Genetic analysis of seed and flower colour in flax (*Linum usitatissimum* L.) and identification of a candidate gene in the *D* locus

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 Biology

University of Saskatchewan

Saskatoon

By

Gurudatt Pavagada Sudarshan

PERMISSION TO USE

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

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

Head of the Department of Biology
University of Saskatchewan
112 Science Place
Saskatoon, Saskatchewan S7N 5E2
Canada

ABSTRACT

Flax (*Linum usitatissimum* L.) is a commercial oilseed crop in Canada. Globally flax is known for industrial oil and fiber. Flaxseed contains Omega 3 fatty acid, lignans like secoisolariciresinol diglucoside (SDG), flavonoids and polysaccharides which offer potential health benefits. Conventional flax cultivars are brown seeded and few mutant lines are yellow seeded. The darkness of seed colour depends on the presence of polymerized proanthocyanidins (PA; condensed tannins) in the seed coat. PAs are the product of the phenylpropanoid pathway. Previous genetic studies by Mittapalli and Rowland (2003) on G1186/94 showed the seed colour trait was governed by the homozygous recessive alleles at *D* locus and the same locus is closely linked to white or pink flower petals. To start with, single seeds of already developed stable recombinant inbred lines (RILs) (of $F_{8:9}$ generation) from a cross of yellow seeded European recessive line (G1186/94) and brown seeded CDC Bethune (popular variety) were grown. In this study, seed colour phenotyping was done by measuring seed colour of each RIL in Red-Green-Blue (RGB) values. To understand the genetic basis of flax seed and flower colour, mapping with single sequence repeats (SSRs) and CAPS (Cleaved Amplified Polymorphic Sequences) markers were used. For the first time, a framework genetic linkage map was constructed from populations of CDC Bethune/ G1186/94 containing 19 linkage groups (LGs). LG 1 with four SSR markers was found to be linked with the seed colour locus *D*. During the fine-mapping, two SSR markers (LuM566 and Lu2351) were found to be linked with the seed colour trait. The *D* locus has been confined in a 2.8 cM region and the closest marker was LuM566 at a distance of 0.6 cM. This was observed to be a stable locus in two growth trials and in different environments with logarithm of odds (LOD) above 39 and more than 84 % of the trait expressed by the major locus in both trials. As there were no recombinants (off types) for flower colour in $F_{8:9}$ plants i.e brown-seeded lines produced blue flowers and yellow-seeded lines produced white flowers, the

same locus holds well for the flower colour trait. The marker associated with seed and flower colour in G1186/94 (European recessive yellow line) was identified and can be used in flax breeding. Additionally, an interesting putative candidate gene of potential significance was identified through genomics assisted gene search from the flax whole genome sequence database. The gene expression analyses showed lower expression of putative flavonoid 3' hydroxylase (F3'H) (a gene involved in flavonoid biosynthesis pathway) in both seed coat and flower petal tissues of G1186/94 as compared to CDC Bethune. Therefore, this study represents the first report on genetic mapping based putative candidate gene finding for recessive yellow seed colour mutation in the *D* locus in flax.

ACKNOWLEDGEMENTS

I am indeed fortunate to have started my M.Sc. programme under the influence of my supervisors; ***Dr. Yangdou Wei****, Department of Biology, University of Saskatchewan, Saskatoon and* ***Dr. Gopalan Selvaraj****, National Research Council of Canada, Saskatoon.*

I express my sincere gratitude for the time Dr. Gopalan Selvaraj devoted to this task and for his constructive criticism, inspiring guidance, scientific freedom, meticulous editorial care, unhesitant and unstinted encouragement and financial support extended to me for the successful completion of this work.

I also express my sincere gratitude to Dr. Yangdou Wei for his support, encouragement and discussions throughout my time as a graduate student.

I thank my advisory committee members ***Drs. Arthur R Davis****, Department of Biology, U of S, Saskatoon and* ***Yong-Bi Fu****, Research Scientist, Agriculture and Agri-Food Canada (AAFC), Saskatoon for their co-operation and guidance to successfully complete my course and research work.*

I sincerely acknowledge the research support of Professor Emeritus Gordon G. Rowland, U of S, Saskatoon for providing me with the starting seed material and Dr. Sylvie Cloutier, AAFC, Winnipeg for her help in accomplishing part of my research work by providing molecular resources.

I thank Dr. Manoj Kulkarni for his advice and for sparing his valuable time for discussions which aided the completion of my thesis. I also thank Dr. Paula Ashe for her help in carrying out bioinformatics work and being a resourceful lab member. I really appreciate Dr. Leonid Akhov, Dr. Sravan Jami, Dr. David Greenshields, and past lab members for being great lab members during my research. I especially thank Hamid for being a wonderful

friend and a lab member by helping and supporting me and making my stay in the lab memorable.

I extend my thanks to Dr. Krishna Gali and Dr. Andrew Sharpe for their timely help and technical support in carrying out genotyping experiments.

I acknowledge all of the members of Dept. of Biology, U of S, Saskatoon and National Research Council of Canada, Saskatoon for the generous support in many occasions.

I remember with pleasure the support and blessings of my parents and everyone in my family who always believed in my abilities and constantly encouraged me in all my endeavors.

I am delighted to thank all my friends for their encouragement and constant support all of the time.

Saskatoon

August, 2013
Pavagada Sudarshan, Gurudatt

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS AND ABBREVIATIONS

±	Plus/minus
°C	Degrees Celsius
µL	Microlitre
χ^2	Chi-square test statistics
2n	Diploid number of chromosomes
AFLP	Amplified fragment length polymorphism
ALA	Alpha-linolenic acid
ANOVA	Analysis of variance
BES	BAC (bacterial artificial chromosome) end sequence
BLAST	Basic Local Alignment Search Tool
bp	Base pair
CAPS	Cleaved Amplified Polymorphic Sequences
cDNA	Complementary Deoxyribonucleic acid
CIM	Composite interval mapping
CTAB	Cetyl Trimethyl Ammonium Bromide
cm	Centimetre
cM	Centimorgan
DAF	Days after flowering
DH	Doubled haploid
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleotide triphosphate
EDTA	Ethylenediaminetetraacetic acid
EST	Expressed Sequence Tag
F_1	First filial generation

F_2	Second filial generation
F_3	Third filial generation
$F_{8:9}$	Eighth filial generation plant and ninth filial generation seed
$F_{9:10}$	Ninth filial generation plant and tenth filial generation seed
Fig.	Figure
FPC	Fingerprint contig
g	Gram
GC	Growth chamber
GH	Greenhouse
HDG	Herbacetin diglucoside
KDG	Kaempferol diglucoside
L	Litre
LG	Linkage group
LOD	Logarithm of the odds
min	Minute
mL	Millilitre
mM	Millimolar
NaCl	Sodium chloride
NaOH	Sodium hydroxide
No.	Number
p	Probability
PAs	Proanthocyanidins
PCR	Polymerase chain reaction
qPCR	Quantitative polymerase chain reaction
qRT	Quantitative Real Time

