Characterizing the Flax Core Collection for Earliness and Canopy Traits

A Thesis Submitted to the College of Graduate Studies and Research

In Partial Fulfillment of the Requirements

For the Degree of Master of Science

In the Department of Plant Sciences

University of Saskatchewan

Saskatoon

By

Tao Zhang

PERMISSION TO USE

In presenting this thesis in partial fulfillment of the requirements for a graduate 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 the professor or professors who supervised my thesis work. 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.

Request 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, SK S7N 5A8

ABSTRACT

Early maturity is an important objective for breeding flax adapted to the Western Canada. Crop canopy traits influence seed yield; however, studying its effects is challenging due to the complexity and limited knowledge of the genetics of this trait. The objectives of this research are : i) to characterize flax accessions from the Canadian gene bank collection for early flowering, maturity and canopy traits; ii) to identify SSR markers associated with plant branching and leaf area index (LAI); iii) to use Structural Equation Modeling (SEM) to identify canopy variables with significant effects on yield.

The flax core collection, consisting of approximately 381 accessions, was grown at the Kernen Crop Research Farm in 2010 and 2011. Additionally, 17 early and 17 late flowering accessions from the flax core collection were screened and their phenotypic responses in both growth chamber and field environments were measured. A large amount of phenotypic diversity was observed in long day and short day environments in these experiments. Some accessions appeared to be more photosensitive, while others were photoperiod insensitive.

The genetic control of canopy traits such as LAI and plant branching were studied using association mapping. Genotyping of the core collection was conducted using 375 SSR markers. Population structure analysis assigned the 381 flax accessions in the core collection into four distinct groups. Model comparison revealed that the mixed linear model reduced spurious marker trait associations. A total of 26 markers were identified to be significantly associated with plant branching and LAI.

The simultaneous examination of crop phenology and canopy traits to seed yield was performed using SEM analysis. The results indicated greater plant stand resulted in higher irradiance absorption and which resulted in greater seed yield. Days to flowering had a significant negative effect on seed yield and growing degree days to

maturity had a significant effect on seed yield. Plant branching and plant height had a positive non-linear effect on seed yield. This study has provided several insights into molecular approaches and statistical methods to improve flax breeding.

ACKNOWLEDGMENTS

I would like to express my deepest appreciation to my graduate supervisor Dr. Helen Booker, for her generous support, continuous guidance and advice, as well as her friendship, patience and understanding in every situation through my studies. A heartfelt thank you goes out to each member of my graduate advisory committee, Dr. Aaron Beattie, Dr. Rosalind Bueckert and Dr. Bruce Coulman for their supervision, motivation and valuable scientific advice throughout my thesis. A special thank you goes to my external examiner, Dr. Axel Diederichsen.

Many thanks to Dr. Eric Lamb, Dr. Yuefeng Ruan and Marwan for their help and guidance to drive this project to a successful end, as well as their help during manuscript preparation. The genotyping of the core collection was conducted in Dr. S. Cloutier's laboratory (A.A.F.C, Winnipeg, MB). I would like to extend my thanks to Dr. S. Cloutier for allowing me to intern and train in microsatellite development. Furthermore, under the tutelage of PhD student Mr. Braulie Soto-Cerda, I learned about association mapping, so that I could conduct the analysis for the canopy traits.

Special thanks to Lester Young, Shannon Froese, Ken Jackle, Barb Boon, Kalya Lindenback and the staff of the University of Saskatchewan Flax Research Field Laboratory for providing field and technical assistance. Thanks also to all my colleagues at the Department of Plant Science. Also, I would like to say many thanks to my friends, Yong Liu, Yining Liu, Ting Chen, Jia Sun, Zhenzhen Li, Hu Wang, Lei Ren, Ke Feng for their friendship and numerous help throughout my study periods.

I want to acknowledge co-leaders Professor Gordon Rowland and Dr. Sylvie Cloutier and other scientists involved in the conceptualization of the Genome Prairie supported Total Utilization Flax Genomics (TUFGEN) project for providing funding for my project. Additionally, I have received financial support from the College of Agriculture and Bioresources, University of Saskatchewan, College of Graduate

Studies (travel award). Finally, I would like to dedicate this thesis to my parents for their life time of support for my education which has brought me success in life.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AFLP: Amplified fragment length polymorphism

ANOVA: Analysis of Variance

EPCs: Expected parameter change

DTF: Days to flowering

DTM: Days to maturity

GDDF: Growing degree days to flowering

GDDM: Growing degree days to maturity

GLM: General linear model

Lavaan: Latent variable analysis

LAI: Leaf Area Index

LD: Linkage disequilibrium

MAD: Modified augmented design

MLM: Mixed linear model

PCoA: Principal coordinate analysis

PCR: Polymerase chain reaction

PGRC: Plant Gene Resources of Canada

PAR: Photosynthetically active radiation

PB: Plant branching

PI: Photoperiod insensitive

PS: Photoperiod sensitive

QTLs: Quantitative trait loci

RCBD: Randomized complete block design

RFLP: Restriction fragment length polymorphism

SEM: Structure equation model

SNP: Single nucleotide polymorphism

SSR: Simple sequence repeats

TSW: Thousand seed weight

CHAPTER 1

1. Introduction

Flax (*Linum usitatissimum* L.) is one of the 'founder crops' that built the foundation for agriculture in the Near East (Zohary, 1999). This plant is a fiber and oil crop which now has a wide range of infraspecific variation after thousands of years of selection (Vavilov, 1926). There is great morphological diversity due to the two distinct uses of flax. Fiber flax varieties tend to be taller and less branched and are usually adapted to cooler climates. Conversely, oil varieties are shorter with more branches and larger seeds. These oil varieties are often seen in warm regions like the Mediterranean area and India, but they are also grown in Canada and the USA (Zohary and Hopf, 2000).

The world germplasm collections maintain about 53,000 flax accessions (Diederichsen and Richards, 2003). Plant Gene Resources of Canada (PGRC) maintains more than 3,300 flax accessions and the flax core world collection consists of approximately 400 of these. The accessions at PGRC originate from 72 countries and represent all regions of historic or present cultivation of flax (Diederichsen and Fu, 2008). Kulpa and Danert (1962) introduced a formal distinction of four convarieties for cultivated flax: (1) dehiscent flax (convar. *crepitans*); (2) fibre flax (convar. *elongatum)*; (3) large seeded flax (convar. *mediterraneum*); and (4) intermediate flax (convar. *usitatissimum*) which is grown for fibre and/or seed production.

Cultivated flax is a summer annual, self-pollinating crop. It usually takes 90-150 days until harvest from seeding. Day-length plays a crucial role in affecting the length of the vegetative period in cultivated flax. Fibre flax varieties have a shorter maturity period than the oilseed ones in northern latitudes. One flax breeding goal is

to develop cultivars with early maturity for northern areas because of the short growing season in these regions (Diederichsen and Richards, 2003). Day-length sensitivity in flax and information on the association of this trait with other agronomic traits will facilitate the breeding of flax cultivars adapted to northern regions.

Association mapping is a method for QTL detection and has been used to find marker-trait associations in genetically diverse populations (Jin *et al.*, 2010), such as, the PGRC flax core collection. Association mapping utilizes the historic patterns of recombination that have developed over generations within the studied population (Cockram *et al*., 2008). The major advantage of this technique is that it does not require time-consuming and expensive development and use of segregating populations to map QTLs (Collard *et al*., 2005). In this study association mapping analysis was used to detect QTLs for plant branching and canopy traits.

Structural equation modeling (SEM) is a method developed from path analysis. It is a powerful statistical approach for analyzing the causal relationships between variables (Grace, 2006). The applications of SEM to crop science have proven to be a useful and powerful tool to understand the relationship between yield and yield components in oat (Lamb *et al*., 2011). The effects of crop phenology, canopy traits and morphological traits on flax yield are not fully understood due to the complexity of their interactions. Structural equation modeling could be used to determine the relationship between crop phenology, canopy traits and morphological traits and seed yield in flax.

Characterization and analysis of flax germplasm will be beneficial for improving crop performance (FAO, 1996). It is also necessary to know the genetic relationships among characterized flax accessions for the purposes of breeding and conservation (Ayad *et al.*, 1997). Traits of interest for breeders and producers include days to flowering (DTF), leaf area index (LAI) and plant branching (PB). Characterization

of the flax core collection for earliness and canopy traits would speed the introgression of these traits from the flax genebank collection into registered varieties for use by Canadian producers. Hence, the present study objectives were to understand the responsiveness of flax accessions to photoperiod and look at the genetic control of canopy traits in flax using association mapping. The second objective was to conduct structural equation modeling analysis on the core collection to determine the inter-relationships of phenology, canopy and morphological traits on yield.

CHAPTER 2

2. Literature review

2.1 Flax production and area

Flax plays an important role in human food due to its health benefits (Harper *et al.*, 2006) and is used as a nutritional supplement in animal feed (Maddock *et al.*, 2005). Flax has been cultivated for production of fiber and oilseed. Canada has been the world's leading producer and exporter of flaxseed since 1994 (Flax Council of Canada, 2007). Additionally, Canada produced an average of 706,000 tonnes annually from the years 1999 to 2009. In 2009, Flax covered about 631,000 hectares in Canada, producing about 861,000 tonnes of seed (Agriculture and Agri-Food Canada, 2010). Most (80-90%) of the flaxseed produced in Canada is exported to Europe, the United States, Japan and South Korea (Flax Council of Canada, 2007). The high demand for flaxseed is mainly due to its industrial application in paints, varnishes and linoleum. Western Canada has become the world's largest producer of high quality flax, because the cool climate is favorable for flax growth and production.

2.2 Centre of origin and domestication

The most likely progenitor of cultivated flax is pale flax (*L. angustifolium* Huds.), based on a phenotypic study (Hammer, 1995). The likely places of origin were the Near East, Southern Europe, or Central Asia (Zohary and Hopf, 2000; Tammes, 1925). An early characterization of flax types, based on agrobotanical characteristics, identified four major regions of flax diversity; the Indian Subcontinent, Abyssinia, the Near East and the Mediterranean (Vavilov, 1926; 1951). The flax types from the Indian Subcontinent and the Abyssinian regions are more morphologically diverse from those from the Near East and Mediterranean regions. Another classification scheme divided cultivated flax into four unique groups: fibre, intermediate,

large-seeded and dehiscent flax (Dillman, 1953; Kulpa and Danert, 1962). All sub-species are predominantly self-pollinated (Zohary and Hopf, 2000), although cross pollination may occur occasionally by insects (Williams, 1988) or through artificial methods.

Flax was grown in Egypt 6000-8000 years ago and is considered to be one of the oldest cultivated species. Flax cultivation reached Western Europe (the Netherlands, Northern France, Belgium and Switzerland) about 7000-5000 years ago (Dewilde, 1983). It is believed that Lois Hebert was the farmer who introduced flax into Canada in 1617 (Prairie Flax Products Dec., 2007). By 1875, European settlers were growing flax in the Canadian prairies. There has been a preference for growing flax either for its fibre or oil since its domestication. It is commonly accepted that the usage as a fibre plant had promoted its domestication (Dillman, 1936). In Europe, flax is mainly grown for its fibre, whereas, in North American flax is produced for its oilseed. Fiber flax grows taller and has an unbranched growth habit; whereas, oil flax is shorter and more branched.

The early and rapid distribution of flax cultivation in the Old World has resulted in a wide range of flax landraces adapted to different environments and uses (Vavilov, 1926). Diverse characteristics are observed between flax cultivars including differences in annual habit, branching habits, non-shattering capsules, self-fertilization and fatty acid profile. Breeding for specific combinations of these characteristics is a primary goal for flax breeders. One challenge is that these traits are the result of the interactions between many inherited and environmental factors.

2.3 Flax cultivar genetic diversity

Flax is a diploid (2n=30) self-pollinating crop plant. Improvement in flax is slow relative to other oilseed crops like soybean and canola because flax covers a smaller growing area, and consequently, receives less investment for development. Genetic

diversity within flax is low (Smýkal *et al*., 2011) and cannot be effectively supplemented by intraspecific hybridization. Additionally, methods for hybrid seed production have not been developed yet. The range of diversity present in Canadian flax can be recognized by comparing them to the word flax collection. These comparisons may identify germplasm with traits not present in Canadian cultivars (Diederichsen, 2001).

Several methods have been used to evaluate genetic variation in flax, such as using morphological characteristics (Diederichsen, 2001; Diederichsen and Raney, 2006; Diederichsen *et al*., 2006), isozymes (Månsby *et al*., 2000) and molecular markers (Fu *et al*., 2002; 2003). DNA molecular markers have become the main methods in the evaluation of genetic variation due to their abundance and environmental insensitivity. The variety of DNA molecular markers applied in flax genetic studies includes random amplified polymorphic DNA (RAPD) (Fu *et al*., 2002, 2003; Fu 2005, 2006; Diederichsen and Fu, 2006), restriction fragment length polymorphism (RFLP) (Oh *et al*., 2002), amplified fragment length polymorphism (AFLP) (Everaert *et al*., 2001) and simple sequence repeats (SSRs) (Cloutier *et al*., 2009; Fu, 2011). SSRs, also called microsatellites, consist of a variable number of tandem repeats. SSR markers are considered one of the best markers for association studies due to their abundance, high number of polymorphic alleles (Huang, 2002); genome-wide distribution and co-dominant inheritance patterns (Collard *et al*., 2005). Recently, Cloutier *et al*. (2009, 2010, 2012) constructed a linkage map of flax based on SSR markers.

Erosion of flax genetic diversity has occurred due to breeding activities (Diederichsen and Richards, 2003). To maintain flax diversity, a core world collection of approximately 400 distinct flax accessions has been selected based on genetic studies (Fu, 2006), phenology, morphological and agronomic traits (Diederichsen and Raney, 2006; Diederichsen *et al*., 2006, Diederichsen *et al*., 2008), as well as significance to Canadian agriculture (Kenaschuk and Rowland, 1995).

Genetic studies of flax accessions archived in the PGRC have identified some likely centers of flax diversity (i.e., North American, European, Asian and Mediterranean) (Diederichsen and Fu, 2008). The limited degree of genetic diversity in contemporary cultivars may limit future breeding efforts due to the reduced level of phenotypic variability. Understanding the genetic control of phenotypic variation, such as yield-related traits and introducing new sources of variation are major goals in genetic studies of flax.

2.4 Leaf canopy studies in flax

Canopy structure is related to foliage elements and their spatial distribution, orientation and size. Leaf area index (LAI) is an important parameter to estimate canopy structure. LAI was defined as the total leaf area per unit of ground area and was first mentioned by Watson (1947). LAI is greatly connected to a plant's primary physiological processes such as light interception, water vapor and CO_2 exchange. LAI can be used to assess potential yield when combined with other descriptors. When light travels through leaves, the light flux at the bottom of the canopy and light flux at the top of the canopy can be estimated. As the plant grows interception of radiation increases (Heath and Hebblethwaite, 1987); however, mutual shading becomes more intense. Mathematical models have been applied to calculate the constant fraction of light intercepted by a unit of leaf area (Newton and Blackman, 1969).

Little is published on the basic crop characteristics of flax that affect yield; such as, canopy expansion and light interception, dry matter production and partitioning; however, strong correlations have been observed in other species. Light use efficiency and harvest index are correlated with linseed production during the reproductive phase under favorable growing conditions (D'Antuono and Rossini, 2006). The relationship between dry matter and final seed yield is strongly correlated with total seasonal intercepted photosynthetically active radiation (PAR) in legumes

species (Ayaz *et al.*, 2004). Selection for seed yield is normally delayed until later generations, at which point selection for complex traits is more effective. The study of the relationship between canopy structure and sunlight helps in the understanding of crop dry matter production, in mathematical terms.

2.5 Early maturity in flax

2.5.1 Importance of early maturity in flax

Flax usually takes 90-150 days to mature (Diederichsen and Richards, 2003). Due to the long time required to reach maturity, flax is not well adapted to the Northern Prairies due to the short growing season and earlier fall frosts. This limits the areas it can be grown in Western Canada. In Southern Saskatchewan, the most probable time at which a killing frost can occur is between September 9 to September 15 (Source: Saskatchewan Crop Insurance). This short growing season (95-125 d) makes it challenging to grow flax in the Northern Prairies (Puvirajah, 2010). Given this risk of early frost affecting seed maturity, it is especially important to develop early-maturing cultivars for the conditions experienced in the Northern Prairies.

2.5.2 Photoperiod sensitivity and photoperiod insensitivity in flax

Cultivated flax is a long day crop (Sizov, 1955). The induction of flowering for photoperiod sensitive (PS) genotypes is long days, while photoperiod insensitive (PI) genotypes flower independent of day length. Photoperiod sensitivity is determined by the delay in flowering under short days compared with long days (Brutch *et al.*, 2008). The degree of photoperiod sensitivity varies greatly among accessions of fiber and oilseed (linseed) flax. In general, fibre flax varieties have a shorter vegetative period than oil seed varieties in northern latitudes (Diederichsen and Richards, 2003).

2.5.3 Theories of flowering

Maturity is quantitatively inherited and little is known about the genetic basis of

early maturity in flax. Vernalization and photoperiod are the two basic physiological systems that control when a crop flowers in response to the environment (Iqbal *et al.*, 2006). Vernalization response of flax which require low temperatures for flowering is not well documented. Photoperiod is a main external factor influencing flowering of crops grown at diverse latitudes (Carder, 1957). Flowering in a short-day plant is accelerated as the photoperiod shortens. In a long day plant flowering is accelerated with lengthening photoperiod (Major, 1980).

A systematic method to describe the photoperiod response of crop plant species can help breeders to develop genotypes that perform well at northern latitudes. In rice, the life cycle is separated into three phases: (l) the vegetative phase, starting at germination; (2) the reproductive phase, starting at floral initiation; and (3) the ripening phase, starting at flowering (Vergara and Chang, 1976). An obvious change could be noted at the transition of each of these phases in rice (Owen, l97l). There is a critical day length requirement that must be met to transition from the basic vegetative phase to the reproductive phase. The optimal photoperiod, for both short day and long day plants, is the critical day length that does not delay flower development. For example, it can be a minimum day length in the long day plants and a maximum day length in the short day plants. The model for photoperiod response of rice can be modified for use on most crop species and these studies are useful for comparing relative photoperiod sensitivity of genotypes (Major, 1980).

The transition of wheat plants to heading is regulated by several gene systems: *Ppd* genes (photoperiod sensitivity), *Vrn* (vernalization duration) and *Eps* (earliness *per se*) (Stel'makh, 1987). The same physiological mechanism and its genetic regulation may be noticed in flax plants.

2.6 Association mapping in flax

2.6.1 Concepts of association mapping and linkage disequilibrium

Association mapping, also known as linkage disequilibrium mapping, is a method for QTL detection. This method consists of finding marker-trait associations in genetically diverse populations (Morton, 2005), such as the PGRC flax core collection. The terms linkage disequilibrium (LD) and association mapping have been used interchangeably in most papers. Association mapping refers to the significance of a marker locus with a phenotype trait, while linkage disequilibrium refers to non-random association between two marker loci (Soto-Cerda and Cloutier, 2012). The difference between association mapping and linkage disequilibrium is that the former is the application of linkage disequilibrium. It means that two markers in LD do not necessarily imply a statistical significance of association.

