic saturation. Until these SNPs are developed and become available, the SSRs remain the best marker system and can even provide indications of regions that could be more intensely mapped by SNPs in the future (Cloutier *et al.*, 2012a).

Since these markers were developed, a consensus genetic (Fig 2.2) and physical map using the markers (Cloutier *et al.*, 2012a) and genomic sequence information (Wang *et al.*, 2012) has been published. This map was generated using three different mapping populations (CDC Bethune/MacBeth, E1474/Viking, and SP2047/UGG5-5). This map has 670 markers anchored to 204 of the 416 fingerprinted contigs of the physical map and the overall map density is 2.0cM for markers that are arranged into 15 linkage groups (Cloutier *et al.*, 2012b). These SSRs can be utilized in many different genetic studies including linkage analysis for the detection of loci responsible for certain traits like seed coat color, or even can be used for marker assisted selection in breeding programs.


LG1
0.0 *Lu2857_0
20.8 *Lu25_187
21.6 *Lu2089_187
22.2 *Lu3235_537
22.9 *Lu2026_560
24.5 *Lu2853_187
26.6 *Lu2861_187
27.3 *Lu2091_0
32.1 *Lu2858_187
36.1 *Lu1066_161_285
36.6 *Lu49B_0
37.2 *Lu3020_286
39.0 *Lu427_161
40.8 *Lu2390_51
41.8 *Lu2807_161
42.4 *Lu2808_161
44.8 *Lu987_0
46.2 *Lu2803_161
46.9 *Lu2802_161 *Lu955_0
47.6 *Lu1160b_566
51.1 *Lu56_0
51.2 *tad2A_325
52.4 *Lu869_0
53.7 *Lu3283_887
58.4 *Lu2392a_51
58.8 *Lu2392b_51
60.5 *Lu2387_51
63.2 *Lu299_51 *Lu866_0
63.3 *Lu796_51
63.9 *Lu2393_51
64.2 *Lu2388_51
65.4 *Lu2053_51
70.2 *Lu3220_512
72.8 *Lu3222_512
74.5 *Lu1160a_566
76.8 *Lu2589_93
78.3 *Lu2597_94
81.6 *Lu2592_93
91.2 *Lu2895b_203
92.0 *Lu2055_119
92.9 *Lu998_275
94.4 *Lu2712_126
98.6 *Lu2698_119
100.9 *Lu868_119
106.2 *Lu2374b_45
109.7 *Lu870_0
111.3 *Lu2010c_0
113.3 *Lu2184a_12
118.8 *Lu2183a_12
121.9 *Lu3053_297
123.5 *Lu2010b_0
125.4 *Lu999_550
129.8 *Lu1148_550
130.8 *Lu3231_550
131.3 *Lu3279_856
132.3 *Lu46_0
132.8 *Lu47_0
136.8 *Lu2687_112
147.1 *Lu114_181
148.8 *Lu981_222
150.0 *Lu943_222
170.0 *Lu2681_0
LG2
0.0 *Lu747b_0
0.1 *Lu2799_160
1.1 *Lu2794_160
11.0 *Lu2795_160
12.7 *Lu2796_160
22.1 *Lu2800_160
28.5 *Lu2113_1
35.1 *Lu129_1
38.9 *Lu906_1
43.4 *Lu2115_1
46.6 *Lu2250_28
50.9 *Lu2247_28
53.5 *Lu2469_108
55.7 *Lu344_28
56.7 *Lu3291_1108
56.8 *Lu2137_4 *Lu2370_44
57.1 *Lu2188_14
57.4 *Lu859_10
58.0 *Lu2135_4 *Lu257_239
58.4 *Lu2139_4
58.7 *Lu910_4
58.9 *Lu3238_568
59.0 *Lu3023_286
59.1 *Lu2144_4 *Lu3022_286 *Lu840_286 *Lu3256_676
59.3 *Lu3269_771
59.7 *Lu2067a_0
59.8 *Lu2145_4 *Lu532_4
60.5 *Lu2457a_66
60.6 *Lu824_4
60.8 *Lu2366_44
61.7 *Lu2959_237
62.9 *Lu926Ba_0
67.0 *Lu128_206
68.3 *Lu209_206
70.0 *Lu2907_206
70.5 *Lu2908_206
71.3 *Lu2909_206
72.4 *Lu2340_39
74.3 *Lu2351_39
78.0 *Lu2349_39
82.0 *Lu2347_39
82.5 *Lu2341_39
83.5 *Lu2346_39
85.9 *Lu125_39
86.4 *Lu3068_313
87.8 *Lu3276_813
89.1 *Lu2344_39
92.5 *Lu2352_39
98.1 *Lu900_39
99.2 *Lu1028_90
105.0 *Lu2027_475 *Lu2007_475
105.8 *Lu2021a_475
106.5 *Lu3206_475
109.2 *Lu324_51
113.3 *Lu2718_134
122.9 *Lu2720_134
137.5 *Lu3205_475
*Lu1115_134
LG3
0.0 *Lu342_101
5.1 *Lu445_542
9.2 *Lu318_413
10.5 *Lu1039_0
12.3 *Lu3024_287
15.4 *Lu1161_101
15.6 *Lu2628_101
18.6 *Lu452_0
20.4 *Lu2625a_101
38.0 *Lu899_150
48.1 *Lu774_150
49.0 *Lu2764_150
49.4 *Lu2767_150
51.0 *Lu821_150
55.7 *Lu3262_729
60.9 *Lu64_0
62.7 *Lu2777b_156
64.4 *Lu373_16
66.0 *Lu139_16
69.3 *Lu787a_0
72.3 *Lu2184_16
73.3 *Lu2689_113
73.6 *Lu2161_7
74.3 *Lu2047_0 *Lu2044_0
74.4 *Lu2040_0 *Lu2049_0 *Lu3117_374
75.2 *Lu2063_0 *Lu2163_7
75.7 *Lu3223_519
76.5 *Lu2164_7
79.1 *Lu3199a_444 *Lu2635_102
82.2 *Lu106_102
82.5 *Lu105_102
83.2 *Lu2633_102
84.8 *Lu2631_102
87.7 *Lu3195_444
91.4 *Lu104_102
92.5 *Lu3111_389
97.0 *Lu3151_401
98.3 *Lu3153_401
98.8 *Lu3152_401
99.5 *Lu933_0
100.1 *Lu638_401 *Lu639_401
103.3 *Lu1144_0
107.0 *Lu2706_119
107.1 *Lu3290_1078
108.1 *dgatA_400
108.9 *Lu2038_0
110.7 *Lu658_400
111.6 *Lu3148_400
113.4 *Lu3150_400
117.4 *Lu3146_400
120.4 *Lu3144_400
121.1 *Lu2838b_177
122.0 *Lu509_118
122.4 *Lu558_118
127.4 *Lu2693_115
127.5 *Lu2775b_156
127.8 *Lu450_115
128.7 *Lu422a_115
LG4
0.0 *Lu3281_865
9.1 *Lu2966_243
17.4 *Lu996_243
27.3 *Lu2006_0 *Lu2004_32 *Lu2002_243
27.4 *Lu2968_243
32.4 *Lu2073_673
34.4 *Lu722B_673
41.8 *Lu2025_324
41.9 *Lu2008_324
42.5 *Lu2059_324
47.2 *Lu2207_487
49.2 *Lu3228_545
50.2 *Lu3229_545
50.9 *Lu2399_55
53.8 *Lu2396_55
53.9 *Lu3252_659
54.1 *Lu3213_487
54.2 *Lu2087_487
57.8 *Lu717_231
58.4 *Lu2397_55
58.8 *Lu2942_231
58.9 *Lu2944_231
59.3 *Lu2940_231
66.4 *Lu2943_231
68.7 *Lu989_675
70.2 *Lu207_675
75.5 *Lu3113_371
76.5 *Lu3116_371
78.2 *Lu2983_261
81.2 *Lu2980_261
82.8 *Lu587_0
83.4 *Lu833_261
84.5 *Lu1049_0
86.6 *Lu2984_261
87.8 *Lu851_261
88.9 *Lu2981_261
89.5 *Lu2239_24
91.8 *Lu2043_261
92.0 *Lu2054_261
94.6 *Lu2237_24
94.9 *Lu2031_24
97.8 *Lu2233_24
102.2 *Lu2230_24
109.0 *Lu919_0
110.4 *Lu2235_24
111.2 *Lu2076_24
119.4 *Lu2286_35
121.2 *Lu2287_35
121.7 *Lu2009_0 *Lu2011_0
121.8 *Lu2024_0


LG5
0.0 *s18B10_0
7.4 *Lu227_130
9.1 *Lu643_130
10.7 *Lu361_130
13.5 *Lu274B_130
24.1 *Lu3201_447
24.2 *Lu223_447
25.9 *Lu2037b_0
28.2 *Lu2014_0
28.9 *Lu2295_36
30.9 *Lu505_0
34.2 *Lu176_36
37.1 *Lu1182_0
42.0 *Lu2288_36
43.1 *Lu2291_36
44.1 *Lu2411_58
45.7 *Lu2292_36
46.9 *Lu744_36
48.3 *Lu2297_36
49.3 *Lu2293_36
50.2 *Lu2704_119
51.2 *dgatB_119
51.6 *Lu2086_0
51.8 *Lu2304_0 *Lu2305_0
51.9 *Lu2752_146
52.2 *Lu3225_537
52.4 *Lu922_0
52.5 *Lu2362_43
52.8 *Lu2365_43
53.2 *Lu2364_43
53.5 *Lu2017_0
53.9 *Lu2466_70
54.7 *Lu2700_119
54.9 *Lu3132_385
55.2 *Lu2884_202 *Lu2889_202
55.3 *Lu2318_0 *Lu2886_202
55.4 *Lu2885_202
55.9 *Lu2883_202
56.1 *Lu2887_202
56.3 *Lu623_202
56.4 *Lu3248_596
56.6 *Lu2890_202
57.1 *Lu41_202
57.6 *Lu164_30
58.0 *Lu2255_30
58.5 *Lu268_30
70.4 *Lu2403_58
73.5 *Lu2408_58
76.2 *Lu738_58
79.1 *Lu652_58
80.7 *Lu2409_58
85.7 *Lu2405_58
98.8 *Lu2512_76
101.6 *Lu968_55
104.8 *Lu330_76
108.0 *Lu2509_76
108.7 *Lu2516a_76
110.4 *Lu682_76
112.4 *Lu2068_76
LG6
0.0 *Lu1006_59
0.9 *Lu1177_0
2.5 *Lu3267_996
6.5 *Lu2420_59
8.7 *Lu2418_59
9.6 *Lu2917b_214
10.4 *Lu2084_59
10.8 *Lu1179_59
16.3 *Lu502_86
26.3 *Lu699_380
26.6 *Lu1094_380
30.7 *Lu1107_197 *Lu944_197
32.3 *Lu2608_97
37.1 *Lu3014_280
39.2 *Lu3013_280
39.8 *Lu2545_81
44.3 *Lu442a_0
45.8 *Lu2169_9
46.6 *Lu2544_81
48.1 *Lu2548_81
48.2 *Lu2972_248
48.6 *Lu1178_27
49.2 *Lu456_775 *Lu2550_81
49.4 *Lu457_775 *Lu2242_27
49.8 *Lu2072_27
49.9 *Lu2039_27
50.8 *Lu2613_97
51.0 *Lu918_0
51.3 *Lu3267_757
51.5 *Lu728b_27
52.2 *Lu69_0 *Lu2542_81
*Lu2549_81 *Lu2543_81
52.6 *Lu2071_81 *Lu2084_81
53.4 *Lu836_0
53.7 *Lu2975_248
54.9 *Lu2974_248
57.5 *Lu2971_248
62.5 *Lu2556_82
62.6 *Lu1002B_82
63.2 *Lu2561b_82
63.4 *Lu2560_82
64.1 *Lu2564_82
67.5 *Lu2565_82
68.6 *Lu2553_82
69.4 *Lu2554_82
72.0 *Lu2555_82
72.6 *Lu2557_82
73.0 *Lu3057a_303
73.7 *Lu60_0
84.1 *Lu861_591
84.8 *Lu1112_171
95.6 *Lu3091_329
107.3 *Lu2078_329_10
LG7
0.0 *Lu260_79
2.7 *Lu2532_79
3.6 *Lu2535_79
5.7 *Lu2536_79
5.8 *Lu2534_79
6.7 *Lu2083_79
7.2 *Lu2540_79
7.9 *Lu402_79
10.6 *Lu675_79
10.8 *Lu2533_79
10.9 *Lu511_79
11.6 *Lu2032_0
13.3 *Lu741_79
13.7 *Lu1146a_79
18.3 *Lu2825b_175
21.1 *Lu2832_175
23.9 *Lu146_175
24.0 *Lu138_175 *Lu151_175
24.9 *Lu235_230
27.3 *Lu2810_166
28.8 *Lu2827_175
38.2 *Lu1055_77
40.2 *Lu3161_436
40.8 *Lu296_436
42.3 *Lu1022_436
42.7 *Lu3160_436
43.4 *Lu672_436
44.1 *Lu3184_436
46.0 *Lu3178b_436
51.7 *Lu2B15_170
52.5 *Lu2812a_170
53.7 *Lu585B_108
54.7 *Lu2065_0
55.6 *Lu2651_108
56.2 *Lu2652_108
56.3 *Lu2654_108
60.8 *Lu557_454
66.8 *Lu3100a_353
71.2 *Lu1124_84
76.6 *Lu2571_84
80.0 *Lu2658_108
80.9 *Lu2648_108
83.8 *tad3A_84
84.7 *Lu44E4_84
90.0 *Lu2003_281
91.6 *Lu566_281
92.1 *Lu449_281
92.5 *Lu3016_281
96.6 *Lu3017_281
99.6 *Lu3266_733
104.3 *Lu58a_257
LG8
0.0 *Lu595_108
12.0 *Lu2561c_82
14.8 *Lu2659_108
18.9 *Lu2649_108
24.8 *Lu2840_178
25.7 *Lu339_178
27.6 *Lu1171_178
29.0 *Lu2561a_82
37.1 *Lu265_736
38.2 *Lu2030_82
40.1 *Lu3057b_303
44.0 *Lu3059_303
45.4 *Lu2563_82
47.9 *Lu3157_405
51.9 *Lu3156_405
53.2 *Lu633_0
53.4 *Lu2428_60
54.7 *Lu2957_235 *Lu2268_31
56.0 *Lu2056_217
56.3 *Lu2923_217
56.6 *Lu2578_88
57.1 *Lu2203_18
64.9 *Lu2103_46
65.3 *Lu857_46
66.1 *Lu2918_215
67.6 *Lu2921_215
70.2 *Lu2082_46
70.6 *Lu2098_46
72.1 *Lu2105_46
72.7 *Lu2102_46
73.0 *Lu2306_46
73.5 *Lu2313_0
73.6 *Lu2312_0 *ysc1_0
*Lu2307_0
74.2 *Lu2316_46
74.3 *Lu928_0 *Lu2106_46
74.8 *Lu2317_0
75.1 *Lu2320_0
75.3 *Lu2326_0
76.4 *Lu2330_0
76.5 *Lu447_215
77.2 *Lu2329_0
78.5 *Lu178_46
79.8 *Lu2424_60
80.8 *Lu2618_98
83.5 *Lu2101_46
84.5 *Lu2431_60
84.6 *Lu2430_60
84.7 *Lu2156_5
85.2 *Lu2429_60
87.9 *Lu3189_441
88.7 *Lu525_60
88.8 *Lu1077_174
90.0 *Lu2425_60
91.0 *Lu2820_174
93.0 *Lu2823_174
93.8 *Lu2822_174
94.2 *Lu3280_864
94.7 *Lu2745_141
95.7 *Lu2587_91
98.0 *Lu2714_130
99.0 *Lu963_91
100.4 *Lu2515_76
103.0 *Lu1103_60
103.9 *Lu2338a_38