QTL	Quantitative trait loci
RAPD	Randomly amplified polymorphic DNA
RFLP	Restriction fragment length polymorphism
RGB	Red Green Blue
RIL	Recombinant inbred line
RNA	Ribonucleic acid
rpm	Revolutions per minute
sec	Second
SCAR	Sequence characterized amplified region
SDG	Secoisolariciresinol diglucoside
SDS	Sodium dodecyl sulfate
SECO	Secoisolariciresinol
SSR	Simple sequence repeat
SRAP	Sequence-related amplified polymorphism
SNP	Single nucleotide polymorphism
Taq polymerase	Polymerase from the bacterial species *Thermus aquaticus*
TAE	Tris Acetic acid EDTA
TE	Tris EDTA
UV	Ultraviolet
YSD	Yellow-seeded degree

1. INTRODUCTION

Flax (*Linum usitatissimum* L.) is a member in the family Linaceae. The name *Linum* originated from *lin* or "thread" and the species name *usitatissimum* is a Latin word meaning "most useful". It is also called flaxseed or linseed when it is used as oilseed and referred to as fiber flax or just flax (in Europe) when it is used for fiber (Vaisey-Genser and Diane, 2003). As the name describes it is one of the important crops in Canada, particularly in Western Canada where it is grown as an annual crop. Both seeds and fiber are commercially valuable products harvested from the plant. However, oilseed flax is mainly grown in North America (Flax Council of Canada, 2012).

Flaxseed oil is used mostly for non-edible purposes because it is rich in alpha-linolenic acid (ALA) (40-65% of fatty acids) and dries very fast when exposed to air. Therefore, the oil is used to produce paints, varnishes, stains, inks and linoleum flooring (Green and Marshall, 1984). In Canada, cultivars intended for food-purpose oil having 5% or less ALA must be yellow seeded and they are called 'solin' type flax (Flax Council of Canada, 2012). In the edible oil market, flax must be yellow seeded as a visible trait to distinguish from the traditional brown flaxseeds. The majority of naturally occurring flax accessions are brown-seeded however, yellow seeded flax accessions are also found (less than 8 % of total accessions in flax world collection) (Diederichsen et al., 2006).

There is accumulation of flavonoids, proanthocyanidins (PAs; condensed tannins), and lignans in flaxseed, especially in its seed coat. Condensed tannins accumulate in pigment cells of the flaxseed-coat and impart colour to the seed (reviewed by Diederichsen and Richards, 2003). Along with PAs, secoisolariciresinol (SECO) and matairesinol are also present in higher quantities and SECO is present in the form of its glucoside in a lignan macromolecule (reviewed

by Mazur and Adlercreutz, 1998; Struijs et al., 2007). Flaxseed also contains the flavonoids herbacetin diglucoside (HDG) and kaempferol diglucoside (KDG). HDG is a part of the lignan molecule present in the seed coat (Qiu et al., 1999; Struijs et al., 2007). The lignans are phyto-chemicals which are polyphenolic substances derived from phenylalanine through the phenylpropanoid pathway. They are formed by dimerization of substituted cinnamic alcohols. Lignins also have the same simple phenolic precursors as lignans but are polymeric compounds with different branching patterns. Flaxseed lignans have various health benefits like breast cancer prevention, diabetes prevention and omega-3-fatty acids prevent cardiac problems (American Cancer Society, 2011; American Diabetes Services, 2012). Therefore, consuming flaxseeds is beneficial for health. The level of PAs in the seed coat determines the colour of the seed; the higher the condensed tannins, the darker will be the seed. PAs are synthesized in the seed coat by the phenylpropanoid pathway followed by the flavonoid pathway controlled at several steps by various catalytic enzymes and transcription factors. These genes regulating the flavonoid pathway are present in various loci in the genome.

Until now, at least four independent loci (*G*, *D*, *B1* and *Y1*) governing seed colour have been identified through genetic studies in flax (Mittapalli and Rowland, 2003). Mutant loci causing yellow seed colour in flax are denoted as *g*, *d*, *b1* and *Y1* and their counter parts (*G*, *D*, *B1* and *y1*) show 'wild-type' seed colour (brown). Recessive yellow seed mutants possess alleles at the *g* and *b1* loci. Variegated seed colour is controlled by a second recessive allele of *b1* referred to as $b1^{vg}$ locus. Dominant yellow lines have the dominant allele at the *Y1* locus and the European recessive yellow line (G1186/94) is conditioned by a homozygous recessive allele at the *D* locus (Mittapalli and Rowland, 2003). It is reported that the *d* allele of the *D* locus is also closely associated with the white petal colour wherein plants with white flowers always possess

the yellow seed colour (reviewed by Mittapalli, 2002). In G1186/94, both traits (yellow seed colour as well as white flower petals) are governed by the homozygous recessive allele '*d*' (Mittapalli and Rowland, 2003).

To obtain better insights into living systems, molecular tools have been generated over a period of time in the biological sciences. Molecular genetic marker system is one among them. For the past decade, microsatellite markers have been used for genetic and physical mapping in flax (Roose-Amsaleg et al., 2006; Ragupathy et al., 2011; Soto-Cerda et al., 2011; Cloutier et al., 2012a) and the details of the studies are described in Chapter 2, Section 2.10. In addition to this technique, a gene expression catalogue and the whole genome sequence of flax are now available to the scientific community (Venglat et al., 2011; TUFGEN, 2012; Wang et al., 2012).

The generation of molecular resources in flax led to various success stories in the past few years. Several quantitative trait loci (QTLs) have been identified for: *Fusarium* wilt resistance (Spielmeyer et al., 1998a; Vromans, 2006), yield related agronomic traits (Vromans, 2006), traits related to fatty acid composition (Cloutier et al., 2011), dominant yellow seed colour gene *Y1* (Cloutier et al., 2011), dominant yellow seed colour trait (*ysc1* gene) in a different population (M. Kulkarni and G. Selvaraj, unpublished) which can be used in breeding and crop improvement programs. Additionally, both a genome-wide physical map and a high density integrated consensus map of 15 linkage groups representing haploid number of chromosomes of flax are published (Ragupathy et al., 2011; Cloutier et al., 2012a). These advancements in flax research will promote use of the diverse molecular resources for further understanding and improvement of desirable traits in future studies.

Seed colour is an important trait in flax as it is known to be linked with other useful traits. Yellow seed colour is positively associated with oil content and also with 1000 seed weight (Culbertson and Kommedahl, 1956; Culbertson et al., 1960; Diederichsen et al., 2006). Yellow-seeded flax cultivars (Foster and Omega) have higher amino-acid content than the brown-seeded cultivar, NorLin (Oomah and Mazza, 1993) and similarly, yellow-seeded lines have 3 to 4 % higher protein content than brown-seeded lines (M. Kulkarni and G. Selvaraj, data unpublished). In addition to these differences, yellow-seeded lines have thin seed coats and clear oil (because of less pigment) compared to brown-seeded flax (reviewed by Sood et al., 2012). Because of these positive associations, the understanding of the genetic basis of the yellow seed colour becomes interesting.