2.6.2 The resolution of association mapping

Association mapping can detect a large number of alleles and enhance mapping resolution (Yu and Buckler, 2006). The resolution of association mapping is mainly determined by the extent and distribution of LD (Myles *et al*., 2009). High LD in a population means a reduced number of markers are required for QTL detection; whereas, low LD requires a large number of markers for the construction a high-resolution map or fine map. In a self-pollinated species such as flax the recombination frequency is relatively low compared to an out-crossing species like maize. Therefore, the estimated LD of flax is relatively high compared with maize. An accurate measure of LD in flax for association studies has not yet been published.

2.6.3 Association mapping population

Association mapping utilizes the historic patterns of recombination that have developed over generations within the studied population (Cockram *et al*., 2008). Usually, a set of breeding lines, cultivars or accessions from a germplasm collection is studied (Stich and Melchinger, 2010). For natural populations or germplasm collections, recombination has occurred for many generations and thus these

collections contain great genetic diversity. QTLs can be identified and then mapped with greater resolution compared to family-based mapping approaches. The mapping population can be divided, by the application of markers, into diverse genotypic groups according to the presence or absence of a particular marker locus (Collard *et al.*, 2005).

One of the disadvantages of association mapping is that the observed LD in a natural population could be caused by non-linkage factors, such as genetic drift, selection and population admixture (Flint-Garcia *et al*., 2003; Mackay and Powell, 2007). As a result, spurious marker-trait associations can be produced. Assessment of the genetic diversity, population structure and linkage disequilibrium (LD) of the target population can lower the chances of developing spurious marker-trait associations.

2.6.4 Marker assisted selection

Molecular markers have been applied in plant research to understand the genetic basis of QTLs. DNA markers that are tightly linked to a QTL can be designed as molecular tools for marker-assisted plant breeding (Ribaut and Hoisington, 1998). Marker-assisted selection is a method by which phenotype selection is based on a genotypic marker (Collard *et al.*, 2005). Marker-assisted selection is particularly efficient for complex traits with low heritability (Knapp, 1998). Another great benefit of marker-assisted selection is that it lowers costs through replacing higher cost methods of phenotypic selection (Peleman and Van der Voort, 2003). Before using marker-assisted selection, important considerations, such as validation of markers and resolution of mapping need to be satisfied through repeated tests.

2.7 Path analysis

Path analysis is a powerful method for partitioning the direct effect of a trait on yield and its indirect effects through other traits. The linear correlation among yield components and yield may cause confusion due to the possible interrelationship

between the components themselves. Structural equation modeling (SEM) is a modern version of path analysis that can be used to study the relationship between intercorrelated variables. There is no previous research using SEM to study flax yield; however, there has been some research using path analysis on flax yield related research.

A positive relationship between seed yield, number of capsules, number of branches and thousand seed weight (TSW) was shown (Chandra, 1977). A positive relationship was also found between seed yield and plant height (Copur *et al.*, 2006). TSW, plant height, number of capsules and number of primary branches have a direct effect on seed yield (Copur *et al.*, 2006), with number of bolls being most highly correlated. Path coefficient analysis and association analysis revealed that the number of capsules, number of primary branches, TSW and plant height are the common causal factors that influence economic yield (Copur *et al.*, 2006). Hence, selection criteria for the improvement of seed yield should be based on the number of primary branches and plant height.

2.8 Objectives

The specific objectives of this study were as follows:

1. To determine the responsiveness of flax accessions to photoperiod and quantify the effect of early and late flowering type in terms of days to flowering and yield in the chamber and field.; to identify accessions that are photoperiod sensitive or photoperiod insensitive.

2. To use association mapping to determine the genetic control of canopy traits and plant branching in flax by genotyping and phenotyping the flax core collection.

3. To determine the relationship of phenology, canopy and morphological traits to yield of the core collection through structural equation modeling.

2.9 Hypothesis

Variation in maturity will be found in the flax core collection that will be useful in breeding flax for the northern grain belt. Quantitative trait loci (QTLs) associated with canopy traits will be detected. Crop phenology and canopy traits will influence seed yield in flax.

CHAPTER 3

3. Characterization of Early and Late Flowering Accessions Under Controlled Growth Chamber Environments and Field Performance

3.1 Abstract

Characterizing flax accessions from gene bank collections is particularly important for plant breeders to meet the challenge of breeding flax varieties adapted to the Northern Prairies. Seventeen early and seventeen late flowering accessions from the flax core collection were screened and studied for their phenotypic responses to photoperiod both in the growth chamber environment and the field. Another goal of the present experiment was to identify accessions that are photoperiod sensitive or photoperiod insensitive. Photoperiod and accessions / cultivar (genotype) are the main factors that affect all measured traits. Results showed that the field performance of the 34 lines was similar to the phenotypic response observed in the long day chamber. Accessions CN98807, CN100828, CN98794, CN101419, CN98370, CN98397, CN97180, CN98135 and CN100559 were determined as the most photoperiod sensitive lines, showing a delay of 40-50 DTF between long and short day chambers. Accessions CN96992, CN98014, CN97530, CN98286, CN98468, CN101610 and CN98150 were considered as the least photoperiod sensitive lines, with about a 20 days difference between long and short photoperiods.

3.2 Introduction

Cultivated flax is a summer annual, self-pollinating crop and is considered a long day (LD) plant (Sizov, 1955). It usually takes 90-150 days to mature (Diederichsen and Richards, 2003). Due to the long time required to reach maturity, flax is not well adapted to the Northern Prairies, which limits its acreage in Western Canada.

Despite this, Western Canada is an important linseed production region (Fig. 3.1). The climate in Western Canada has a risk of early fall frost, the probable time for the first frost in southern Saskatchewan is between September 9 and September 15 (Source: Saskatchewan Crop Insurance). Therefore, the short growing season (95-125 d) makes it challenging to grow flax in Western Canada (Puvirajah, 2010). Given the risk of early frost affecting seed maturity, it is especially important to develop early-maturing cultivars for the conditions experienced in the Northern Prairies.

Fig. 3.1. Traditional growing areas for linseed in Western Canada (Used with permission from the Flax Council of Canada).

The diversity of photoperiod sensitivity in flax can be utilized for flax breeding. Photoperiod sensitive genotypes planted early may not flower until the day length reaches a critical level. This may cause a longer period of vegetative growth, but may also delay flowering until a hotter period of the summer. Weak photoperiod sensitivity is always associated with early flowering in most plant species (Jackson, 2009). Conversely, photoperiod insensitive cultivars are expected to flower independent of day length and thus may be expected to flower earlier in response to

early planting. Therefore, both photoperiod sensitive genotypes and photoperiod insensitive genotypes have their respective roles in a breeding program.

Flax is an old crop and cultivated on most continents which represent different climate zones. To comprehensively understand the life cycle and physiological characteristics, accessions classified as early or late maturing were selected from the PGRC core collection. The growth and development phase of flax is controlled by its photoperiod response, together with its interaction with growth temperatures, playing a major role in the adaptation of flax in various environments. As temperatures fluctuate seasonally, growing degree days are used to predict plant development based on actual temperature. A better understanding of the underlying genetic control of agronomically important traits in relation to photoperiod response may aid in breeding for early maturity. A field study and growth chamber experiments were conducted:

1. To determine the responsiveness of flax accessions to photoperiod;
2. To quantify the effect of early and late flowering types in terms of days to flowering and height in the chamber and field；
3. To identify accessions that are more photoperiod sensitive or least photoperiod insensitive.

3.3 Materials and Methods

3.3.1 Collections and experimental setup

The lines used in this study were obtained from Plant Gene Resources of Canada (PGRC). Seventeen early-flowering accessions and seventeen later-flowering accessions were selected and compared with two checks (CDC Bethune and Flanders). Table 3.1 listed all the accessions with their accession name, country origin, botanical convariety and reason selected. The collection included fibre flax cultivars, linseed cultivars, some unknown purpose types and breeding material originating from different locations.

Table 3.1. List of accessions used in this study.

CN number	Cultivar	Origin	Convarieties	Reason selected *
98807	028-7	France	Not determined	Early
97530	No name	Russia	*Elongatum*	Early fibre flax
98440	N.P. 109	India	*Usitatissimum*	Early
98974	N.P. 118	India	*Mediterraneum*	Early and high TSW
97529	No name	Russia	*Usitatissimum*	Early
98237	24836	Pakistan	*Usitatissimum*	Early
98973	N.P. 117	India	*Mediterraneum*	Early and high TSW
98370	N.P. 37	India	*Mediterraneum*	Early and high TSW
97350	De metcha 1-3-3 Vilm	France	*Usitatissimum*	Early
98794	Lino de Cabiro	France	Not determined	Early
98397	N.P. 65	India	*Usitatissimum*	Early
98683	Mapum M.A.	Czech Republic	*Elongatum*	Early
98014	10451/46	Argentina	*Usitatissimum*	Early
98398	N.P. 66	India	*Usitatissimum*	Early
98468	N.P. (RR.) 407	India	*Usitatissimum*	Early
98135	Indian Type 6	India	*Usitatissimum*	Early
98286	Mapun	Hungary	*Elongatum*	Early fibre flax
97180	Sorth Behbehan	Iran	*Elongatum*	Late
96992	No name	Ethiopia	*Usitatissimum*	Late
97004	No name	Ethiopia	*Usitatissimum*	Late
98150	Z 11637	The Netherlands	*Elongatum*	Late
100837	Noname	Turkey	*Crepitans*	Late
101421	Noname	China	*Elongatum*	Late
100828	Winterlein	Turkey	*Usitatissimum*	Late
101052	L-93-2	China	*Elongatum*	Late
101416	No name	China	*Elongatum*	Late
101419	No name	China	*Elongatum*	Late
97129	No name	Iran	*Usitatissimum*	Late
101559	Sel. of CIli-2283 (C4)	Canada	*Usitatissimum*	Late
101572	Sel. of CIli-2560 (C4)	Canada	*Usitatissimum*	Late
96991	No name	Ethiopia	*Usitatissimum*	Late
40081	Natasja	The Netherlands	*Usitatissimum*	Late
101610	Sel VIR-2404 (Sterile)	Canada	*Usitatissimum*	Late
97584	Minn. Sel. Winona x 770B F5	USA	*Usitatissimum*	Late

* Diederichsen *et al.* 2012.

An experiment with two randomized blocks was grown in each controlled environment growth chamber to investigate the responses of the 36 lines to photoperiod. In chamber 1, plants were grown under long-days (16/8h day/night cycle). In chamber 2, plants were grown under short-days (10/14h day/night cycle). The temperature was maintained at 25°C/day time and 17°C/night time in both chambers. All accessions were grown with four plants per pot, two pots per accession in each chamber. Plants were watered with tap water as needed and were treated with biological agents to control thrips. The controlled environment experiment was repeated once.

A field test was conducted in 2011 to study the agronomic traits of early and later flowering accessions and check cultivars grown in the controlled environment experiment. The site at the Kernen Crop Research Farm was a clay loam (Dark Brown Chernozem) (Lat. 52º09´N; Long. 106º33´). The accessions and checks used in the chamber experiment were planted in the field in a randomized complete block design (RCBD) with two replications. The seeding date was May 24 and harvested on Sep. 19.

3.3.2 Data collection

In the growth chamber, observations were taken on the days to flowering, plant height, yield and aerial biomass. The number of days between emergence and first anthesis was measured for each pot. A pot was determined to have reached first anthesis once a single plant started flowering. Photoperiod sensitivity was determined by the delay of flowering under the short day environment in comparison with the long day environment for all the accessions. The average duration of this delay was also calculated for each accession. Total height was measured after flowering. Seeds were collected from all plants; the average seed weight per plant was calculated and designated as seed yield. After harvest, the plants were dried down for two weeks at room temperature. The total seed weight per plant along with

above ground biomass per plant (as dry weight) was determined as the aerial biomass. Harvest index was calculated as: % HI = (seed weight / aerial biomass) x 100.

In the field experiment, days to flowering and days to maturity were observed. The number of calendar days was converted to Growing Degree Days by summing the average daily temperature. To calculate the Growing Degree Days of flax accessions grown in the field, the flax base temperature was determined as 0°C (Miller *et al.* 2001). The yield was calculated based on final seed weight per plot and it was converted to kg/ha according to the total harvested area.

3.3.3 Statistical models and data analyses

Analysis of variance (ANOVA) for all trait data in the controlled environment study was done using PROC MIXED of SAS (SAS Institute Incorporated, Cary, North Carolina, USA). The two years of chamber data were analyzed together. Photoperiod and accession were considered as the fixed effects, while year and year (block) were considered as random effects. PROC MIXED was used and LSMEANS were estimated. The flowering type was treated as a fixed effect. To quantify the effect of early flowering type, the group of early accessions was compared to accessions having early and late type. Similarly, the effect of late flowering type was estimated by comparing the group of late accessions to accessions having early and late type. For the field experimental data, a similar ANOVA was conducted. Accession was considered as a fixed effect and block was a random effect. Contrast analysis was used to compare all traits using the early and late genotype as the classification variable. Photoperiod response was considered as the difference in days to flowering between short and long photoperiod treatments.

3.4 Results

3.4.1 Growth chamber analyses

The ANOVA showed that the main effect, photoperiod, was highly significant ($p < 0.01$) for DTF, height, yield and harvest index. It was also significant for aerial biomass ($p < 0.05$) (Table 3.2). The effect of genotype (accession) was highly significant ($p < 0.001$) for all measured traits. The interaction between photoperiod and genotype only showed a significant effect for DTF and plant height ($p < 0.05$). Other measured traits including seed yield, aerial biomass and harvest index were not observed to be significantly effect by the interaction between photoperiod and genotype.

Table 3.2. Analysis of variance for testing the effect of photoperiod and accessions on DTF, plant height, yield, aerial biomass and harvest index for accessions grown under short-days (10/14h day/night cycle) and long-days (16/8h day/night cycle) in the controlled environment chamber.

Source	*Significance of fixed effects (Pr>F)*				
	DTF	**Height**	**Yield**	**Aerial biomass**	**Harvest index**
Photoperiod	<.0001	<.0001	<.0001	0.0211	0.0002
Accession	<.0001	<.0001	<.0001	<.0001	<.0001
Photoperiod × Accession	<.0001	0.0105	0.99	0.8381	0.9577

The least square means (LS means) of all the accessions and two checks for days to flowering and height were presented in Table 3.3 and Table 3.4 based on two controlled environment experiments. In Table 3.3, all the accessions and checks were ranked from low to high based on the days to flowering difference between the two photoperiod treatments. The last two accessions CN97004 and CN96991 did not flower under the short photoperiod condition. The flax accessions in this study exhibited a wide range of DTF. DTF under the long photoperiod environment ranged

from 31 days to 82 days after sowing. DTF under the short photoperiod environment ranged from 53 days to 99 days. A short photoperiod delayed DTF in all accessions and varieties. Some accessions did not flower under short days. Plant height was also affected by photoperiod regime. For the long day photoperiod, plant height showed a range from 58cm to 151cm. Plant height ranged from 71cm to 159 cm under the short day photoperiod. The shorter photoperiod usually resulted in taller plants.

Table 3.3. Means days to flowering of flax accessions and two checks ('CDC Bethune' and 'Flanders') grown under two different photoperiods in controlled environment chambers.

Accessions/Cultivars (Type)	*DTF (short)*	*DTF (long)*	*Delay*
CN96992 (Late)	82	66	17
CDC Bethune (Check)	69	49	21
CN98014 (Early)	67	46	21
CN98286 (Early)	61	40	21
CN97530 (Early)	60	39	22
CN101610 (Late)	71	49	22
CN98468 (Early)	53	31	23
CN97529 (Early)	60	37	23
CN98150 (Late)	72	48	23
CN100837 (Late)	81	56	26
CN97350 (Early)	72	45	27
CN101572 (Late)	80	52	28
CN101052 (Late)	82	53	29
CN97584 (Late)	82	52	30
CN98440 (Early)	67	38	30
CN98683 (Early)	72	42	31
CN98973 (Early)	74	40	34
CN98974 (Early)	75	40	35
CN98237 (Early)	79	44	35
CN98398 (Early)	71	36	36
CN100828 (Late)	87	51	36
CN101416 (Late)	86	49	37
CN101421 (Late)	87	49	38
CN97180 (Late)	89	52	38
CN101559 (Late)	93	56	38
Flanders (Check)	89	51	38
CN98794 (Early)	73	35	39
CN98370 (Early)	79	41	39
CN98135 (Early)	76	37	40
CN40081 (Late)	91	50	41
CN98397 (Early)	76	34	42
CN101419 (Late)	88	46	43
CN97129 (Late)	99	51	48
CN98807 (Early)	86	31	55
CN97004 (Late)		57	57
CN96991 (Late)		82	82

Table 3.4. Means for the plant height of flax accessions and two checks ('CDC Bethune' and 'Flanders') grown under two different photoperiods in controlled environment chambers.

Accessions/Cultivars (Type)	*Height (short)*	*Height (long)*	*Differences*
CN96992 (Late)	98.5	115	16.5
CN101416 (Late)	153	160.5	6.5
CN101052 (Late)	148	151	3
CN101421 (Late)	159	161	2
CN98150 (Late)	155	151	4
CN101419 (Late)	161.5	157	4.5
CDC Bethune (Check)	130	125.5	4.5
CN97004 (Late)	117	111.5	5.5
CN98440 (Early)	99	93	6
CN98468 (Early)	71	65	6
CN97529 (Early)	111	104	7
CN101559 (Late)	142.5	134	8.5
CN97350 (Early)	100	91	9
CN98683 (Early)	125	114	11
CN101572 (Late)	148.5	137	11.5
Flanders (Check)	119	107	12
CN98237 (Early)	88	75.5	12.5
CN40081 (Late)	129.5	117	12.5
CN98286 (Early)	132.5	118	14.5
CN100837 (Late)	136.5	119	17.5
CN97129 (Late)	109	91	18
CN97584 (Late)	126.5	109	17.5
CN98014 (Early)	109	91	18
CN98794 (Early)	103	84.5	18.5
CN97530 (Early)	133	113	20
CN98398 (Early)	85	64	21
CN98370 (Early)	79	58	21
CN98973 (Early)	83	62	21
CN101610 (Late)	99.5	76.5	23
CN97180 (Late)	156	128	28
CN98135 (Early)	88	60	28
CN100828 (Late)	118.5	90	28.5
CN98397 (Early)	96	67	29
CN96991 (Late)	119	87.5	31.5
CN98974 (Early)	84	46	38
CN98807 (Early)	132	84	48

The measurement of photoperiod sensitivity was indicated by the reduction in the number of days to flowering between two conditions. The two experiments conducted at different times were analyzed separately and average mean values were also calculated. A total of nine most photoperiod sensitive flax accessions were selected based on the two controlled environment experiments. The most photoperiod sensitive line among all these accessions was the accession CN98007, which showed the largest average days to flowering variation (54.5 days) (Table 3.5) between long and short photoperiod, followed by the accession CN101419, which has an average of 42.5 days difference (Table 3.5). In contrast, the least photoperiod sensitive line was the accession CN96992, which showed only a 16.5 day difference. Accessions CN98014, CN97530 and CN98286 had a 21.5 day average difference between long and short photoperiod (Table 3.6) while flax accession CN101610, which belonged to the late flowering type showed 22 days average difference.

Table 3.5. More photoperiod sensitive of flax accessions grown under short (10/14h day/night cycle) and long day (16/8h day/night cycle) controlled environments.