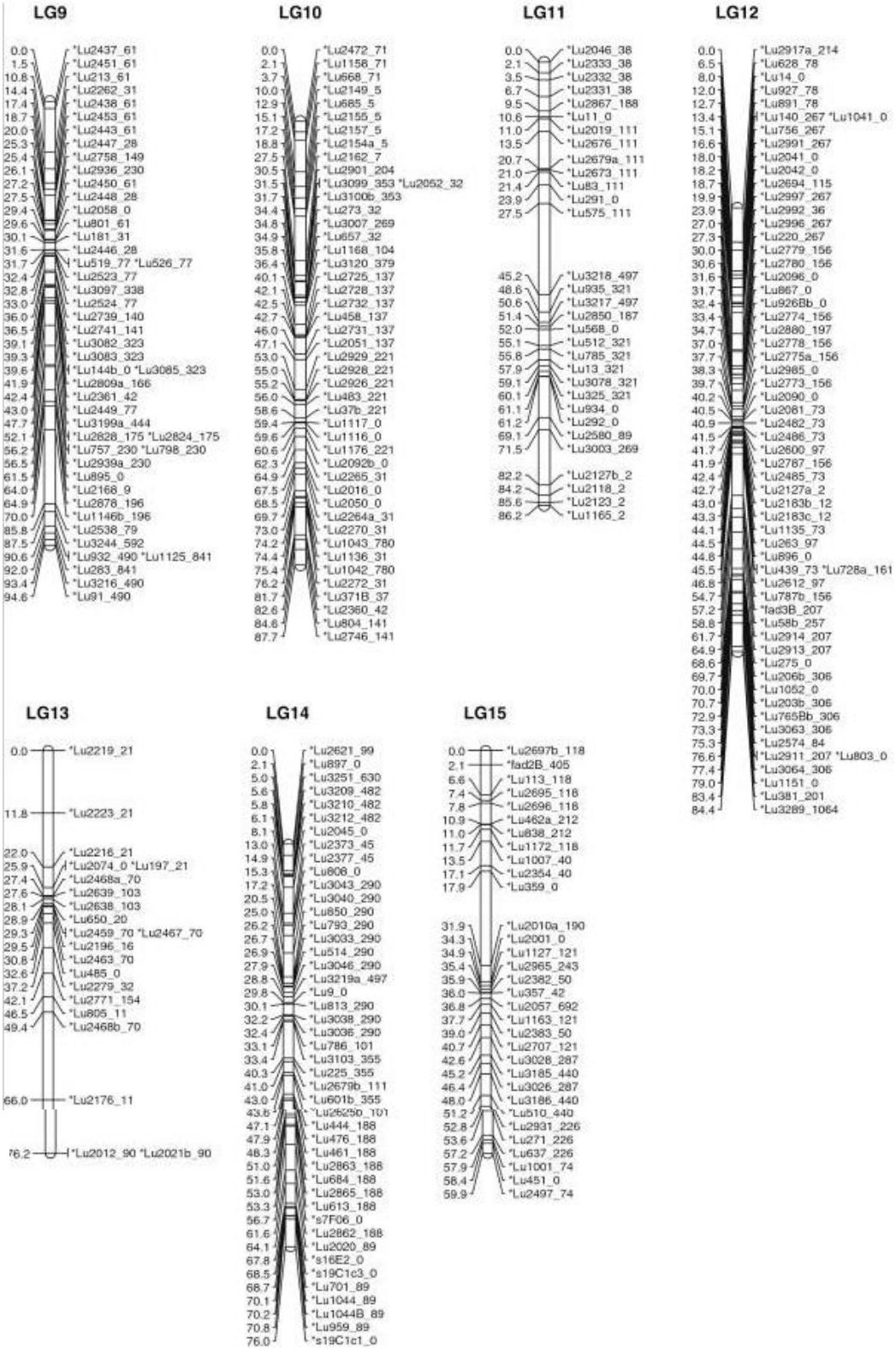

LG9
0.0 *Lu2437_61
1.5 *Lu2451_61
10.8 *Lu213_61
14.4 *Lu2262_31
17.4 *Lu2438_61
18.7 *Lu2453_61
20.0 *Lu2443_61
25.3 *Lu2447_28
25.4 *Lu2758_149
26.1 *Lu2936_230
27.2 *Lu2450_61
27.5 *Lu2448_28
29.4 *Lu2058_0
29.6 *Lu801_61
30.1 *Lu181_31
31.6 *Lu2446_28
31.7 *Lu519_77 *Lu526_77
32.4 *Lu2523_77
32.8 *Lu3097_338
33.0 *Lu2524_77
36.0 *Lu2739_140
36.5 *Lu2741_141
39.1 *Lu3082_323
39.3 *Lu3083_323
39.6 *Lu144b_0 *Lu3085_323
41.9 *Lu2809a_166
42.4 *Lu2361_42
43.0 *Lu2449_77
47.7 *Lu3199a_444
52.1 *Lu2628_175 *Lu2824_175
56.2 *Lu757_230 *Lu798_230
56.5 *Lu2939a_230
61.5 *Lu895_0
64.0 *Lu2168_9
64.9 *Lu2878_196
70.0 *Lu1146b_196
85.8 *Lu2538_79
87.5 *Lu3244_592
90.6 *Lu932_490 *Lu1125_841
92.0 *Lu283_841
93.4 *Lu3216_490
94.6 *Lu91_490
LG10
0.0 *Lu2472_71
2.1 *Lu1158_71
3.7 *Lu668_71
10.0 *Lu2149_5
12.9 *Lu685_5
15.1 *Lu2155_5
17.2 *Lu2157_5
18.8 *Lu2154a_5
27.5 *Lu2162_7
30.5 *Lu2901_204
31.5 *Lu3099_353 *Lu2052_32
31.7 *Lu3100b_353
34.4 *Lu273_32
34.8 *Lu3007_269
34.9 *Lu657_32
35.8 *Lu1168_104
36.4 *Lu3120_379
40.1 *Lu2725_137
42.1 *Lu2728_137
42.5 *Lu2732_137
42.7 *Lu458_137
46.0 *Lu2731_137
47.1 *Lu2051_137
53.0 *Lu2929_221
55.0 *Lu2928_221
55.2 *Lu2926_221
56.0 *Lu483_221
58.6 *Lu37b_221
59.4 *Lu1117_0
59.6 *Lu1116_0
60.6 *Lu1176_221
62.3 *Lu2092b_0
64.9 *Lu2265_31
67.5 *Lu2016_0
68.5 *Lu2050_0
69.7 *Lu2264a_31
73.0 *Lu2270_31
74.2 *Lu1043_780
74.4 *Lu1136_31
75.4 *Lu1042_780
76.2 *Lu2272_31
81.7 *Lu371B_37
82.6 *Lu2360_42
84.6 *Lu804_141
87.7 *Lu2746_141
LG11
0.0 *Lu2046_38
2.1 *Lu2333_38
3.5 *Lu2332_38
6.7 *Lu2331_38
9.5 *Lu2867_188
10.6 *Lu11_0
11.0 *Lu2019_111
13.5 *Lu2676_111
20.7 *Lu2679a_111
21.0 *Lu2673_111
21.4 *Lu83_111
23.9 *Lu291_0
27.5 *Lu575_111
45.2 *Lu3218_497
48.6 *Lu935_321
50.6 *Lu3217_497
51.4 *Lu2850_187
52.0 *Lu568_0
55.1 *Lu512_321
55.8 *Lu785_321
57.9 *Lu13_321
59.1 *Lu3078_321
60.1 *Lu325_321
61.1 *Lu934_0
61.2 *Lu292_0
69.1 *Lu2580_89
71.5 *Lu3003_269
82.2 *Lu2127b_2
84.2 *Lu2118_2
85.6 *Lu2123_2
86.2 *Lu1165_2
LG12
0.0 *Lu2917a_214
6.5 *Lu628_78
8.0 *Lu14_0
12.0 *Lu927_78
12.7 *Lu891_78
13.4 *Lu140_267 *Lu1041_0
15.1 *Lu756_267
16.6 *Lu2991_267
18.0 *Lu2041_0
18.2 *Lu2042_0
18.7 *Lu2694_115
19.9 *Lu2997_267
23.9 *Lu2992_36
27.0 *Lu2996_267
27.3 *Lu220_267
30.0 *Lu2779_156
30.6 *Lu2780_156
31.6 *Lu2096_0
31.7 *Lu867_0
32.4 *Lu926Bb_0
33.4 *Lu2774_156
34.7 *Lu2880_197
37.0 *Lu2778_156
37.7 *Lu2775a_156
38.3 *Lu2985_0
39.7 *Lu2773_156
40.2 *Lu2090_0
40.5 *Lu2081_73
40.9 *Lu2482_73
41.5 *Lu2486_73
41.7 *Lu2600_97
41.9 *Lu2787_156
42.4 *Lu2485_73
42.7 *Lu2127a_2
43.0 *Lu2183b_12
43.3 *Lu2183c_12
44.1 *Lu1135_73
44.5 *Lu263_97
44.8 *Lu896_0
45.5 *Lu439_73 *Lu728a_161
46.8 *Lu2612_97
54.7 *Lu787b_156
57.2 *fad3B_207
58.8 *Lu58b_257
61.7 *Lu2914_207
64.9 *Lu2913_207
68.6 *Lu275_0
69.7 *Lu206b_306
70.0 *Lu1052_0
70.7 *Lu203b_306
72.9 *Lu765Bb_306
73.3 *Lu3063_306
75.3 *Lu2574_84
76.6 *Lu2911_207 *Lu803_0
77.4 *Lu3064_306
79.0 *Lu1151_0
83.4 *Lu381_201
84.4 *Lu3289_1064
LG13
0.0 *Lu2219_21
11.8 *Lu2223_21
22.0 *Lu2216_21
25.9 *Lu2074_0 *Lu197_21
27.4 *Lu2468a_70
27.6 *Lu2639_103
28.1 *Lu2638_103
28.9 *Lu650_20
29.3 *Lu2459_70 *Lu2467_70
29.5 *Lu2196_16
30.8 *Lu2463_70
32.6 *Lu485_0
37.2 *Lu2279_32
42.1 *Lu2771_154
46.5 *Lu805_11
49.4 *Lu2468b_70
66.0 *Lu2176_11
76.2 *Lu2012_90 *Lu2021b_90
LG14
0.0 *Lu2621_99
2.1 *Lu897_0
5.0 *Lu3251_630
5.6 *Lu3209_482
5.8 *Lu3210_482
6.1 *Lu3212_482
8.1 *Lu2045_0
13.0 *Lu2373_45
14.9 *Lu2377_45
15.3 *Lu808_0
17.2 *Lu3043_290
20.5 *Lu3040_290
25.0 *Lu850_290
26.2 *Lu793_290
26.7 *Lu3033_290
26.9 *Lu514_290
27.9 *Lu3046_290
28.8 *Lu3219a_497
29.8 *Lu9_0
30.1 *Lu813_290
32.2 *Lu3038_290
32.4 *Lu3036_290
33.1 *Lu786_101
33.4 *Lu3103_355
40.3 *Lu225_355
41.0 *Lu2679b_111
43.0 *Lu601b_355
43.6 *Lu2625b_101
47.1 *Lu444_188
47.9 *Lu476_188
48.3 *Lu461_188
51.0 *Lu2863_188
51.6 *Lu684_188
53.0 *Lu2865_188
53.3 *Lu613_188
56.7 *s7F06_0
61.6 *Lu2862_188
64.1 *Lu2020_89
67.8 *s16E2_0
68.5 *s19C1c3_0
68.7 *Lu701_89
70.1 *Lu1044_89
70.2 *Lu1044B_89
70.8 *Lu959_89
76.0 *s19C1c1_0
LG15
0.0 *Lu2697b_118
2.1 *fad2B_405
6.6 *Lu113_118
7.4 *Lu2695_118
7.8 *Lu2696_118
10.9 *Lu462a_212
11.0 *Lu838_212
11.7 *Lu1172_118
13.5 *Lu1007_40
17.1 *Lu2354_40
17.9 *Lu359_0
31.9 *Lu2010a_190
34.3 *Lu2001_0
34.9 *Lu1127_121
35.4 *Lu2965_243
35.9 *Lu2382_50
36.0 *Lu357_42
36.8 *Lu2057_692
37.7 *Lu1163_121
39.0 *Lu2383_50
40.7 *Lu2707_121
42.6 *Lu3028_287
45.2 *Lu3185_440
46.4 *Lu3026_287
48.0 *Lu3186_440
51.2 *Lu510_440
52.8 *Lu2931_226
53.6 *Lu271_226
57.2 *Lu637_226
57.9 *Lu1001_74
58.4 *Lu451_0
59.9 *Lu2497_74

Figure 2.2. The genetic map for flax made by combining three linkage maps that contained a total of 770 markers on 15 linkage groups. The total length of the genetic map is 1,551 cM with a mean marker density of 2.0 cM.

2.9 Bulked Segregant Analysis

Bulked segregant analysis (BSA) was developed to rapidly identify markers linked to any specific trait (Michelmore *et al.,* 1991). To perform this method you can use a recombinant inbred line (RIL), double haploid (DH), backcross (BC) or an F2 population that is segregating for a certain trait. This trait could either be monogenic or quantitatively inherited (Varshney *et al.*, 2009). The premise of this technique is to establish two pools of DNA that contain individuals that are genetically similar at the locus of interest within the pool, genetically dissimilar amongst the pools and all other unlinked regions are arbitrary (Michelmore *et al.*, 1991). In monogenic traits you choose individuals for the pools based on versions of the trait present, for example, brown or yellow seed coat color. But for quantitatively inherited traits, you choose individuals that represent the tails of the traits normal distribution curve.

BSA can be used with large populations or populations with distorted segregation to identify putative markers between the pooled genotypes (Varshney *et al.,* 2009). These markers can then be used in traditional linkage analysis on a larger population to determine recombination frequency and therefore genetic distance from the locus of interest. But this procedure is most effective using markers of high density (Ehrenreich *et al.,* 2009). However, markers far away from the locus of interest are detected with decreasing frequency the further it is away, up to a 25 cM window on either side of the locus, at which point linkage cannot be accurately detected (Michelmore *et al.,* 1991).

Using BSA can help to rapidly detect molecular markers linked to seed coat color loci in flax populations. Pools of DNA made from brown and yellow or variegated seed can be screened to identify the location of the loci involved in seed coat color.

Chapter 3 The Inheritance of Seed Coat Color Conditioned by the *b1* Locus

3.1 Abstract

By understanding the inheritance behavior of seed coat color conditioned by the *b1* locus breeders could use variegated and yellow seed coat colors as a marker for flax varieties with unique combinations of quality traits. The objective of this study was to determine how stably inherited seed coat color conditioned by this locus is as well as to confirm the allelic nature of the two known recessive alleles of this locus, b^{1vg} and b^{1}. To conduct these experiments observations of seed coat color from one generation to the next as well as population exploration of recombinant inbred line (RIL) populations and an allelism test were performed.

Spearman's rank correlations for the CDC Bethune x M96006 and the CDC Bethune x USDA-ARS Crystal RIL populations showed that seed coat color conditioned by this locus was stably inherited from one generation to the next and that there was no transgressive segregation of seed coat color (0.853; $P<0.001$ and 0.999; $P<0.001$). A 3:1 ratio of brown to yellow seed coat color was shown through population exploration in one population, which suggests that either there are two genes responsible for yellow seed coat color in the USDA-ARS Crystal parental line, the USDA-ARS Crystal parent of the population was a heterozygote at the *b1* locus, or selection against lines with yellow seed coat color occurred. The results of the allelism test confirmed that the b^{1vg} and b^{1} alleles do not complement each other and are therefore allelic and that the variegated seed coat color phenotype conditioned by the b^{1vg} allele is dominant over the yellow seed coat color phenotype conditioned by the b^{1} allele.

The stability of seed coat color conditioned by this locus makes it suitable to use the *b1* locus as a phenotypic marker. Additionally, the information provided about the allelic nature of these alleles and their dominance relationship between them will allow flax breeders to use these seed coat color phenotypes as markers.

3.2 Introduction

Flax is an important crop to Canada, especially to Saskatchewan, where the province produces and exports 150-180 million dollars worth of flax annually (SaskFlax 2013). Canada chiefly produces Linseed-type flax that is subdivided into two market classes, brown and yellow seeded flax (Canadian Grain Commission 2012). Previously, the Canadian Grain Commission required yellow seed coat color to be used only as a phenotypic marker for Solin, a crop bred from flax to have less than five percent linolenic fatty acid content. However, recently this requirement has been removed. By understanding the inheritance of seed coat color and choosing a stably inherited color, seed coat color can be used by flax breeders as markers for specialty lines with specific combinations of traits.

Flax is a self-pollinating diploid plant with 15 sets of chromosomes (Lin & Paterson, 2011). There are four known independently inherited loci that condition seed coat color in flax; *Y1*, *g*, *d* and *b1*, but the locations within the genome, genes involved or even the behavior and inheritance of the seed coat colors conditioned by these loci is not known or fully understood (Mittapalli & Rowland, 2003). This study focusses on the *b1* locus and its three known alleles, B^1, b^{1vg} and b^1 where brown is the dominant phenotype and the recessive b^{1vg} and b^1 alleles result in variegated and yellow seed coat color phenotypes respectively. Additionally, there is a known semi-lethality associated with the b^1 allele of the *b1* locus(Mittapalli & Rowland, 2003) that is not fully understood but results in a reduction in the number of individuals found within the yellow seed coat color phenotypic class.

The following research was conducted with the objective of showing how stably inherited seed coat color conditioned by the *b1* locus is as well as confirming the allelic nature of b^{1vg} and b^1.

3.3 Materials and Methods

3.3.1 Plant Materials and Field Experiment

To test the inheritance of seed coat color conditioned by the *b1* locus two recombinant inbred line (RIL) populations were grown in separate field tests in both 2012 and 2013 and seed coat color data was collected. The populations were developed by crossing CDC Bethune (Rowland *et al.*, 2001), a brown seeded variety, to M96006 (Saeidi & Rowland, 1997) a line with variegated seed coat color or to USDA-ARS Crystal (Arny, 1944) a yellow seeded variety (Table 3.1). For both populations reciprocal crosses were also performed but no viable individuals were produced from the USDA-ARS Crystal x CDC Bethune cross. Following the crosses, these populations were grown by single seed descent to the F_6 generation in the growth chamber. The $F_{6:7}$ generation and $F_{6:8}$ were grown under field conditions in 2012 and 2013 respectively at the Kernen Crop Research Farm (KCRF) (52^o N, 106°32'W) (Table 3.2). The parental lines, CDC Bethune with M96006 or CDC Bethune with USDA-ARS Crystal, as well as CDC Sorrel were grown as check varieties for each population.

The $F_{6:7}$ and $F_{6:8}$ populations were grown as replicated (n=2) hand planted hills in square or rectangular lattice designs based on their number of entries. There were 50 seeds planted into each hill. The CDC Bethune x USDA-ARS Crystal population had 144 entries and was grown in a square lattice design. Therefore, the CDC Bethune x USDA-ARS Crystal population had 12 plots in each iBlock and 12 iBlocks in each repetition. The CDC Bethune x M96006 population had 110 entries and was grown in a rectangular lattice design with 10 plots in each iBlock and 11 iBlocks in each repetition. All plots were hand harvested at maturity and threshed using a vogel thresher.

Table 3.1. Description of the flax parental lines used to generate the CDC Bethune x M96006 and the CDC Bethune x USDA-ARS Crystal recombinant inbred line (RIL) seed coat color populations.

Parental Line	Pedigree	Seed Coat Color	Proposed Alleles	Reference
CDC Bethune	Norman/FP857	Brown	B^1B^1	(Rowland *et al.*, 2002)
USDA-ARS Crystal	Bison/770B	Yellow	b^1b^1	(Arny, 1944)
M96006	1536/1747/4/1747/1929//E67/1747/3/E1306	Variegated	$b^{1vg}b^{1vg}$	(Saeidi & Rowland, 1997)

Table 3.2. Climate data for the 2012 and 2013 growing seasons at Kernen Crop Research Farm (KCRF) (52ºN, 106º32'W) in Saskatoon, Saskatchewan.

		April	May	June	July	August	September	October
Mean Temperature (ºC)	**2012**	4.25	9.75	15.6	19.6	17.2	13.2	2.17
	2013	-2.33	12.3	15.6	17.7	18.1	15.3	3.40
	30 year average	3.9	11.5	16.2	18.6	17.4	11.2	4.8
Total Precipitation (mm)	**2012**	34.4	143	97.0	82.8	11.2	0.80	11.0
	2013	9.20	19.4	123	40.2	11.4	16.6	5.60
	30 year average	19.7	44.2	63.4	58.0	36.8	32.1	16.9

3.3.2 Allelism Test

To confirm the allelic nature of b^1 and b^{1vg} (Mittapalli & Rowland, 2003) an allelism test was performed. USDA-ARS Crystal (b^1b^1) was crossed to M96006 ($b^{1vg}b^{1vg}$) and the reciprocal cross was also performed. Six F_1 plants from each cross were grown in the growth chamber. The F_2 and F_3 seed phenotypes which, is a reflection of the previous generation's genotype, was observed. The F_1, F_2 and F_3 seed were scored for seed coat color using the scale as mentioned in table 3.3. Since both of these alleles were proposed to be recessive, it was expected that if they were not allelic that complementation would occur and the dominant brown phenotype would be present in the F_2 seed produced by the F_1 plant. Recall that the tissues of the seed coat are products of the mother plant of the previous generation. If no complementation occurred and another phenotype other than brown was observed, they were considered to be allelic.

Table 3.3. Plant Genome Resource Canada's scale for scoring flax seed coat color phenotypes and their corresponding Munsell Color System values.

Scale Number	Seed Coat Color	Munsell Value
1	Light Brown	7.5YR5/6
2	Medium Brown	7.5YR4/6
3	Dark Brown	7.5YR3/2
4	Yellow	2.5Y6/6
5	Olive	5Y5/6
6	Variegated	-

3.3.3 Statistical Analysis

Spearman rank correlations were performed for seed coat color between the $F_{6:7}$ and $F_{6:8}$ generations that were grown in 2012 and 2013 respectively to demonstrate that individual lines do not change phenotypic class from one generation to the next. Chi-squared tests were performed to test the 1:1 and 3:1 brown seed coat color to yellow seed coat color populations ratios expected of recombinant inbred line populations segregating for one and two genes respectively. A chi-squared test was also used to test whether the ratio of variegated to yellow seed coat color from the allelism test fit a 3:1 simple dominance model.

3.4 Results and Discussion

3.4.1 Inheritance of Seed Coat Color

The parental lines of the CDC Bethune x M96006 and CDC Bethune x USDA-ARS Crystal recombinant inbred line (RIL) mapping populations were chosen for their differences in seed coat color which were observed in the 2012 and 2013 growing seasons where CDC Bethune yielded brown seed, M96006 yielded variegated seed and USDA-ARS Crystal yielded yellow seeds (Table 3.4). The seed coat colors of all of the lines in the $F_{6:7}$ generation directly reflected the parental seed coat colors where there was no transgressive segregation and only two phenotypic classes were seen in each population. The $F_{6:8}$ generation seed produced by the $F_{6:7}$ generation field test had seed coat colors reflective of the previous generation (Table 3.5); where, the seed coat color that was harvested had a Spearman rank correlation factor of 0.853 ($P<0.001$) between the $F_{6:7}$ and $F_{6:8}$ generations in the CDC Bethune x M96006 population and 0.999 ($P<0.001$) in the CDC Bethune x USDA-ARS Crystal population. The Spearman rank correlation factor for the CDC Bethune x M96006 population is much lower than 1, likely due to errors in observation during scoring as variegated seed coat is highly influenced by maturity of the seed where immature seeds with a brown seed coat can appear variegated. These high correlations suggest that the environment does not influence the expression of seed coat color conditioned by the *b1* locus and that seed coat color is stably inherited from generation to generation.

Table 3.4. Seed coat color scores based on Munsell values for the parental check lines in the $F_{6:7}$ and $F_{6:8}$ recombinant inbred line (RIL) mapping populations grown in 2012 and 2013 respectively, at the Kernen Crop Research Farm (KCRF) where 1-light brown, 2-medium brown, 3-dark brown, 4-yellow, 5-olive, and 6-variegated.

Parental Line	**$F_{6:7}$**	**$F_{6:8}$**
CDC Bethune	2	2
M96006	6	6
USDA-ARS Crystal	4	4

Table 3.5. The Spearman rank correlations use for discrete ordinal traits to show the correlation of seed coat color from the $F_{6:7}$ and $F_{6:8}$ generation CDC Bethune x M96006 and CDC Bethune x USDA ARS Crystal recombinant inbred line mapping populations grown in 2012 and 2013 respectively at the Kernen Crop Research Farm (KCRF).

Population	**Spearman Rank Correlation for Seed Coat color**	
CDC Bethune x M96006	$F_{6:8}$	$F_{6:7}$ 0.853***
CDC Bethune x USDA-ARS Crystal	$F_{6:8}$	$F_{6:7}$ 0.999***

*P<0.05, **P<0.01, ***P<0.001

3.4.2 Population Exploration

Upon further investigation of the CDC Bethune x M96006 and CDC Bethune x USDA-ARS Crystal populations used in this study it was found that seed coat color was not always segregating at ratios that were consistent with recombinant inbred line populations segregating for a monogenic trait. By the F_6 generation the majority of the heterozygous class should be eliminated and a segregation ratio of ~1:1 of brown to yellow or brown to variegated seed coat colors should be observed if conditioned by a single gene. A Yate's corrected chi-squared test was performed to see if a single gene model that would result in a 1:1 ratio fit for these populations (Table 3.6 and 3.7). The Yate's correction was applied to correct for the single degree of freedom.

The null hypothesis of this test was that brown and variegated seed coat colors in the CDC Bethune x M96006 population followed a 1 brown:1 variegated seed coat color ratio that is expected when it is segregating for a single gene. When the one gene model was fit using the Yate's corrected chi-squared test, the results were not significant and the null hypothesis was not rejected. Therefore, this population segregates for seed coat color conditioned by a single gene.

This second test had a similar null hypothesis that brown and yellow seed coat colors in the CDC Bethune x USDA-ARS Crystal population followed the 1 brown:1 yellow seed coat color ratio expected of a single gene model. When this model was fit using a Yate's corrected chi-squared test, the result was significant. Therefore, the null hypothesis was rejected, suggesting that there is some source of distortion occurring in this population. The cause of this distortion could be due to one of three reasons. The first is that either of the CDC Bethune or USDA-ARS Crystal parental lines used for crossing could have been heterozygous for this *b1* locus, however the selfed seed of the parental lines did not show segregation. Secondly, it could be caused by selection events, where the lines were grown 41-75 plants per pot with minimal water such that only two seed bolls would be collected per plant. These restrictive conditions could have selected against lines with yellow seed coats as they are previously known to have a lower germination rate and reduced vigour in comparison to lines with brown seed coats (Culbertson & Kommendahl, 1956; Saeidi & Rowland, 1999a; Saeidi & Rowland, 1999b). Additionally, selection could have occurred caused by reduced germination and vigour, or even semi-lethality all

previously noted to be associated with yellow seed coat color (Culbertson & Kommendahl, 1956; Mittapalli & Rowland, 2003; Saeidi & Rowland, 1999a; Saeidi & Rowland, 1999b; Shaw *et al.*, 1931). Thirdly, this distortion could be due to a second gene contributing to seed coat color determination of the USDA-ARS Crystal parent. This last reason would mean that the CDC Bethune x USDA-ARS Crystal population is segregating for two seed coat color genes instead of just the one as in the CDC Bethune x M96006 population. To test the third hypothesis another Yate's corrected chi-squared test was performed to fit a two-gene model (Table 3.8).