Detailed genetic studies for seed colour trait in flax have been established by Mittapalli and Rowland (2003). However, the information on the genetic basis or the genomic region governing the seed-coat colour in the European recessive yellow line G1186/94 was not known and investigation was needed to increase the understanding. Hence, the objective of this research was to characterize the genomic region at the *D* locus governing the seed-coat colour and flower colour in the European recessive yellow line G1186/94. The graphical representation of the summary of the flow of work during the study is shown in Figure 1.1.

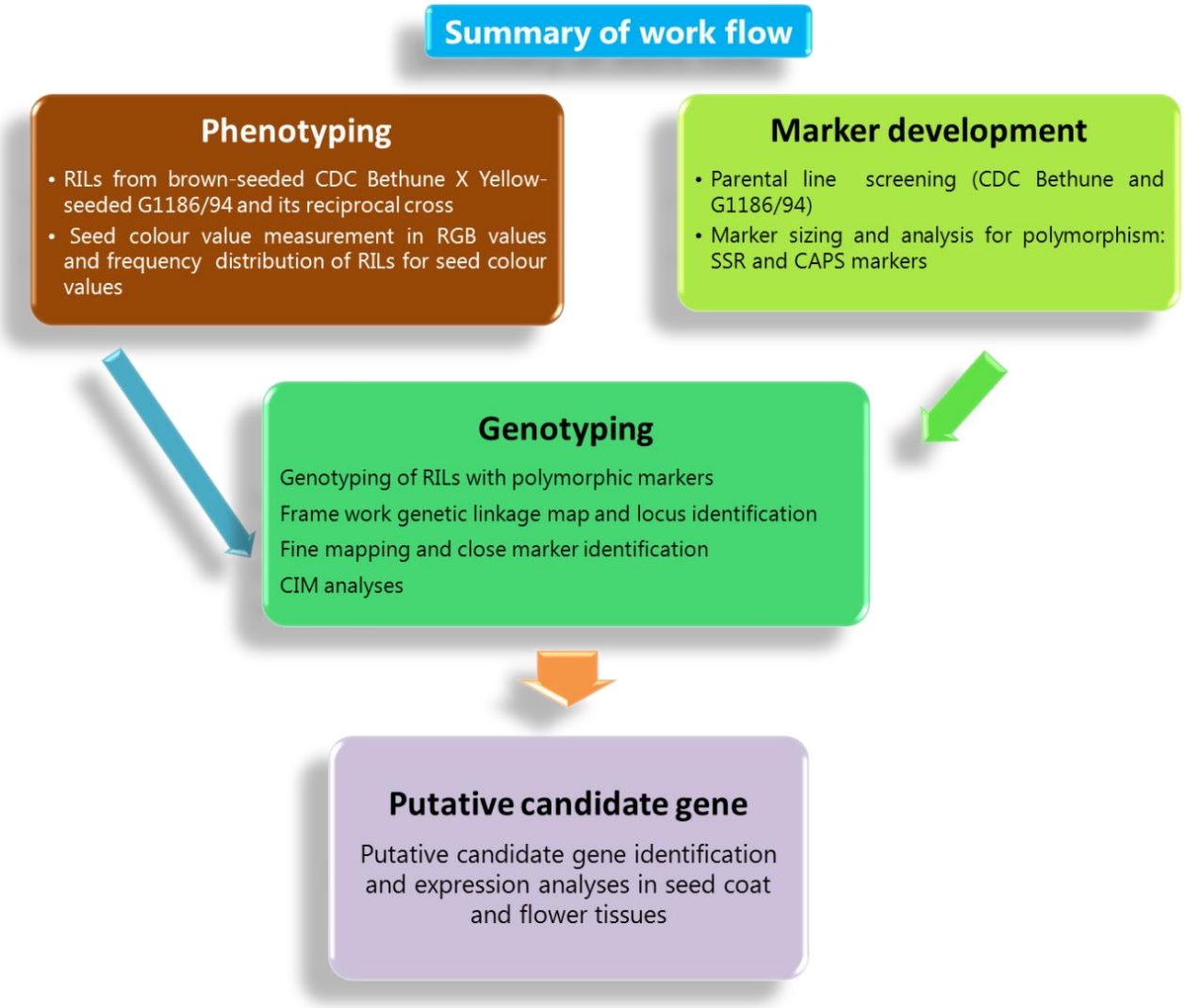


Figure 1.1: An overview of the research conducted during the study. The research started with single seed descent of RILs followed by phenotyping, marker screening, genotyping and linkage analyses, fine-mapping and eventually identifying the putative candidate gene.

2. REVIEW OF LITERATURE

This literature survey will review briefly flax as a crop, its general and botanical description, production throughout the world as well as in Canada, breeding and potential uses of the crop. Since the primary interest of this thesis is seed colour genetics of flax, flaxseed, genetics of seed colour, flavonoid biosynthesis and genetic marker based studies are also reviewed.

2.1 Flax: general description of the plant

Flax (*Linum usitatissimum* L.) is a member of the genus *Linum* in the family Linaceae. Flax is an annual plant generally growing to a height ranging from 40 to 91 cm, depending on various agronomic and environmental conditions (Flax Council of Canada, 2012). The flax plant requires 90 to 125 days to complete the growing cycle in Canada, which includes 45 to 60 days of vegetative period, 15 to 25 days of flowering period and 30 to 40 days of maturation period depending on the genotype and environmental conditions (Flax Council of Canada, 2012). It is a self-pollinated crop with a very low rate of outcrossing of 0 to 5% (Dillman, 1938).

2.2 Botanical description

Flax has a long and branched taproot system which grows more than one metre deep and side branches extend up to 30 cm (Diederichsen and Richards, 2003). The leading shoot is slender and erect and lateral branching occurs at the base of the stem. Several secondary basal horizontal sprouts arise when the primary shoot is injured and branching also increases with an increase in soil fertility (Dillman and Brinsmade, 1938; Diederichsen and Richards, 2003). On the other hand, fiber flax requires dense planting because this causes suppression of lateral branching (Diederichsen and Richards, 2003). Phyllotaxy (leaf arrangement) shows alternate

arrangement and the leaves are sessile and are three-veined. Smaller leaves are linear whereas the larger ones are linear-lanceolate. The dimensions of the leaves vary between 1.5 to 5.5 cm in length and 0.3 to 1.3 cm in width. The senescence of leaves occurs during the harvest maturity of the plant (Diederichsen and Richards, 2003).