Flax accessions	*Type*	*Difference in Time to Flowering (days)*		
		2010	**2011**	**Average**
CN98807	Early*	52.5	56.5	54.5
CN100828	Late**	34	38	36
CN98794	Early	37	40	38.5
CN101419	Late	40.5	44.5	42.5
CN98370	Early	41.5	36	38
CN101559	Late	37	38.5	38
CN98397	Early	36	48	42
CN97180	Late	34.5	40.5	37.5
CN98135	Early	35	44	39.5

*Early type was the group of early flowering accessions selected from core collection

** Late type was the group of late flowering accessions selected from core collection

Table 3.6. Least photoperiod sensitive of flax accessions grown under short (10/14h day/night cycle) and long day (16/8h day/night cycle) controlled environments.

Flax accessions	*Type*	*Difference in Time to Flowering (days)*		
		2010	**2011**	**Average**
CN96992	Late**	16.5	-	16.5
CN98014	Early*	17.5	25	21.5
CN97530	Early	24	19	21.5
CN98286	Early	23.5	19	21.5
CN98468	Early	28	17	22.5
CN101610	Late	24.5	19.5	22
CN98150	Late	24.5	22	23.5

*Early type was the group of early flowering accessions selected from core collection

** Late type was the group of late flowering accessions selected from core collection

To estimate the effect of treatment based on flowering type, the least square means (LS means) for agronomic traits are presented in Table 3.7 based on two controlled environment experiments. The early flowering type accessions flowered 13 days earlier than the late flowering type accessions. Short photoperiod delayed the days to flowering for both types by an average of 32 days. Early flowering type accessions had reduced plant height compared to late type accessions in both photoperiods. Accessions grown in the long photoperiod were shorter compared to the short photoperiod treatment.

Table 3.7. Means of days to flowering and height for 34 flax accessions grown as a randomized complete block design with four replications under two photoperiods (16 h long day and 10 h short day).

Type	*Photoperiod*	*Days to flowering (d)*	*Height (cm)*
Early		55	91
Late		68	131
	Short	78	121
	Long	46	109
Early	**Long**	38 D	82 D
Late	**Long**	52 C	125 B
Early	**Short**	71 B	101 C
Late	**Short**	84 A	136 A

A-D within variable, means followed by the different letters have significant differences at the $p \leq 0.05$ level using the Least Significant Difference (LSD).

3.4.2 Field analyses

Analysis of variance (Table 3.8) showed accession (genotype) highly affected all measured variables, such as DTF, plant height, plant branching, DTM, yield, growing degree days to flowering and growing degree days to maturity.

Table 3.8. Analysis of variance for testing the effect of accession (genotype) on DTF, plant height, yield, plant branching (PB), DTM, growing degree days to flowering (GDDF) and growing degree days to maturity (GDDM) grown at the KCRF in 2011.

Source	*Significance of fixed effects (Pr>F)*						
	DTF	**Height**	**Yield**	**PB**	**DTM**	**GDDF**	**GDDM**
Accession	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001

Analysis of variance (Table 3.9) showed convarieties highly affected plant height, plant branching, DTM and growing degree days to maturity, but other measured traits including DTF, yield and growing degree days to flowering were not observed to be significantly effect by the convarieties.

Table 3.9. Analysis of variance for testing the effect of convarieties on DTF, plant height, yield, plant branching (PB), DTM, growing degree days to flowering (GDDF) and growing degree days to maturity (GDDM) grown at the KCRF in 2011.

Source	*Significance of fixed effects (Pr>F)*						
	DTF	**Height**	**Yield**	**PB**	**DTM**	**GDDF**	**GDDM**
Convarieties	0.601	<.0001	0.829	<.0001	0.002	0.65	0.0022

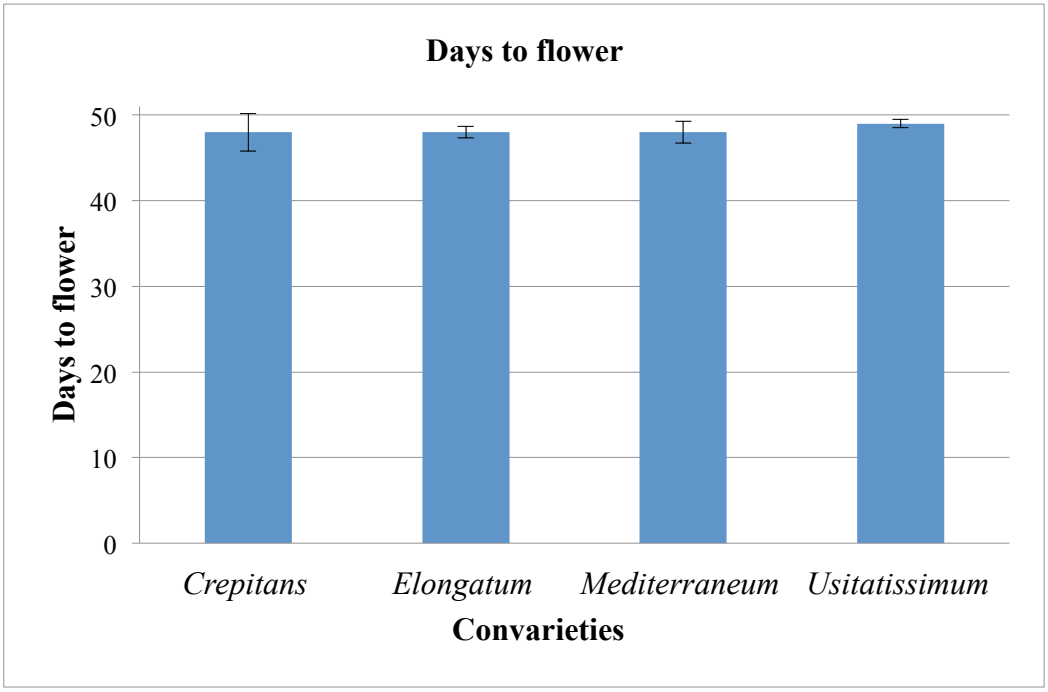

Fig. 3.2. Average Days to flowering differences among four different flax types among 34 accessions grown at the KCRF in 2011. Error bars on the graphs indicate ± standard error.

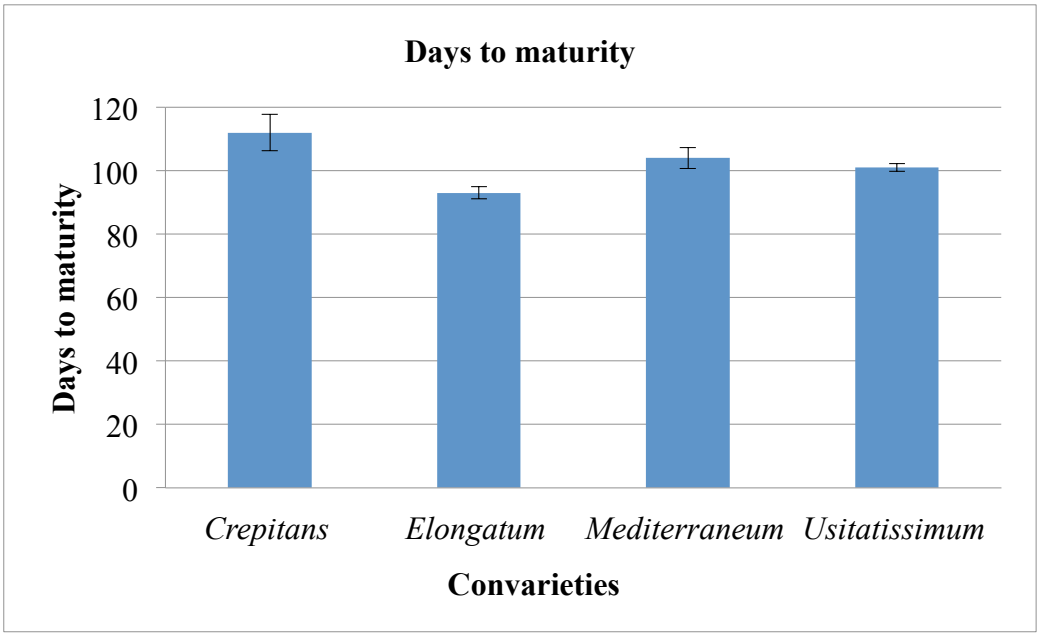

Fig. 3.3. Average Days to maturity differences among four different flax types among 34 accessions grown at the KCRF in 2011. Error bars on the graphs indicate ± standard error.

The average days to flowering and days to maturity were calculated in terms of four different flax types (*crepitans*, *elongatum*, *mediterraneum* and *usitatissimum*). The detail information about four covarieties was presented by Diederichsen and Richards (2003). There was no clear difference for the DTF in the field test among the flax convarieties (Fig. 3.2). However, great differences were seen in DTM among the four different flax types (Fig. 3.3). The dehiscent flax (*crepitans*) required the most days (112 days) to reach maturity. The large seeded flax (*mediterraneum*) required 104 days for maturity. The fibre flax (*elongatum*) took the least days (93 days) to mature. The intermediate flax (*usitatissimum*) required a moderate number of days (101 days) for ripening of bolls.

3.4.3 Growth chamber versus field performance

A clear difference could be observed for the DTF between the short day controlled environment experiment and the field test (Fig. 3.4). The flowering pattern was comparable between the long day controlled environment and field experiment, because the field site (Lat. 52º09´N; Long. 106º33´) in Saskatoon has a summer day length reaching a maximum of about 16 h. As a whole, the early types flowered about nine days earlier in the long day growth chamber experiment when compared to the field test. The check CDC-Bethune showed only a 0.5 days difference between long day controlled environment experiment and the field test. Early types flowered earlier than late types in all three environments. Even under field conditions, early flowing types were about three days earlier than the late flowering types (Fig. 3.4). The late type flax accessions flowered earlier under field conditions compared to the long day controlled environment conditions. It takes about 50 days for late types to flower in the field and 52 days in long day chamber. This may be due to differences between long day field and chamber conditions, such as, light intensity and temperature. The maximum light intensity in the field under cloudless skies reaches about 2000 $\mu mol \cdot m^{-2} \cdot s^{-1}$, but the light intensity in the chamber was approximately 500 $\mu mol \cdot m^{-2} \cdot s^{-1}$.

Later flowering flax types grew taller than early types in the growth chamber experiments and field trial (Fig. 3.4). Both early and late types grew taller in the growth chamber than under field conditions. The check cultivars CDC-Bethune and Flanders grew significantly taller when grown in the growth chamber. In the field the early types grew an average of 44cm taller and the late types grew an average of 70cm taller.

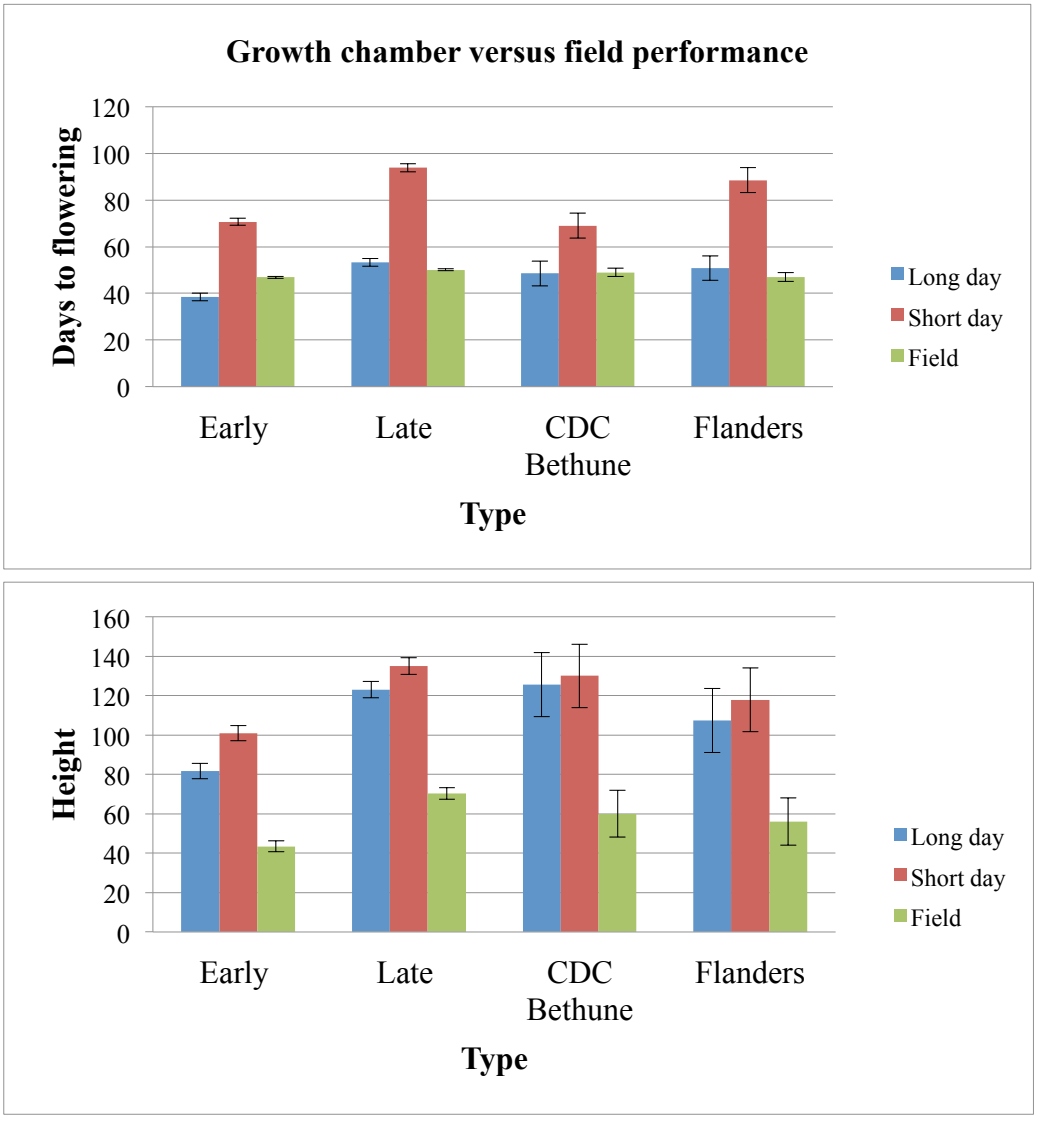

Fig. 3.4. Differences between long day and short day treatments compared to field tests for days to flowering (up) and plant height (down) among 36 flax accessions. Early type was group of early flowering accessions and late type was group of late flowering accessions selected from core collection. Error bars on the graphs indicate ± standard error.

3.5 Discussion

Results from the growth chamber experiments indicated that DTF, plant height and yield were highly affected by photoperiod and accession. All the accessions/cultivars showed earlier flowering under the long day treatment. This study confirmed that flax is a long-day plant and that flowering is accelerated under long day conditions. This study showed that there was considerable variability in the response of genotypes to photoperiod.

The most photoperiod sensitive lines among all these accessions were CN98807, CN100828, CN98794, CN101419, CN98370, CN98397, CN97180, CN98135 and CN100559. They showed about 40-50 days to flowering variation between long and short photoperiod. The first three accessions were early flowering types and the rest were later flowering types. The early flowering types with extreme photosensitivity will have potential for the northern adaptation. Genotypes with extreme sensitivity to photoperiod can be adapted to a specific region (Wallace, 1991).

Accessions CN96992, CN98014, CN97530, CN98286, CN98468, CN101610 and CN98150 were considered as the least photoperiod sensitive lines, with about a 20 day difference between long and short photoperiods. Four of them were early flowering types and three accessions, CN96992, CN101610 and CN98150, were later flowering types. Photoperiod insensitive accessions have been introduced in some species to broaden adaptation and insensitivity is usually related to extremely early lines (Lawn, 1989). In the case study here, the least photoperiod sensitive accessions were early flowering type. Therefore, there was potential for exploiting photoperiod insensitive accessions in flax to breed for early maturity.

The DTF difference between the same accessions grown under controlled environment conditions from year to year (i.e. replicate) was not consistent. The seed source used in two years study originated from the PGRC and increased at the

Morden Research Center, Monden, Manitoba. The variability between phenology of the same accession grown under similar conditions might be due to different conditions in the chambers in 2010 and 2011, such as temperature or light intensity fluctuations. It should be noted that the chamber lighting system were replaced for the second year (second replicate) of this study.

All the accessions/cultivars grew taller under the short day treatment compared to the long day treatment. The plants flowered later under short days and the vegetative period was extended. Early and late types grew taller in the growth chamber than under field conditions. This observation could be explained by the plant response to higher light intensity and fluctuating temperature and wind under field conditions.

Generally, early flowering leads to early maturity in the growth development of a crop. However, flax is a dual purpose crop (fibre and linseed). The time length required from green capsule to maturity was different between fibre types and linseed under field conditions. Among the accessions studied, there were four different convarieties (Table 3.1). The large seeded flax (convar, *mediterraneum*) took longer for the ripening of the boll. There were some large seeded types and linseed flax among the early types. The larger the seed, the longer the drying period required. It makes sense that some large seeded early types took longer to reach maturity. Conversely, there was more fibre type flax (convar, *elongatum*) among the late flowering accessions.

Comparisons between the growth chamber experiment and field test showed similarity with the long day treatment in the growth chamber. This is due to the high latitude location of Saskatoon. The length of photoperiod usually attained is about 16 h during the growing season. However, the variation in days to flowering in these two environments may be due to the light intensity and temperature difference between the growth chamber and field environments. Temperature is an important factor that affects rate of plant development and time to flowering.

CHAPTER 4

4. Association Mapping of Canopy Traits and Plant Branching in the Flax Core Collection

4.1 Abstract

Association mapping is a method for detection of gene effects based on linkage disequilibrium (LD). The flax core collection consists of 381 accessions which were selected by Plant Gene Resources of Canada (PGRC). Canopy traits like Leaf Area Index (LAI) are associated with crop yield. Plant branching is an important trait to classify variation in cultivated flax. 375 SSR markers were used to conduct association mapping analysis. The population structure analysis assigned 381 flax accessions into four groups. Model comparison revealed that the mixed linear model reduced spurious marker-trait associations. In total, 16 markers were identified to be significantly associated with plant branching and 18 markers were identified to be significantly associated with LAI. However, there were no marker trait associations that were consistently identified across years. There were 10 markers associated with both plant branching and LAI. Additionally, they were located on four linkage groups in both years.

4.2 Introduction

Flax (*Linum usitatissimum* L., 2n = 30) is one of the 'founder crops' that built the foundation for agriculture in the classic 'old world' (Zohary, 1999). Flax is known as a fibre and oilseed crop and has been cultivated for both purposes (Dillman, 1953). Plant branching is an important trait to classify cultivated flax. Plant branching prolongs vegetative grow and creates multiple sites on the plant for seed production (Dun *et al*., 2006). Fibre flax and seed flax (linseed) are the two divergent forms in commercial production. Less branching is desired in fibre flax, because less branching

allows for better quality longer fibers. For oil production, more branching is desired for increased seed set. A positive relationship has been observed between seed yield and number of branches (Chandra, 1977; Copur *et al*., 2006). Branching in the upper parts of the stem is mainly determined by genotype (Diederichsen and Richards, 2003). Dissecting the genetic basis of plant branching is of importance in flax breeding. However, studying plant branching using molecular mapping has not been reported in flax.

Association mapping is known as linkage disequilibrium (LD) mapping and is a method for QTL detection that is used to find marker-trait associations in genetically diverse populations (Morton, 2005), such as the PGRC flax core collection. Association mapping is becoming a useful method for the dissection of complex genetic traits (Churchill *et al*., 2004). Leaf area index and plant branching are two complex genetic traits; therefore, association mapping is well suited to this research. The major advantage is that it does not require the time-consuming and expensive development and use of segregating populations to map QTLs (Collard *et al*., 2005).