The null hypothesis of this last Yate's corrected chi-squared test was this population followed the 3:1 brown to yellow seed coat color ratio that is expected of a two-gene model. The result was not statistically significant and this null hypothesis was not rejected. Therefore, this population fits a two-gene model, which suggests that the USDA-ARS Crystal parent actually has two genes responsible for seed coat color. Further evidence for this could be found in the future by screening the entire population with a marker perfectly linked with the *b1* locus once one is developed. If the marker appears in a 3:1 ratio, then it would not support this two-gene model; instead, it would support the theories that selection events against yellow seed coat lines or heterogeneity of the *b1* locus of one of the parental lines caused the 3:1 phenotypic ratio that was previously observed. However, if the marker segregates at a 1:1 ratio then it would support the two-gene theory or the theory that the USDA-ARS Crystal parental line was heterozygous for the *b1* locus (Fig 3.1).

Table 3.6 Chi-squared test to fit a single gene model of the 1:1 brown seed coat to variegated seed coat color distribution expected of the 104 F_6 individuals in the CDC Bethune x M96006 recombinant inbred line (RIL) population where CDC Bethune (B^1B^1) and M96006 ($b^{1vg}b^{1vg}$) have brown and variegated seed coat colors respectively.

	Brown (B^1B^1)	**Variegated ($b^{1vg}b^{1vg}$)**
Observed	56.0	48.0
Expected	52.4	51.6

Chi-squared 0.1302, df 1, p-value 0.7182
*P<0.05, **P<0.01, ***P<0.001

Table 3.7. Chi-squared test to fit a single gene model of the 1:1 brown seed coat to yellow seed coat color distribution expected of the 104 F_6 individuals in the CDC Bethune x USDA-ARS Crystal recombinant inbred line (RIL) population where CDC Bethune (B^1B^1) and USDA-ARS Crystal (b^1b^1) have brown and yellow seed coat colors respectively.

	Brown (B^1B^1)	**Yellow (b^1b^1)**
Observed	80.0	24.0
Expected	52.4	51.6

Chi-squared 14.7034, df 1, p-value 0.0001258***
*P<0.05, **P<0.01, ***P<0.001

Table 3.8. Chi-squared test to fit a two gene model of the 3:1 brown seed coat to yellow seed coat color distribution expected of the 104 F_6 individuals in the CDC Bethune x USDA-ARS Crystal recombinant inbred line (RIL) population where CDC Bethune (B^1B^1) and USDA-ARS Crystal (b^1b^1) have brown and yellow seed coat colors respectively.

	Brown (B^1B^1)	**Yellow (b^1b^1)**
Observed	80.0	24.0
Expected	78.0	26.0

Chi squared 0.0263, df 1, p-value 0.8711
*P<0.05, **P<0.01, ***P<0.001

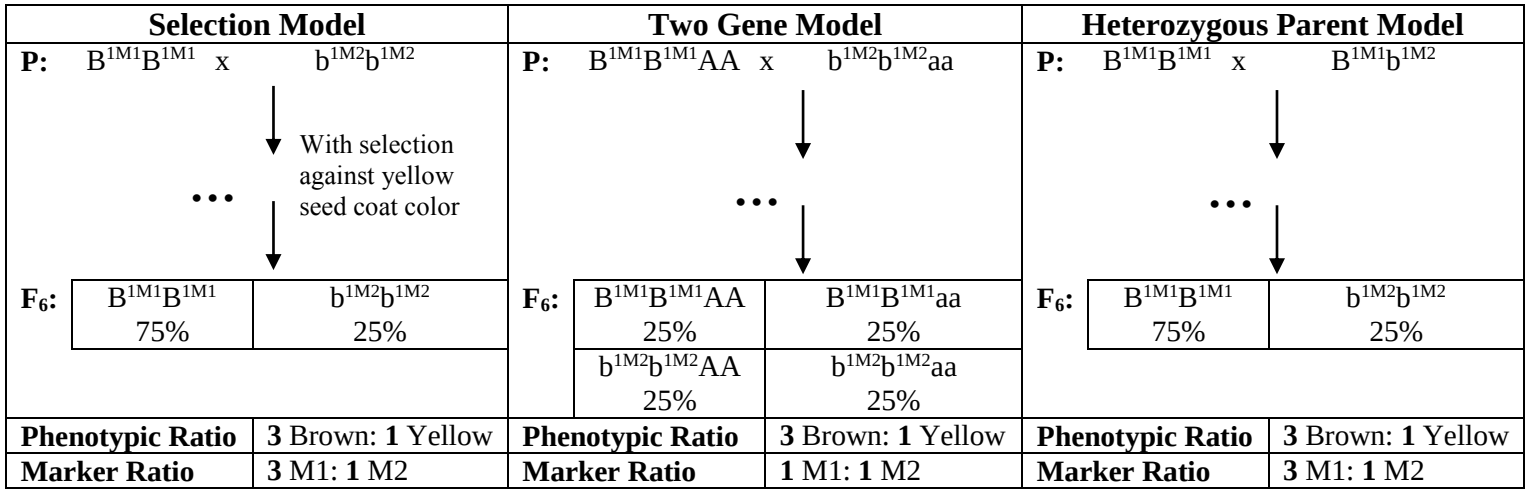


Figure 3.1. Models explaining possible causes of the 3 brown: 1 yellow phenotypic ratio observed in the $F_{6:7}$ generation of the CDC Bethune x USDA-ARS Crystal population. The alleles of the two genes are represented as B^1 and *A* where brown is dominant over yellow for both genes and where, when either gene is homozygous recessive it is epistatic to the other locus carrying dominant alleles. The hypothetical marker "M" is used to represent a marker linked with the B^1 gene where the dominant allele is linked with the "M1" polymorph of the marker and the recessive allele with the "M2". It is shown that if the distortion of the ratio was caused by selection or if the USDA-ARS Crystal parent was heterozygous a 3 M1: 1 M2 ratio would be observed in the F_6 generation. However, if a second gene was the cause the marker ratio would be 1 M1: 1 M2.

3.4.3 Allelism Test

An allelism test was performed to confirm the suggested allelic nature of the b^1 and b^{1vg} alleles (Mittapalli & Rowland, 2003). It should be noted that seed coats are made of maternal tissue and seed coat color reflects the previous generation's genotype, therefore segregation would not be seen until the F_3 seed. As seen in figure 3.2 all of the F_2 seeds from each of the six F_1 plants of the cross and the six F_1 plants from the reciprocal cross all yielded seeds with a variegated seed coat color. There was no complementation and therefore, b^1 and b^{1vg} are alleles of the same locus. These results also show that the variegated allele is dominant over the yellow allele while we already know that both are recessive to the brown allele. Additionally, in figure 3.2, the F_3 seed is the first generation to see segregation in the seed coat color. This was expected because as previously mentioned seed coats are made up of F_2 maternal tissue. However, 427 plants yielded variegated seed while 53 yielded yellow and therefore, a 7:1 variegated to yellow seed coat color ratio was seen in this generation which does not fit the expected 3:1 ratio of simple dominance. This ratio of a greater number of variegated versus yellow individuals is likely due to the b^1 allele, which conditions yellow seed coat color, has previously been shown to be associated with semi-lethality (Mittapalli & Rowland, 2003; Shaw *et al.*, 1931). It should be noted that all plants with yellow seed coats also had white star shaped flowers, while those with brown seed coats had blue bowl shaped flowers.

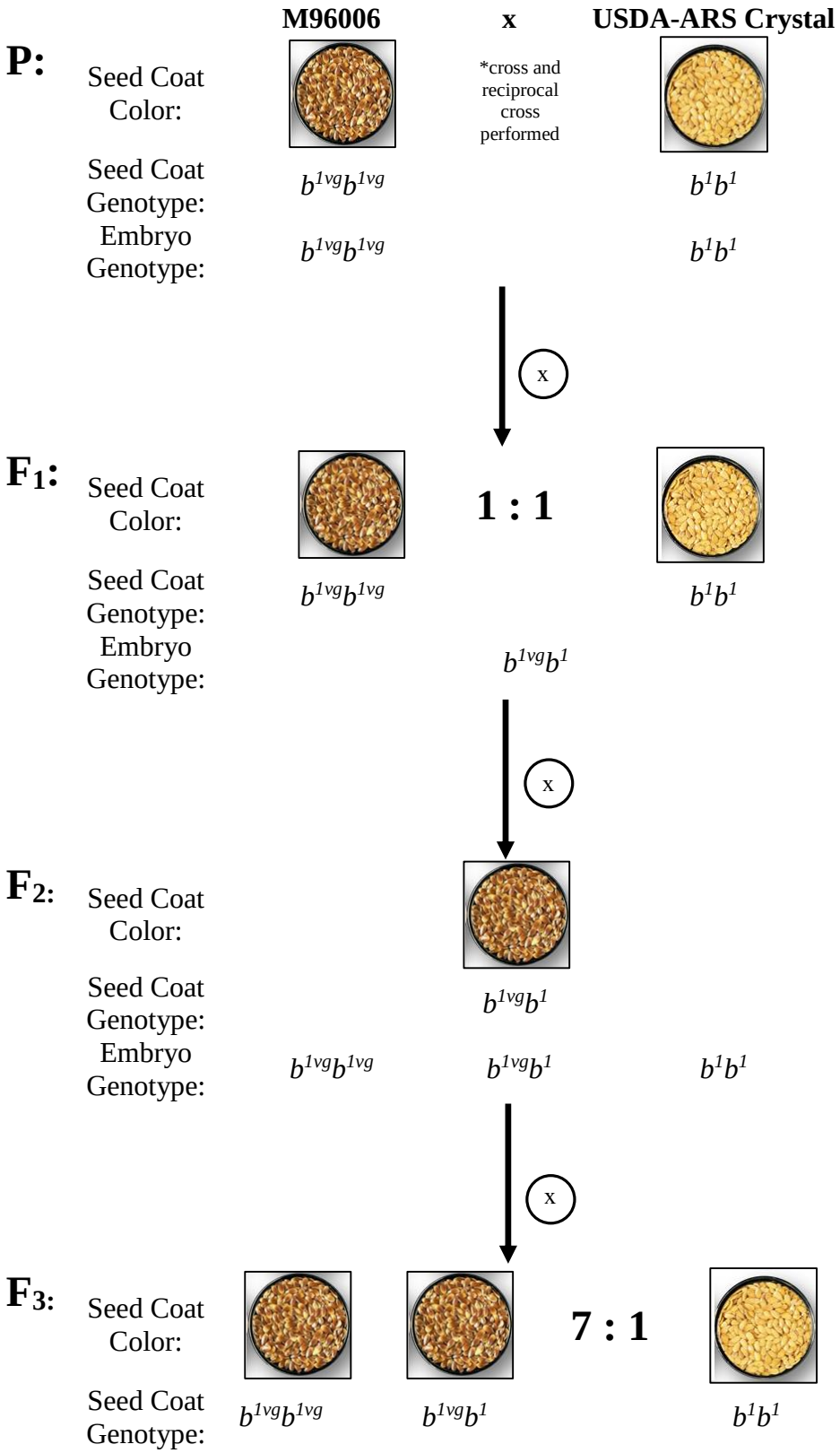

M96006
x
USDA-ARS Crystal
P:
Seed Coat Color:
*cross and reciprocal cross performed
Seed Coat Genotype:
$b^{1vg}b^{1vg}$
$b^{1}b^{1}$
Embryo Genotype:
$b^{1vg}b^{1vg}$
$b^{1}b^{1}$
x
F_1:
Seed Coat Color:
1 : 1
Seed Coat Genotype:
$b^{1vg}b^{1vg}$
$b^{1}b^{1}$
Embryo Genotype:
$b^{1vg}b^{1}$
x
F_2:
Seed Coat Color:
Seed Coat Genotype:
$b^{1vg}b^{1}$
Embryo Genotype:
$b^{1vg}b^{1vg}$
$b^{1vg}b^{1}$
$b^{1}b^{1}$
x
F_3:
Seed Coat Color:
7 : 1
Seed Coat Genotype:
$b^{1vg}b^{1vg}$
$b^{1vg}b^{1}$
$b^{1}b^{1}$

Figure 3.2. The allelism test confirming the allelic nature of the b^1 and b^{1vg} alleles. The yellow homozygous recessive USDA-ARS Crystal (b^1b^1) parent was crossed to variegated homozygous recessive M96006 ($b^{1vg}b^{1vg}$) parent, the reciprocal cross was also performed. No complementation was observed in either the F_2 or F_3 seed. Additionally, the variegated seed coat color was dominant to the yellow. Seed coat color in the F_3 generation did not appear in the 3:1 ratio of variegated to yellow as expected, this was likely due to semi-lethal action of the b^1 allele. Note: embryo and seed coat genotypes are shown as seed coats are made of maternal tissue and therefore are reflective of the previous generation's genotype.

3.5 Conclusions and Future Research

This study demonstrated that seed coat color conditioned by the *b1* locus is stably inherited; whereas lines remained the same seed coat color from one generation to the next. It was also shown that there was no transgressive segregation in seed coat color conditioned by this locus; whereas only the parental type phenotypic classes were observed within the RIL populations.

This study confirmed that seed coat color in the CDC Bethune x M96006 population is conditioned by a single locus as shown by the 1:1 ratio of individuals with brown seed coat color to individuals with variegated seed coat color. However, it was shown that the CDC Bethune x USDA-ARS Crystal population fit a 3:1 ratio suggesting that either two genes are responsible for conditioning seed coat color in the USDA-ARS Crystal parental line, the USDA-ARS Crystal parent of the population was a heterozygote at the *b1* locus or selection against lines with yellow seed coat color occurred. Future screening of this population with a marker linked to the *b1* locus should be done in order to understand the reason for this 3:1 ratio where if the marker appears in the population at a 1:1 ratio a second gene is likely the cause and if the marker appears at a 3:1 ratio selection or a heterozygous parental lines is likely the cause. However, this skewing of 3:1 brown to yellow seed coat color makes the CDC Bethune x USDA-ARS Crystal mapping population less amenable for future genetic studies and it is suggested that the CDC Bethune x M96006 population continue be used to identify the location of the *b1* locus.

An allelism test confirmed the allelic nature of the b^{1vg} and b^{1} alleles. This allelism test also showed for the first time that the variegated b^{1vg} allele if dominant over the yellow b^{1} allele where the heterozygous class all appears variegated. However, in the F_3 generation a 7:1 ratio of variegated to yellow seed coat color was observed which doesn't fit a simple dominance model but it is suggested that this is likely due to semi-lethality known to be associated with the b^{1} allele.

Chapter 4 The Relationships of Flax Traits with Seed Coat Color

4.1 Abstract

It is important for plant breeders to be aware of undesirable traits that may be inadvertently selected for when breeding for seed coat color. The objectives of this study were to determine if flower color, flower shape, grain yield, thousand seed weight, oil content, or oil profile have a relationship with seed coat color. To do this, data was collected from two recombinant inbred line (RIL) mapping populations grown in a replicated lattice design under field conditions. Each of these populations segregated for one of the two different alleles known to condition seed coat color at the *b1* locus.

This study showed that the *b1* locus was found to have a pleiotropic effect on flower color and flower shape where seeds with yellow seed coat color produced plants with white star shaped flowers and seeds with brown seed coat color had blue bowl shaped flowers. These results suggest that the *b1* locus either conditions the expression of enzymes upstream of where the flavonoid biosynthesis pathway diverges into the seed coat color and flower color branches, or it codes for a transcription factor that influences a suite of genes in both branches of the flavonoid pathway. Additionally, it was found in the CDC Bethune x M96006 population that grain yield, thousand seed weight, oil content and oil profile all independently assorted with seed coat color but in the CDC Bethune x USDA-ARS Crystal population linolenic fatty acid content was associated with seed coat color suggesting that one or more of the *FAD3* genes, known to condition linolenic fatty acid content, or their modifying genes are linked or are otherwise associated with the *b1* locus.

4.2 Introduction

Linseed-type flax is an important crop to Canada, especially to Saskatchewan where 70% of total Canadian and 25-30% of the total global flax is produced (Food and Agriculture Organization of the United Nations, 2013). Yellow seed coat color was previously used as a phenotypic marker to identify Solin flax. The Solin oil profile contains less than five percent linolenic fatty acid content compared to the greater than fifty percent that is required of traditional flaxseed oil. Due to this fact, there are grading regulations by the Canadian Grain Commissions that prevent the mixing of more than two percent yellow seeds in a shipment of brown seeded flax. Recently, the requirement of yellow seed coat color marking a Solin oil profile has been removed due to a lack of demand for this type of oil derived from flax. The grading regulations however, are still in effect despite seed coat color independently assorting from oil profile in flax.

There are four independently inherited loci that are known to be responsible for conditioning seed coat color, but their position in the genome and therefore linkage with other trait loci are unknown or not fully understood (Cloutier *et al.*, 2011; Mittapalli & Rowland, 2003). It is important to understand these associations because if two traits travel together, in either coupling or repulsion, indirect and sometimes undesirable selection can occur. It can also give evidence that seeds with yellow and brown seed coat colors are not compositionally different from each other, and therefore grade reductions should not occur.

In flax, and in other species like the *Brassicaceae* family (Simbaya *et al.*, 1995), grain yield has been found to be influenced by seed coat color (Culbertson & Kommendahl, 1956). In flax, lines with yellow seed coat color were found to yield lower than brown ones due to lower seed vigor, poorer stands (Culbertson & Kommendahl, 1956) and lower germination rate (Saeidi & Rowland, 1999b), where the opposite was found to be true within the *Brassicaceae* family. No previous studies have been published on the relationship between variegated seed coat color and flax grain yield.

Some published work has characterized the relationship between oil profile and oil content and seed coat color. Palmitic fatty acid content was found to segregate independently of variegated seed coat color (Saeidi & Rowland, 1997), which is conditioned by the b^{1vg} allele (Mittapalli & Rowland, 2003). Oil content on the other hand has been found in both the *Brassicaceae* family and in flax to be influenced by seed coat

color, where lines with a yellow seed coat tended to have higher oil content levels (Culbertson & Kommendahl, 1956; Saeidi & Rowland, 1999b; Simbaya *et al.*, 1995).

Yellow seed coat color is a result of a loss of pigments in the inner integument layers of the seed coat. These pigments are a product of the flavonoid biosynthesis pathway. This pathway is also known to be responsible for production of pigments in the other parts of the plant, such as the flowers (Lepeniec *et al.*, 2006; Taylor & Grotewold, 2005; Winkel-Shirley, 2001). Therefore it is reasonable that the *b1* seed coat color locus has been found to be pleiotropic to seed coat color, flower color and flower shape (Barnes *et al.*, 1960; Culbertson & Kommendahl, 1956), where lines that are homozygous for the recessive b^1 allele have yellow seeds and crimped white flowers. Seed coat color has also been found in previous studies to be associated with germination, seed vigor and yield. It is important to understand and be aware of these relationships, so that trait combinations can be fixed in future lines, and undesirable traits are not indirectly selected during the selection of a desired one.

The following research was conducted with the objective of identifying relationships between seed coat color and grain yield, thousand seed weight, oil content and oil profile in flax.

4.3 Materials and Methods

4.3.1 Plant Materials and Field Experiment

Two recombinant inbred line (RIL) populations grown in separate field tests in both 2012 and 2013 were used to collect phenotypic data for the trait relationship studies with seed coat color. The populations were developed by crossing CDC Bethune (Rowland *et al.*, 2001), a brown seeded variety, to M96006 (Saeidi & Rowland, 1997) a line with variegated seed coat color or to USDA-ARS Crystal (Arny, 1944) a yellow seeded variety (Table 3.1). For both populations reciprocal crosses were also performed but no viable individuals were produced from the USDA-ARS Crystal x CDC Bethune cross. Following the crosses, these populations were grown by single seed descent to the F_6 generation in the growth chamber. The $F_{6:7}$ generation and $F_{6:8}$ were grown under field conditions in 2012 and 2013 respectively at the Kernen Crop Research Farm (KCRF) (52^o N, 106°32'W) for the collection of phenotypic data used in this study, climate data for the field seasons is shown in table 3.2. The parental lines, CDC Bethune with M96006 or CDC Bethune with USDA-ARS Crystal, as well as CDC Sorrel were grown as check varieties for each population.

The $F_{6:7}$ and $F_{6:8}$ populations were grown as replicated (n=2) hand planted hills in square or rectangular lattice designs based on their number of entries. There were 50 seeds planted into each hill. The CDC Bethune x USDA-ARS Crystal population had 144 entries and was grown in a square lattice design. Therefore, the CDC Bethune x USDA-ARS Crystal population had 12 plots in each iBlock and 12 iBlocks in each repetition. The CDC Bethune x M96006 population had 110 entries and was grown in a rectangular lattice design with 10 plots in each iBlock and 11 iBlocks in each repetition. All plots were hand harvested at maturity and threshed using a vogel thresher.

4.3.2 Measuring the Qualitatively Inherited Traits

Seed Coat Color

Seed coat color measurements for this study were performed using a visual 6-point scale developed using the Munsell color system (Diederichsen & Richards, 2003) as shown in table 3.3. Seed coat color was assessed post-harvest after samples were cleaned with a column blower.