The flowering occurs at the terminal position of branches. The flowers are generally regular, hypogynous and tetracyclic. The flowers are pentamerous with a radially symmetrical arrangement (Hayward, 1938; Schewe et al., 2011). The petals of flowers are found in different forms i.e., funnel, disk-shaped, star-shaped and tubular and the diameter of the corolla varies between 1 to 3.2 cm (Dillman, 1938; Diederichsen and Richards, 2003). Colour of the petals varies from light blue, blue, dark blue, violet, to red-violet, pink, white and further variations may also possible (Hayward, 1938; Diederichsen and Richards, 2003). The petals have veins and mostly they are coloured. The flowers open during mornings and the petals drop off by noon (Diederichsen and Richards, 2003). The stamens are placed at the base of the flower where nectar is secreted from the five small, flat pits of flask-shaped nectaries having stomata (Schewe et al., 2011) and are arranged alternately with respect to petals and appear as a thick ring. Generally, the anthers and stamens have similar colour as petals and they vary in colour (Dexter, 2009). The anthers and pollen in both blue and white flowers of fiber flax are blue; however, pollen is occasionally yellow in linseed flaxes, and anther walls might be white or yellow (Hayward, 1938). The anthers are introrse (when flowers close during noon, the filaments are bent and shed pollen inwards) and cause self-pollination. A photograph of a fully opened flower of CDC Bethune cultivar is shown in Figure 2.1. The ovary is pentalocular and each locule has two chambers separated by a false septum and each chamber contains an ovule (Hayward, 1938; Schewe et al., 2011).

Figure 2.1: A fully opened flower of a flax cultivar (CDC Bethune) showing the five blue petals with their dark blue veins and blue anthers.

The flax fruit is either a spherical or an egg-shaped capsule (Hayward, 1938) which is also commonly called a 'boll' (Beard and Comstock, 1980). Each boll has five carpels and ten lodicules and each lodicule is filled with seed. The completely matured boll is either fully closed or marginally opened along the septa, depending on the genotype. A photograph of flax capsules of line G1186/94 is shown in Figure 2.2. The bolls in oilseed type are not opened whereas they are slightly opened in fiber flax. The seeds are flattened and are ovoid or oblong elliptic, pointed at the tip, curved at the base, and 3.3 to 5 mm long. Based on the genotype, 1000 seed weight varies between 4 to 13 g and seed colour also varies from dark brown to yellow and olive (Diederichsen and Richards, 2003). It is observed that generally oilseed flax seeds are larger than those of fiber flax (Hayward, 1938).

Figure 2.2: A photograph showing flax capsules of the European recessive yellow line G1186/94.

2.3 Origin of flax crop and current cultivation

Flax is one of the ancient crops known to be cultivated some 10,000 years ago (reviewed by Allaby et al., 2005). Flax fibers found in Dzudzuana Cave, Georgia were around 30,000 years old indicating the use of flax in prehistoric periods (Kvavadze et al., 2009). *Linum angustifolium* Huds. is the ancestor of the common cultivated flax, *Linum usitatissimum* L. (Hayward, 1938). The botanical origin of flax is known to be in the Indian subcontinent because of the greater biodiversity with in the genus *Linum*. Domestication and cultivation began in the Fertile Crescent (Syria, Turkey and Iran) as larger seeds were found in Syria, Iraq and other parts of the Mesopotamian basin indicating the evolution of an irrigation-based agriculture system (reviewed by Vaisey-Genser and Morris, 2003). The domestication of flax is also observed on the Indian subcontinent near the Mediterranean Sea and this region is known to have high biological diversity of genus *Linum* (as reviewed by Fu, 2005).

Flax can be grown both as a single and a dual purpose crop for its oil (commonly called linseed) or bast fiber (also called fiber flax), or both. Flax is mainly grown for its seed oil in North America (Flax Council of Canada, 2012). Canada presently is the highest producer and exporter of linseed, which is mainly exported to European countries and the United States of America. Oilseed flax in Canada is produced mainly in Western Canada, primarily contributed by Saskatchewan, Manitoba and Alberta. In 2011, the flax crop was grown on 353,000 hectares with a production of 423,000 metric tonnes with an average yield of 1,200 kg ha^{-1}; and the bulk of it (404,000 metric tonnes) was exported (Statistics Canada, 2012). Even though Canada is a leader in world flax production, flax is considered a minor crop in Canada when compared to wheat (8,269,000 ha) or canola (6,848,000 ha) (Statistics Canada, 2012). A graphical representation of production statistics of linseed in the top ten countries in 2011 is shown in Figure 2.3.

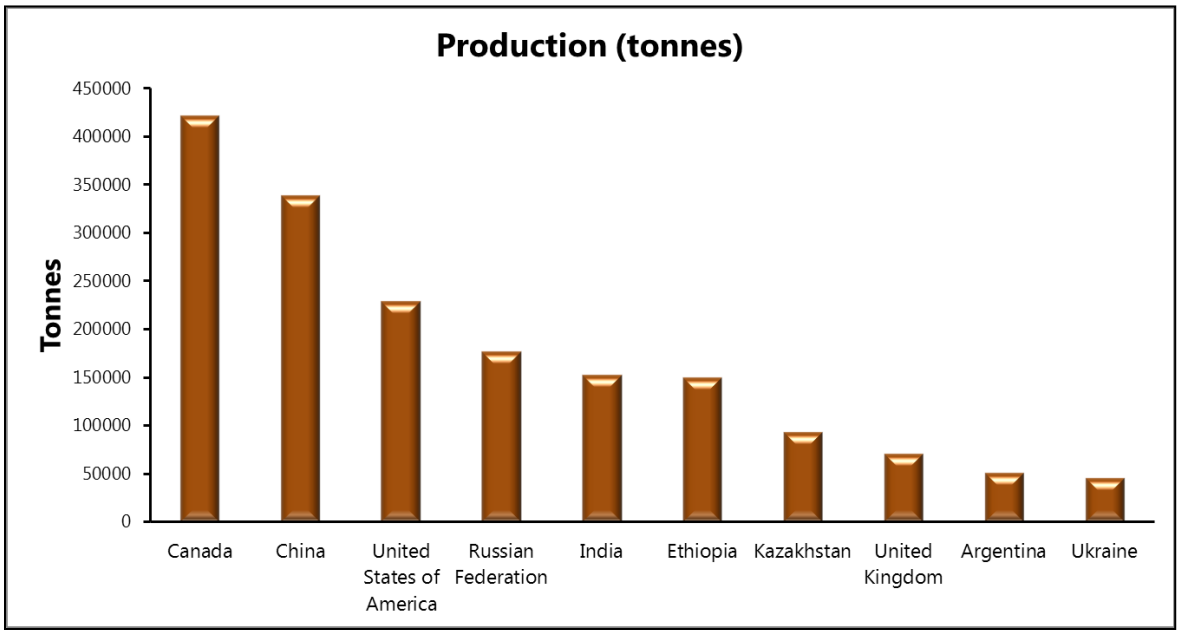


Figure 2.3: A graph showing the production of linseed in the top ten countries in 2011. Canada is a leading producer with 423,000 tonnes (the data is from FAOSTAT, 2013).