Spurious associations can be produced between markers and phenotype in association analysis. To solve this problem, population structure information and kinship information among individuals are two important components in accurately estimating marker trait associations. Two widely used software packages for association mapping analysis are STRUCTURE (Pritchard *et al.*, 2010) and TASSEL (Buckler *et al*., 2009). STRUCTURE runs through a model-based clustering method using genotype data. This program will identify the presence of population structure and distinct genetic sub-populations then assign a proportion of each individual's genome to each sub-population. TASSEL is a powerful statistical method to find QTL and it is done by incorporating inferred ancestry coefficients of the individuals (Q matrix) across the sub-populations as a covariate in the association mapping analysis (Vinod, 2011).

Plant branching and canopy traits like leaf area index are complex genetic traits related to yield. In this study, the variation of LAI and plant branching in the flax core collection were examined in the field. Therefore, the objectives of this research were:

1. To genotype and phenotype the flax core collection for canopy traits and plant branching;
2. To use association mapping to determine marker associations with these traits.

4.3 Material and Methods

4.3.1 Mapping Population

All 381 accessions representing the core collection from Plant Gene Resources of Canada (PGRC) were selected by the curator Dr. Axel Diederichsen for this study. The core collection covers most (>95%) of the diversity estimated to be present in cultivated flax (Diederichsen and Fu, 2008).

4.3.2 Phenotypic Evaluation

Experimental plots were planted at the Kernen Crop Research Farm, Saskatoon, Saskatchewan, Canada (Lat. 52º09´N; Long. 106º33´) on clay loamy, dark brown chernozemic soil. In the 2010 growing season, the seeding date was June 14. Accessions were planted in 3 row plots of 3.66 m length at 17.5 cm spacing by seeding 13 grams of seed per plot. In the 2011 growing season, the seeding date was May 24. Accessions were planted in 6 row plots using a 6-row Wintersteiger small plot seeder. A modified augmented design (MAD) was used. MAD is widely used in testing large numbers of lines without replication (Schaalje *et al*., 1987). Control plots are introduced to adjust for environmental heterogeneity and the final adjusted mean values were used for further analysis, in this study, CDC Bethune, Macbeth and AC Hanley were used as control plots.

To better understand the impact of environment, rainfall and Growing-degree days

(GDD) in the two growing seasons were compared. Growing-degree days were calculated as follows: GDD = (max. daily temp. + min. daily temp) / 2 - base temperature and were accumulated starting on June 14th and May 24th for the 2010 and 2011 growing seasons, respectively. A base temperature of 0 °C was used (Miller *et al*., 2001).

4.3.3 Canopy Traits

Leaf area index (LAI) is defined as the leaf area per unit area of ground. It is a very important canopy variable (Lane *et al*., 2000) and is measured to estimate canopy biomass and density. The device used to estimate LAI measurement was an AccuPAR Ceptometer (Decagon Devices, Pullman, WA). The AccuPAR calculates LAI by comparing the above and below-canopy photosynthetically active radiation (PAR), as well as other variables that relate to the canopy architecture and position of the sun, then converting these parameters to LAI using the standard equation below:

$$L = \frac{\left[\left(1 - \frac{1}{2K}\right) f_b - 1\right] \ln \hat{o}}{A\left(1 - 0.47 f_b\right)} \qquad [1]$$

in which L is the LAI, f_b is the fractional of beam, which means the ratio of direct beam radiation coming directly from the sun to radiation coming from all ambient sources. Where $A = 0.283 + 0.785a - 0.159a^2$, and a is the leaf absorptivity in the PAR band (AccuPAR assumes $_a = 0.9$ in LAI sampling routines). $\hat{o}$ is the ratio of below canopy PAR measurements to the above canopy measurements. K is the extinction coefficient for the canopy, which depends on the leaf angle distribution of canopy elements (also known as x) and the zenith angle of the probe ($\grave{e}$). Once you set your location, data and time, the AccuPAR will automatically calculates both the zenith angle and fractional beam. The leaf angle distribution of flax was set to be 0.4 based on the measurement of Newton and Blackman (1970). The measurements to estimate LAI were made in the early season of canopy development and after flowering. The growth stage of every plot was recorded at the time LAI was

estimated.

The PAR interception efficiency, also known as interception absorption, is the proportion of PAR intercepted by the crop. It was measured as: above canopy PAR– below canopy PAR/above PAR. For the PAR measurement, one above-canopy measurement of total PAR was recorded and three below-canopy measurements were taken in each plot. Due to the width of the plot, the AccuPAR probe was placed parallel to the plot to cover the whole sensor.

4.3.4 Plant Branching

As more branches means more light is intercepted, plant branching is closely related to canopy traits. Plant branching was scored based on the description in Fig. 4.1. while the plant height was measured. More specifically, if the plant branched from the central part of the stem, it was scored as a two. If the plant branched only near the top or all the way from the bottom, it was scored as a six.

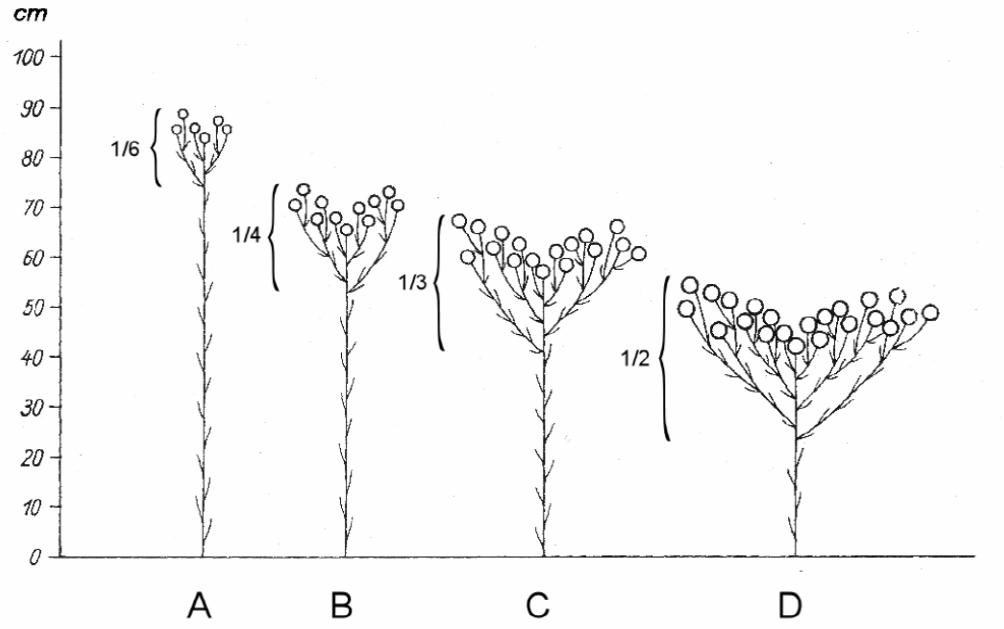

Fig. 4.1. Plant branching score as a fraction of the stem with side branches. Basal branches are not considered. **(**Used with permission from Diederichsen and Richards, 2003, Diederichsen and Fu, 2006. The original figure was adapted from Kulpa and Danert, 1962.).

4.3.5 Genotyping

Leaf tissues (100 mg) from one month old single plants were snap-frozen in liquid nitrogen and lyophilized. Genomic DNA was extracted with the DNeasy 96 plant kit (Qiagen, Missisauga, ON, Canada) according to the manufacturer's instructions, quantified by fluorometry and diluted to a final concentration of 6 ng/µL.

Amplification and analysis of SSRs were performed according to the method of Cloutier *et al.* (2009). A total of 30 ng of genomic DNA from each of the 384 flax core collection accessions was used as template for SSR amplification by polymerase chain reaction (PCR). A final volume of 10 µL per reaction. PCR recipes and program were the same as Huang *et al.* (2006) (5 min at 94°C, followed by 35 cycles with 30 sec at 94°C, 30 sec at 60°C, 90 sec at 72°C and a final extension step

of 10 min at 72°C). The product from single FAM-labeled, HEX-labeled and NED-labeled reactions (2 µL of each) were pooled and mixed with 24 µL of water. An aliquot of 2 µL of the pooled products were mixed with 3.9 µL of Hi-Di formamide and 0.1 µL of Genescan ROX-500 standard (Applied Biosystem, Foster City, CA), denatured 5 min at 95°C and chilled on ice for 5 min before being resolved on an ABI 3100 or 3130xl DNA analyzer (Applied Biosystems). Output files were analyzed by GeneScan (Applied Biosystems) and subsequently imported into Genographer or, alternatively, the ‘‘.fsa’’ files were directly imported into Genographer (Benham *et al*., 1999) as modified by T. Banks for SSR data resolved on ABI DNA analyzers (http://sourceforge.net/projects/genographer). The three labeled reactions were transformed into independent gel-like images. Fragment sizes were estimated using the GeneScan ROX-500 internal size standard and recorded for each accession. Amplicons larger than 450 bp were rerun on the ABI analyzer using the MapMarker®1000 (BioVentures Inc., Murfreesboro, TN) internal size standard. The marker information for the flax core collection was kindly provided by Dr. Sylvie Cloutier (Agriculture, Agri-Food, Canada, and Winnipeg, MB, Canada).

4.4 Data Analysis

4.4.1 Phenotypic Data

Analysis of variance was performed according to the non-replicated modified augmented design (MAD) to use control plots to adjust for environmental heterogeneity (Lin and Poushinky, 1985). The statistical software Agrobase Gen 11 (Agronomix software Inc., Winnipeg Manitoba, Canada) was used. All the phenotypic data were adjusted using the control plots through row-column (method 1) or covariance (method 3) (May *et al*. 1989). The method to compare the adjustment methods is to look at the relative efficiencies for Method 1 and Method 3. A higher relative efficiency means a greater reduction in experimental error after adjustments. Therefore, the method with the greatest relative efficiency was selected. The

association analysis was conducted based on the phenotypic data in 2010 and 2011 separately.

4.4.2 Analysis of Population Structure and Kinship

Population structure was calculated using the software package STRUCTURE (Pritchard and Wen 2004, http://pritch.bsd.uchicago.edu/software.html) in its revised version 2.2 (Falush *et al*., 2003, 2007). The program was run initially by setting K (the number of group in a population) from 1 to 10 using an admixture model, with a 10,000 burn-in time and 100,000 iterations of Markov chain convergence for each run. Every run used five independent replications for each K. Delta K was plotted to identify the most likely number of clusters (K). Delta K is an ad hoc quantity related to the second order change of the log probability of data with respect to the number of clusters inferred by STRUCTURE (Evanno *et al*., 2005).

The kinship coefficient matrix (K matrix) was estimated using software SPAGeDi (Hardy and Vekemans, 2002). Kinship coefficients lower than zero was set to zero and values of "2" were added between same individuals.

In order to visualize the distribution of individual genotypes using the complete SSR dataset, a principal coordinate analysis (PCoA) was conducted using GenAlEx (Peakall and Smouse, 2007).

4.4.3 Association analysis

Association between markers and leaf area index (LAI) and plant branching were tested in TASSEL (trait analysis by association, evolution and linkage) version 2.1 as described by Yu *et al*. (2006). Three different models were applied in the association analysis.

1. The naïve model, the simplest model to explain phenotypic variation with marker alleles.

2. The general linear model (GLM) which includes the population structure (Q matrix) as fixed covariates.
3. The mixed linear model (MLM) which includes population structure (Q matrix) as a fixed effect and the K matrix of pairwise kinship coefficients (K) as a random effect.

To assess the impact of population structure and kinship in these models, cumulative distributions of P-values for both models were calculated and compared to the naïve model across all loci. Distributions were summarized across canopy traits and plant branching.

P-values were adjusted using the Bonferroni correction. Associations between marker and trait were considered significant where $p \leq 0.05$/ marker number.

4.5 Results

4.5.1 Phenotypic variation

The frequency distributions of plant branching and leaf area index were different across years (Fig. 4.2). Wide variation in plant branching for accessions was exhibited in the 2010 growing season, as the distribution looks flattened (Fig. 4.2). Variation of plant branching in 2011 growing season showed a normal distribution. This result showed the impact of environment on plant branching in flax. Most accessions in the flax core collection had scores of 3 and 4 for plant branching in season 2011 (Fig. 4.2). For LAI, growing season 2011 exhibited higher values than in 2010. In both years, the LAI in flax core collection showed a normal distribution. The most frequent value of leaf area index was 3 in 2010 and 4 in 2011. The check cultivar CDC Bethune had a LAI of 4 in the 2011 and 2012 growing seasons.

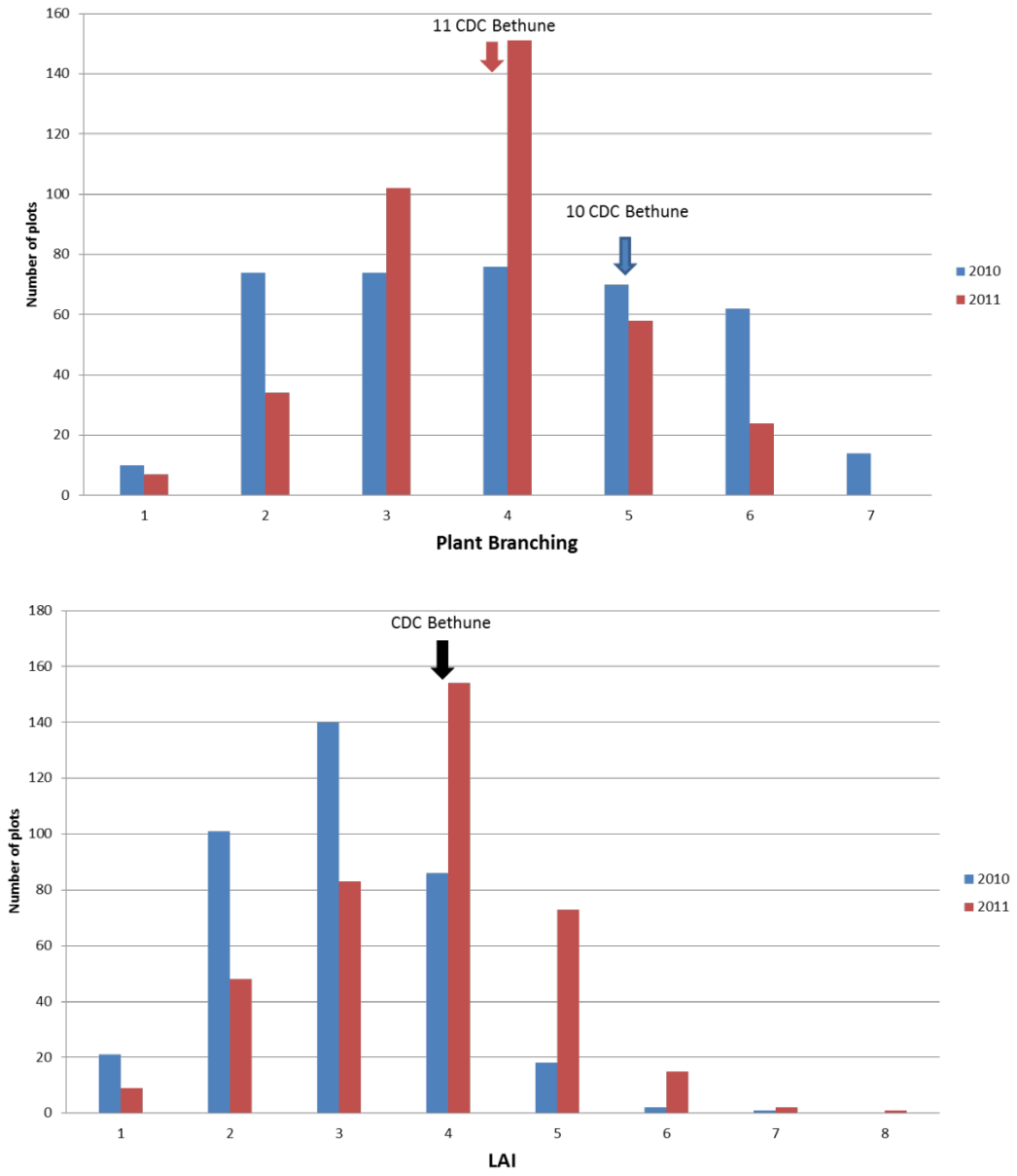

Fig. 4.2. Frequency distribution of plant branching and leaf area index in the flax core collection in growing seasons 2010 and 2011. The control cultivar CDC Bethune was indicated.

4.5.2 Population Structure and Kinship

A total of 375 SSR markers were used to study the population structure. All 375 SSR markers had at least one polymorphism within the 381 flax core collection lines examined. The 381 accessions were assigned to four clusters in the STRUCTURE analysis model as plotting Delta K against each K showed a clear peak in the value of

Delta K at four. Also, there was no change in Delta K after that. The results suggested the flax accessions could be grouped into four subpopulations and this value corresponded to the four major centers of origin or diversity of cultivated flax.

The four subpopulation clusters were denoted as C1, C2, C3 and C4 (Fig. 4.3), consisting of 74, 55, 119, 113 accessions respectively. Most American accessions were grouped into C1 and had previously been assigned to the intermediate flax group (convar. *usitatissimum*). Most Asian accessions were grouped into C2 and are mainly the intermediate group. Most Mediterranean accessions were grouped into C3 and all the five dehiscent flax (convar. *crepitans*) were included in this cluster. Finally, most Europe accessions were captured into C4 and it contained the most fibre flax (convar. *elongatum*) (Table 4.1).

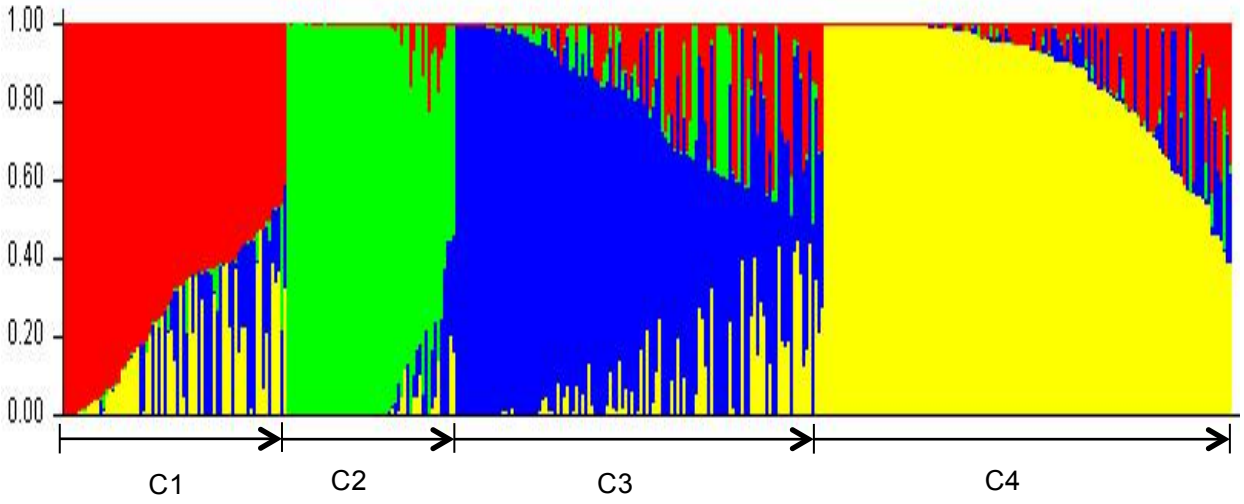

Fig. 4.3. Identification of population structure in 381 flax accessions based on 375 SSR markers. Each accession is represented by a thin vertical line. The accessions can be partitioned into K=4 colored segments that represent the estimated membership probabilities (Q) of the individual to the clusters. Each accession was assigned to a cluster if the estimated membership probability is > 0.5. If an accession could be assigned to more than one cluster, it was placed in the cluster with the highest membership probability.