Flower Color and Shape

Flower color and shape were both assessed visually during the time between 5% and 95% plot flowering. All measurements pertaining to flowers were taken after 9 am but before noon in order to maintain consistency in the measurements as flowers open up, appear lighter and senesce as the day progresses. Flower color was measured using a 6-point scale used by the Plant Genome Resources Canada (PGRC) for cultivated flax, as well flower shape was measured using a 4-point scale modified from one used by the PGRC (Diederichsen & Richards, 2003) shown in table 4.1.

Table 4.1. Flax flower color and shape measurement scales for scoring the different morphological phenotypes.

Character	Scale	Phenotype
Flower Color	1	White
	2	Light Blue
	3	Blue
	4	Dark Blue
	5	Pink
	6	Violet
Flower Shape	1	Tube
	2	Funnel
	3	Bowl
	4	Star

4.3.3 Measuring Yield Component Traits

Grain Yield

Each plot was hand harvested and threshed with a vogel thresher using consistent fan speed for all plots of both populations. The samples were then fine cleaned using a column blower with consistent wind speed that was optimized to limit seed loss. Remaining debris was picked out of samples manually and the resultant sample was weighed with a bench top balance to two decimal places.

Thousand Seed Weight

Thousand seed weight was measured using a seed counter to dispense 250 seeds, weighing the resultant subsample with a bench top balance to two decimal places and then multiplying the value by four. This method was found to be more accurate than using the 1000 slotted trays. It was also found to be just as accurate but less time consuming than using the seed counter to dispense 1000 seeds based on a comparison of 10 samples.

4.3.4 NIR Spectroscopy to Measure Percent Oil Content

Near infrared radiation spectroscopy (NIR) was performed on samples to estimate the percent oil content. The seed sat on the lab bench top for a minimum of two weeks in order to equilibrate seed moisture of all samples to the lab conditions as seed moisture differences can affect NIR results. A NIRS 6500 (Foss, Hillerød, Denmark) was fitted with a tungsten-halogen lamp and a silicon/lead sulfide detector to detect transmission. This method was used because it is a way of quickly and non-destructively predicting the oil content. In order to increase the accuracy of the percent oil content measurement, crude fat was extracted and measured according to an ANKOM protocol (ANKOM Technology, Macedon, York, USA) from 30 samples that represented the oil content range for each population. An R^2 as well as a correction factor for the NIR data was calculated for each population.

4.3.5 Gas Chromatography to Measure Parameters of Oil Profile

Samples were ground using a GenoGrinder 2010 (OPS Diagnostics, Lebanon, New Jersey, USA) at 1400rpm for 4 min. Fatty acid methyl ester (FAME) gas chromatography (GC) was performed to measure the relative percentages of palmitic (16:0), stearic (18:0),

oleic (18:1), linoleic (18:2) and linolenic (18:3) fatty acid proportions. The gas chromatograph used a flame ionization detector, hydrogen as the carrier gas of the mobile phase and (50% cyanopropyl) methylpolysiloxane as the stationary phase.

4.3.6 Statistical Analysis

Data was analyzed using the SAS 9.3 statistical software package. Traits with homogeneity of variance within the data between the 2012 and 2013 growing seasons were combined while traits that failed the Levene's homogeneity of variance tests were analyzed separately.

To detect relationships between seed coat color with flower color and flower shape chi-squared tests for the determination of relationships between non-ordinal categorical traits was performed. A Yate's correction (Yates, 1934) was used to correct for only one degree of freedom.

Pooled and Satterthwaite two sample t-tests were performed to compare the performances of the parental lines of each population, where pooled was used for traits with homogeneous variance and Satterthwaite was used for traits with heterogeneous variance. Pooled and Satterthwaite two sample t-tests and frequency distributions using least significant means were generated to compare the performances of individuals with brown seed coat color to those with either variegated or yellow seed coat colors and to determine whether seed coat color independently assorted with the trait. Variance estimates of the random effects, F-values of the fixed effects and broad sense heritabilities were calculated for thousand seed weight, grain yield, oil content and all the parameters of oil profile using PROC MIXED. Seed coat color and genotype nested within seed coat color fixed effects were used to partition the variance to demonstrate if seed coat color and genotype had a significant effect on grain yield, thousand seed weight, oil content and oil profile. Broad sense heritabilities for the these traits were also estimated using the formula $H^2 = V_g/V_p$ (Reid *et al.*, 2009).

4.4 Results and Discussion

4.4.1 Relationship between Seed Coat Color and Flower Color and Shape

Both qualitatively and quantitatively inherited traits were examined for their relationship to seed coat color in the populations. Of the qualitative traits flower color and flower shape were measured in addition to seed coat color. All three of these traits were measured visually using discrete scales (Table 3.3 and 4.1).

In the CDC Bethune x M96006 population the parental lines had the same flower color and flower shape, all blue bowls, despite segregating for the variegated and brown seed coat colors (Table 3.7). The RILs were all observed to have blue bowl shaped flowers like the parents as expected. In the CDC Bethune x USDA-ARS Crystal population the parental lines had different flower morphology: CDC Bethune had blue bowl shaped flowers and USDA-ARS Crystal had white star shaped flowers (Table 4.2). The RILs were found to segregate for these phenotypes with no intermediate types. A chi-squared test was performed to test independence between seed coat color and flower color or flower shape to determine if there is a relationship between non-ordinal categorical traits. It was found that both flower color and flower shape had a significant relationship ($P<0.001$) (Table 4.3) with seed coat color; whereas, all yellow seeded lines had white star shaped flowers and all brown seeded lines had blue bowl shaped flowers. This is consistent with results found by Barnes et al. (1960) who concluded that the *b1* locus was pleiotropic to seed coat color, flower color and flower shape.

These results are interesting because seed coat color and flower color are known in *Arabidopsis* to be products of two different branches of the flavonoid biosynthesis pathway (Debeaujon *et al.*, 2003; Lepeniec *et al.*, 2006; Winkel-Shirley, 2001). This shows that the *b1* locus likely harbors a gene that acts prior to where the pathway diverges into the two different branches, or it codes for a transcription factor that alters the expression of genes within both branches of the pathway. From these findings possible future hypotheses to test would be that the *b1* locus conditions the expression of flavonoid biosynthesis enzymes upstream of where the pathway diverges into the seed coat color and flower color branches, or that the *b1* locus codes for a transcription factor that influences a suite of genes in both branches.

Table 4.2. Flower color° and flower shape¥ scores for the parental lines of the recombinant inbred line (RIL) mapping populations grown at the Kernen Crop Research Farm (KCRF).

Parental Line	Flower Color	Flower Shape
CDC Bethune	3	3
M96006	3	3
USDA-ARS Crystal	1	4

° 1-White, 2-Light Blue, 3-Blue, 4-Dark Blue, 5-Pink, 6-Violet

¥ 1-Tube, 2-Funnel, 3-Bowl, 4-Star

Table 4.3. A 1:1 Chi-squared test of flax seed coat color to flower color and flower shape to determine the presence of a relationship between non-ordinal categorical traits in the CDC Bethune x USDA-ARS Crystal recombinant inbred line (RIL) mapping population.

Relationship	**χ^2**	**df**	**P-value**
Seed coat color and flower color	268.0	1	<0.001***
Seed coat color and flower shape	268.0	1	<0.001***

*P<0.05, **P<0.01, ***P<0.001

4.4.2 Mean Comparisons, Variance Components Analysis and Heritability of the Quantitatively Inherited Traits

In table 4.4 the pooled and Satterthwaite two sample t-test results for the parental lines are listed for grain yield, thousand seed weight, oil content and oil profile in the CDC Bethune x M96006 RIL population that is segregating for the b^{1vg} seed coat color allele. The mean grain yield for the parental lines CDC Bethune, with brown seed coat color and M96006 with variegated, was significantly different (T-value=2.51; P<0.05), where CDC Bethune averaged 16.1g more seed per hill than M96006. However, when all the RIL individuals of each seed coat color are compared (Table 4.6) there was no significant difference between the average grain yields of lines with brown seed coat color versus ones with variegated. This shows that while the population is segregating for the grain yield trait, grain yield is independently assorting with seed coat color. In table 4.8 the variance components estimates for the random effects and F-values for fixed effects of all the previously mentioned traits for this population are shown. The variation found within grain yield in this population was significantly explained (F-value=4.07; P<0.05) by seed coat color, which contradicted the results from table 4.6. These results could be due to the relationship being too weak to be detected with the t-test, further data collection over years and/or increasing the population size would help to determine if grain yield is associated with seed coat color in this population. In figure 4.1 the frequency distribution of grain yield for lines in this population with brown and variegated seed coat colors shows that there are individuals of both seed coat colors at the upper and lower extremes. This shows regardless of whether there is a significant relationship between seed coat color and grain yield, lines of either seed coat color can be bred to be high yielding. Additionally, grain yield was estimated to have a low heritability at 0.0792 in this population so it will be difficult to determine lines from this population to be stably high yielding across environments and generations (Table. 4.8). This heritability estimate is extremely low and it is likely influenced by grain yield being measured from hill plots, grain yield should be measured on larger scale plots in order to get a more accurate heritability estimate.

In table 4.5 the pooled and Satterthwaite two sample t-test results of the means of the parental lines are listed for the traits mentioned in the previous paragraph for the CDC Bethune x USDA-ARS Crystal population that is segregating for the b^{1} seed coat color

allele. The mean grain yield between the parental lines CDC Bethune having a brown seed coat color, and USDA-ARS Crystal with yellow, was significantly different (T-value=3.16; P<0.01) where CDC Bethune averaged 18.8g more seed per hill than USDA-ARS Crystal. In contrast to the CDC Bethune x M96006 population, table 4.7 shows that when all the individual RILs of either brown or yellow seed coat color are compared in this population grain yield is significantly different (T-value=5.61; P<0.001) between the two where lines with a brown seed coat color yielded 7.1g more seed per hill than the lines with yellow seed coat color suggesting that grain yield is associated seed coat color in this population. The variance components of the random effects and F-values of the fixed effects in table 4.10 provides further evidence that grain yield is associated with seed coat color in this population where the variance found in grain yield was significantly (F-value=70.3; P<0.001) explained by seed coat color. These results support the findings of previous studies where lines with yellow seed coats tended to have lower grain yields (Culbertson & Kommendahl, 1956), which was attributed to lower seed vigor, poorer stands (Saeidi & Rowland, 1999b) and lower germination rate (Saeidi & Rowland, 1999a). While grain yield is associated with seed coat color in this population the frequency distribution of grain yield for both the brown and yellow seed coat colors found in figure 4.3 shows that there are high and low yielding lines of both seed coat color phenotypes and therefore, lines of either seed coat color can be bred to be high yielding. Table 4.9 shows that grain yield was estimated to have a low heritability at 0.123 so it will be difficult to select a line from this population that is consistently high yielding across environments and generations.

Table 4.4. The pooled two sample t-tests for traits with equal variances and Satterthwaite two sample t-tests for those with unequal variances to compare means for grain yield, thousand seed weight, oil content (Oil), palmitic (C16:0), stearic (C18:0), oleic (C18:1), linoleic (C18:2) and linolenic fatty acid content (C18:3) for CDC Bethune and M96006, with brown and variegated seed coat colors respectively, parental lines of the CDC Bethune x M96006 recombinant inbred line (RIL) mapping population grown in 2012 and 2013 respectively, at the Kernen Crop Research Farm (KCRF).

Trait	Method	Mean		|T-value|
		CDC Bethune	M96006	
Grain yield (g)	Pooled	49.1	33.0	2.51*
Thousand seed weight (g)	Pooled	6.14	5.11	6.62***
Oil (%)	Pooled	42.2	39.6	8.25***
C16:0 (%)	Pooled	5.4	19.0	60.8***
C18:0 (%)	Satterthwaite	3.74	3.05	9.92***
C18:1 (%)	Satterthwaite	24.6	16.2	6.77***
C18:2 (%)	Pooled	16.5	57.0	50.5***
C18:3 (%)	Pooled	49.7	2.85	39.9***

*P<0.05, **P<0.01, ***P<0.001

Table 4.5. The pooled two sample t-tests for traits with equal variances and Satterthwaite two sample t-tests for those with unequal variances to compare means for grain yield, thousand seed weight, oil content (Oil), palmitic (C16:0), stearic (C18:0), oleic (C18:1), linoleic (C18:2) and linolenic fatty acid content (C18:3) for CDC Bethune and USDA-ARS Crystal, with brown and yellow seed coat colors respectively, parental lines of the CDC Bethune x USDA-ARS Crystal recombinant inbred line (RIL) mapping population grown in 2012 and 2013 respectively, at the Kernen Crop Research Farm (KCRF).

Trait	Method	Mean		|T-value|
		CDC Bethune	USDA-ARS Crystal	
Grain yield (g)	Pooled	48.8	29.9	3.16**
Thousand seed weight (g)	Pooled	6.01	6.54	2.65*
Oil (%)	Pooled	42.2	39.1	12.7***
C16:0 (%)	Pooled	5.37	6.39	7.92***
C18:0 (%)	Pooled	3.60	2.83	7.07***
C18:1 (%)	Satterthwaite	24.7	16.3	7.74***
C18:2 (%)	Pooled	16.0	15.5	3.66**
C18:3 (%)	Satterthwaite	50.3	59.0	7.28***

*P<0.05, **P<0.01, ***P<0.001

Table 4.6. The pooled two sample t-tests for traits with equal variances and Satterthwaite two sample t-tests for those with unequal variances to compare means for grain yield,

thousand seed weight, oil content (Oil), palmitic (C16:0), stearic (C18:0), oleic (C18:1), linoleic (C18:2) and linolenic fatty acid content (C18:3) for lines with brown seed coat color versus ones with variegated in the $F_{6:7}$ and the $F_{6:8}$ CDC Bethune x M96006 recombinant inbred line (RIL) mapping population grown in 2012 and 2013 respectively, at the Kernen Crop Research Farm (KCRF).

		Mean		
Trait	**Method**	**Brown**	**Variegated**	**|T-value|**
Grain yield (g)	Pooled	42.7	40.2	1.71
Thousand seed weight (g)	Pooled	5.58	5.61	0.71*
Oil (%)	Pooled	41.1	42.0	5.63***
C16:0 (%)	Pooled	10.5	9.82	1.28
C18:0 (%)	Satterthwaite	3.47	3.34	2.96**
C18:1 (%)	Satterthwaite	19.9	18.8	3.50**
C18:2 (%)	Pooled	33.7	33.1	0.34
C18:3 (%)	Pooled	31.7	34.3	1.36

*P<0.05, **P<0.01, ***P<0.001

Table 4.7. The pooled two sample t-tests for traits with equal variances and Satterthwaite two sample t-tests for those with unequal variances to compare means for grain yield, thousand seed weight, oil content (Oil), palmitic (C16:0), stearic (C18:0), oleic (C18:1), linoleic (C18:2) and linolenic fatty acid content (C18:3) for lines with brown seed coat color versus ones with yellow in the $F_{6:7}$ and the $F_{6:8}$ CDC Bethune x USDA-ARS Crystal recombinant inbred line (RIL) mapping population grown in 2012 and 2013 respectively, at the Kernen Crop Research Farm (KCRF).

		Mean		
Trait	**Method**	**Brown**	**Yellow**	**|T-value|**
Grain yield (g)	Pooled	44.3	37.2	5.61***
Thousand seed weight (g)	Pooled	6.39	6.40	0.31
Oil (%)	Pooled	40.1	40.4	2.40*
C16:0 (%)	Pooled	5.91	5.91	0.07
C18:0 (%)	Pooled	3.26	3.10	3.52**
C18:1 (%)	Satterthwaite	22.7	18.1	17.7***
C18:2 (%)	Pooled	16.6	15.8	8.75***
C18:3 (%)	Satterthwaite	51.5	57.1	18.8***

*P<0.05, **P<0.01, ***P<0.001

Table 4.8. Variance estimates for random effects, F-values for fixed effects from analysis of variance (ANOVA) using PROC MIXED of grain yield (Yield), thousand seed weight (TSW), oil content (Oil), palmitic (C16:0), linoleic (C18:2) and linolenic fatty acid content (C18:3) in the $F_{6:7}$ and the $F_{6:8}$ CDC Bethune x M96006 recombinant inbred line (RIL) mapping population grown in 2012 and 2013 respectively, at the Kernen Crop Research Farm (KCRF).

Source	Yield (g)	TSW (g)	Oil (%)	C16:0 (%)	C18:2 (%)	C18:3 (%)
Random effects Variance estimates						
Year	208	0.0506	0.646	0.1128	0.0746	5.56
Block(year)	33.9	0.0248	0.0188	0	0.0353	0.0653
iBlock(block*year)	22.9***	0.0247***	0.0195*	0.00449	0	0
Genotype*year	0	0.0992**	0.148***	0.201***	2.18***	4.82***
Residual	45.4***	0.0213***	0.126***	0.296***	2.78***	3.96***
Fixed effects F-values						
Seed color	4.07*	0.69	115***	39.2***	5.61*	36.4***
Genotype(seed color)	2.92***	6.74***	17.2***	147***	163***	96.3***
H^2	0.0792***	0.349***	0.688***	0.977***	0.983***	0.961***

*P<0.05, **P<0.01, ***P<0.001, ($H^2= V_g/V_p$)

Table 4.9. Variance estimates for random effects, F-values for fixed effects from analysis of variance (ANOVA) using PROC MIXED of stearic (18:0) and oleic fatty acid content (18:2) in the $F_{6:7}$ and the $F_{6:8}$ CDC Bethune x M96006 recombinant inbred line (RIL) mapping population grown in 2012 and 2013 respectively, at the Kernen Crop Research Farm (KCRF). These traits were separated by year due to heterogeneity in the variances between field seasons.

	2012		2013	
Source	C18:0 (%)	C18:1 (%)	C18:0 (%)	C18:1 (%)
Random effects Variance estimates				
Block	0.000235	0.448	0.00274	0
iBlock	0.0107*	0.198	0.0000180	0.165
Residual	0.0240***	0.964***	0.0210***	1.26***
Fixed effects F-values				
Seed color	19.7***	15.57***	43.7***	41.0***
Genotype(seed color)	10.3***	4.25***	19.4***	9.91***
H^2	0.779***	0.515***	0.903***	0.812***

*P<0.05, **P<0.01, ***P<0.001, ($H^2= V_g/V_p$)

Table 4.10. Variance estimates for random effects, F-values for fixed effects from analysis of variance (ANOVA) using PROC MIXED of grain yield (Yield), thousand seed weight (TSW), oil content (Oil), palmitic (C16:0), stearic (18:0) and linoleic fatty acid content (C18:2) in the $F_{6:7}$ and the $F_{6:8}$ CDC Bethune x USDA-ARS Crystal recombinant inbred line (RIL) mapping population grown in 2012 and 2013 respectively, at the Kernen Crop Research Farm (KCRF).

Source	Yield (g)	TSW (g)	Oil (%)	C16:0 (%)	C18:0 (%)	C18:2 (%)
Random effects Variance estimates						
Year	178	0.0544	0.709	0.0056	0.0765	0.00251
Block(year)	5.45	0.00524	0	0.00829	0.000209	0
iBlock(block*year)	24.3***	0.0118**	0.0132	0.00757*	0.00649**	0.000915
Genotype*year	14.1**	0.0148*	0.367***	0	0.00788**	0.178***
Residual	47.4***	0.0818***	0.253**	0.0789***	0.0362***	0.200***
Fixed effects F-values						
Seed color	70.3***	0.280	11.3***	1.37	65.3***	151***
Genotype(seed color)	2.37***	6.43***	5.61***	8.72***	14.7***	7.04***
H^2	0.123***	0.485***	0.468***	0.620***	0.606***	0.725***

*P<0.05, **P<0.01, ***P<0.001, ($H^2= V_g/V_p$)

Table 4.11. Variance estimates for random effects, F-values for fixed effects from analysis of variance (ANOVA) using PROC MIXED of oleic (C18:1) and linolenic fatty acid content (C18:3) in the $F_{6:7}$ and the $F_{6:8}$ CDC Bethune x USDA-ARS Crystal recombinant inbred line (RIL) mapping population grown in 2012 and 2013 respectively, at the Kernen Crop Research Farm (KCRF). These traits were separated by year due to heterogeneity in the variances between field seasons.