2.4 Uses of flax

2.4.1 Industrial uses of flax

Flax as a crop has different uses: raw material for different industries, food products, and animal feed supplement. Flax is an industry friendly raw material. Flaxseed is crushed to produce linseed oil. This oil is then used to manufacture various environment friendly products such as paints, varnishes, linoleum flooring and stains. In the same way, stem fiber is used in manufacturing of papers, clothes and, recently, fiber is being used as the interior panelling of some cars (Flax Council of Canada, 2012). Because of intensive breeding programs and technological improvements, production of fiber flax is mainly concentrated in Europe and China (Dexter, 2009), with China being the highest producer of fiber flax (116940 tonnes) in 2010 (FAOSTAT, 2013). The stem fibers which are present between the outer epidermis and inner woody tissues are extracted through the process of retting. These fibers, especially long-lined fibers, are very suitable for the textile industry because they are flexible, lustrous and soft so are used in high value linen products (Sharma and Sumere, 1992; Dexter, 2009).

2.4.2 Flax in the food industry

Flax is a recent favourite in the food industry in addition to the industrial and ancient food uses because of its nutraceutical properties. Flaxseed is being used in functional food markets, where seeds are being used for fortification of food products as a health benefit ingredient (Flax Council of Canada, 2012). Flaxseed, with its high levels of alpha-linolenic acid (ALA), an omega-3 fatty acid, good amount of fiber, mucilage (long chain carbohydrates), and cancer-fighting lignans, is a healthy food supplement (Flax Council of Canada, 2012).

Flax seeds normally have a very high content (45% to 65%) of ALA among the total fatty acids. This trait causes rapid drying of the oil upon exposure to the air which is very favourable for the production of paints, varnishes and other similar products like ink and linoleum (Green and Marshall, 1984). During the 1960s the demand for edible oil increased radically thereby increasing the demand for development of varieties with altered oil profile suitable for the edible oil market (Cullis and Kole, 2007). A mutant flax line with decreased ALA content (around 29 %) as compared to normal 43 % of fatty acids was developed from the Glenelg cultivar using gamma rays and ethyl methanesulphonate (EMS) (Green and Marshall, 1984). This decrease in ALA was not enough for edible purposes. Later, two stable 2 % ALA lines were obtained from the cross between two mutant lines, M1589 and M1722. In another study, the McGregor cultivar was subjected to EMS mutation and a stable flax line (in M_4 generation) with only 2 % of ALA was selected (Rowland, 1991). The low ALA trait is controlled by two recessive genes at independent loci as tested for goodness of fit to a 15:1 in F_2 populations of crosses M1722 x M1589 and E1747 x McGregor and its reciprocal cross (Green, 1986; Rowland, 1991; Rowland et al., 1995). In addition to decreased ALA, changes in other traits were also observed such as increased plant height, albinism, floral abnormalities and yellow seed colour (Green and Marshall, 1984). The yellow seeded, low ALA flax lines (Green, 1986; Rowland, 1991) are known as 'solin' (trade name Linola^{TM}) in Canada. Here 'sol' refers to sun and light and 'lin' is for linseed (Mittapalli, 2002). Solin oil has less than 5% ALA and is suitable for cooking purposes. In the Canadian market, all registered brown-seeded varieties have high ALA. So, solin varieties must have yellow seed colour as a visible distinguishing trait (Flax Council of Canada, 2012).

Apart from its oil and fiber, flax is also known for secondary metabolites. Flavonoids and lignans are major constituents of flaxseed. Among lignans, secoisolariciresinol (SECO) and matairesinol were found to be in higher levels in flaxseed, along with herbacetin diglucoside (HDG) (Mazur et al., 1996; Struijs et al., 2007; Muir, 2009). Secoisolariciresinol diglucoside (SDG) mainly accumulates in the outer integument of the seed as shown by immunolocalization studies (Attoumbre et al., 2010). SDG is known to have different biological activities, such as prevention of cancer as demonstrated by its antitumorigenic effect in rats (Thompson et al., 1996) and also its antioxidant property (as reviewed by Charles, 2013). Dietary supplementation of flax lignan reduced the risk of cardiovascular diseases in Western population (Peterson et al., 2010). Because of various health benefits, the demand for whole and milled flaxseed, and cold-pressed flaxseed oil is increasing (Flax Council of Canada, 2012).

2.4.3 Flax as a feed

In addition to industry and food usage of flax, the crop has been a promising contributor in the poultry and animal feed industries. When animal feeds were supplemented with flaxseeds, there was improvement in animal health and performance. In the pet food industry, flax is being used to solve the digestive and skin problems in cats and dogs (Flax Council of Canada, 2012). In poultry, laying hens, when fed with 10 to 20% flax in their rations, produce eggs with increased amount of omega-3 fatty acids and decreased amount of saturated fatty acids. The eggs are now being produced in the USA and Canada and are called “Omega eggs” (Cullis and Kole, 2007).

2.5 Flaxseed and its seed coat

Seed is one of the economically important parts in flax. Flaxseed consists of the seed coat (testa), the endosperm and the embryo. The flax seed coat is composed of six different layers which, starting from the outermost to innermost layer are: cuticula, epidermis, ring or round cells, sclerenchyma fibers, transversal or cross cells, and pigment cells. The cuticula is the outermost layer of the seed coat, which is glossy and slightly wavy (Diederichsen and Richards, 2003).

A single layer of epidermal cells (mucilage cells) covers the seed. The main function of these cells is to play an important role in effective germination by attaching seed to the soil upon water absorption (Boesewinkel, 1980). Mucilage is produced by epidermal cells and is deposited in the extracellular space facing outside of the seed just below the outer cell wall. Consumption of flaxseeds have health benefits which are attributed to the mucilage as this protects the epithelia of the digestive system (Diederichsen and Richards, 2003).

The next region consists of one to five layers of round parenchyma cells, called ring cells or round cells. Ring cells are developed from the inner layer of the outer integument (Boesewinkel, 1980). Round cells have conspicuous intercellular spaces. The round appearance of these cells is due to uneven thickening of their cell walls. The ring cells may contain condensed tannin-like substances and sometimes chlorophyll (Diederichsen and Richards, 2003).

Next, sclerenchyma fibers are found in a single layer which varies from 16 to 25 µm in thickness. These cells are generally colourless but can be occasionally dark yellow. Below that, transversal-cells are present in at least two layers. They are also called cross cells because of their uneven orientation. Next, square shaped pigment-cells are found as the innermost testal

layer. These isodiametric cells contain condensed tannins and other pigments. Because of the presence of the tannic pigments, the seed looks brown-yellow in colour. The pigment cells are often absent in yellow-seeded flax lines however, if they are present, the cells do not contain any pigments. Seed colour is influenced by the cotyledons in the absence of pigments, normally yellowish or white in colour (Diederichsen and Richards, 2003).