Table 4.1. Composition of subpopulations C1-C4, with number of accessions per origin and each convariety.

Subpopulation	Total no. accessions	Origin				Infraspecific groups (convarieties)			
		Asian	Mediterranean	Europe	American	*Usitatissimum*	*Mediterraneum*	*Elongatum*	*Crepitans*
C1	74	2	2	10	60	61	2	11	0
C2	55	37	1	5	12	49	6	0	0
C3	119	23	23	42	31	90	15	9	5
C4	133	11	8	83	31	71	0	62	0

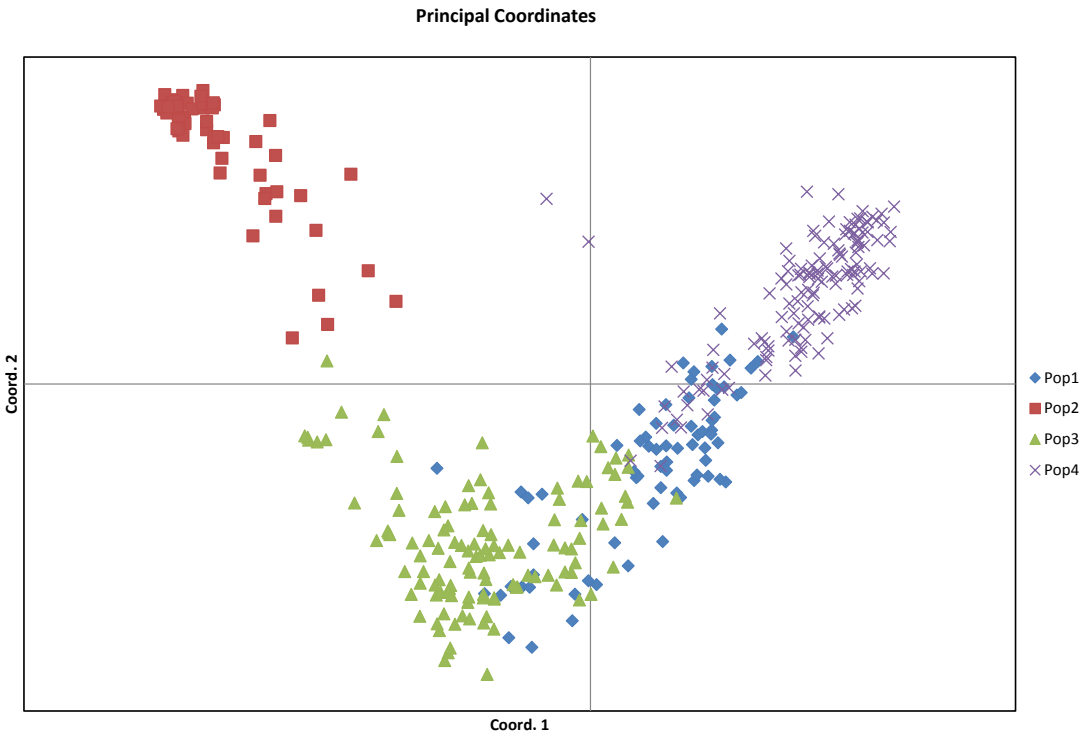

Fig. 4.4. Principal coordinate analysis (PCoA) of 381 flax core accessions. The different colors represent the four subpopulations inferred by structure analysis.

Principal coordinate analysis (PCoA) of 381 flax core accessions was conducted to give a visual pattern of genetic relationship among the flax core collection. The different colors represented the four subpopulations inferred by STRUCTURE

analysis (Fig. 4.4). It was clear there were four different genetic groups in this core collection and that subpopulation C2 was separated from the rest.

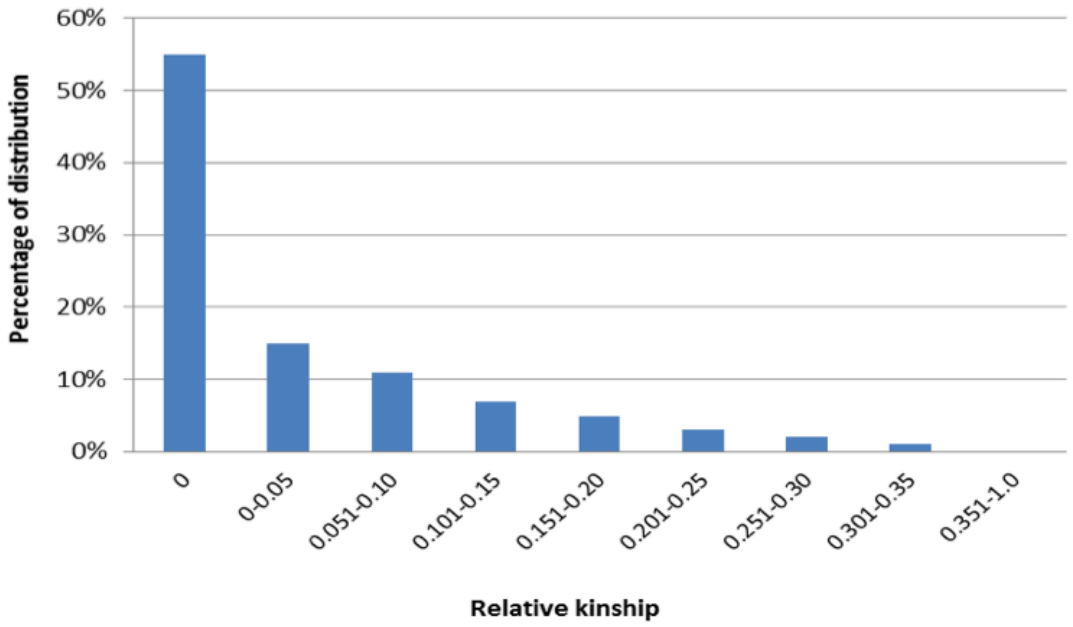

Fig. 4.5. The distributions of pair-wise kinship coefficients for 381 diverse flax accessions.

There was very little relative kinship (K) in the core collection (Fig. 4.5). Approximately 70% of the pair-wise kinship estimates ranged from 0 to 0.05. More than 20% of the pair-wise kinship estimates were from around 0.05 to 0.2, indicating some familial relationships. Less than 10% of the pair-wise kinship was from 0.2 to 0.35, representing strong relationships.

4.5.3 Model comparison and marker-trait association

Mean leaf area index and plant branching for each accession was measured. Association mapping was done through three different models based on the mean traits across two years. The cumulative distributions of P values for both traits were plotted (Fig. 4.6). The blue curves correspond to naïve tests of association without correction for population structure. The cumulative distribution of P values for the Q and Q + K models were strongly skewed towards significance compared to the naïve

model. The Q + K model corrected the skew to a greater degree than the Q model, indicating that the former model has a lower Type I error probability (lower chance of false positives). Therefore, the third model (using Q+K in a MLM) was adopted for association mapping.

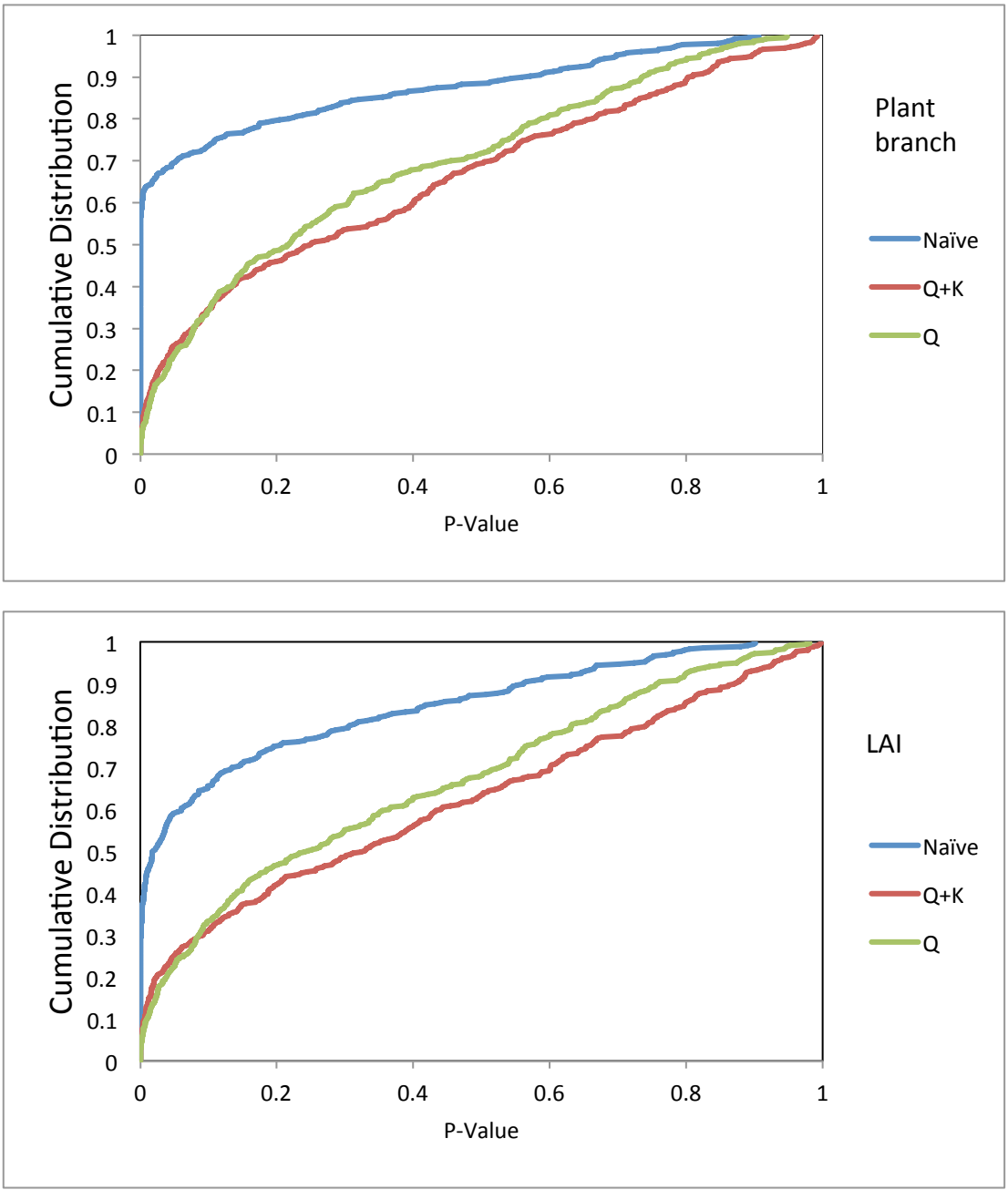

Fig. 4.6. Cumulative distribution of P values for three different models. The naïve model without structure control is compared to GLM (Q) and MLM (Q+K). Distributions were summarized across LAI (bottom) and plant branching (top).

Association mapping analysis was conducted separately for the 2010 and 2011 data due to the large differences observed in branching number and LAI (Fig 4.2). Twenty-six SSR markers linked to plant branching and LAI were detected at the $p \leq 0.05$ probability level with Bonferroni correction in the flax core collection. In 2010, two markers were identified to be associated with plant branching ($p \leq 0.05$). In 2011, there were 14 significant markers associated with plant branching (Table 4.2). For LAI, a total of 18 markers were identified to be significant in both years. There were 7 markers in 2010 and 11 markers in 2011 (Table 4.2). However, there were no markers that were associated with plant branching in both years. Similarly, no markers significantly associated with LAI were observed in both 2010 and 2011.

The corresponding linkage group of each marker was shown in Table 4.2. It is interesting to find that several markers were significantly associated with both plant branching and LAI over both years (Lu125, Lu140, Lu2044, Lu2047, Lu2313, Lu2773, Lu2055, Lu3189). Both marker Lu140 and marker Lu2773 showed strong association with plant branching and LAI in 2011. These two markers were separated by ~26 cM on linkage group (LG) 12 (Cloutier *et al*, 2012). Markers Lu2044 and Lu2047 were adjacent on LG 3, separated by less than 1 cM. In addition, markers Lu2313 and Lu3189, separated by ~14 cM on LG 8, were associated with plant branching and LAI. Thus it was most likely that QTLs for plant branching and LAI are located on LGs 3, 8 and 12. The results indicated that some correlation might be found between plant branching and LAI. Other markers to consider in this part were Lu447, Lu2103 and Lu2317 on LG 8. They were all located between Lu2313 and Lu3189 and significant in association analysis. Marker Lu2991 was 3 cM away from Lu140 on LG 12. Lu638 and Lu342 were also significant in association analysis and located on LG 3, but a little further away from 2044 and 2047. A minor QTL might be Lu125 and Lu532, both on LG 2, but 26 cM apart. A detail integrated consensus genetic and physical maps of flax can be found in Cloutier et al. (2012).

Table 4.2. Association of SSR markers with plant branching and leaf area index (LAI) in 2010 and 2011 for the flax core accession.

Trait	Marker	Linkage group	cM	2010	2011
Plant branching	Lu2103	8	64.9	***	
	Lu805	13	46.5	*	
	Lu2055	1	92		***
	Lu532	2	59.8		***
	Lu125	2	85.9		***
	Lu2044	3	74.3		**
	Lu2047	3	74.4		***
	Lu638	3	100.1		***
	Lu505	5	30.9		***
	Lu2313	8	73.5		**
	Lu3189	8	87.9		*
	Lu458	10	42.7		***
	Lu140	12	13.4		***
	Lu2991	12	16.6		***
	Lu2773	12	39.7		***
	Lu462a	15	10.9		***
Leaf area index	Lu342	3	0	***	
	Lu176	5	34.2	**	
	Lu2103	8	64.9	***	
	Lu2317	8	74.8	*	
	Lu447	8	76.5	*	
	Lu785	11	55.8	***	
	Lu514	14	26.9	**	
	Lu2055	1	92		***
	Lu532	2	59.8		***
	Lu125	2	85.9		***
	Lu2044	3	74.3		*
	Lu2047	3	74.4		***
	Lu638	3	100.1		***
	Lu2313	8	73.5		**
	Lu3189	8	87.9		*
	Lu140	12	13.4		***
	Lu2991	12	16.6		***
	Lu2773	12	39.7		***

*** P<0.001/375, Strong association

** 0.001/375<P<0.01/375, Moderate association

* 0.01/375<P<0.05/375, Weak association

Table 4.3. Monthly growing season precipitation (mm) received at the Kernen Crop Science Research Farm in 2010 and 2011. The 30-year average is presented for comparison.

Year	Seeding date	Harvest date	May	June	July	August	Sept	TOTAL
2010	June 14th	Oct.17th	120	150	91	58	100	519
2011	May 24th	Sep.19th	26	119	96	37	10	388
30 year average			42	71	61	38	30	242

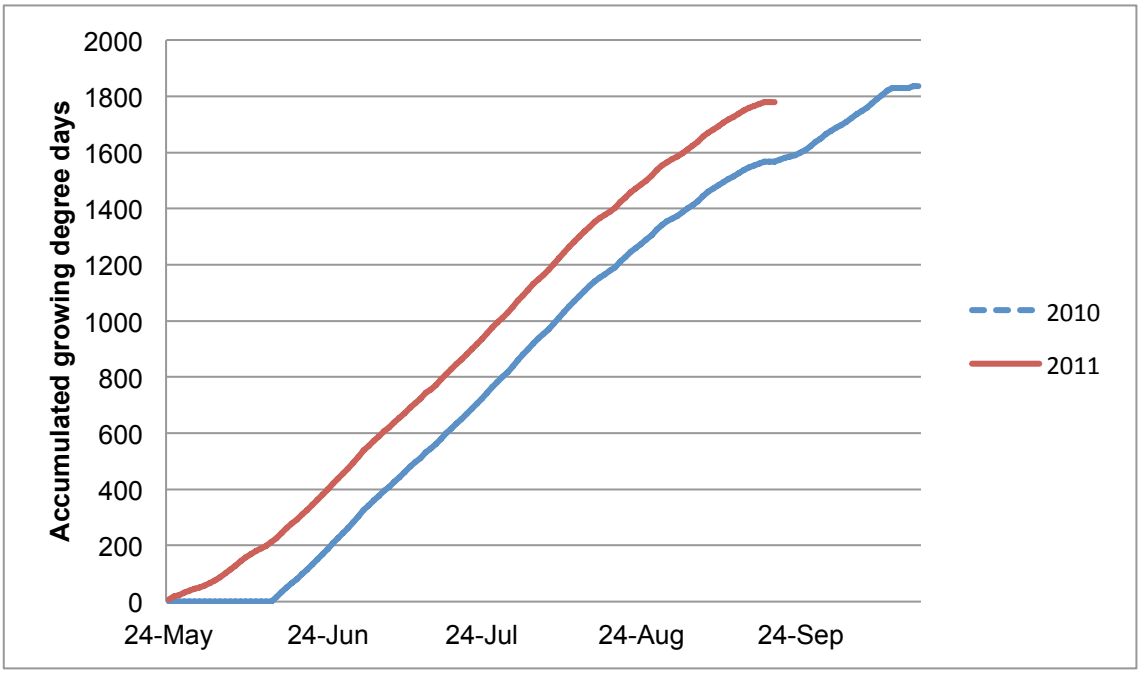

Fig. 4.7. Accumulated growing degree days received at Kernen Crops Research Farm in 2010 and 2011 from seeding to harvest.

4.6 Discussion

The flax core collection was well suited good population for association mapping analysis, because great phenotypic variation among accessions for all the traits

evaluated was observed in this core collection. Challenges exist when assessing complex traits such as branch number and LAI due to the interaction between environment and genes, as observed with the 2010 and 2011 data. The same accessions within the core collection often showed different phenotypic values in 2010 and 2011. Due to the different environmental conditions in the two years, a consistent distribution of plant branch and leaf area index was not observed between years.

Population structure could be a source of Type I error in association mapping analysis, so it is important to correct the analysis for bias introduced by population structure. In this study, the 381 flax accessions were divided into four groups based on the STRUCTURE analysis. Principal coordinate analysis (PCoA) showed a similar result regarding population structure in this collection. The results obtained from these two separate analyses support each other. The STRUCTURE classification agreed with a previous flax genetic study (Diederichsen and Fu, 2008). Moreover, the four group population structure was observed to be related to geographic origins. The distinct geographic origins reflect differences in ecological environments and could partially explain the morphological characterization into different convarieties.

In the association analysis, three models were examined. The cumulative distributions of p-values of each model were compared to find the curve which was the least skewed towards significance. The Q and Q + K models showed a similar trend in reducing the skew towards significance, possibly due to the low degree of kinship observed between accessions in this core collection. The mixed linear model approach is particularly useful in reducing the spurious associations resulting from population structures (Zhao *et al*., 2007). In addition, determination of the kinship matrix (K) improves the reliability of association mapping (Zhao *et al*., 2007). Therefore, the mixed linear model, which included Q and K, was used here for association mapping.