	2012		2013	
Source	C18:1 (%)	C18:3 (%)	C18:1 (%)	C18:3 (%)
Random effects Variance estimates				
Block	0.0150	0.0867	0	0.986
iBlock	0.123	0.162	0	0
Residual	1.947***	2.90***	8.38***	27.2***
Fixed effects F-values				
Seed color	401***	480***	312***	212***
Genotype(seed color)	2.56***	2.97***	1.43*	1.48*
H^2	0.692***	0.727***	0.622***	0.476***

*P<0.05, **P<0.01, ***P<0.001, ($H^2= V_g/V_p$)

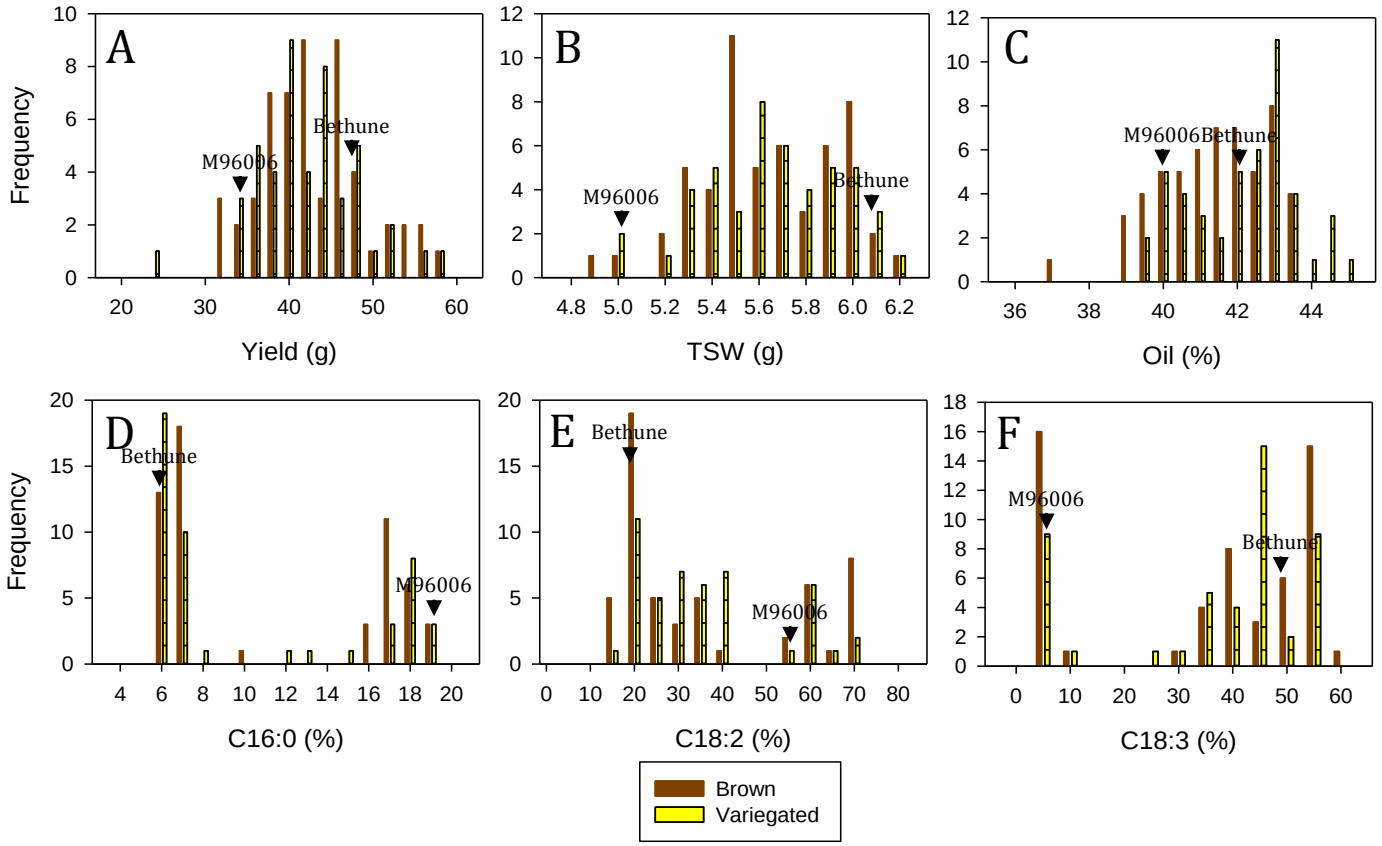


Figure 4.1. The frequency distributions based on least squares means of lines with variegated and brown seed coat color for (A) grain yield, (B) thousand seed weight, (C) oil content, (D) palmitic fatty acid content, (E) linoleic fatty acid content and (F) linolenic fatty acid content of the CDC Bethune x M96006 recombinant inbred line (RIL) mapping population grown in 2012 and 2013 respectively, at the Kernen Crop Research Farm (KCRF).

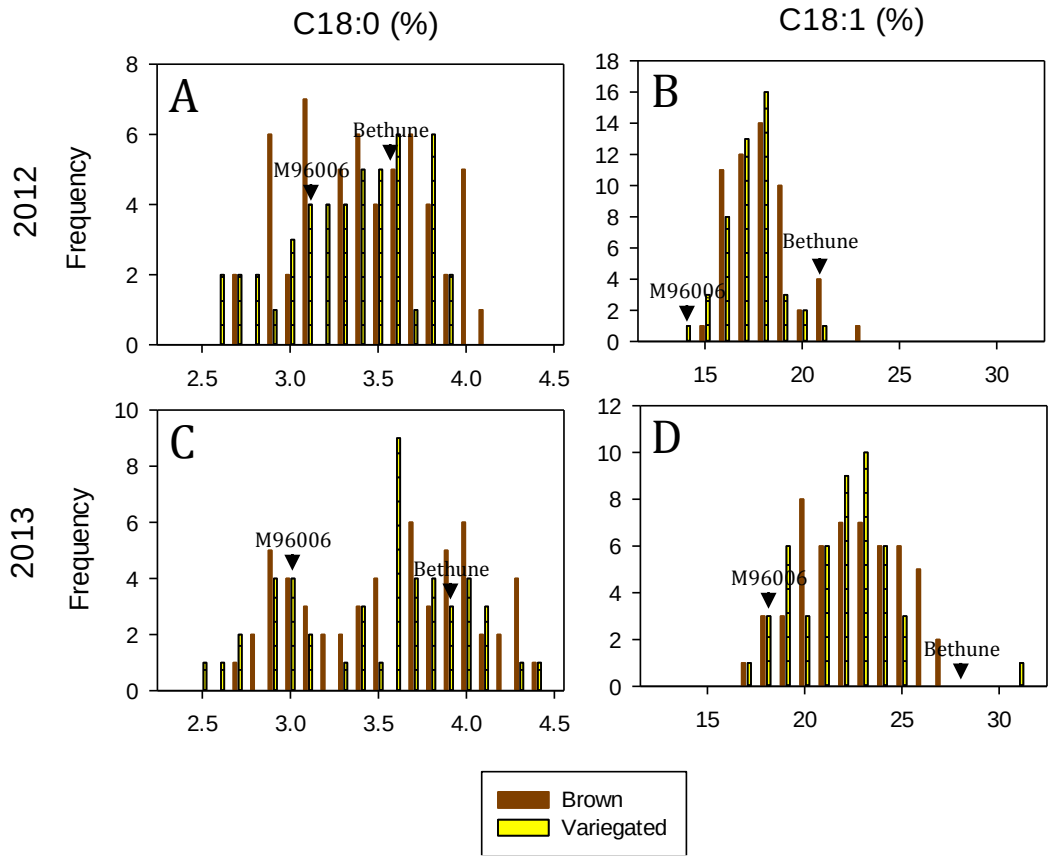


Figure 4.2. The frequency distributions based on least squares means of lines with variegated and brown seed coat color for (A) stearic fatty acid content in 2012, (B) stearic fatty acid content in 2013, (C) oleic fatty acid content in 2013 and (D) oleic fatty acid content in 2013 of the CDC Bethune x M96006 recombinant inbred line (RIL) mapping population, grown at the Kernen Crop Research Farm (KCRF). These traits were separated by year due to heterogeneity in the variances between field seasons.

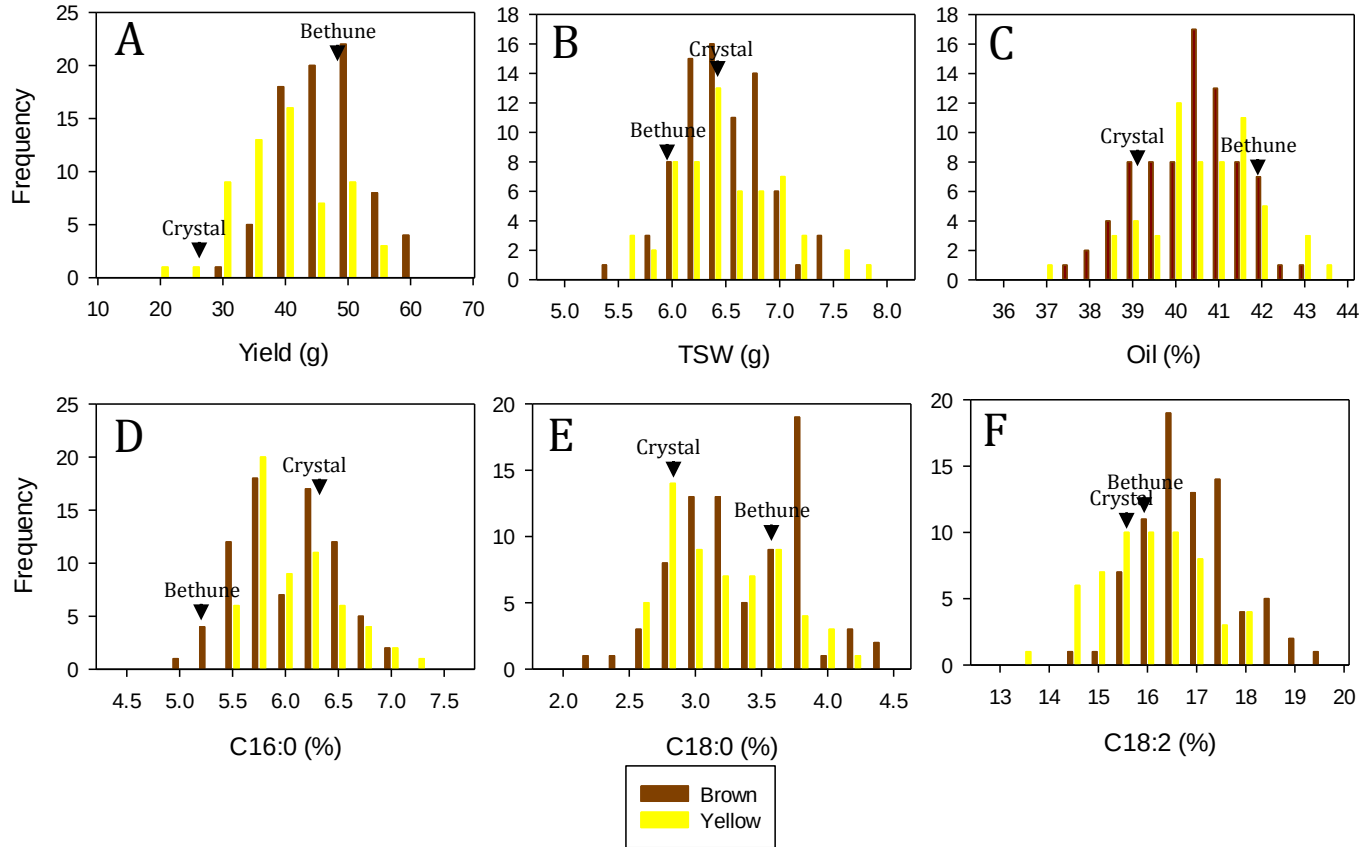


Figure 4.3. The frequency distributions based on least squares means of lines with yellow and brown seed coat color for (A) grain yield, (B) thousand seed weight, (C) oil content, (D) palmitic fatty acid content, (E) stearic fatty acid content and (F) linoleic fatty acid content of the CDC Bethune x USDA-ARS Crystal recombinant inbred line (RIL) mapping population in 2012 and 2013 respectively, grown at the Kernen Crop Research Farm (KCRF).

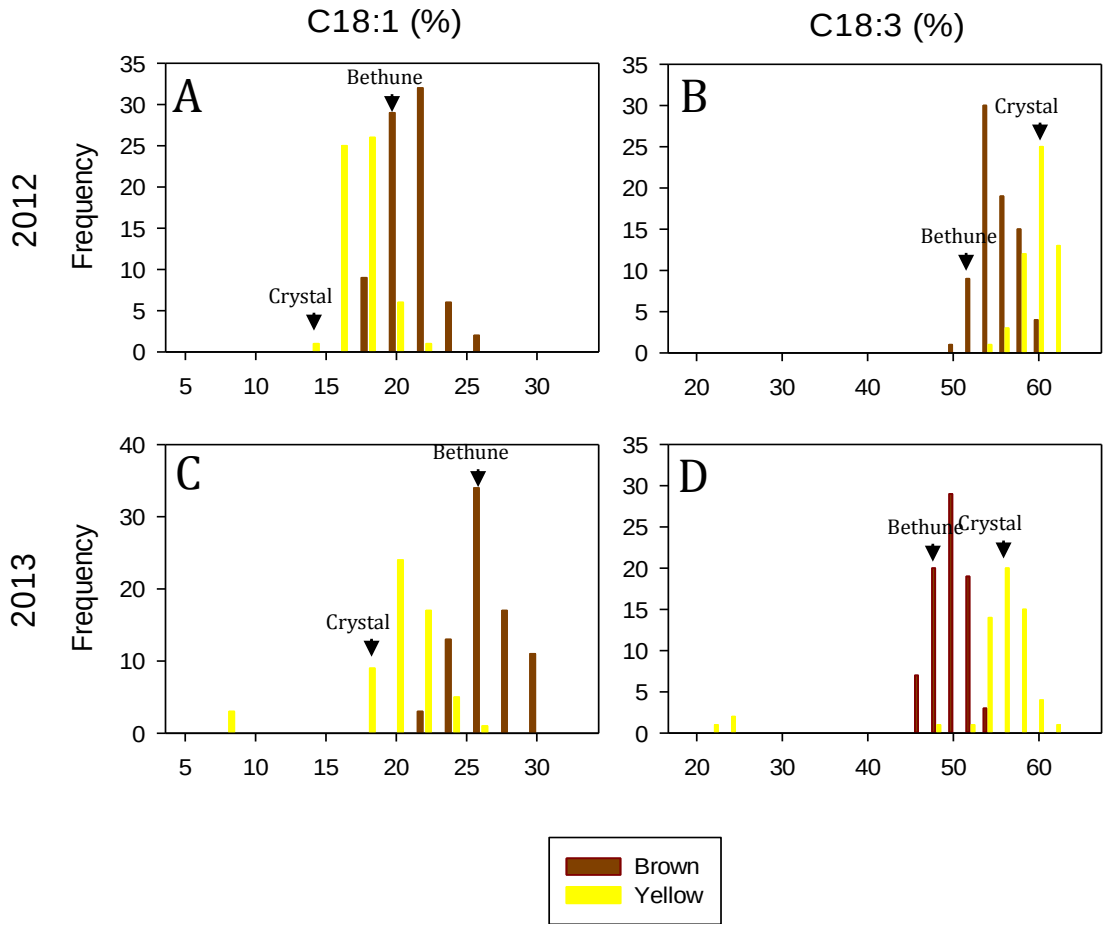


Figure 4.4. The frequency distributions based on least squares means of lines with yellow and brown seed coat color for (A) oleic fatty acid content in 2012, (B) oleic fatty acid content in 2013, (C) linolenic fatty acid content in 2013 and (D) linolenic fatty acid content in 2013 of the CDC Bethune x USDA-ARS Crystal recombinant inbred line (RIL) mapping population, grown at the Kernen Crop Research Farm (KCRF). These traits were separated by year due to heterogeneity in the variances between field seasons**.**

As seen in table 4.4 the mean thousand seed weight of the two parental lines for the CDC Bethune x M96006 population are significantly different (T-value=6.62; P<0.001), where the mean for CDC Bethune with brown seed was 6.14g for a thousand seeds and M96006 with variegated seed was 5.11g. When the individual RILs of the population were compared it showed that lines with a variegated seed coat color had a mean of 5.61g for a thousand seeds which was found to be significantly higher (T-value=0.71; P<0.05) than the mean of 5.58g for lines with brown seed coat color (Table 4.6) suggesting that seed weight is associated with seed coat color in this population. In contrast to these results, the F-values for thousand seed weight in table 4.8 show that there was not a significant relationship between thousand seed weight and seed coat color which contradict the previous findings. The difference in means between thousand seed weight for lines with brown and variegated seed coat colors may be a false positive result or it may just be a weak relationship that was not detected by the variance components analysis. Further data collection or expansion of the test population will help to determine if thousand seed weight is associated with seed coat color. Figure 4.1 shows that regardless of whether thousand seed weight is associated with seed coat color it is possible to breed lines of either seed coat color with high or low seed weight as there are individuals of both seed coat colors in the high and low ends of the frequency distribution. Thousand seed weight was estimated to be moderately heritable at 0.349 and therefore lines selected from this population will have moderately stable seed weights across environments and generations (Table 4.8).

In table 4.5 the mean thousand seed weight of the two parental lines for the CDC Bethune x USDA-ARS Crystal population were found significantly different (T-value=2.65; P<0.05), where the mean for CDC Bethune with brown seed was 6.01g for a thousand seeds and USDA-ARS Crystal with yellow seed was 6.54g. But when the individual RILs of the population are compared it showed that there were no significant differences between the means of lines with brown and yellow seed coat color phenotypes (Table 4.7). Table 3.15 shows similar results where seed coat color was found to not significantly explain the variance found within thousand seed weight. These findings suggest that thousand seed weight is not associated with seed coat color in this population. These results in addition to the frequency distributions in figure 4.3 show that lines of either seed coat color can be bred to have high or low thousand seed weight. Furthermore,

thousand seed weight was estimated to be moderately heritable at 0.485 and therefore selected lines from this population will have moderately stable thousand seed weight across environments and generations (Table 4.10).

The mean oil content of each parental line for the CDC Bethune x M96006 population shown in table 4.4, CDC Bethune with brown seed coat color had 42.2 percent oil content in the seed was found to be significantly higher (T-value=8.25; P<0.001) than M96006 with variegated seed coat color that had 39.6 percent. When the means of the individual RILs were compared it was shown that lines with variegated seed coat color actually had significantly higher (T-value=5.63; P<0.001) oil content where variegated seed coat colored lines had 42.0 percent and brown seed coat color lines had 41.1 percent. Additionally, in table 4.8 the F-value for seed coat color was significant for the oil content trait (F-value=115; P<0.001), these results suggest that oil content is associated with seed coat color in this population where lines with variegated seed coat color have more seed oil. In the frequency distribution (Figure 4.1) of oil content for brown and variegated seed coat color phenotypes it is shown that there are lines of either seed coat color that had high and low oil content. This shows that even though oil content is associated with seed coat color it is still possible to breed lines of either seed coat color to have high oil content. Additionally, the heritability for oil content in this population is estimated to be high at 0.688 (Table 4.8) which shows that lines from this population can be chosen to have high oil content that is stable across environments and generations (Table 4.8).

As shown in table 4.5 the mean oil content for each parental line in the CDC Bethune x USDA-ARS Crystal were significantly different (T-value=12.7; P<0.001), where brown seeded CDC Bethune had 42.2 percent seed oil content and yellow seeded USDA-ARS Crystal had 39.1. When the individual RILs of each seed coat color were compared in table 4.7 lines with yellow seed coat color had 0.3 percent higher seed oil content which was found to be significantly different (T-value=2.40; P<0.05) then seeds with brown seed coat color. This provides evidence that seed coat color is associated with seed oil content in this population, further evidence of this is that seed coat color significantly explained (F-value=11.3; P<0.001) the variance found within oil content (Table 4.10). Despite this association it is possible to breed a line with high oil content with either seed coat color from this population, as shown in figure 4.3 where there are lines of

both colors in the high end of the frequency distribution. Additionally, the heritability of seed oil content within this population is moderate (0.468) so lines selected to have high oil content with either seed coat color in this population will be stable across environments and generations (Table 4.10).

The mean palmitic fatty acid content between each parental line of the CDC Bethune x M96006 population was significantly different (T-value=60.8; P<0.001), where brown seeded CDC Bethune had 5.4 percent of the total seed oil content made up of palmitic fatty acid and variegated seeded M96006 had 16.0 percent (Table 4.4) this difference is due to M96006 being bred to be high in palmitic fatty acid (Saeidi & Rowland, 1997). The t-tests in table 4.6 between the mean of the lines from either seed coat color of this population show that there was no significant difference in palmitic fatty acid content, which supports the previous findings where high palmitic fatty acid content segregated independently from the variegated seed coat color trait (Saeidi & Rowland, 1997). This suggests there is no association between seed coat color and palmitic fatty acid content in this population. However, in table 4.8 the F-value measuring how much variance within palmitic fatty acid content is explained by seed coat color was found to be highly significant (F-value=39.2; P<0.001) suggesting that there is a relationship between the two traits. In the frequency distribution found in figure 4.1 the curves for individuals of either seed coat color completely overlap for palmitic fatty acid content showing that regardless of a relationship there are individuals that can be selected with either high or low palmitic fatty acid content of either seed coat color. Additionally, as shown in figure 4.1 the frequency distribution is bimodal showing that the difference in palmitic fatty acid content is likely controlled by a major gene with minor genes with small effect in this population (Fehr, 1993). The extremely high heritability (0.977) of palmitic fatty acid content provides further evidence that a single gene is responsible for palmitic fatty acid content in this population, it also shows that lines from this population can be selected to have consistently high or low palmitic fatty acid content over environments and generations (Table 4.8).

As shown in table 4.5 the mean palmitic fatty acid content between the parental lines of the CDC Bethune x USDA-ARS Crystal population are 5.67 percent of the total seed oil content for the brown seeded CDC Bethune parent and 6.39 for the yellow seeded USDA-ARS Crystal parent which were significantly different (T-value=7.92; P<0.001)

from each other. When the means of the individual RILs of each seed coat color were compared lines with both brown and yellow seed coat color had a mean palmitic fatty acid content of 5.91 percent, this shows that palmitic fatty acid content is segregating independently of seed coat color and therefore, the traits are not associated with each other (Table 4.7). Not surprisingly, similar results were found in table 4.10 where the F-value for how well seed coat color explains palmitic fatty acid content was found to be not significant. Additionally, the frequency distribution in figure 4.3 for the palmitic fatty acid content of individual RILs separated by seed coat color show that there is complete overlap where there are lines of either seed coat color in the high or low palmitic fatty acid content extremes of the distribution. All of this evidence shows that palmitic fatty acid content is not associated with either seed coat color in this population. The heritability of palmitic fatty acid content in this population was found to be high (0.620), which shows that lines from this population will have stable palmitic fatty acid content over environments and generations (Table 4.10).