Endosperm is located below the multilayered seed coat. This layer contains oil and protein and occupies one third or less of the seed volume. The embryo has two cotyledons which are surrounded by the endosperm. The cotyledons fill more than two thirds of the total seed volume and are a major oil storage tissue (Diederichsen and Richards, 2003).

2.6 Flaxseed colour genetics

Studies on seed colour genetics have been done for ninety years. One of the first classical genetic studies on flaxseed colour was performed by Tammes (1922, 1928). She observed that seed colour in flax is primarily governed by three genes or factors, and she designated them as *G*, *D*, and *B'*. Factor *G* was designated as the basic factor whereas *D* and *B'* were considered as modifying factors. It was hypothesized that the seed colour is brown only in the presence of *G* and colourless in its absence (*g*). When factor *G* was present along with factor *D*, then the seed colour became greyish-green, because *D* was acting as an inhibitory-factor. On the other hand, when *B'* was present along with the basic factor (*G*), there was no effect of inhibition by *B'* and the seeds were brown. However, in the presence of all three factors, the seeds were brown because *B'* had overcome the inhibitory action of *D* (Tammes, 1922). Later it was found that factor *B'* consisted of two alleles *B1* and *B2* which were not involved in governing seed colour (Tammes, 1928). Then it was observed that light coloured flaxseed is actually due to the

visibility of the yellow cotyledons through the transparent seed coat (Tammes, 1928). The appearance of coloured (darker) seed is due to the presence of pigments in the inner layer of inner integument (pigment-cells) (Hayward, 1938). The cavities of the pigment cells are filled with tannic substances which imparts colour to the seed coat (Boesewinkel, 1980).

A brief review by Mittapalli and Rowland (2003) describes that the studies by Shaw et al. (1931) on Indian linseed indicated that there are three factors (*G*, *M* and *D*) governing seed colour in flax. It was observed that *G* produced a grey colour in the seed coat whereas *M*, in association with *D*, produced fawn colour. However, in the presence of all three factors, the fawn colour changed to brown. Yellow was considered as the primary seed colour (Shaw et al., 1931). Independent studies suggested that the yellow seed colour trait in flax cultivar Crystal is governed by a recessive allele at the *B1* locus, which also has a pleiotropic effect on crimped flower petals (Culbertson and Kommedahl, 1956). Seed colour in flax was governed by two or three pairs of complementary genes as yellow seed was produced in the presence of homozygous recessive alleles at any of the three loci and brown seed was produced in the presence of at least one dominant allele at all three loci (Barnes et al., 1960). It is reported that the three basic loci, *G*, *D*, and *B1* are responsible for governing seed colour in flax seed with the exception of solin yellow flax lines which are determined by a separate locus (Mittapalli and Rowland, 2003).

The summary of genetic studies conducted by Mittapalli and Rowland (2003) to understand the allelic-gene relationship of different yellow flax lines is reviewed here. Eleven flax lines, namely four spontaneous recessive yellow seed mutants, two dominant yellow lines, one variegated line, three brown lines, and one recessive yellow European line (G1186/94) were crossed in 49 possible combinations. Test crosses were also performed with four lines possessing the four different seed colour genes (*G, D, B1* and *Y1*). The results suggest that there are at least

four different loci, *G*, *D*, *B1* and *Y1* governing the seed colour in flax. Spontaneous recessive yellow seed mutants possess the recessive allele at the *G* locus; dominant yellow lines have dominant allele at the *Y1* locus; the variegated line is controlled by an alternative form of recessive allele ($b1^{vg}$) at the *B1* locus; and the European yellow line is conditioned by a recessive allele (*d*) at the *D* locus. The yellow seed colour was observed in the presence of the dominant allele at the *Y1* locus and homozygous recessive alleles of either or both the genes (*g* and *d*) also made seeds yellow. Variegated seeds were produced in the presence of a homozygous recessive allele at the *B1* locus even when a dominant allele at the *G* locus was present. Yellow seed was produced in the presence of a homozygous recessive allele of the *G* locus irrespective of the allelic status at the *B1* locus. Yellow seed was observed in the presence of homozygous recessive allele *d* irrespective of the allelic status of $B1^{vg}$ (variegated *B1*). The recessive loci (*g*, *d* and $b1^{vg}$), when present in a homozygous recessive condition, were epistatic to the loci containing dominant alleles. All four loci controlling seed colour were inherited independently except for a probable weak linkage between *G* and $B1^{vg}$.

2.7 Genetics of seed colour in European recessive yellow line (G1186/94)

When G1186/94 (female) was crossed with brown lines (male), F_1 plants produced brown seeds (F_2 seeds). That confirmed that brown seed colour was dominant over yellow seed colour. A segregation ratio of 3:1 brown as to yellow in F_3 seeds (F_2 plants) suggested the inheritance of a single gene model. As colour of seed is due to accumulation of pigments in the seed coat tissue and seed coat is maternally governed, genetics of the seed coat is always one generation behind the embryo. The effect of genetics of seed colour in seed coat can be seen in the succeeding generation. Therefore, segregation of the seed colour trait was observed in F_3 seeds produced by F_2 plants. A graphical representation of the cross between G1186/94 (female) and CDC Bethune

(male) showing the seed colour inheritance is given in Figure 2.4. The locus is designated as *D* and the recessive allele is *d* which makes the seed yellow (Mittapalli and Rowland, 2003). Henceforth, the first parent mentioned in a cross is always female and the latter is male.