Based on genome wide association analysis with 375 SSR loci and the separate year phenotypic traits, a total of 26 SSR markers were identified as being associated with plant branching and leaf area index at the $p \leq 0.05$ probability level with the Bonferroni correction. There were no markers that showed consistent association across both years; however, there were markers that showed association between traits. These associations could be due to the complexity and interrelatedness of the two traits measured. Plant branching and leaf area index were both impacted by crop development. Their phenotypic expression could be affected by environmental effects or other correlated traits such as plant height. Great environmental variation was observed between 2010 and 2011 (Table 4.3, Fig. 4.7). In the 2010 growing season, the field received a much higher amount of rainfall. In addition, fewer accumulated GDD were recorded in 2010 compared to 2011. These weather conditions would have slowed crop development and significantly impacted canopy traits (Fig. 4.2). Additionally, due to wet conditions seeding was delayed in 2010 (Table 4.3). Seeding date has been shown to impact crop development in flax (Lafond *et al*., 2008).

This study demonstrated that determining consistent marker-trait associations with complex traits, such as plant branching and LAI, requires replicated multi-location multi-year experiments. Another way to find marker associations with these complex traits is to study their correlation with other architectural traits such as plant height. Once a strong correlation between two traits is confirmed, significant marker associations with this other trait can be determined and used in a screening program. Correlation between related traits and yield in flax were analyzed through structural equation modeling in chapter five.

Association mapping is a powerful tool for identifying the association between DNA markers and phenotypic traits (Zhao *et al.*, 2007). This study showed that canopy traits were highly influenced by environment, making the identification of markers associated with these traits challenging.

CHAPTER 5

5. The Relationship between Crop Phenology, Canopy, Branching Characteristics and Seed Yield of the Core Collection through Structural Equation Modeling

5.1 Abstract

Flax seed yield is a complex characteristic and is a result of the inter-relationship between many plant traits. SEM is an extremely flexible statistical method used for determining the relationships between variables. Observed and composite variables (crop phenology, canopy traits and morphological traits) in structural equation models were used to determine how these crop characteristics relate to seed yield. The results indicated greater plant stand resulted in higher absorbed irradiance and higher absorbed irradiance resulted in greater seed yield. Days to flowering had a significant negative effect on seed yield and growing degree days to maturity had a significant effect on seed yield. Plant branching and plant height had a positive non-linear effect on seed yield.

5.2 Introduction

Flax (*Linum usitatissimum* L.) is grown in Western Canada mainly for its seed. Yield improvement is always a major aim in a breeding program; however, yield is a complex trait and is a result of the inter-relationship between many plant characters. To be precise, flax yield is related to plant density, number of capsules per plant and weight of seed per capsule (Copur *et al*., 2006). Yield is a polygenic trait and is also greatly affected by environment. Yield related genes are affected to a different degree by environment in any given year; therefore, selection based on yield alone is not effective over a longer period (Tadesse *et al*., 2009). Ford (1964) illustrated that

indirect selection through yield components in flax is more effective.

The growth of a crop is determined by its ability to capture light and the efficiency of conversion of intercepted light into biomass (Confalone *et al*., 2010). Particularly, the dry matter accumulation of a crop can be estimated as the outcome of three terms: the incident photosynthetically active radiation (PAR) per unit of soil surface, the proportion of PAR intercepted by the crop (PAR interception efficiency) and the production of dry matter per unit of PAR intercepted (PAR use efficiency). A method for analyzing crop growth based on these three terms was studied by Monteith (1994) and some simple crop models were developed. The PAR interception efficiency is affected by the leaf area of the plant population, the leaf structure and inclination of the canopy (Bergamaschi *et al*., 2010).

Considerable controversy has arisen concerning the validity of a single relationship between solar radiation intercepted and crop growth. Kiniry *et al*. (1989) suggested that there is a linear relationship between seasonal biomass accumulation and cumulative intercepted radiation in grain crops. A linear relationship has been found in potato between tuber yield and leaf area index when the leaf area index value was above three (Bremner and Radley, 1966). Other studies have found that significant linear yield losses occurred in soybean when the leaf area index value was reduced to below 3.5-4.0 by manual defoliation (Malone *et al.*, 2002). The effects of canopy traits on flax yield have not been adequately documented.

Previous research on flax has shown a positive relationship between seed yield, number of capsules, number of branches and 1000 seed weight (Chandra, 1977). A positive relationship was also found between seed yield and plant height (Copur *et al.*, 2006). One thousand seed weight, plant height, number of capsules and number of primary branches have a direct effect on seed yield (Copur *et al.*, 2006). Number of bolls is most highly correlated to seed yield, followed by number of primary branches, 1000 seed weight and plant height (Copur *et al.*, 2006).

Yield is a quantitative trait and is highly influenced by both genotype and environmental conditions. It is therefore important to understand the environmental effects on yield. The seed yield of linseed varies due to different weather patterns and soil types; whereas, plant density has little effect on seed yield (Casa *et al.*, 1999). Casa *et al*. (1999) found that flax compensates for reduced stand densities mainly through increasing the number of capsules per plant. Temperature affects the rate of crop development in a positive manner; however, excessively high temperatures during flowering limits flax seed production due to reduced seed and boll number (Cross *et al.*, 2003).

Path analysis has frequently been used to study the relationship between crop yield and yield components (Dewey and Lu, 1959). Structural equation modeling (SEM) is the method developed from path analysis. Recent examples of this type of analysis in plant sciences can be found in Guillen-Portal et al. (2006) and Lamb *et al.* (2011). The applications of SEM to crop science have proven to be a useful and powerful tool to understand the relationship between yield and yield components (Grace, 2006; Lamb *et al*., 2011). The hypothesis tested in these studies is that crop phenology, canopy traits and plant branching will influence seed yield in flax. More specifically, that higher irradiance absorption will be positively correlated with yield potential of flax accessions. However, there is no evidence to conclude that SEM was useful for looking at canopy traits. The reason why I used SEM is that it allows for evaluation of model fits that permit the application of a causally structured theory. SEM helps us to incorporate this theory hypothesis information into the multiple regressions. The objective of this study was to use SEM to determine the relationship of crop phenology, canopy traits and morphological traits in flax and how these crop characteristics relate to seed yield.

5.3 Material and Methods

5.3.1 Experiment material and design

See section 4.3.1 and 4.3.2 for experimental set up. The experimental plots from 2011 were used. Data collected for this experiment were: plant height, plant branching, days to flowering, growing degree days to maturity, irradiance absorption and seed yield. Another important traits measured here was plant stand which could be treated as a early season vigor. Flax normally takes about 5 days to emerge after seeding. Plant stand was measured 10-15 days after seeding. The scale used is shown below:

Plant Stand: 1= 10% of row had adequate plant stand
5= 50% of row had adequate plant stand
10= 100% of row had adequate plant stand.

5.3.2 SEM software

Lavaan (latent variable analysis) is a free, open source package for SEM implemented in the R system for statistical computing (R Development Core Team 2012). This statistical software is used to estimate a large variety of multivariate statistical models, including path analysis, confirmatory factor analysis, SEM and growth curve models (Rosseel, 2012).

5.3.3 The initial model development

To determine the relationships among experimental variables, a data screening test was applied. It is often found that plant height has a humped shape effect on yield – i.e. the highest yields were produced by plants with intermediate height (Fig. 5.1a). Similarly, plant branching had the same effects (Fig. 5.1b). To address this issue a composite variable in structure equation model was introduced (personnel communication, E. Lamb).

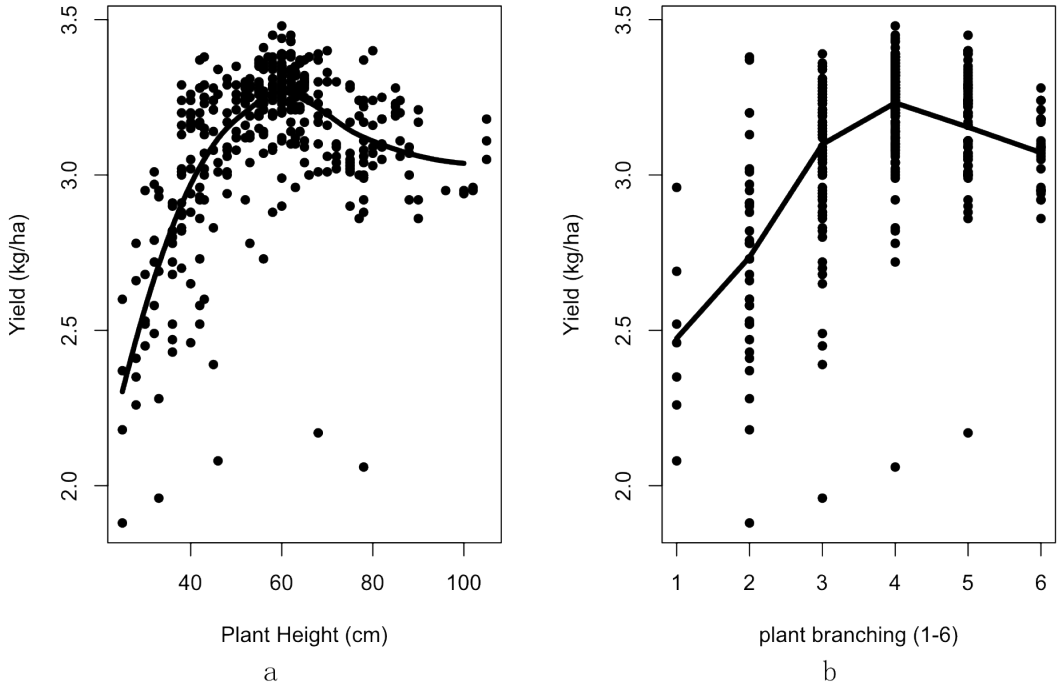

Fig. 5.1. A hump shaped relationship between plant height and yield (a) and a hump relationship between plant branching and yield (b) for the flax core collection in 2011.

An initial path model was developed for the flax core collection (Fig. 5.2). Plant branching and plant height were treated as two composite variables. All variables were assumed to have a direct causal path to yield. Plant stand indicated early season vigor of the crop and was assumed to have a direct causal path to yield. The phenology trait days to flowering was added into the model based on this hypotheses. Irradiance absorption known as canopy absorption showed a direct relationship to yield. Finally, growing degree days to maturity (GDDM) is the sum of GDD from planting to maturity and was also assumed to have a direct causal path to yield. The plot yield was ln-transformed to linearize the relationship with the other variables.

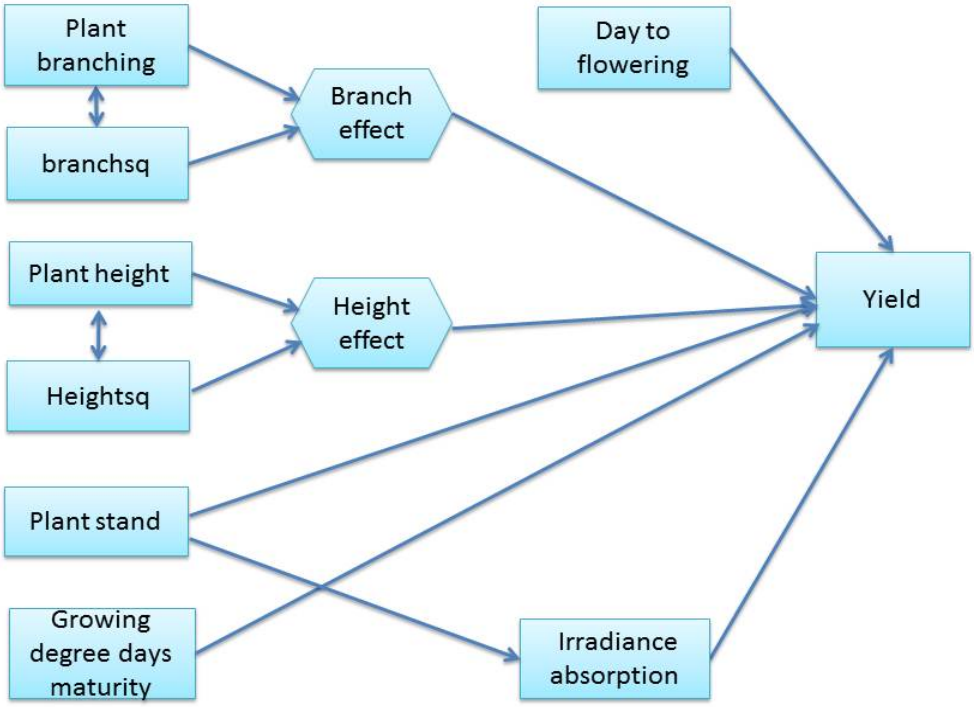

Fig. 5.2. An initial path model. Rectangles are used to indicate observed variables. Hexagons are used to indicate composite variables. Single-headed arrows indicate a causal relationship where a change in the variable at the tail is a direct cause of changes in the variable at the head. Double-headed arrows indicate an unresolved covariance between two variables. (χ_1^2=36.877 and P<0.001).

5.3.4 Composite Variables

Composite variables are a useful extension of the SEM framework. It has received more and more attention due to the limitations of latent variables (Grace, 2006; Grace and Bollen, 2008). The creation of a composite variable implies combining different variables together into a single causal effect and evaluating the relative contribution of each of those variables within a model. Plant branching and plant height were represented as composite variables. Here, plant height was squared to create as new variable, then transformed to a new composite model (Fig. 5.3). To capture that relationship with a composite variable, one would have a 'height effect' affected by plant height and height square. This height effect would then have a direct effect to yield (Fig. 5.3). The same creation principle was adopted for the

variable plant branching.

Composite variables are useful for collecting information about multiple aspects as a single effect. Their use is particularly powerful when representing nonlinear relationships, incorporating complex treatment effects and adding degrees of freedom for model fitting. Both latent and composite variables represent unmeasured quantities, but there is no error variance in the composite variable. Furthermore, a composite variable is associated with "indicators" actually driving the variable, rather than having the unmeasured variable causing the expression of its indicator (Grace, 2006).

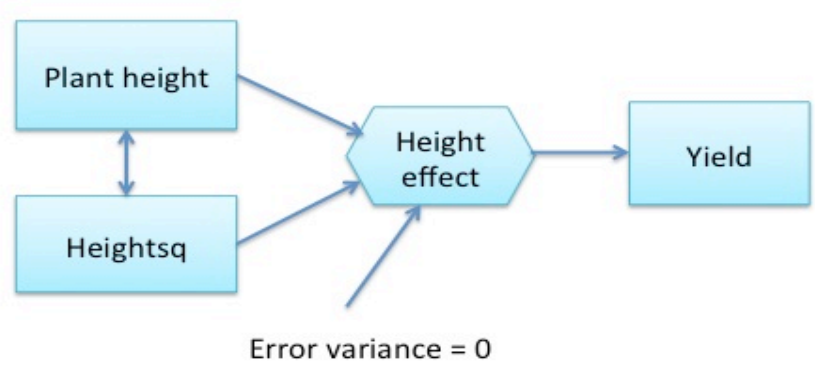

Fig. 5.3. A Height Composite Model. Response variables act like latent variable indicators.

5.3.5 Evaluating Model and modification

Model evaluation and modification is necessary after development of the initial model and typically uses the paths from the initial model. Structural equation models were evaluated by assessing the fit between expectations and data, or the comparative fit between alternative models. The method used to test the model fit was the chi-square test. The amount of difference between expected and observed covariance matrices was indicated by a chi-square value. Little difference was

indicated when a chi-square value close to zero, which indicated a better model fit.

The models in this study were fitted using the lavaan software package (Rosseel, 2012). The overall chi-square value (χ^2) was determined as the first step of evaluating the model. The χ^2 provides the primary information about measure of overall model fit. The initial model (Fig. 5.2) had a χ_1^2=36.877 and P<0.001. A satisfactory model should have a non-significant χ^2 value (with probability >0.05). Therefore, the model was considered an inadequate fit to the data. The next step taken was to examine the modification indices and their corresponding expected parameter changes. Modification indices provide the information about the improvement in fit when adding an additional path to the model.

Examination of modification indices indicated that several paths should be added to improve model fit, such as, path from height effect to IA, path from branching effect to IA, path from DTF to IA and path from GDDM to IA. Based on biological significance and model modification indexes, all four paths mentioned above were added to the model (Fig 5.4)

The modified model has a χ_2^2=2.394, P=0.302. This evaluation indicated that the composite variable model adequately fitted the dataset. The composite variable model with the new paths was developed (Fig. 5.4). In the composite variable model, both plant branching and plant height were treated as composite variables. Two indicators of composite variables were measured and the composite variables had an error variance of zero.

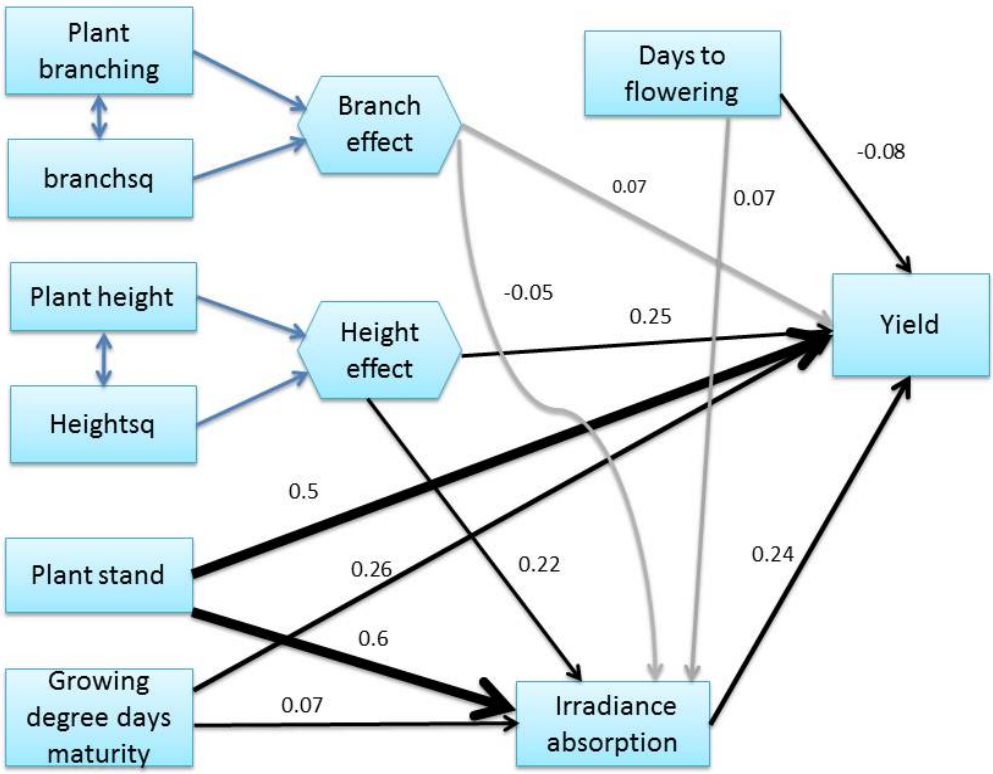

Fig. 5.4. A fitted composite variable model with two composite variables. Rectangles are used to indicate observed variables. Hexagons are used to indicate composite variables. Standardized path coefficients are displayed for significant (P≤0.05) paths. Non-significant paths (P>0.05) are in gray. Path width is proportional to the magnitude of the standardized coefficients. (χ_2^2=2.394, P=0.302).