The mean stearic fatty acid content between the parental lines for the CDC Bethune x M96006 population were found to be significantly different from each other (T-value=9.92; P<0.001), where CDC Bethune with brown seed coat color had 3.74 percent stearic fatty acid of the total seed oil content and M96006 with variegated seed coat color had 3.05 percent (Table 4.4). When the individual RILs of either brown or variegated seed coat color were compared from this population the mean stearic fatty acid content was significantly different (T-value=2.96; P<0.001) where lines with a brown seed coat color had a mean of 3.47 and lines with variegated seed coat color had a mean of 3.34 (Table 4.6). Additionally, as seen in table 4.9 seed coat color significantly explained the variance in stearic fatty acid content between the lines of this population when grown in 2012 (F-value=19.7; P<0.001) and 2013 (F-value=43.7; P<0.001). All of the above is evidence that stearic fatty acid content is associated with seed coat color in this population where lines with brown seed coat color tend to have more stearic fatty acid than seeds with a variegated seed coat color. However, in the frequency distributions in figure 4.2 there is complete overlap of individuals of either seed coat color at the high and low ends of the distribution showing that lines of either seed coat color can be bred to have high or low stearic fatty acid content in this population. The heritability of stearic fatty acid content was high both

in 2012 (0.779) and 2013 (0.903) showing that lines can be selected from this population to have consistent stearic fatty acid content over different generations and environments (Table 4.9).

As shown in table 4.5 the mean stearic fatty acid content of the two parental lines for the CDC Bethune x USDA-ARS Crystal population were significantly different (T-value=7.07; P<0.001) where CDC Bethune with brown seed coat color had a mean of 3.60 percent stearic fatty acid content of the total seed oil content and USDA-ARS Crystal with yellow seed coat color had 2.83. The means of the individual RILs of each seed coat color in this population were significantly different (T-value=3.52; P<0.01) where lines with brown seed coat color had a mean of 3.26 while those with yellow seed coat color had a mean of 3.10 percent stearic fatty acid content (Table 4.7). Additionally, in table 4.10 seed coat color significantly explained the variance found in stearic fatty acid content within the population (F-value=65.3; P<0.001). All of this is evidence that stearic fatty acid content is associated with seed coat color where lines with brown seed coat color tend to have a higher stearic fatty acid content than those with a yellow seed coat color. However, in the frequency distribution for stearic fatty acid content in figure 4.3 there is complete overlap of individuals of either seed coat color that have high and low levels of stearic fatty acid content showing that lines can be bred from this population that have low or high levels of stearic fatty acid with either seed coat color. The heritability of stearic acid in this population is high at 0.606 showing that lines will have consistent stearic fatty acid content over different generations and environments (Table 4.10).

The mean oleic fatty acid content was significantly different between the parental lines of the CDC Bethune x M96006 population (T-value=6.77, P<0.001), where CDC Bethune with brown seed coat color had a mean of 24.6 percent oleic fatty acid of the total seed oil and M96006 with variegated seed coat color had 16.2 percent (Table 4.4). When the individual lines of this population were compared the mean oleic fatty acid content for lines with brown seed coat color were significantly higher (T-value=3.50, P<0.01) at 19.9 percent, while those with variegated seed coat color had 18.8 percent (Table 4.6). Similarly, in table 3.14 the F-value of how well the variance of oleic fatty acid content is explained by seed coat color was significant both in 2012 (F-value=15.57, P<0.001) and 2013 (F-value=41.0, P<0.001). These results show that oleic fatty acid content is associated with

seed coat color where lines with a brown seed coat color tend to have higher oil content then lines with variegated seed coat color. However, as shown in figure 4.2 the frequency distribution of oleic fatty acid content both in 2012 and 2013 there are lines of either seed coat color that had high or low levels of oleic fatty acid content. This in addition to the high heritability of oleic fatty acid content in 2013 (0.515) and 2013 (0.812) show that lines of either seed coat color from this population can be bred to have stable high or low oleic fatty acid content across environments and generations (Table 4.9).

As shown in table 4.5 the mean oleic fatty acid content is significantly different between the parental lines of the CDC Bethune x USDA-ARS Crystal population (T-value=7.74, P<0.001), where CDC Bethune with brown seed coat color had 24.7 percent oleic fatty acid of its total seed oil content and USDA-ARS Crystal with yellow seed coat color had 16.3 percent. When the individual RILs of this population for either seed coat color were compared it was shown in table 4.7 that lines with brown seed coat color had significantly higher mean oleic fatty acid content at 22.7 percent while those with a yellow seed coat color had 18.1 percent (T-value=17.7, P<0.001). Similarly, as shown in table 4.11 the F-values for how well the variance within oleic fatty acid content is explained by seed coat color was significant both in 2012 (F-value=401, P<0.001) and 2013 (F-value=312, P<0.001). These results show that oleic fatty acid content in this population is associated with seed coat color where lines with brown seed coat color tend to have higher oleic fatty acid content then lines with yellow seed coat color. However, as shown in the frequency distributions for oleic fatty acid content in 2012 and 2013 between the two seed coat colors of this population, there are some lines of either seed coat color with high or low oleic fatty acid content levels. These results in addition to the high heritability of oleic fatty acid content both in 2012 (0.692) and 2013 (0.622) show that lines of either seed coat color can be bred to have high or low oleic fatty acid content that remains stable over environments and generations (Table 4.9).

The mean linoleic fatty acid content was significantly different between the parental lines of the CDC Bethune x M96006 population (T-value=50.5; P<0.001), where CDC Bethune with brown seed coat color had a mean of 16.5 percent of total seed oil content and M96006 with a variegated seed coat color had a mean of 57.0 percent (Table 4.4). When the individual RILs of both seed coat colors were compared from this population it

was found that there was no significant difference between the mean linoleic fatty acid content (Table 4.6), which suggests that there is no association between seed coat color and linoleic fatty acid content. However, in table 4.8 it was found that linoleic fatty acid content in this population was significantly explained by seed coat color (F-value=5.61; $P<0.05$) which indicates that there may be an association between these two traits in this population. It is possible that the relationship may be weak and therefore not detected by the T-test in table 4.6, further data collection over years and different environments will help elucidate if seed coat color is associated with linoleic fatty acid content in this population. Regardless, of this potential association it is possible to breed lines from this population with either seed coat color that have high or low linoleic fatty acid content because, as shown in the frequency distribution of linoleic fatty acid in figure 4.1, there are lines of both seed coat colors in the high and low ends of the distribution. Additionally, linoleic fatty acid content had a very high heritability in this population (0.983) so these lines will have consistent linoleic fatty acid content across environments and generations (Table 4.8).

As shown in table 4.5 the mean linoleic fatty acid content between the two parental lines of the CDC Bethune x USDA-ARS Crystal population were significantly different (T-value=3.66; $P<0.01$), where CDC Bethune with brown seed coat color had a mean of 16.0 percent linoleic fatty acid content of the total seed oil and USDA-ARS Crystal with yellow seed coat color had 15.5 percent. When the mean linoleic fatty acid content for the individual RILs either seed coat color were compared for this population they were significantly different (T-value=8.75; $P<0.001$), where lines with a brown seed coat color had a mean of 16.6 percent and lines with yellow seed coat color had 15.8 percent (Table 4.7). Similarly, as shown in table 4.10 the F-value for how well seed coat color explains the variance found within linoleic fatty acid content was significant (F-value=151; $P<0.001$). These results show that seed coat color is associated with linoleic fatty acid content in this population where lines with a brown seed coat color tend to have higher linoleic fatty acid content then those with yellow seed coat color. Despite this association it is possible to breed lines of either seed coat color from this population to have either high or low linoleic fatty acid content as shown in the frequency distribution in figure 4.3, where there are lines of either seed coat color in the high or low end of the distribution. The

heritability of linoleic fatty acid content was calculated to be high (0.725) so it will remain consistent in lines from this population over different environments and generations.

The mean linolenic fatty acid content between the parental lines of the CDC Bethune x M96006 population were significantly different (T-value 39.9; P<0.001), where CDC Bethune with brown seed coat color had a mean of 49.7 percent of the total seed oil and M96006 with variegated seed coat color had a mean of 2.85 percent (Table 4.4). When means for the individual RILs of either seed coat color from this population were compared there was no significant difference between them suggesting that there is no association between seed coat color and linolenic fatty acid content within this population (Table 4.6). However, the F-value for how well seed coat color explains the variance found within linolenic fatty acid content within this population was significant (F-value=36.4; P<0.001), suggesting that linolenic fatty acid content has an association with seed coat color, where lines with brown seed coat color tend to have less linolenic fatty acid than lines with variegated seed coat color. These contradictory results are likely due to an association that was too weak to be detected by the t-test in table 4.6 and further data collection over environments and generations will help to determine if these traits are associated. Regardless of whether an association is found, it is possible to breed lines from this population with either seed coat color that are high or low in linolenic fatty acid content as there were lines of either seed coat color in both extremes of its frequency distribution (Figure 4.1). Additionally, the heritability of linolenic fatty acid content was very high (0.961) and therefore lines selected from this population will have consistent linolenic fatty acid content over different environments and generations.

In table 4.5 the mean linolenic fatty acid content for both parental lines of the CDC Bethune x USDA-ARS Crystal population are compared and were found to be significantly different (T-value=7.28; P<0.001), where CDC Bethune with a brown seed coat color had a mean of 50.3 percent linolenic fatty acid content of the total seed oil and USDA-ARS Crystal with a yellow seed coat color had 59.0 percent. Similar results were seen when the individual RILS from this population of either seed coat color were compared there was a significant difference between the means (T-value=18.8; P<0.001), where lines with a brown seed coat color had a mean of 51.5 percent linolenic fatty acid of the total seed oil and lines with a yellow seed coat color had a mean of 57.1 percent (Table 4.7).

Additionally, it was found that the variance found within linolenic fatty acid content for the individual lines was significantly explained by seed coat color in both in 2012 (F-value=480; P<0.001) and 2013 (F-value=212; P<0.001). These results all show that linolenic fatty acid content is associated with seed coat color where lines with a yellow seed coat tend to have higher linolenic fatty acid content. In the frequency distributions in figure 4.4 when grown in both 2012 and 2013 lines with very high linolenic fatty acid content all had yellow seed coat color showing that it would not be possible to select lines from this population that have high linolenic fatty acid content with brown seed coat color. This suggests that a gene responsible for linolenic fatty acid content is linked or otherwise associated to the *b1* seed coat color locus. Conversely, the CDC Bethune x M96006 population showed that the b^{1vg} allele was not associated with linolenic fatty acid content. These results may be due to M96006 derived from McGregor flax possessing different alleles for the *FAD3* genes or their modifying genes known to be responsible for linolenic fatty acid content in flax than USDA-ARS Crystal (Kenaschuk & Hoes, 1986). There 15 and 18 known alleles for *FAD3a* and *FAD3b* genes that result in 6 and 7 different isoforms respectively, which in addition to modifying gene effects can result in variable linolenic fatty acid content level from one genotype to the next (Thambugala & Cloutier, 2014; Thambugala *et al.*, 2013). The broad sense heritability for this trait was high in 2012 (0.727) and moderate in 2013 (0.476), showing that lines selected from this population will have consistent levels of linolenic fatty acid content over different environments and generations.

These results are preliminary as this test was performed in two station-year environments. Repeating these experiments in more environments and/or expanding the test populations should be done in order to achieve more confidence in the results. Additionally, another field test should be done to directly compare differences between lines with variegated seed coats and yellow seed coats which would allow plant breeders to choose the best seed coat color to use as a phenotypic marker for important seed quality traits.

4.5 Conclusions and Future Research

This study demonstrated that in the CDC Bethune x USDA Crystal population the *b1* locus which conditions seed coat color is pleiotropic to flower color and flower shape; whereas, all lines with yellow seed coats had white star shaped flowers and all lines with brown seed coats had blue bowl shaped flowers.

This study looked for the first time at the relationships between the variegated seed coat color conditioned by the b^{1vg} allele and its relationship with grain yield, thousand seed weight, oil content and oil profile traits. It was shown that in the CDC Bethune x M96006 population seed coat color was associated with grain yield, oil content, palmitic, stearic, oleic, linoleic and linolenic fatty acid traits but that it independently assorted with grain yield, thousand seed weight, oil content, palmitic, stearic, oleic, linoleic and linolenic fatty acid traits. Additionally, a bimodal distribution for the palmitic fatty acid trait was observed showing that there is likely one major gene responsible for the percent palmitic fatty acid of the total seed oil content within this population.

It was shown that in the CDC Bethune x USDA-ARS Crystal population seed coat color was associated with grain yield, oil content, stearic, oleic, linoleic and linolenic fatty acid contents but that it independently assorted with grain yield, thousand seed weight, oil content, palmitic, stearic, oleic and linoleic fatty acid contents but not with linolenic fatty acid content showing that it will not be possible to select lines of brown seed coat color from this population to have as high linolenic fatty acid content as the yellow seeded lines with the highest linolenic fatty acid content from this population.

In the future another field test should be done to directly compare differences between lines with variegated seed coats and yellow seed coats which would allow plant breeders to choose the best seed coat color to use as a phenotypic marker for important seed quality traits.

Chapter 5 Linkage Analysis of the *b1* Seed Coat Color Locus

5.1 Abstract

By determining the location of the *b1* locus and identifying linked markers, it will allow breeders the use of seed coat color as a marker for flax varieties with unique combinations of quality traits and to perform confirmatory screening of breeding lines and of parental lines for genes of interest. The objective of this study was to identify linked markers with the *b1* locus and determine the map position of *b1*. To conduct these experiments bulked segregant analysis and screening with putatively linked markers was performed to identify linked markers with the *b1* locus.

Through bulked segregant analysis (BSA) 18 markers were identified to be putatively linked and were screened onto a recombinant inbred line (RIL) population which showed that Lu456 from linkage group (LG) 6 was not assorting independently with seed coat color (χ^2=3.90; P<0.05). Another marker from LG6 called LuM157 was the only other marker with a near significant result as supporting evidence that *b1* is on LG6.

These results show that more individuals should be used for future bulk segregant analysis (BSA) studies in order to accurately discriminate between markers that are and are not linked. The linkage with Lu456 suggests that the locus is present on LG6 but better marker coverage of markers polymorphic between the seed coat color parents on this linkage group would be necessary for finer mapping.

5.2 Introduction

Canada is the number one producer of flax globally, and Canadians produced over 700,000 tonnes as an annual average for the period from 2000-2010 (Food and Agriculture Organization of the United Nations, 2013). The export of Canadian Linseed has been valued at between 150-180 million dollars (SaskFlax 2013). Canadian production is almost entirely focused on Linseed-type flax. Linseed-type flax, which is grown for its seed oil, has until recently been divided into two classes: yellow and brown. Yellow seed coat color was used as a visual marker to distinguish a type of flax seed called Solin from traditional brown seed coat flax. Solin was required to contain less than five percent linolenic fatty acid; whereas, traditional brown seed flax contained greater than 50 percent (Mittapalli & Rowland, 2003). With increasing demand of high omega-3 flaxseed in the human and animal health markets and no market demand for Solin flaxseed, the yellow seed coat is no longer required to be a marker for the low linolenic fatty acid seed trait. There is however a demand to use seed coat color as a phenotypic marker for unique combinations of flax quality traits such as yellow seed coat color marking lines with high grain yield, seed oil content and linolenic fatty acid content.

Flax is a self-pollinating diploid plant with 15 sets of chromosomes (Lin & Paterson, 2011). It is known that flax has four independently inherited loci called *Y1*, *g*, *d*, and *b1* that are responsible for determining seed coat color. At the *Y1* locus there is a dominant yellow allele while the *g*, *d* and *b1* loci all have recessive yellow alleles (Mittapalli & Rowland, 2003). In all these cases the alternative phenotype is always brown. There has also been suggested to be a second recessive allele for the *b1* locus called b^{1vg}. This allele conferred a variegated or mottled seed coat color (Saeidi & Rowland, 1997).

There is not any published information on the genes responsible for seed coat color at each of the above loci, or even information about the flax biosynthetic pathway responsible for the production of seed coat pigment. There is however, information available in *Arabidopsis.* There are 26 genes (Table 2.1) known to play various roles in the *Arabidopsis* flavonoid biosynthesis pathway (Fig. 2.1) (Debeaujon *et al.*, 2003; Lepeniec *et al.*, 2006; Winkel-Shirley, 2001). This pathway results in the pigmentation of seeds and flowers (Winkel-Shirley, 2001). In seeds it is the accumulation and oxidation of condensed tannins that result in the brown color of mature seeds (Debeaujon *et al.*, 2003; Lepeniec *et*

al., 2006). When one of the 26 known genes is mutated, it disrupts the pathway and alternative seed coat colors result (Winkel-Shirley, 2001). These genes all represent potential orthologs of genes present in each of the four flax seed coat color loci.

To detect the location of the seed coat color loci in flax, linkage analysis using molecular markers that are anchored to a physical map can be performed. Until recently there have been very few molecular markers available in flax to perform linkage analysis (Cloutier *et al.*, 2012b). Recently, there have been tremendous advancements in flax genomic resources due to the International Total Utilization Flax Genome (Tufgen) project. The primary marker type that was developed through this project were SSRs, and since their development a consensus genetic and physical map has been created (Cloutier *et al.*, 2012a; Wang *et al.*, 2012). Using these SSR markers linkage analysis can be performed to identify SSR markers linked to specific genomic regions that control a gene of interest (Michelmore *et al.*, 1991). The integrated consensus genetic and physical map (Cloutier *et al.*, 2012b) can be used to determine the specific genomic region of markers that are identified to be linked to the *b1* seed coat color locus.

To perform BSA you must first have a recombinant inbred line (RIL), double haploid (DH), backcross (BC) or F_2 population that is segregating for a certain trait. This trait could either be monogenic or quantitatively inherited (Varshney *et al.*, 2009). The premise of this technique is to establish two pools of DNA that contain individuals that are genetically similar at the locus of interest within the pool, genetically dissimilar amongst the pools and all other unlinked regions are arbitrary (Michelmore *et al.*, 1991). In monogenic traits you choose individuals for the pools based on the two versions or extremes of the trait present; for example, brown or yellow seed coat color. But for quantitatively inherited traits you choose individuals that represent the tails of the traits normal distribution curve. The putatively linked markers that are identified can then be used to screen the entire populations in order to confirm linkage.

There are several advantages to identifying the position of the seed coat color loci as well as markers that are linked to it. This information can be used to identify the causal genes of seed coat color in flax; to breed for unique combinations of seed coat color with other quality traits through marker assisted selection; to screen early generations in a

breeding program for the desired seed coat color genes, and also to screen breeding lines to confirm that they contain the desired seed coat color genes.

The following research was conducted with the objective of identifying SSR markers that are linked with the *b1* locus and which linkage group contains this locus.

5.3 Materials and Methods

5.3.1 DNA Extraction

One individual plant of each line from the $F_{6:7}$ generation of the CDC Bethune x M96006 RIL population was grown in the growth chamber for leaf tissue collection. Collected tissue (~100mg) was frozen at -80ºC, ground using liquid nitrogen and total genomic DNA was extracted using a cetyltrimethyl ammonium bromide (CTAB) protocol (Doyle & Doyle, 1987). Samples were quantified using a NanoVue spectrophotometer (GE Life Sciences, Baie d'Urfe, Quebec, Canada) and diluted to a final concentration of 100ng/µl.

5.3.2 Simple Sequence Repeat (SSR) Molecular Markers

The primer pairs of 192 EST-derived SSR markers and six genomic SSR markers known to be polymorphic between the parental lines of this study were kindly provided by Drs. Sylvie Cloutier (Agriculture and Agri-Food Canada) and Manoj Kulkarni (Saskatoon National Research Council) respectively. Additionally, Dr. Cloutier also provided the integrated consensus genetic and physical map that contained the EST-derived markers. The polymerase chain reaction (PCR) was done in a 10µl final reaction volume containing 150ng of genomic DNA, 1µl 10x GenScript PCR buffer, 0.175µl of 5U/µl GenScript Taq DNA polymerase (GenScript USA Inc., Piscataway, New Jersey, USA), 0.4µl of 5mM dNTPs, 0.5µl of 1pM/µl forward primer with M13 tag, 0.5µl of 10pM/µl reverse primer with M13 tag and 0.18µl of 10mM M13 FAM, HEX or NED fluorescently labeled primer (Applied Biosystems, Life Technologies Inc., Burlington, Ontario, Canada)(Schuelke, 2000). The reaction was conducted in an Eppendorf Mastercycler Pro384 thermocycler (Eppendorf AG, Hamburg, Germany). The thermocycling program used consisted of denaturation at 94°C for 3min, followed by 4 cycles of 30 sec at 94°C, 50 sec at 56°C, 55 sec at 72°C, 30 sec at 94°C, 55 sec at 72°C, 30 sec at 94°C, 50 sec at 52°C, 55 sec at 72°C, 30 sec at 94°C, 50 sec at 50°C and 55 sec at 72°C, then 30 cycles of 30 sec at 94°C, 50 sec at 53°C, and 55 sec at 72°C, with a final cycle of 15 min at 72°C.