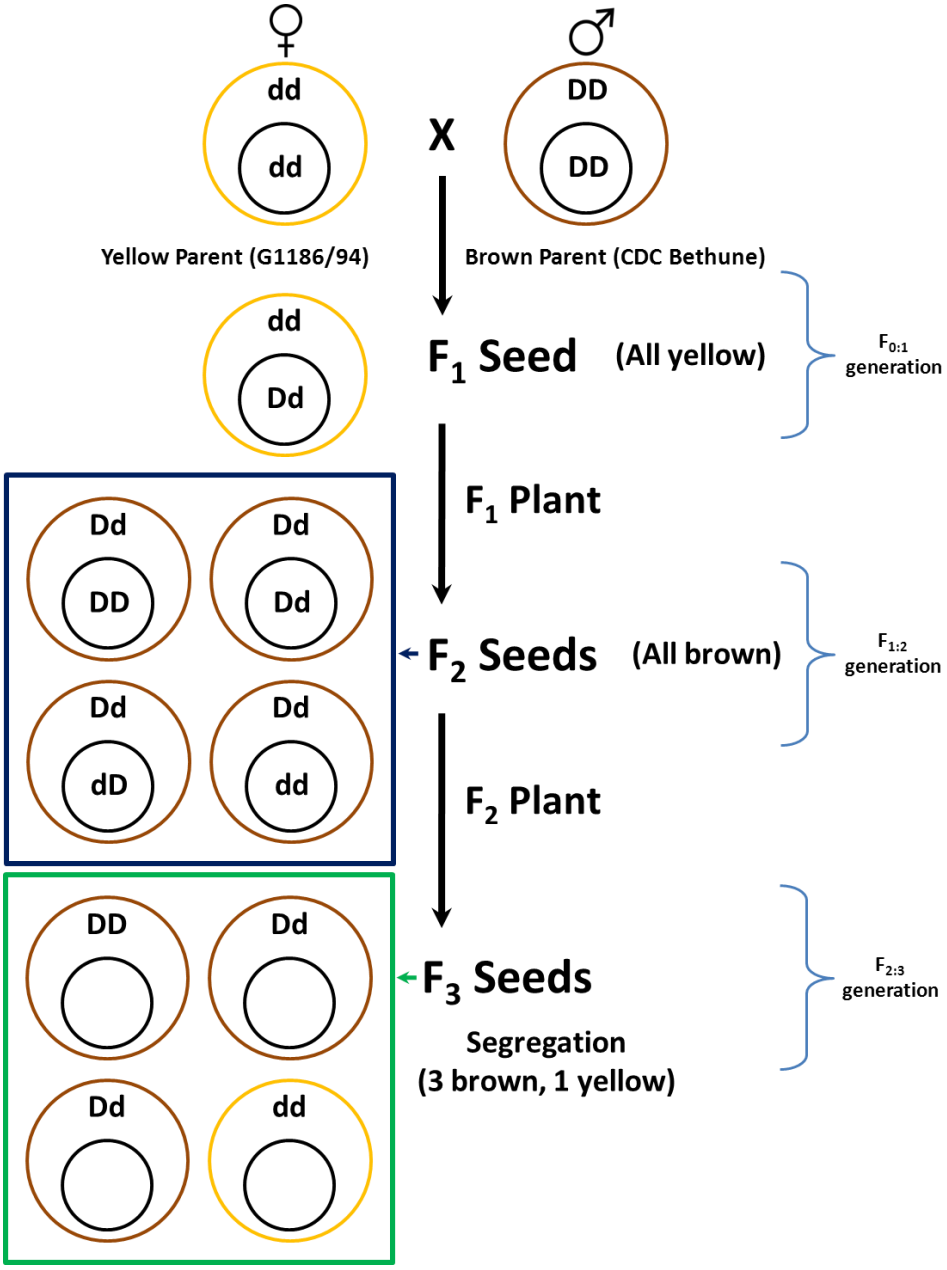


Figure 2.4: A schematic representation of inheritance of seed colour trait through seed coat in the cross G1186/94 X CDC Bethune. As the genetics of seed coat is one generation behind the embryo, phenotyping for seed colour is always done for seeds of succeeding generation. Here, F_3 seeds show segregation for seed colour.

F_1 seeds of the cross between dominant yellow line (CPI84495) and G1186/94 were yellow. F_2 plants (F_3 seeds) showed a 13:3 yellow as to brown seed ratio indicating the two gene model having two epistatic genes, one dominant and one recessive, controlling the yellow seed colour in the parental lines. It was concluded that yellow seed is produced in the presence of dominant allele of *Y1* and also in the absence of *Y1* but in the presence of homozygous recessive alleles at the *D* locus. However, brown seed is produced only in the absence of the *Y1* gene but in the presence of at least a single dominant allele of the *d* gene (Mittapalli and Rowland, 2003). In the F_3 population of this same cross, all homozygous recessive lines for the *D* locus produced white or pink flower petals indicating the tight linkage of flower colour with the *D* locus (Mittapalli, 2002). He reviewed that a yellow seeded line (Bolley Golden) possessing the *d* allele had a close association with the white flower colour. Thus, the recessive allele at the *D* locus regulates both yellow seed colour and white flower colour (Mittapalli, 2002).

In a G1186/94 X M96006 (variegated yellow parent) cross, the seeds of F_1 plants (F_2 seeds) were brown. F_2 (F_3 seeds) phenotypic data showed 9 brown, 3 variegated and 4 yellow seed coloured plants suggesting the regulation of seed colour is due to independent recessive genes. Yellow seed is produced in the presence of a homozygous recessive allele at the *D* locus irrespective of allelic status of the *b1* gene. Variegated seed is produced only in the presence of a homozygous recessive allele of the *b1* gene and at least one dominant allele at the *D* locus (Mittapalli and Rowland, 2003).

In the cross between CDC recessive yellow (YSED2 or YSED4 or S95407 or S96071) X G1186/94 (European recessive yellow), the seeds from F_1 plants (F_2 seeds) were observed to be brown seeded. F_2 plants (F_3 seeds) gave segregation of 9 brown: 7 yellow phenotypic ratio, suggesting the existence of two independent recessive genes for yellow seed colour. This

explained that the presence of homozygous recessive alleles of either or both genes (*G-dd*, *ggD-*, or *ggdd*) cause the production of yellow seeds. Brown seed is produced only when a dominant brown allele at both loci (*G-D-*) is present. These results confirm the complementary effect of a pair of genes for brown seed colour. The results are agreeing with the previous studies and observations pertaining to the *D* locus (Mittapalli and Rowland, 2003).

In summary, based on the above mentioned studies conducted by Mittapalli and Rowland (2003), it can be concluded that a recessive mutation at the *D* locus is independently inherited and regulates yellow seed colour and white or pink flower petal colour in G1186/94.

2.8 Flavonoids

2.8.1 General description and functions in plants

Flavonoids are the low molecular weight and diverse secondary metabolites derived from the phenylpropanoids. These phenylpropanoids are the derivatives of precursor molecules amino acid L-phenylalanine and malonyl-coenzyme A (Winkel-Shirley, 2001). Flavonoid biosynthesis and associated metabolic pathways have been vastly studied in the plant system (Winkel-Shirley, 2001; Abeynayake et al., 2012). Flavonoids consist of different subclasses, such as anthocyanins, flavonols, proanthocyanidins (PAs; also called as condensed tannins), chalcones, flavones and flavandiols (Winkel-Shirley, 2001; Routaboul et al., 2006).

Flavonoids and their derivatives have many physiological roles in plants. One of the main roles is to influence auxin transport. Other roles include defense against microbes and insects, controlling the levels of reactive oxygen species (ROS) during oxidative stress, allelopathy, protection from ultra violet (UV) radiation by absorption; and finally, anthocyanins impart colours to flowers that attract pollinators (Winkel-Shirley, 2001; Buer et al., 2010). Some other

functions of flavonoids that were recently observed in legumes are as follows. They are involved in signalling in legume-bacterium symbiosis, they play a vital role in root nodule formation and influence the development of *in vitro* root formation through accumulation in progenitor cells of various organs (Buer et al., 2010).