Unstandardized path coefficients, standardized path coefficient and tests of coefficient significance were shown in Table 5.1. Unstandardized coefficient represents the slope of the relationship (i.e. the mean response), it represented the effect of a change on other variables in absolute terms based on raw data. The standardized coefficient reflected the square root of the variance explained in the response variable in standard deviation units (Grace, 2005). Standardized path coefficient corresponds to the effect-size estimate. Standardized path coefficients with absolute values less than 0.1 could be viewed have a "small" effect, values around 0.3 show a "medium" effect and values greater than 0.5 indicate a "large" effect. Typically, both unstandardized and standardized types of coefficients were

presented due to the different application and interpretation (Grace, 2006). The Z-value followed the Wald test and was used to test the hypothesis of whether the unstandardized coefficient was significantly different from zero. Most of the paths in this model were indicated significant based on the P values shown in Table 5.1. The non-significant paths were still kept according to the biological significance.

Table 5.1. Parameter estimates in the initial structure equation model. Including the unstandardized path coefficients, the standard error (SE), Z-value (based on the Wald test), P-value on test of path coefficient significance and standardized path coefficient. Std. means variables are standardized. Plant stand (PS) and Irradiation Absorption (IA)

Paths	*Unstd. Estimate*	*SE*	*Z-value*	*P*	*Std. Estimate*
Branch → Branch effect	1	-	-	-	4.706
Branch square → Branch effect	-0.138	0.015	-9.416	<0.001	-4.981
Height → Height effect	1	-	-	-	5.160
Height square → Height effect	-0.007	0	-32.647	<0.001	-4.713
PS → Yield	0.076	0.007	10.371	<0.001	0.507
DTF→ Yield	-0.011	0.004	-2.446	0.014	-0.081
IA → Yield	0.005	0.001	5.513	<0.001	0.244
GDDM → Yield	0.001	<0.001	7.216	<0.001	0.262
Branch effect → Yield	0.083	0.047	1.762	0.078	0.073
Height effect → Yield	0.022	0.004	6.166	<0.001	0.254
PS → IA	4.318	0.330	13.097	<0.001	0.603
Branch effect → IA	-2.712	2.440	-1.111	0.266	-0.049
Height effect → IA	0.893	0.182	4.914	<0.001	0.219
GDDM → IA	-0.008	0.004	-2.041	0.041	-0.084
DTF → IA	0.462	0.238	1.941	0.052	0.072

5.4 Agronomic Interpretation

The structural equation models in this study summarized how crop phenology, canopy and branching characteristics affected seed yield in the flax core collection. The significant relationships between two variables in this model were plotted in Fig. 5.5. A curvilinear function line was fitted in the plots. A clear direct significant

positive relationship between plant stand and final seed yield was observed. Plant stand indicated early season vigor in each plot. As plant stand increased so does plant density increased. This confirmed our hypothesis that higher plant stand resulted in higher yield.

The direct positive relationship between irradiance absorption and seed yield was also expected. The irradiance absorption is the PAR absorbed by the crop. The higher irradiance absorption could produce higher crop dry matter. This relationship proved our previous hypothesis that higher absorbed irradiance resulted in higher yield. Of the phenology traits, the days to flowering had a significant negative effect on seed yield. As higher yielding plots were observed in early flowering flax accessions in the flax core collection.

The plant branching characteristic did not show a significant effect on seed yield and irradiance absorption. These paths were non-significant in the composite variables (Fig. 5.4 and Table 5.1). The non-significant paths were kept in the models, as it preserved the original theorized model. A statistically non-significant path may have been the result of insufficient statistical power, but its removal from a model could be theoretically misleading (Grace 2006). A hypothesized relation that was not shown in the current model does not indicate there is no true relationship. Therefore, it is important to conduct further studies to confirm any other relationships added based on modification indices.

For plant height, a significant effect on seed yield was observed, but a non-linear relationship on seed yield was noted. This was because the highest yielding accessions were usually found at a median height plant. Plant stand had a significant positive effect on absorbed irradiance. As initial plant stand increases so does mid-season irradiance absorption.

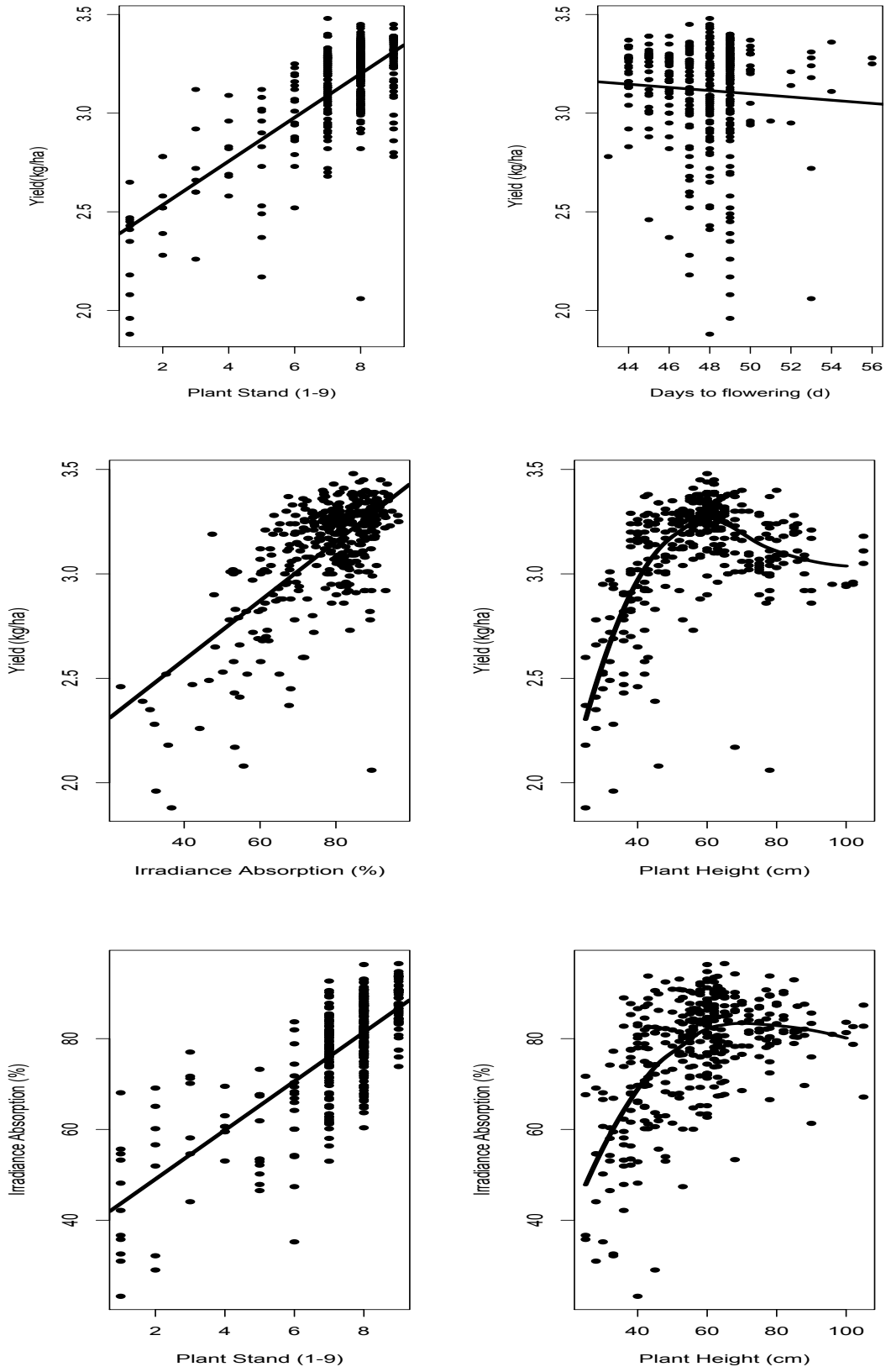

Fig. 5.5. Bivariate scatter plots between variables with significant relationship in the composite model.

5.5 Discussion

It is critical to develop the theoretical models prior to SEM data analysis. The direction of path coefficients indicated effects on yield. For example, early season greater plant stand indicated higher mid-season absorbed irradiance and higher absorbed irradiance indicated higher seed yield (Table 5.1). Therefore, it scoring for early season vigor can select higher irradiance absorption flax types to increase seed yield. The days to flowering had a negative effect on seed yield, which implies selecting early flowering types may also be related to higher seed yield. It is important to introduce earlier flowering coupled with higher seed yield in Western Canada due to the short growing season. Plant height is another important trait identified in this study related to seed yield. The selection criterion of median plant height (60-70 cm) may result in an increase seed yield.

The results of the SEM analysis could serve to support or refute previous research. Previous research using path analysis shows a positive relationship between flax seed yield, number of branches and, plant height (Copur *et al.*, 2006), this was supported by the SEM analysis in this study (Fig. 5.5). The SEM analysis is an extremely flexible statistical method in determining the relationships between variables. Direct as well as indirect relationships between variables can be specified and estimated using SEM.

This study demonstrated that it was feasible to run SEM analysis on a large dataset. SEM is an appropriate analysis tool when you face a challenging research problem that requires the modeling of numerous, complex, interrelated variables. The final model suggested here had a good fit to the analyzed data. SEM was best suited for the purpose to understand the processes or principles underlying the relationships among variables. In particular, SEM is well suited to data arising from observed (measured) and unobserved (latent) variables.

With SEM methodology, models were specified a priori and tested for acceptable fit with chi-square and several fit indices. Statistical support (e.g., change in χ^2 values) was supplemented by a theoretical and biological rationale for adding paths. For example, adding paths among variables not originally hypothesized and statistical supported should be discussed explicitly to balance the model fit. Grace (2006) describes subsequent models after modification as exploratory models. The modified model may no longer act as a priori theoretical hypothesis, but an exploratory analysis that requires additional testing using independent data.

The possible enhancements of seed yield through yield related traits was set as primary target for genetic improvement in oilseed flax. This requires understanding the amount of correlation among various yield contributing characters. In the current study, the phenotypic correlations were analyzed by SEM analysis, which identified that plant stand, plant height and irradiance absorption were highly correlation with seed yield. This type of analysis will assist plant breeders to identify the characters that could serve as selection criteria in oilseed flax breeding programs.

CHAPTER 6

6. General discussion, conclusion and future research

Early maturity is a significant objective of Western Canadian flax breeding programs, especially for the higher northern latitude flax growing region. Earliness ensures timely crop harvest and avoids frost damage affecting seed quality and yield. Photoperiod has been considered the most important factor affecting flowering time in plants. Therefore, it is important to know the photoperiod sensitivity of the genotypes for specific growing locations. Another constraint faced by flax breeders aiming to improve the earliness of flax is the lack of knowledge about the genetic control of these traits. These traits were hypothesized to display quantitative inheritance, which was strongly influenced by environmental factors. Quantitative inheritance is the result of multiple gene action. Characterization and identification of the genetic and environmental control of these traits is important for effective plant breeding.

The present investigation was carried out on flax germplasm representing the flax core collection assembled by Plant Gene Resources of Canada (PGRC). This core collection consists of 381 accessions and represents most of the genetic diversity within the whole PGRC flax collection. Characterization and evaluation of the flax core collection is important as this knowledge can be used to incorporate accessions with key traits into flax breeding programs. The objective of this research was to determine the diversity of the core collection and identify accessions that were photoperiod sensitive or insensitive. This knowledge will help plant breeders improve their strategies to breed for early maturity.

In chapter 3, accessions with the smallest differences between long day and short day photoperiod treatments, accessions CN96992, CN98014, CN97530, CN98286,

CN98468, CN101610 and CN98150, were identified as the least photosensitive lines. Identification of photoperiod insensitive accessions could be used for broadening adaptation of this crop species. Accessions CN98807, CN98370, CN98794, CN98397 and CN98135 were identified as the most photosensitive lines. These accessions were all early flowering types and may have potential for improving earliness if incorporated into a future breeding program selecting for this trait.

It was evident that the growth chamber could be used to assess the response to field conditions related to photoperiod sensitivity. The flowering pattern between the long day controlled environment and field site was similar, because the field site (Lat. 52º09´N; Long. 106º33´) in Saskatoon has summer days reaching about 16 h day length.

This thesis also described the application of molecular markers and association mapping analysis in flax (Chapter 4). One of the objectives of this study was to obtain insight in the structure of the flax core collection. Furthermore, for the first time there some useful markers was identified for future mapping of QTLs towards the objective of making marker-assisted breeding an option in flax breeding for canopy related traits.

By means of STRUCTURE analysis and PCoA cluster analysis, the 381 flax accessions were divided into four groups. Moreover, the groups differentiated were related to geographic origins. Given the clear structured population, it was important to select the right model to conduct the association analysis. False positive associations of markers with traits could happen if a model was selected without considering the impact of population structure or familial relatedness on the traits within the population. This study adopted the mixed linear model (Q+K) for analysis. In a previous study in a structured maize population, the mixed model (Q+K) also had a significant improvement in goodness of fit compared with that of the simple linear, Q, or K models (Yu *et al*., 2006).

Previously, little was known about the association of SSR markers with plant branching and canopy-related traits in flax. 26 SSR markers linked to plant branching and LAI in the flax core collection were identified. It was interesting to see that three pairs of markers, Lu140 and Lu2773, Lu2044 and Lu2047, Lu 2313 and Lu3189, were linked to both traits. The results indicated plant branching and LAI might be closely related to each other. Potential QTLs may be found within these three pairs of markers. The results of this study suggest it is feasible to use SSR markers for a genome-wide association mapping analysis. To increase mapping accuracy of flax core collection, further research is required to select candidate genes underlying traits and to conduct multi-location experiments.

This thesis described the application of SEM analysis to model the relationship between yield and canopy traits. In chapter 5, an ideal structure equation model was fitted to the analyzed data according to the chi-square and several fit indices. SEM analysis of canopy and agronomic traits on seed yield demonstrated that plant stand, DTF, IA, GDDM and plant height, had the most significant direct effects on seed yield. Other traits such as plant branching had non-significant direct effects on seed yield. Most of the results revealed from the SEM confirmed our previous hypothesis. The positive associations of such traits with yield indicated especially that such characteristics could be used for indirect selection for seed yield in oilseed flax. Also, the successful application of SEM confirmed that it was a flexible and powerful statistical methodology useful for the study of relationships between measured variables. For future yield related analysis, SEM will play an important role.

Throughout the chapters in this thesis, the diversity of the flax core collection had been demonstrated. It allowed breeders to better understand the potential of these genetic resources. To ensure long-term sustainability of flax production, the use of characterization and evaluation data would enhance the overall utility of flax core collection. The results of this thesis will assist flax breeding.

References

Ayaz, S., McKenzie, B.A., McNeil, D.L. and Hill, D.G. 2004. Light interception and utilization of four grain legumes sown at different plant populations and depths. Journal of Agricultural Science. 142:297-308.

Benham, J., Jeung, J.U., Jasieniuk, M., Kanazin, V. and Blake, T. 1999. Genographer: a graphical tool for automated fluorescent AFLP and microsatellite analysis. J. Agric. Genom. 4. http://wheat.pw.usda.gov/jag/.

Bergamaschi, H., Dalmago, G.A., Bergonci J.I, Bianchi C.A.M., Heckler B.M.M., Comiran, F. 2010. Intercepted solar radiation by maize crops subjected to different tillage systems and water availability levels. Pesq. Agropec. Bras., Brasilia, 45:1331-1341.

Bremner, P.M. and Radley, R.W., 1966. Studies in potato agronomy. II. The effects of variety and time of planting on growth, development and yield. J. Agric. Set. Camb. 66:253-62.

Brutch, N., Koshkin, V., Matvienko, I., Matvienko, I. *et al.* 2008. Influence of low temperatures and short photoperiod on the time of flowering in flax. International Conference on Flax and Other Bast Plants. ID No. 18.

Buckler, E., Casstevens, T., Bradbury, P. and Zhang, Z. 2009. Trait Analysis by association , Evolution and Linkage (TASSEL): User Manual. Cornell University.

Carder, A.C. 1957. Growth and development of some field crops as influenced by climatic phenomena at two diverse latitudes. Can. J. plant Sci. 37: 392-406.

Casa, R., Russell, G., Lo Cascio, B. and Rossini, F. 1999. Environmental effects on linseed yield and growth of flax at different seed rates. European Journal of Agronomy. 11:267-277.

Chandra, S., 1977. Use of index selection method in improvement of yield linseed (*Linum usitatissimum* L.) Plant Breed. Abs., 47:994.

Churchill, G., Airey, D.C., Allayee, H., Angel, J.M., Attie, A.D. *et al.* 2004. The Collaborative Cross, a community resource for the genetic analysis of complex traits. Nature Genet. 36:1133-1137.

Cloutier, S., Niu, Z., Datla, R. and Duguid, S. 2009. Development and analysis of EST-SSRs for flax (*Linum usitatissimum* L.). Theor. Appl. Genet. 119:53-63.

Cloutier, S., Ragupathy, R., Niu, Z. and Duguid, S. 2010. SSR-based linkage map of flax (*Linum usitatissimum* L.) and mapping of QTLs underlying fatty acid composition traits. Mol. Breeding. doi: 10.1007/s11032-010-9494-1.

Cloutier, S., Ragupathy, R., Miranda, E., Radovanovic, N., Reimer, E., Walichnowski, A., Ward, K., Rowland, G., Duguid, S., Banik, M. 2012. Integrated consensus genetic and physical maps of flax (*Linum usitatissimum* L.). Theor. Appl. Genet. doi: 10.1007/s00122-012-1953-0.

Cockram, J., White, J., Leigh, F.J., Lea, V.J., Chiapparino, E., Laurie, D.A., Mackay, I.J., Powell, W. and O'Sullivan, D.M. 2008. Association mapping of partitioning loci in barley. BMC Genet. 9:16.

Collard, B.C.Y., Jahufer, M.Z.Z., Brouwer, J.B. and Pang, E.C.K. 2005. An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: the basic concepts. Euphytica. 142:169-196.

Confalone, A., Lizaso, J.I., Ruiz-Nogueira, B., Lopez-Cedron F.-X. and Sau, F. 2010. Growth, PAR use efficiency and yield components of field-grown *Vicia faba* L. under different temperature and photoperiod regimes. Field Crops Research. 115 (2): 140-148.

Copur, O., Gur, M.A., Karakus, M. and Demirel, U. 2006. Determination of correlation and path analysis among yield components and seed yield in oil flax varieties (*Linum usitatissimum* L.). J. Biol. Sci. 6:738-743.

Cross, R.H., McKay, S.A.B., McHughen, A.G. and Bonham-Smith, P.C. 2003. Heat stress effects on reproduction and seed set in *Linum usitatissimum* L. (flax). Plant Cell and Environment. 26:1013-1020.

D'Antuono, L.F. and Rossini, F. 2006. Yield potential and ecophysiological traits of the altamurano linseed (Linum usitatissimum L.), a landrace of southern Italy. Genetic Resources and Crop Evolution. 53:65-75.

Dewey, D. R. and Lu, K. H. 1959. A correlation and path- coefficient analysis of components of crested wheatgrass seed production. Agron. J. 51:515-518.

Dewilde, B. 1983. 20 eeuwen vlas in Vlaanderen: 439 pp.

Diederichsen, A., Kusters, P.M., Kessler, D., Bainas, Z. and Gugel, R. K. 2012. Assembling a core collection from the flax world collection maintained by Plant Gene Resources of Canada. Genet. Resour. Crop Evol. Doi:10.1007/s10722-012-9936-1.

Diederichsen, A. 2001. Comparison of genetic diversity of flax (*Linum usitatissimum* L.) between Canadian cultivars and a word collection. Plant Breed. 120:360-362.