PCR products were electrophoresed on a 2% (w/v) agarose gel to confirm amplification and a 1 in 10 dilution was used upon pooling the PCR products of different

dyes. After the dilution and pooling was done 9µl of Hi-Di Formamide and 0.1µl of ROX labeled size standard were added (Applied Biosystems, Life Technologies Inc., Burlington, Ontario, Canada) and the samples were incubated in the thermocycler at 95°C for 10 min. The migration of the fragments were visualized with an ABI 3130xl using a standard run protocol with pop7 polymer in 33cm capillaries (Applied Biosystems, Life Technologies Inc., Burlington, Ontario, Canada) and analyzed using GeneMapper version 4 software (Applied Biosystems, Life Technologies Inc., Burlington, Ontario, Canada).

5.3.3 Bulked Segregant Analysis

Bulked segregant analysis (BSA) was performed by pooling DNA from eight individual RILs of each parental phenotype separately. These pools were then screened with all 198 SSR markers according to the above protocols. The results were scored for putative linkage of the markers to seed coat color.

5.3.4 Whole Population Screening

The whole population was screened with all 18 markers that were identified to be putatively linked through BSA. Sixteen of these markers were EST-derived simple sequence repeat (SSR) DNA markers provided by Cloutier et al. (2012) and two were genomic SSR DNA markers provided by Kulkarni et al. (unpublished). A simple chi-squared test was performed between each of the markers and seed coat color testing a 1:1 ratio to determine if there was dependence between the marker and the phenotype. Any marker with a significant relationship was determined to not be independently assorting and therefore linked.

5.4 Results and Discussion

5.4.1 Bulked Segregant Analysis (BSA)

Bulked segregant analysis was performed using the CDC Bethune x M96006 population to determine linkage of a molecular marker with a gene that conditions seed coat color at the *b1* locus. The CDC Bethune x USDA-ARS Crystal population was not used for confirmation of the results of this analysis because the likelihood of a second gene being involved in seed coat color determination would confound the results where the phenotype would not seem to segregate with the DNA marker during marker validation. Performing BSA on a population that is segregating for two-genes controlling the trait of interest will cause all DNA markers to appear unlinked if the two genes are not on the same linkage group. All 198 DNA markers available were screened onto the two DNA bulks, one for brown seed coat lines and one for variegated.

The results of the bulked segregant analysis (BSA) were not very clear. A result was considered positive when the intensity of the DNA marker's peak of the yellow seed coat color polymorph was at least two times larger than the brown marker version in the yellow DNA bulk and vice versa. There were positive results of 18 potentially linked DNA markers throughout all linkage groups except linkage groups 1, 3, 7 and 15. There cannot be loci for seed coat color on all of these LGs, because as seen previously it was confirmed that this population has only one gene controlling seed coat color determination. To clarify which LG the locus was on these DNA markers were screened onto the entire CDC Bethune x M96006 mapping population.

5.4.2 Whole Population Screening

All 18 of the DNA markers identified as likely candidates of being linked to the *b1* locus were screened onto the entire CDC Bethune x M96006 population. This was done to clarify which linkage group the locus was present on. A simple chi-squared test was performed using the SAS 9.3 statistical software package. This was done to determine if the marker-trait segregation ratio differed significantly from the 1:1 that would occur if the DNA marker was segregating randomly against seed coat color conditioned by the *b1* locus. Table 5.1 shows the chi-squared results. Markers with a significant relationship with seed coat color were determined to be linked.

These results suggest a linkage of Lu456 to seed coat color on linkage group six according to the Cloutier et al. (2012) consensus map. Further evidence that this locus is on LG6 is shown by the LuM157 marker of LG6 having a low p-value (0.0754) despite not being considered significant at the alpha level 0.05. More markers that are currently being developed should be used in the future, and this experiment should be repeated screening LG6 more intensely in order to conclusively determine the location of the *b1* locus. It is suggested that BSA works best with high density coverage (Ehrenreich *et al.*, 2009) where the best results are achieved when the marker is within 15 cM of the locus. Anything greater than 15 cM achieves results of less and less certainty (Michelmore *et al.*, 1991). This means that there should be polymorphic markers available approximately every 15 cM rather than the greater than 50 cM that is currently available between these seed coat color parents. Further marker development work should be done to develop more markers that are polymorphic between lines of different seed coat colors.

Additionally, there were several "false positives" identified in bulked segregant analysis that appeared to be linked with the *b1* locus. Bulked segregant analysis identifies linkage based on the absence of recombination between the specific seed coat color allele and the marker based on the allele-marker pairing from the parental line. It is possible that these bulks contain individuals that did not recombine despite the locus being far from the marker. It is also possible that individuals within the bulks had a double cross over event. Both of these scenarios make the bulks appear to only have parental type allele-marker pairings, and therefore, the marker appeared to be linked in the bulks but unlinked when actual whole population screening was performed. In the future these false positives can be reduced by increasing the number of individuals in the bulks, which would prevent these chance events from confounding the results of the bulked segregant analysis.

It is important to provide plant breeders with perfectly linked markers for the *b1* locus so that marker assisted selection can be used to combine unique combinations of flax quality traits with seed coat color. This way seed coat color can be used as a phenotypic marker for specialty flax varieties. It is also important for researchers to know the position of the *b1* locus so that characterization of the gene involved can be performed. This information can also be useful for plant breeders to screen breeding lines to confirm genetic backgrounds, or to screen early generations for specific seed coat color genes.

Table 5.1. The contingency table and chi-squared tests to test a 1:1 relationship between seed coat color and 16 EST-derived SSR (expressed sequence tag derived simple sequence repeat) markers developed by Cloutier et al. (2012b) and two genomic SSR markers developed by Manoj Kulkarni (unpublished) for the CDC Bethune x M96006 recombinant inbred line population. If the marker ratios for the seed coat colors differed significantly from 1:1 then it was considered linked. See table A.1. in appendix II for marker information and table A.2. for marker scoring data.

Marker	LG	Brown Seeds Allele A vs. B	Variegated Seeds Allele A vs. B	χ^2	df	P-value
Lu628	2	26:22	21:24	0.523	1	0.470
Lu56	4	24:21	28:16	0.972	1	0.324
Lu2183	4	23:22	19:21	0.111	1	0.740
Lu658	5	29:14	31:8	1.51	1	0.219
Lu933	5	28:15	27:17	0.132	1	0.717
Lu2628	5	20:24	15:29	1.19	1	0.276
Lu2767	5	26:22	27:19	0.196	1	0.658
Lu456	6	26:22	16:31	3.90	1	0.0483*
LuM157	6	13:31	22:24	3.16	1	0.0754
LuM280	6	25:20	17:21	0.965	1	0.326
Lu566	8	28:21	23:23	0.487	1	0.485
Lu672	8	18:26	16:25	0.0314	1	0.859
Lu805	10	29:20	23:23	0.808	1	0.369
Lu458	11	28:21	24:23	0.357	1	0.550
Lu850	12	24:25	18:29	1.11	1	0.292
Lu2625	12	19:21	21:22	0.932	1	0.334
Lu526	13	15:25	22:21	1.57	1	0.211
Lu291	14	24:15	14:22	0.718	1	0.397

*P<0.05

5.5 Conclusions and Future Research

Seed coat color conditioned by the *b1* locus remained stable from one generation to the next and there was not transgressive segregation observed within the study populations. The allelism test confirmed that b^1 and b^{1vg} are two alleles of the same locus, as shown by all F_2 seeds having a variegated seed coat. BSA of 198 available markers identified 18 markers present on 13 of the 15 flax linkage groups. But since there is only one gene responsible for seed coat color determination in the CDC Bethune x M96006 population, it was only possible for the locus to be on one LG. This study showed that seed coat color segregated independently of most markers except for Lu456 on LG6, which appears to be linked with the seed coat color trait and therefore the *b1* locus (χ^2=3.90; P<0.05). More markers that are currently being developed should be used in the future, and this experiment should be repeated screening LG6 more intensely in order to conclusively determine the location of the *b1* locus. Additionally, the data collected within this project from the relationship studies between seed coat color and the other flax traits examined can be used in the future in conjunction with future marker data to identify QTLs for each of those traits.

By identifying perfectly linked markers and the position of the *b1* locus in the genome, plant breeders will be provided with powerful tools for marker-assisted selection of seed coat color so that lines can be culled at the seedling stage. Researchers will also be able to use this information for the characterization of the gene involved at the *b1* locus. This information can lead to the combination of seed coat color with other useful combinations of flax quality traits, the screening of breeding lines to confirm their genetic background, as well as the screening of earlier generation lines for seed coat color genes of interest.

Chapter 6 General Discussion

6.1 Discussion

This study showed that in the CDC Bethune x M96006 population and the CDC Bethune x USDA-ARS Crystal population, both segregating for the *b1* seed coat color conditioning locus, seed coat color remained consistent from generation to generation. This showed that there was very little influence on seed coat color by the environment and thus allowed a discrete scale to be used for phenotypic measurement and data analysis. Additionally, it was found that the *b1* locus had a pleiotropic effect on seed coat color, flower color and shape where all seeds with yellow seed coat color had white star shaped flowers and all seeds with a brown or variegated seed coat color had blue bowl shaped flowers. This suggests that the *b1* locus harbors genes that act on the flavonoid biosynthesis pathway prior to where the seed coat color and flower characteristic branches diverge, or that it codes for a transcription factor that alters expression of genes on both branches of this pathway.

This study looked for the first time at the relationship of variegated seed coat color, conditioned by the recessive b^{1vg} allele of the *b1* locus, with grain yield, thousand seed weight, seed oil content and palmitic, stearic, oleic, linoleic and linolenic fatty acid contents using the CDC Bethune x M96006 population. Associations between seed coat color and seed oil content, stearic and oleic fatty acid contents were detected where lines with variegated seeds had higher seed oil content but lower levels of stearic and oleic fatty acids then lines with brown seed coat color. However, while there were associations between seed coat color with these traits there were lines in this population with variegated seed coat color that had low seed oil content, and high stearic and oleic fatty acid content and vice versa where lines with brown seed coat color had high seed oil content and low stearic and oleic fatty acid content. This shows that all traits independently assorted and therefore, it is possible to develop lines from this population to have different combinations of these traits with seed coat color. This suggests that variegated seed coat color would be a good candidate for Plant Breeders to use as a phenotypic marker. Within this population a bimodal frequency distribution was observed for palmitic fatty acid content showing that there is one major gene responsible for production of palmitic fatty acid.

Additionally, the relationship of yellow seed coat color, conditioned by the recessive b^1 allele of the *b1* locus, with grain yield, thousand seed weight, seed oil content, palmitic, stearic, oleic, linoleic and linolenic fatty acid contents was examined using the CDC Bethune x USDA-ARS Crystal population. Putative associations were detected between seed coat color and grain yield, seed oil content, stearic, oleic, linoleic and linolenic fatty acid contents where lines with yellow seed coat color had lower grain yield, stearic, oleic and linoleic fatty acid contents and higher seed oil content and linolenic fatty acid content. However, there proved to be independent assortment observed between seed coat color and grain yield, seed oil content, stearic, oleic and linoleic fatty acid contents but there was not between seed coat color and linolenic fatty acid content where there were no lines with brown seed coat color that had higher linolenic fatty acid content than the yellow lines. This suggests that the *b1* locus may be linked or otherwise associated with a *FAD3* gene as the *FAD3a* and *FAD3b* genes are known to be responsible for linolenic fatty acid production in flax seed by desaturating linoleic fatty acid (Vrinten *et al.*, 2003). A third FAD3 paralog, *FAD3c*, has also been identified (Stefunova & Lancikova, 2013), but it was determined not to have a role in linolenic fatty acid accumulation within the seed of flax (Banik *et al.*, 2011).

The two previously mentioned CDC Bethune x M96006 and CDC Bethune x USDA-ARS Crystal RIL populations were used to determine the location of the *b1* locus. Population exploration was performed and determined that the CDC Bethune x USDA-ARS Crystal population segregated at a 3:1 ratio of brown to yellow seed coat colored lines in the F_6 generation. This might suggest that there are likely two genes responsible for yellow seed coat color determination in the USDA-ARS Crystal parent and both have to be homozygous recessive for the yellow phenotype to be expressed. It might also suggest that a selection event against yellow seeded lines occurred, or that the CDC Bethune parental line was heterozygous at the *b1* locus. This population can be screened with a perfectly linked marker to the *b1* locus in order to determine which of these three scenarios are responsible for the 3:1 ratio of seed coat colors in this population.

An allelism test of M96006 x USDA-ARS Crystal and the reciprocal was also performed to confirm the allelic nature of the b^1 and b^{1vg} alleles. There was no segregation of seed coat colors in the F_2 generation (Note: seed coat of F_2 seed is F_1 tissue) showing

there was no complementation and that b^1 and b^{1vg} are in fact two different alleles of the same locus. In this generation all seeds were variegated showing that the b^{1vg} allele is dominant to the b^1 allele. In the F_3 generation, there was a 7:1 variegated to yellow seed coat color ratio which did not fit the expected 3:1 ratio of simple dominance, which is likely do to semi-lethality associated with the *b1* locus (Mittapalli & Rowland, 2003; Shaw *et al.*, 1931). The same pleiotropic effect previously mentioned of the *b1* locus on seed coat color, flower color and flower shape was observed in this experiment where all plants with white star shaped flowers yielded seeds with yellow seeds coats and all plants with blue bowl shaped flowers yielded variegated seeds with variegated seed coats (Barnes *et al.*, 1960; Culbertson & Kommendahl, 1956).

SSR markers (18) were chosen using bulked segregant analysis on the CDC Bethune x M96006 population and chi-squared tests testing independence of each marker with seed coat color after being screened on the entire population found that the Lu456 marker of LG6 may be linked with the *b1* locus ($p<0.05$) (Table 4.4). In order to confirm this result LG6 should be screened with more markers. However, as previously mentioned there is evidence that a *FAD3* gene or one of its modifying genes is linked with the *b1* locus. The position of *FAD3a* and *FAD3b*, has been shown since the data was collected for this study to be on LG7 and LG12 respectively (You *et al.*, 2014). It is possible that modification genes are responsible for the increased linolenic fatty acid content of the yellow seeded lines in the CDC Bethune x USDA-ARS Crystal population.

6.2 Conclusions

Through this study it was determined that it is possible to develop varieties from these populations with any combination of grain yield, thousand seed weight, seed oil content, stearic, oleic, linoleic or linolenic fatty acid contents with variegated seed coat color from the CDC Bethune x M96006 population. Additionally, seed coat color did not independently assort with linolenic fatty acid in the CDC Bethune x USDA-ARS Crystal population segregating for the b^1 recessive allele of the *b1* locus. This suggests that the *b1* locus could be linked or otherwise associated with a *FAD3* gene. Additionally, it was shown that the *b1* locus, which conditions seed coat color, is pleiotropic to flower color and flower shape, so when breeding from these populations all lines with variegated seed coat colors will have blue bell shaped flowers and those with yellow seed coats will have white star shaped flowers.

Furthermore, this study confirmed the allelic nature of b^1 and b^{1vg}. A marker linked with this locus was identified which suggests that the *b1* locus is on LG6. Further marker development and screening should be performed to confirm this result.

It is important to gain information about the relationship of seed coat color to other important flax quality and agronomic traits as well as develop and identify perfectly linked markers to seed coat color loci. By doing this breeders can use marker assisted selection to breed varieties of flax that would use seed coat color as a phenotypic marker for different combinations of grain yield, thousand seed weight, seed oil content, or oil profile traits.

6.3 Future Research Directions

There are several directions in which future research stemming from this project could take. First and foremost work should be done to develop a better coverage of markers that are polymorphic between the seed coat color parents. Next generation sequencing will be a useful tool for the discovery of new DNA markers. Once more markers are available screening should be performed to identify the *b1* locus such that it can be more accurately mapped. Due to the presence of an apparent linkage between the *b1* locus and seed coat color it is suggested to focus screening onto LG6. Additionally, once a perfectly linked marker is identified with the *b1* locus the CDC Bethune x USDA-Crystal population can be screened to determine whether or not the population skewing observed in the F6:7 was due to selection event, heterozygous parental lines or a second gene in the USDA-ARS Crystal parent. If the marker presents itself in the same 3:1 of brown to yellow ratio that the phenotype does then the skewing was caused by selection, but if the marker appears in a 1:1 ratio it would provide further evidence supporting the two-gene theory. Furthermore, once a more significantly linked marker is found, the marker sequence can be can be searched using a basic local alignment search tool (BLAST) within the reference genome to determine candidate genes within an interval around the linked marker.

Screening with a larger amount of markers that span the entire genome would also allow for QTL detection of the quantitatively inherited traits that were observed through this study. Marker data can be combined with the field observations of each of the traits to detect loci involved in the expression of them. To perform QTL analysis it would also be useful for the field observations to be repeated in more environments over additional years. These replications will allow for more accurate QTL detection. Additional replication would also strengthen the results of the seed coat color-trait relationship studies because currently the environmental component of the variance estimates is likely artificially low due to measurements being taken from one field experiment in two growing seasons. Additional field experiments should be performed comparing the performance of the lines with variegated seed coats to the lines with yellow seed coats directly through the establishment of a variegated x yellow RIL population. This would allow plant breeders to choose the seed coat color with the best performance within the traits of interest to be used as a phenotypic marker.

LITERATURE CITED

Afzal Naz, R. M. (1976). Inheritance of seed color in linseed (*Linum usitatissimum* L.) in Pakistan. *Journal of Agricultural Research, 14*, 13-21.

Allaby, R. G., Peterson, G. W., Merriweather, A., & Fu, Y. B. (2005). Evidence of the domestication history of flax (*Linum usitatissimum*) from genetic diversity of the *sad2* locus. *Theoretical and Applied Genetics, 112*, 58-65.

Arny, A. C. (1944). Registration of Improved Flax Varieties, II. *Journal of the American Society of Agronomy, 36*, 454-457.

Banik, M., Duguid, S., & Cloutier, S. (2011). Transcript profiling and gene characterization of three fatty acid desaturase genes in high, moderate and low linolenic acid genotypes of flax (*Linum usitatissimum* L.). *Genome, 54*, 471-483.

Barnes, D. K., Culbertson, J. O., & Lambert, J. W. (1960). Inheritance of seed and flower color in flax. *Agronomy Journal, 52*, 456-459.

Beard, B. H., & Comstock, V. E. (1965). Flax genetics and gene symbolism. *Crop Science, 5*, 151-155.

Bouman, F. (1975). Integument initiation and testa development in some *Cruciferae*. *Botanical Journal of the Linnean Society, 70*, 213-229.

Buer, C. S., Imin, N., & Djordjevic, M. A. (2010). Flavonoids: New Roles for Old Molecules. *Journal of Integrative Plant Biology, 52*(1), 98-111.

Canadian Grain Commission. (2012). Official Grain Grading Guide. Retrieved January 30, 2013, from http://www.grainscanada.gc.ca/oggg-gocg/ggg-gcg-eng.htm

Cloutier, S., Miranda, E., Ward, K., Radovanovic, N., Reimer, E., Walichnowski, A., Datla, R., Rowland, G., Duguid, S., & Ragupathy, R. (2012a). Simple sequence repeat marker development from bacterial artificial chromosome end sequences and expressed sequence tags of flax (Linum usitatissimum L.). *Theoretical and Applied Genetics, 125*(4), 685-694.

Cloutier, S., Niu, Z., Datla, R., & Duguid, S. (2009). Development and analysis of EST-SSRs for flax (*Linum usitatissimum* L.). *Theoretical and Applied Genetics, 119*, 53-63.

Cloutier, S., Ragupathy, R., Miranda, E., Radovanovic, N., Reimer, E., Walichnowski, A., Ward, K., Rowland, G., Duguid, S., & Banik, M. (2012b). Integrated consensus genetic and physical maps of flax (Linum usitatissimum L.). *Theoretical and Applied Genetics, 125*(8), 1783-1795.

Cloutier, S., Ragupathy, R., Zhixia, N., & Duguid, S. (2011). SSR-based linkage map of flax (*Linum usitatissimum* L.) and mapping of QTLs underlying fatty acid composition traits. *Molecular Breeding, 28*(4), 437-451.

Culbertson, J. O., & Kommendahl, T. (1956). The effect of seed coat color upon agronomic and chemical characters and seed injury in flax. *Agronomy Journal, 48*, 25-28.

Cunnane, S. C., Ganguli, S., Menard, C., Liede, A. C., Hamadeh, M. J., Chen, Z. Y., Wolever, T. M., & Jenkins, D. J. (1993). High alpha-linolenic acid flaxseed (*Linum usitatissimum*): some nutritional properties in humans. *British Journal of Nutrition, 69*(2), 443-453.

Debeaujon, I., Leon-Kloosterziel, K. M., & Koorneef, M. (2000). Influence of the testa on seed dormancy, germination, and longevity in *Arabidopsis*. *Plant Physiology, 122,* 403-413.

Debeaujon, I., Nesi, N., Perez, P., Devic, M., Grandjean, O., Caboche, M., & Lepiniec, L. (2003). Proanthocyanidin-accumulating cells in *Arabidopsis* testa: regulation of differentiation and role in seed development. *Plant Cell, 15,* 2514-2531.

Diederichsen, A., & Fu, Y. B. (2006). Phenotypic and molecular (RAPD) differentiation of four infraspecific groups of cultivated flax (*Linum usitatissimum* L. subsp. *usitatissimum*). *Genetic Resources and Crop Evolution, 53*, 77-90.

Diederichsen, A., & Hammer, K. (1995). Variation of cultivated flax (*Linum usitatissimum* L. subsp. *usitatissimum*) and its wild progenitor pale flax (subsp. *angustifolium* (Huds.) Thell.). *Genetic Resources and Crop Evolution, 42*(3), 263-272.