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

Diederichsen, A. and Fu, Y.B. 2006. Phenotypic and molecular (RAPD) differentiation of four infraspecific groups of cultivated flax (*Linum usitatissimum* L. subp. *usitatissimum*). Genet. Resour. Crop Evol. 53:77-90.

Diederichsen, A. and Fu, Y.B. 2008. Flax Genetic Diversity as the raw Material for Future Success. Paper presented at the International Conference on Flax and Other Bast Plants July 20-23, 2008, Saskatoon, SK, Canada.

Diederichsen, A. and Raney, J.P. 2006. Seed colour, seed weight and seed oil content in *Linum usitatissimum* accessions held by Plant Gene Resources of Canada. Plant Breed. 125:372-377.

Diederichsen, A., Raney, J.P. and Duguid, S.D. 2006. Variation of mucilage in flax seed and its relationship with other seed characters. Crop Sci. 46:365-371.

Dillman, A.C. 1936. Improvement in flax. Year book of Agriculture. USDA, Washington D.C. pp. 745-784.

Dillman, A.C. 1953. Classification of flax varieties, 1946. USDA, Technical Bulletin No.1054, USDA, Washington.

Dun, E.A., Ferguson, B.J. and Beveridge, C.A. 2006. Apical dominance and shoot branching. Divergent opinions or divergent mechanisms? Plant Physiol. 142:812-819.

Evanno, G., Regnaut, S. and Goudet, J. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol. Ecol. 14: 2611-2620.

Everaert, I., Riek, J.D., Loose, M.D., Waes, J.V. and Bockstaele, E.V. 2001. Most similar variety grouping for distinctness evaluation of flax and linseed (*Linum usitatissimum* L.) varieties by means of AFLP and morphological data. Plant Var. Seeds. 14:69-87.

Falush, D., Stephens, M. and Pritchard, J.K. 2003. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies.

Genetics. 164:1567-1587.

Falush, D., Stephens, M. and Pritchard, J.K. 2007. Inference of population structure using multilocus genotype data: dominant markers and null alleles. Mol. Ecol. Notes 7:574-578.

Flint-Garcia, S.A., Thornsberry, J.M. and Buckler, E.S. 2003. Structure of linkage disequilibrium in plants. Annu. Rev. Plant Biol. 54:357-374.

Ford, J.H. 1964. The influence of time of flowering on seed development of flax. Crop Sci. 4:52-54.

Fu, Y.B. 2005. Geographic patterns of RAPD variations in cultivated flax. Crop Sci. 45:1048-1091.

Fu, Y.B. 2006. Redundancy and distinctness in flax germplasm as revealed by RAPD dissimilarity. Plant Genet. Resour. 4:117-124.

Fu, Y.B., Peterson, G., Diederichsen, A. and Richards, K.W. 2002. RAPD analysis of genetic relationships of seven flax species in the genus *Linum* L. Genet. Res. Crop Evol. 49:253-259.

Fu, Y.B., Rowland, G.G., Duguid, S.D. and Richards, K.W. 2003. RAPD analysis of 54 North American flax cultivars. Crop Sci. 43:1510-1515.

Fu, Y.B. 2011. Genetic evidence for early flax domestication with capsular dehiscence. Genet. Resour. Crop Evol. doi: 10.1007/s10722-010-9650-9.

Grace, J.B. 2006. Structural equation modeling and natural systems. Cambridge University Press, Cambridge, UK.

Grace, J.B. and Bollen, K.A. 2005. Interpreting the results from multiple regression and structural equation models. Bull. Ecol. Soc. Am. 86:283-295.

Grace, J.B. and Bollen, K.A. 2008. Representing general theoretical concepts in structural equation models: the role of composite variables. Environ. Ecol. Stat. 15: 191-213.

Guillen-Portal, F.R., Stougaard, R.N., Xue, Q. and Eskridge, K.M. 2006. Compensatory mechanisms associated with the effect of spring wheat seed size on wild oat competition. Crop Sci. 46:935-945.

Harper, C.R., Edwards, M.J. and DeFilipis, et al. 2006. Flaxseed oil Increases the Plasma Concentrations of Cardioprotective (n-3) Fatty Acids in Humans. J. Nutr.

136:83.

Heath, M.C. and Hebblethwaite, P.D. 1987. Precision drilling combing peas (Pisum sativum L.) of contrasting leaf types at varying densities. J. Agric. Sci. 108: 425-430.

Hedrick, P.W. 1987. Gametic disequilibrium measures: proceed with caution. Genetics. 117:331-341.

Huang, Q., Borner, A., Roder, S. and Ganal, W. 2002. Assessing genetic diversity of wheat (*Triticum aestivum* L.) germplasm using microsatellite markers. Theor. Appl. Genet. 105:699-707.

Huang, X.Q., Cloutier, S., Lycar, L., Radovanovic, N., Humphreys, D.G., Noll, J.S., Somers, D.J. and Brown, P.D. 2006. Molecular detection of QTLs for agronomic and quality traits in a doubled haploid population derived from two Canadian wheat (*Triticum aestivum* L.). Theor. Appl. Genet. 113:753-766.

Jackson, S.D. 2009. Plant responses to photoperiod. New Phytologist. 181(3):517-531.

Jin, L., Lu, Y., Xiao, P., Sun, M., Corke, H. and Bao, J.S. 2010. Genetic diversity and population structure of a diverse set of rice germplasm for association mapping. Theor Appl Genet.121:475–487.

Kenaschuk, E.O. and Rowland, G.G. 1995. Flax. In: A. E. Slinkard and D. R. Knott (eds.), Harvest of Gold: The History of Field Crop Breeding in Canada, 173-176. University of Saskatchewan, Saskatoon.

Kiniry, J.R., Jones, C.A., O'Toole, J.C., Blanchet, R., Cabelguenne, M. and Spanel, D.A., 1989. Radiation-use efficiency in biomass accumulation prior to grain-filling for five grain-crop species. Field Crops Res. 20, 51-64.

Knapp, S.J. 1998. Marker-assisted selection as a strategy for increasing the probability of selecting superior genotypes. Crop Science 38: 1164-1174.

Kulpa, W. and Danert, S. 1962. Zur Systematik von *Linum usitatissimum* L. [On the systematics of *Linum usitatissimum* L.]. Kulturpflanze. Beiheft 3:341-388.

Lafond, G.P., Irvine, R.B., Johnston, A.M., McAndrew, D.W., Shirtliffe, S.J. and Stevenson, F.C. 2008. Impact of agronomic factors on seed yield formation and quality in flax. Canadian Journal of Plant Science, 88(3):485-500.

Lamb, E.G., Shirtliffe, S.J. and May, W.E. 2011. Structural equation modeling in the plant sciences: An example using yield components in oat. Can. J. Plant Sci. 91:

603-619.

Lane, D.R., Coffin, D.P. and Lauenroth, W.K. 2000. Changes in grassland canopy structure across a precipitation gradient. J. Veg. Sci. 11:259-268.

Lawn, R.J. 1989. Agronomic and Physiological Constraints to the Productivity of Tropical Grain Legumes and Prospects for Improvement. Experimental Agriculture 25:509-528.

Lin, C.S. and Poushinsky, G. 1985. A modified augmented design (type 2) for rectangular plots. Canadian Journal of Plant Science 65:743-749.

Mackay, I., Powell, W. 2007. Methods for linkage disequilibrium mapping in crops. Trends Plant Sci. 12:57-63.

Maddock, T.D., Anderson, V.L. and Landry G.P. 2005. Using flax in livestock diets. AS-1283.

Major, D.J. 1980. Photoperiod response characteristics controlling flowering of nine crop species. Can. J. plant Sci. 60:777-784.

Malone, S., Herbert, D.A. and Holshouser, D.L. 2002. Relationship between Leaf Area Index and yield in Double-crop and full-season soybean systems. J. Ecol. Entomol. 95(5):945-951.

Månsby, E., Díaz, O. von and Bothmer, R. 2009. Preliminary study of genetic diversity in Swedish Flax (*Linum usitatissimum*). Gent. Resour. Crop Evol. 47:417-424.

May, K.W., Kozub, G.C. and Schaalje, G.B. 1989. Field evaluation of modified augmented design (Type 2) for screening barley lines. Can. J. Plant Sci. 69:9-15.

Miller, P., Lenier, W. and Brandt, S. 2001. Using growing degree days to predict plant stages. Montana Guide Fact Sheet MT 200103 AG 7/2001. Montana State University Extension Service, Bozeman, Montana.

Monteith, J.L. 1994. Principles of resource capture by crop stands. In: Monteith JL, Scott. R.K., Unsworth, M.H. (Eds.), Resource Capture by Crops. Nottingham University Press, Nottingham, UK, pp.1-15.

Morton, N.E. 2005. Linkage disequilibrium maps and association mapping. J. Clin. Invest. 115:1425-1430.

Myles, S., Peiffer, J., Brown, P.J., Ersoz, E.S., Zhang, Z., Costich, D.E. and

Bukler, E.S. 2009. Association mapping: Critical considerations shift from genotyping to experimental design. The Plant Cell. 21:2194-2202.

Newton, J.E. and Blackman, G.E. 1970. The penetration of solar radiation through leaf canopies of different structure. Annals of Botany. 34:329-348.

Oh, T.J., Gorman, M. and Cullis, C.A. 2002. RFLP and RAPD mapping in flax (*L. usitatissimum*). Theor. Appl. Genet. 101:590-593.

Owen, P.C. 1971. The effects of temperature on the growth and development of rice. Field Crop Abstr. 24:l-8.

Peakall, R. and Smouse, P.E. 2007. GENALEX 6.1: genetic analysis in Excel. Population genetic software for teaching and research. Australian National University, Canberra. Available via http://www.anu.edu.au/BoZo/GenAlEx.

Peleman, J.D. and van der Voort, J.R. 2003. Breeding by Design. Trends in Plant Science 8: 330-334.

Pritchard, J.K., Stephens, M. and Donnelly, P. 2000. Inference of population structure using multilocus genotype data. Genetics. 155:945-959.

Puvirajah, A.S. 2010. Quality of western Canadian flaxseed. Canadian Grain commission. ISSN, 1700-2087.

Ragupathy, R., Rathinavelu, R., Cloutier, S. 2011. Physical mapping and BAC-end sequence analysis provide initial insights into the flax (Linum usitatissimu L.) genome. BMC Genomics. 12:217. doi:10.1186/1471-2164-12-217.

Robertson, M.J., Holland, J.F., Cawley, S., Bambach, R., Cocks, B., Watkinson, A.R. 2001. Phenology of canola cultivars in the northern region and implications for frost risk. 10th Australian Agronomy Conference, Hobart.

Rosseel, Y. 2012. lavaan: An R package for structural equation modeling. Journal of Statistical Software. 48(2): 1-36.

Schaalje, G.B., Lynch, D.R., Kozub, G.C. 1987. Field evaluation of a modified augmented design for early stage selection involving a large number of test lines without replication. Potato Res. 30:35-45.

Sizov, I.A. 1955. Flax. Selhozgiz, Moscow. pp.97-101.

Soto-Cerda, B.J. and Cloutier, S. 2012. Association Mapping in Plant Genomes, Genetic Diversity in Plants, Mahmut Çalişkan (Ed.), ISBN: 978-953-51-0185-7,

InTech, doi: 10.5772/33005.

Stel'makh, A.F., Avsenin, V.I., Kucherov, V.A. and Voronin, A.N., 1987. Study of Genetic Systems Vrn and Ppd in Common Wheat, Voprosy genetiki i selektsii zernovykh kul'tur KOTs SEV (Problems of Genetics and Breeding of Cereals, Coordination Center, Comecon), Odessa (USSR), NIIR Prague–Ruzyne (Czech SSR), issue 3, pp.125-132.

Stich, B. and Melchinger, A.E. 2010. An introduction to association mapping in plants. CAB Reviews: Perspectives in Agriculture, Veterinary Science, Nutrition and Natural Resources. 5, No. 039.

Smýkal, P., Bačová, N., Kalendar, R., Corander, J., Schulman, A. and Pavelek, M. 2011. Genetic diversity of cultivated flax (Linum usitatissimum L.) germplasm assessed by retrotransposon- based markers. Theor. Appl. Genet. 122:1385-1397.

Tadesse, T., Singh, H. and Weyessa, B. 2009. Correlation and Path Coefficient Analysis among Seed Yield Traits and Oil Content in Ethiopian Linseed Germplasm. Int. J. Sustain. Crop Prod. 4:8-16.

Tammes, T. 1928. The genetics of the genus Linum. P. 1-34 in Bibliographia Genetica, Lotsy, J.P. and W.A. Goddijn (eds.). Martinus Nijhoff, The Netherlands.

Vavilov, N.I. 1926. Studies on the origin of cultivated plants. Inst. Bot. Appl. et d' Amelior. des Plantes. State Press, Leningrad. (In Russian and English).

Vavilov, N.I. 1951. The origin, variation, immunity and breeding of cultivated plants. Chron. Bot. 13:1–366.

Vergara, B.S. and Chang, T.T. l976. The flowering response of the rice plant to photoperiod: A review of the literature (3rd ed.). Int. Rice Res. Inst., Tech. Bull. 8.

Vinod, K.K. 2011. Structured association mapping using STRUCTURE and TASSEL. Advanced faculty training on "Impact of genomics in crop improvement: Perceived and achieved", Jan 20 - Feb 9, 2011, Centre for Advanced Faculty Training in Genetics and Plant Breeding, Tamil Nadu Agricultural University, Coimbatore.

Wallace, D.H. 1991. Photoperiod, temperature and genotype interaction effects on days and nodes required for flower of bean. J. Am. Soc. Hortic. Sci. 116:534-543.

Watson, D.J. 1947. Comparative physiological studies in growth of field crops. I. Variation in net assimilation rate and leaf area between species and varieties, and within and between years. Anatomy and Botany. 11:41-76.

Williams, I.H. 1988. The pollination of linseed and flax. Bee World. 69: 145-152.

Yu, J. and Buckler, E.S. 2006. Genetic association mapping and genome organization of maize. Curr. Opin. Biotechno. 17:1-6.

Yu, J., Pressoir, G., Briggs, W.H., Vroh, B.I., Yamasaki, M., Doebley, J.F., McMullen, M.D., Gaut, B.S., Nielsen, D.M., Holland, J.B., Kresovich, S. and Buckler, E.S. 2006. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat. Genet. 38:203-208.

Zhao, K., Aranzana, M.J., Kim, S., Lister, C. and Shindo, C., et al. 2007. An *Arabidopsis* Example of Association Mapping in Structured Samples. PLoS Genet 3(1): e4. doi:10.1371/journal.pgen.0030004.

Zohary, D. 1999. Monophyletic and polyphyletic origin of the crops on which agriculture was formed in the Near East. Genet. Resear. Crop Evol. 46:133-142.

Zohary, D. and Hopf, M. 2000. Domestication of plants in the old world: the origin and spread of cultivated plants in West Asia, Europe and the Nile Valley. 3rd ed. Oxford Univ. Press.

Appendices

Appendix 1. SAS code for ANOVA analysis the chamber experiment data.

```
PROC MIXED Data=chamber ratio covtest cl;

class year block ID MT PP;

model DTF= MT PP MT*PP/ddfm=satterth;
random year    year(block)    year*MT*PP;
lsmeans MT PP MT*PP/pdiff;
ods output LSmeans =DTF;
RUN;
```

R code for ANOVA Analysis the canopy data in the field.

```
canopy <- read.table("E://canopy.txt",header=T,sep="\t", quote="")
attach(canopy)
names(canopy)

library(lme4)

model1<-lme(DTF~convarieties,random=~1|block,na.action=na.omit)
anova(model1)

model2<-lme(Maturity~convarieties,random=~1|block,na.action=na.omit)
anova(model2)

model3<-lme(Yield~convarieties,random=~1|block,na.action=na.omit)
anova(model3)

model4<-lme(Height~convarieties,random=~1|block,na.action=na.omit)
anova(model4)

model5<-lme(PB~convarieties,random=~1|block,na.action=na.omit)
anova(model5)

model6<-lme(GDDF~convarieties,random=~1|block,na.action=na.omit)
anova(model6)

model7<-lme(GDDM~convarieties,random=~1|block,na.action=na.omit)
anova(model7)
```

Appendix 2. R code and output for the Structure equation modeling in flax core collection.

```
core11 <- read.table("core11.txt",header=T,sep="\t",quote="")
attach(core11)
library(lavaan)

core11$branchsq <- core11$branch^2
core11$heightsq<- core11$height^2
model.lav6<-'brancheffect ~ 1*branch + branchsq
brancheffect ~~ 0*brancheffect
heighteffect ~ 1*height + heightsq
heighteffect ~~ 0*heighteffect
brancheffect=~0
heighteffect=~0
yield~PS+ia+brancheffect+heighteffect+DTF+GDDM
ia~PS+brancheffect+heighteffect+DTF+GDDM'

fit6 <- sem(model.lav6, data=core11, std.lv=T)
summary(fit6, standardized=TRUE, fit.measures=TRUE)
```

lavaan (0.4-14) converged normally after 99 iterations

Number of observations	390
Estimator	ML
Minimum Function Chi-square	2.394
Degrees of freedom	2
P-value	0.302

Chi-square test baseline model:

Minimum Function Chi-square	829.409
Degrees of freedom	15
P-value	0.000

Full model versus baseline model:

Comparative Fit Index (CFI)	1.000
Tucker-Lewis Index (TLI)	0.996

Loglikelihood and Information Criteria:

Loglikelihood user model (H0)	-10381.404
Loglikelihood unrestricted model (H1)	-10380.207
Number of free parameters	15
Akaike (AIC)	20792.808
Bayesian (BIC)	20852.300
Sample-size adjusted Bayesian (BIC)	20804.706

Root Mean Square Error of Approximation:

RMSEA	0.022	
90 Percent Confidence Interval	0.000	0.106
P-value RMSEA <= 0.05	0.589	

Standardized Root Mean Square Residual:

SRMR	0.006

Parameter estimates:

Information	Expected
Standard Errors	Standard

	Estimate	Std.err	Z-value	P(>|z|)	Std.lv	Std.all
Latent variables:						
brancheffect =~						
brancheffect	0.000				0.000	0.000
heighteffect =~						
heighteffect	0.000				0.000	0.000
Regressions:						
brancheffect ~						
branch	1.000				4.387	4.706
branchsq	-0.138	0.015	-9.416	0.000	-0.605	-4.981
heighteffect ~						
height	1.000				0.326	5.160
heightsq	-0.007	0.000	-32.647	0.000	-0.002	-4.713
yield ~						
PS	0.076	0.007	10.371	0.000	0.076	0.507
ia	0.005	0.001	5.513	0.000	0.005	0.244
brancheffect	0.083	0.047	1.762	0.078	0.019	0.073
heighteffect	0.022	0.004	6.166	0.000	0.066	0.254
DTF	-0.011	0.004	-2.446	0.014	-0.011	-0.081
GDDM	0.001	0.000	7.216	0.000	0.001	0.262

ia ~						
PS	4.318	0.330	13.097	0.000	4.318	0.603
brancheffect	-2.712	2.440	-1.111	0.266	-0.618	-0.049
heighteffect	0.893	0.182	4.914	0.000	2.740	0.219
DTF	0.462	0.238	1.941	0.052	0.462	0.072
GDDM	-0.008	0.004	-2.041	0.041	-0.008	-0.084
Variances:						
brancheffect	0.000				0.000	0.000
heighteffect	0.000				0.000	0.000
yield	0.021	0.001			0.021	0.303
ia	61.707	4.419			61.707	0.395