Diederichsen, A., & Richards, K. W. (2003). Cultivated flax and the genus *Linum* L. - taxonomy and germplasm conservation. In A. Muir & N. Wescott (Eds.), *Flax, the Genus Linum* (pp. 22-54). London, UK: Taylor & Francis.

Doyle, J. J., & Doyle, J. L. (1987). DNA isolation from small amounts of plant tissue. *Phytochemical Bulletin, 19*, 11-15.

Ehrenreich, I. M., Gerke, J. P., & Kruglyak, L. (2009). Genetic dissection of complex traits in yeast: insights from studies of gene expression and other phenotypes in the byxrm cross. *Cold Spring Harbor Symposia on Quantitative Biology, 74*, 145-153.

Fehr, W. R. (1993). Chapter 3 Genetic Principles. *Principles of Cultivar Development: Theory and Technique* (Vol. 3). New York, USA: Macmillan USA.

Food and Agriculture Organization of the United Nations. (2013). FAOSTAT. Retrieved January 30, 2013, from http://faostat3.fao.org/home/index.html#VISUALIZE

Helbaek, H. (1959). Domestication of food plants in the Old World. *Science, 130*, 365-372.

Kenaschuk, E. O., & Hoes, J. A. (1986). McGregor Flax. *Canadian Journal of Plant Science, 66*(1), 175-176.

Kuehni, R. G. (2002). The early development of the Munsell system. *Color research and application, 27*(1), 20-27.

Lepeniec, L., Debeaujon, I., Routaboul, J. M., Baudry, A., Pourcel, L., Nesi, N., & Caboche, M. (2006). Genetics and Biochemistry of Seed Flavonoids. *Annual Review of Plant Biology, 57*, 405-430.

Lin, L., & Paterson, A. H. (2011). Size variation in homologous segments across divergent plant genomes. *Mobile Genetic Elements, 1*, 92-96.

Linum.ca. (2011). Total Utilization Flax Genomics - Tufgen. Retrieved January 30, 2013, from http://carpel.ccis.ualberta.ca/TUFGEN/index.html#a

Ma, F., Cholewa, E., Mohamed, T., Peterson, C. A., & Gijzen, M. (2004). Cracks in the palisade cuticle of soybean seed coats correlate with their permeability to water. *Annals of Botany, 94*, 213-228.

Michelmore, R. W., Paran, I., & Kesseli, R. V. (1991). Identification of markers linked to disease-resistance genes by bulked segregant analysis: a rapid method to detect

markers in specific genomic regions by using segregating populations. *Proceedings of the National Academy of Sciences, 88*, 9828-9832.

Mittapalli, O., & Rowland, G. (2003). Inheritance of Seed Color in Flax. *Crop Science, 43*(6), 1945-1951.

Moise, J. A., Han, S., Gudynaite-Savitch, L., Johnson, D. A., & Miki, B. L. A. (2005). Seed coats: structure, development, composition, and biotechnology. *In Vitro Cellular Developments Biology - Plant, 41*, 620-644.

Powell, W., Morganate, M., Andre, C., Hanafey, M., Vogel, J., Tingey, S., & Rafalski, A. (1996). The comparison of RFLP, RAPD, AFLP, and SSR (microsatellite) markers for germplasm analysis. *Molecular Breeding*(2), 225-238.

Reid, T., Navabi, A., Cahill, J., Salmon, D., & Spaner, D. (2009). A genetic analysis of wheat competitive ability in spring wheat. *Canadian Journal of Plant Science, 89*(4), 591-599.

Ross, J. A., & Kasum, C. M. (2002). Dietary flavonoids: bioavailability, metabolic effects, and safety. *Anuual Review of Nutrition, 22*, 19-34.

Rowland, G. G., Hormis, Y. A., & Rashid, K. Y. (2001). CDC Bethune flax. *Canadian Journal of Plant Science, 82*, 101-102.

Rowland, G. G., Hormis, Y. A., & Rashid, K. Y. (2002). CDC Bethune flax. *Canadian Journal of Plant Science, 82*(1), 101-102.

Ryan, C. D., & Smyth, S. J. (2012). Economic implications of low level presence in a zero-tolerance European import market: the case of Canadian Triffid flax. *AgBio Forum, 15*(1), 21-30.

Saeidi, G., & Rowland, G. G. (1997). The inheritance of variegated seed color and palmitic acid in flax. *Journal of Heredity, 88*(6), 466-468.

Saeidi, G., & Rowland, G. G. (1999a). The effect of temperature, seed colour and linolenic acid concentration on germination and seed vigour in flax. *Canadian Journal of Plant Science, 79*(3), 315-319.

Saeidi, G., & Rowland, G. G. (1999b). Seed colour and linolenic acid effects on agronomic traits in flax. *Canadian Journal of Plant Science, 79*(4), 521-526.

Saskatchewan Flax Development Comission. (n.d). SaskFlax. Retrieved January 30, 2013, from http://www.saskflax.com/index.html

Schuelke, M. (2000). An economic method for the fluorescent labeling of PCR fragments. *Nature Biotechnology, 18*(2), 233-234.

Shaw, F. J., Khan, A. R., & Alam, M. (1931). Studies in Indian Oilseeds. V. The inheritance of characters in Indian linseed. *Indian Journal of Agricultural Science, 1*, 1-57.

Simbaya, J., Slominsky, B. A., Rakow, G., Campbell, L. D., Downey, R. K., & Bell, J. M. (1995). Quality characteristics of yellow-seeded *Brassica* seed meals: Protein, carbohydrates, and dietary fiber components. *Journal of Agricultural and Food Chemistry, 43*, 2062-2066.

Stefunova, V., & Lancikova, V. (2013). Screening of Common Flax *FAD* Genes by PCR. *Journal of Microbiology, Biotechnology and Food Sciences, 2013*(2 (Special issue 1)), 1665-1676.

Tammes, T. (1922). Genetic analysis, schemes of co-operation and multiple allelemorphs of *Linum usitatissimum. Journal of Genetics, 12*, 19-46.

Taylor, L. P., & Grotewold, E. (2005). Flavonoids as developmental regulators. *Current Opinion in Plant Biology, 8,* 317-323.

Thambugala, D., & Cloutier, S. (2014). Fatty acid composition and desaturase gene expression in flax (*Linum usitatissimum* L.). *Journal of Applied Genetics, 55,* 423-432.

Thambugala, D., Duguid, S., Loewen, E., Rowland, G., Booker, H., You, F. M., & Cloutier, S. (2013). Genetic variation of six desaturase genes in flax and their impact on fatty acid composition. *Theoretical and Applied Genetics, 126,* 2627-2641.

van Zeist, W., & Bakker-Heers, J. A. H. (1975). Evidence for linseed cultivation before 6000 B.C. *Journal of Archaeological Science, 2*, 215-219.

Varshney, R. K., Hoisington, D. A., Nayak, S. N., & Graner, A. (2009). Molecular plant breeding: Methodology and achievements. *Methods in Molecular Biology, 513,* 283-304.

Vaughan, J. G., & Whitehouse, J. M. (1971). Seed structure and the taxonomy of the *Cruciferae*. *Botanical Journal of the Linnean Society, 64*, 383-409.

Vrinten, P., Hu, Z., Munchinsky, M. A., Rowland, G., & Qui, X. (2003). Two *FAD3* desaturase genes control the level of linolenic acid in flax seed. *Plant Physiology, 139*, 79-87.

Wang, Z., Hobson, N., Galindo, L., Zhu, S., Shi, D., McDill, J., Yang, L., Hawkins, S., Neutelings, G., Datla, R., Lambert, G., Galbraith, D. W., Grassa, C. J., Geraldes, A., Cronk, Q. C., Cullis, C., Dash, P. K., Kumar, P. A., Cloutier, S., Sharpe, A. G., Wong, G. K. S., Wang, J., & Deyholos, M. K. (2012). The genome of flax

(Linum usitatissimum) assembled de novo from short shotgun sequence reads. *Plant Journal, 72*(3), 461-473.

Weber, H., Borisjuk, L., & Wobus, U. (2005). Molecular physiology of legume seed development. *Annual Review of Plant Biology, 56*, 253-279.

Wiesnerova, D., & Wiesner, I. (2008). Computer image analysis of seed chape and seed color for flax cultivar description. *Computers and Electronics in Agriculture, 61*, 126-135.

Winkel-Shirley, B. (2001). Flavonoid Biosynthesis. A Colorful Model for Genetics, Biochemistry, Cell Biology, and Biotechnology. *Plant Physiology, 126*, 485-493.

Yates, F. (1934). Contingency tables involving small numbers and the χ^2 test. *Supplement to the Journal of the Royal Statistical Society, 1*(2 (1934)), 217-235.

You, F. M., Li, P., Kumar, S., Rajupathy, R., Li, Z., Fu, Y., & Cloutier, S. (2014). Genome-wide identification and characterization of the gene families controlling fatty acid biosynthesis in flax (*Linum usitatissimum* L.). *Proteomics & Bioinformatics, 7*(10), 310-326.

APPENDIX I – Laboratory Protocols

Plant DNA Extraction

The plants were grown in the growth chamber under a controlled environment. The tissues of the top two centimeters of a tiller were snipped off the plant and placed into labeled test tubes. The tissue was then placed in a -80^{o}C freezer for the tissues to freeze. The tissue samples were then ground with new sterile pestles for each sample and kept frozen using liquid nitrogen. The method that was used for genomic DNA extraction was the CTAB method (Doyle & Doyle 1987). After tissues were ground 500µl of CTAB DNA extraction buffer was added and the tubes were incubated at 57^{o}C for a half hour. After this half hour 1µl of RNase A was added and the tubes were incubated for another hour, mixing once after thirty minutes. Once the tubes were finished incubating 500µl of chloroform was added and the tubes were mixed by gently shaking. The tubes were then centrifuged for 7 min at 16 rcf. The aqueous phase was then pipetted into a fresh centrifuge tube. In order to obtain as much DNA as possible another 500µl of CTAB buffer was added to the tube with the used chloroform and the interphase. These tubes were shaken, centrifuged a second time and the aqueous phase was pipetted off into a second fresh centrifuge tube. To each of these tubes 32µl of ice cold ammonium acetate and 233µl of ice cold isopropanol were added. These tubes were then inverted 40 times and stored in the freezer overnight. In the morning the tubes were centrifuged for 3 min at 16 rcf and the supernatant was discarded. To wash the pellet 700µl of 70% ethanol was added, the tubes were inverted 10 times, centrifuged for 1 min at 16 rcf and then the supernatant was discarded. The pellets were washed again but this time with 700µl of 90% ethanol, the tubes were then inverted 10 times centrifuged for 5 min at 16 rcf and the supernatant was discarded. The tubes were then put in a 40^{o}C incubator until all ethanol had been evaporated. Once dry the pellets were resuspended in 50µl of TE and DNA was quantified using a NanoVue spectrophotometer calibrated using water and TE buffer.

APPENDIX II – Raw Marker Data

Table A.1. The raw marker data for the screening of 18 markers identified through bulked segregant analysis (BSA) onto the CDC Bethune x M96006 population. Markers were scored for A) CDC Bethune parental type or B) M96006 parental type and left blank if heterozygous or indeterminate.

		LuM157	LuM280	Lu2625	Lu805	Lu850	Lu458	Lu526	Lu2183	Lu566	Lu291	Lu628	Lu456	Lu672	Lu658	Lu2767	Lu2628	Lu56	Lu933	Color
GENOTYPE	**101313**	A	A	A	A	A	A	-	B	A	-	A	B	B	A	A	A	A	-	A
	101314	A	B	-	A	B	A	-	B	B	-	B	B	A	A	B	B	B	-	A
	101315	B	A	A	A	B	A	A	B	A	B	A	A	B	A	A	A	A	-	B
	101316	A	B	A	A	A	A	-	B	B	-	A	B	A	-	A	A	A	A	B
	101317	A	A	-	A	B	A	-	B	B	-	-	B	B	A	A	A	A	-	A
	101318	B	B	A	B	A	A	-	A	A	A	B	A	B	B	A	A	A	B	B
	101319	A	B	A	A	B	A	B	B	A	A	B	B	A	B	A	A	B	A	A
	101320	B	A	B	A	A	A	B	B	A	-	B	A	-	A	A	B	A	A	B
	101321	-	B	A	A	B	A	B	B	A	A	A	-	-	B	B	A	A	B	A
	101322	-	A	B	A	B	B	A	B	A	A	B	B	B	B	B	B	A	B	A
	101323	A	B	A	A	B	B	-	B	B	-	A	B	-	-	B	B	B	B	B
	101324	A	B	B	B	B	A	A	A	A	B	A	B	B	-	B	B	-	A	B
	101325	B	A	A	B	B	B	B	-	B	A	A	A	A	A	A	A	B	-	B
	101326	B	A	A	B	A	A	A	-	B	B	B	B	A	A	A	A	A	A	B

101328	B	B	B	B	B	A	A	A	A	A	B	A	B	A	B	B	B	A	B
101329	B	-	A	B	A	A	A	A	B	B	B	A	B	A	A	A	B	A	A
101330	B	A	A	B	B	A	A	A	B	A	A	B	B	B	A	B	A	A	A
101331	B	B	A	A	B	A	B	B	B	B	A	B	B	A	A	A	A	A	B
101332	A	-	A	B	A	A	B	A	B	A	A	A	B	A	B	B	A	A	A
101333	A	A	A	B	B	A	B	A	A	A	A	B	A	A	B	A	A	A	A
101334	A	B	B	B	B	A	A	A	A	A	A	B	B	B	A	B	A	A	A
101335	A	A	A	A	B	A	A	B	B	A	B	B	B	A	A	A	B	A	B
101336	A	B	B	A	B	B	A	B	A	B	A	B	B	A	A	A	A	B	B
101337	A	B	B	B	A	B	A	A	A	B	B	A	-	-	B	B	-	B	B
101338	-	A	-	B	B	A	A	A	A	-	B	A	B	A	B	A	B	A	A
101339	B	A	-	B	A	A	B	A	A	A	A	B	A	A	B	B	A	A	B
101341	B	B	B	B	A	A	A	A	A	-	B	B	A	A	A	B	B	A	B
101342	A	A	B	A	B	A	B	B	A	B	B	B	A	A	B	B	A	A	B
101343	B	B	B	A	B	A	B	B	B	B	B	B	B	B	B	B	A	B	B
101344	A	B	-	A	B	A	B	-	B	B	B	B	B	A	A	A	B	A	B
101345	B	B	A	A	B	B	A	B	B	B	A	B	-	A	A	B	A	A	B
101346	A	A	B	A	B	B	B	B	B	B	B	B	B	B	B	-	A	-	A
101347	A	-	B	A	B	B	A	B	A	B	B	B	B	A	B	B	B	B	B
101348	A	A	A	B	A	A	B	A	A	B	B	A	B	A	B	B	B	A	A
101349	A	B	B	A	B	B	A	B	A	B	-	B	B	A	B	B	B	A	B
101350	A	A	A	B	B	B	A	A	A	A	A	B	A	B	B	A	A	B	A

101351	B	B	A	B	B	A	B	A	A	-	A	A	A	A	A	A	A	A	A
101352	B	A	-	B	A	A	B	A	A	A	B	B	-	A	B	B	A	A	A
101353	A	A	A	A	B	B	B	B	A	B	A	A	A	-	A	A	A	B	A
101355	A	B	A	A	A	B	B	B	B	B	B	B	B	-	A	A	A	B	A
101356	A	A	A	A	A	B	A	B	B	B	B	A	B	A	A	A	A	A	B
101357	BA	B	B	B	A	B	B	A	B	B	B	A	-	B	B	B	A	A	A
101358	B	A	B	A	B	A	-	B	A	B	A	B	A	A	A	B	-	-	A
101359	B	-	B	A	A	B	B	B	B	B	A	A	B	A	A	-	A	A	A
101360	A	A	A	B	B	A	-	A	B	B	B	B	B	A	A	A	A	A	A
101361	A	B	B	A	A	A	-	B	B	A	B	A	B	A	B	-	A	A	A
101362	A	A	A	A	B	B	A	B	A	A	B	B	B	A	A	B	A	A	B
101363	A	A	B	B	A	B	B	A	A	B	B	A	B	A	A	B	A	A	B
101364	A	A	B	B	B	A	B	A	B	A	B	B	A	B	A	B	A	B	A
101365	B	B	B	B	B	A	B	A	A	A	A	A	B	B	A	A	B	B	A
101366	B	A	A	B	B	A	-	A	A	A	B	A	A	A	B	B	B	B	A
101367	A	B	B	A	A	A	B	B	A	A	B	A	B	B	B	B	B	B	B
101368	A	B	B	A	A	A	A	B	A	-	A	A	B	B	A	B	B	B	A
101369	B	A	B	A	B	A	A	B	A	A	A	B	B	B	B	B	B	B	A
101370	B	A	B	A	A	B	B	B	A	B	A	A	B	A	-	A	A	B	A
101371	A	A	B	B	A	A	B	A	B	-	B	A	A	A	B	B	A	A	B
101372	A	-	B	B	B	B	B	A	B	B	B	B	B	A	B	B	B	B	B
101373	A	-	-	B	B	B	B	A	A	A	A	B	B	B	A	A	B	B	A

101374	A	-	A	B	B	B	B	A	A	B	A	B	B	A	A	A	B	B	B
101375	B	-	B	A	B	B	B	B	A	B	A	B	B	A	A	B	A	B	B
101377	A	-	-	B	A	B	A	-	A	-	B	A	A	B	A	B	A	B	B
101379	B	-	-	B	A	B	-	A	A	-	B	A	B	A	A	B	A	A	A
101380	B	-	A	B	A	B	A	A	B	A	A	A	B	A	B	A	A	A	A
101381	A	-	B	B	A	B	B	A	A	A	B	A	A	A	B	B	B	B	A
101382	-	B	A	B	B	A	A	A	-	A	A	B	A	A	A	B	A	A	B
101385	A	B	A	B	A	A	B	A	B	A	A	A	A	B	A	B	A	A	B
101386	A	B	A	A	B	B	A	B	B	-	-	B	-	A	A	-	A	B	B
101387	B	A	B	B	B	A	B	A	B	A	A	B	B	A	B	B	B	A	B
101388	A	B	B	B	A	B	A	A	B	B	B	A	A	A	-	B	A	A	B
101389	A	B	-	A	B	A	B	B	A	A	A	B	A	A	A	A	B	A	A
101390	A	B	A	A	A	A	B	B	B	B	A	B	B	A	A	A	A	A	B
101391	A	B	B	A	B	B	A	B	A	A	A	A	A	-	A	A	B	A	A
101392	A	B	B	A	A	B	A	A	B	B	B	A	B	-	B	B	B	A	A
101393	B	A	-	A	A	B	-	A	A	-	A	A	B	B	A	B	B	B	B
101395	B	B	B	A	B	A	B	-	B	A	A	B	B	A	B	B	-	A	A
101396	B	A	A	A	A	B	A	B	A	-	B	A	B	A	A	B	-	A	A
101397	A	A	A	A	B	B	B	-	B	A	B	B	A	A	A	B	B	A	A
101398	B	A	A	A	A	B	A	B	A	B	B	A	A	B	A	A	A	B	B
101399	A	A	A	B	A	B	B	A	B	B	A	A	A	A	A	A	B	-	A
101400	A	B	B	A	A	A	A	B	A	A	B	A	B	-	B	B	A	-	B

101402	B	B	B	A	A	B	A	B	A	A	A	A	A	A	A	B	B	A	A
101403	A	A	-	A	A	A	B	-	A	B	B	A	A	B	A	-	B	B	A
101404	-	B	A	B	B	B	B	A	B	A	A	B	B	-	A	B	B	B	B
101405	A	B	B	A	A	B	B	B	B	A	B	A	B	-	B	B	B	B	A
101406	A	A	B	A	A	A	A	A	B	B	A	A	-	-	A	B	-	A	A
101407	A	A	-	A	A	A	-	-	A	-	A	B	-	A	B	-	B	A	A
101408	A	B	B	-	B	B	A	-	B	-	A	B	B	A	B	B	A	B	B
101409	B	A	B	A	B	A	A	-	B	-	A	B	A	A	B	B	-	A	B
101410	A	B	B	A	A	B	B	B	B	B	B	A	A	A	B	B	A	A	A
101411	A	B	A	B	B	A	A	-	A	A	B	B	A	A	A	A	A	A	B
101412	B	A	A	B	B	B	B	A	A	-	B	B	-	A	A	-	B	A	B
101414	B	A	A	B	B	A	B	A	B	A	A	B	A	-	B	A	A	A	B
101415	B	-	B	A	B	B	B	B	B	-	A	B	B	A	B	B	B	A	A
101416	A	A	B	B	B	B	B	A	B	B	A	B	A	-	B	-	A	A	B
CDC Bethune	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
M96006	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B

APPENDIX III – SAS Code

```
/*Variance Components Calculation*/
proc mixed data=sasuser.Population4combined covtest;
class seed_clr genotype year rep irep;
model plot_yld= seed_clr genotype(seed_clr);
random year rep(year) irep(rep*year) genotype*year;
run;

/*Broad Sense Heritability*/
proc mixed data=sasuser.Population4combined covtest;
class genotype year rep irep;
model plot_yld=;
random genotype rep(year) year irep(rep*year)
genotype*year;
run;

/*LS means*/
proc mixed data=sasuser.Population4combined;
class YEAR REP iREP GENOTYPE;
model plot_yld = GENOTYPE;
random year REP(year) iREP(REP*year) genotype*year;
lsmeans GENOTYPE;
run;

/*T-tests*/
Proc ttest data=sasuser.pop4parents;
Class name;
Var plot_yld;
run;
```