Flavonoids also have several health benefits for animals as a nutraceutical. They are important dietary components which prevent and treat disease. These compounds have anti-oxidative, anti-tumor, anti-inflammatory, anti-proliferative and metal chelating properties. Flavonoid rich diets protect against cardiovascular diseases by inhibiting lipid peroxidation (Lin et al., 2006; Buer et al., 2010).

2.8.2 Flavonoid and proanthocyanidin biosynthesis

Flavonoid and PA biosynthesis pathway in the seed coat have been studied widely in several plant systems such as *Arabidopsis* (transparent testa mutants), *Glycine max, Medicago, Desmodium uncinatum, Lotus corniculatus* and *Brassica* (Abeynayake et al., 2012). Flavonoid biosynthetic enzymes have been mostly found to be operating in enzyme complexes and they are generally located in the cytoplasm. End products of flavonoid synthesis are transported to several extracellular or subcellular locations. However, vacuoles generally act as a storage organelle for the flavonoids involved in colouration (Davies and Schwinn, 2006).

PAs (condensed tannins) are monomeric or oligomeric end products of the flavonoid pathway and the composition varies with the plant species. PAs consist of catechin (10 to 30% of extension units), epicatechin, gallocatechin (GC) and epigallocatechin (ECG) (Tian et al., 2008). PAs are synthesised in the cytosol and stored in vacuoles after their glycosylation (Tanner et al., 2003). The seed coat of *Arabidopsis* contains PAs completely consisting of epicatechin units

(Routaboul et al., 2006) and PAs in *Medicago truncatula* also have a similar profile where PAs are mainly composed of epicatechin units with very low levels of epigallocatechin or gallocatechin and epiafzelechin or afzelechin (Pang et al., 2007). This indicates low levels of delphinidin and pelargonidin. PAs and anthocyanins serve several functions in plants such as flower pigmentation to attract pollinators, defense against pathogens and pests and protection against abiotic stresses. Recently, efforts have being made to genetically engineer plants to enhance anthocyanins in flower colours, fruits and vegetables and to increase PAs in fruits and in plants for defense against biotic and abiotic stresses and to increase forage quality (Dixon et al., 2012).

In the biosynthetic pathway, flavonoid precursors (phenylpropanoids) are formed and they subsequently branch into PA biosynthesis. A comprehensive review on flavonoid biosynthesis in plants is published in a book chapter by Davies and Schwinn (2006). A recent review on flavonoid biosynthesis in *Arabidopsis* has been published by Saito et al. (2013). First, *trans*-cinnamic acid is formed from the precursor phenylalanine catalyzed by phenylalanine ammonialyase (PAL). Then reactions are catalysed by cinnamate 4-hydroxylase (C4H) and 4-coumarate:CoA ligase (4CL) where *p*-coumaric acid and 4-coumaroyl CoA are formed from respective enzymes. In few species, caffeoyl-CoA and feruloyl-CoA may also be formed after 4-coumaroyl CoA. Malonyl-CoA is synthesized from acetyl-CoA by acetyl-CoA carboxylase (ACC). Chalcone synthase (CHS) is the first committed enzyme in flavonoid biosynthesis which combines 4-coumaroyl-CoA and malonyl-CoA to form naringenin chalcone in a condensation reaction. It is the first chalcone to be formed in the phenylpropanoid pathway. Then, naringenin is formed by chalcone isomerase (CHI) through the stereospecific cyclization of naringenin chalcone. Flavanone 3-hydroxylase (F3H) is a 2-oxoglutarate-dependent dioxygenase that

catalyzes the formation of dihydroflavonol (dihydrokaempferol) via oxygenation at the 3-position of naringenin. Flavonoid 3' hydroxylase (F3'H) is a cytochrome P450 monooxygenase that hydroxylates at the 3'-position of the B-ring of either dihydrokaempferol or kaempferol to form dihydroquercetin and quercetin, respectively (as reviewed by Saito et al., 2013).

Anthocyanins are the bridging molecules between chalcones and PAs. The key enzyme involved in anthocyanin production is dihydroflavonol 4-reductase (DFR) that catalyzes the conversion of (2R,3R)-trans-dihydroflavonols to (2R,3S,4S)-flavan-2,3-trans-3,4-cis-diols (leucoanthocyanidins) (reviewed by Davies and Schwinn, 2006). Further, anthocyanins lead to the branching of PA biosynthesis. Two precursor groups, 2,3-*trans*-flavan-ols (catechin) and 2,3-*cis*-flavan-ols (epicatechin), are involved in PA biosynthesis and are synthesised by two different pathways. In one pathway, leucoanthocyanidin can be converted into catechin by leucoanthocyanidin reductase (LAR) and in another pathway, conversion of leucoanthocyanidin to coloured anthocyanidin can be catalysed by anthocyanidin synthase (ANS). Later, anthocyanidin acts as a precursor for anthocyanidin reductase (ANR) to produce epicatechin. In the process of understanding the PA biosynthetic pathway, several genes such as *LAR*, *ANS* and *ANR* have been identified and characterised from different plant species (Tanner et al., 2003; Xie et al., 2003; Pang et al., 2007; Tian et al., 2008). Several reports on transcription factors (R2R3 MYB and WD40 protein) as the key regulators in PA biosynthesis and their use in plant genetic engineering have been published (reviewed by Dixon et al., 2012). In *M truncatula*, glycosylation of epicatechin to epicatechin 3'-O-glucoside is catalysed by glycosyltransferase UGT72L1. This glycosylated PA is preferred by vacuolar multidrug and toxic compound extrusion (MATE) transporters. However, the polymerization of condensed tannins is not clear at the later part of the pathway (Dixon et al., 2012).

2.8.3 Distribution of flavonoids and proanthocyanidins in flax

Flaxseed consists of flavonoids and lignans (polyphenolic compounds that are derived from phenylalanine by way of dimerization of substituted cinnamic alcohols). One of the first reports suggested that SECO and matairesinol were observed in higher quantities in flaxseed among the lignans or phytoestrogens (Mazur et al., 1996; Muir, 2009). Secoisolariciresinol was later found in the form of secoisolariciresinol diglucoside (SDG) as a part of lignan polymer (Johnsson et al., 2000). Flavonoids, like herbacetin diglucoside (HDG) and kaempferol diglucoside (KDG), were also found in flaxseed (Qiu et al., 1999). Later it was found that HDG, which is derived from herbacetin, is also a part of the lignan macromolecule (Struijs et al., 2007). Herbacetin is a flavonol molecule and it is synthesized from the flavonoid pathway in flaxseed (Struijs et al., 2007). SDG in defatted flaxseed flour is found to be up to 2.4% (w/w) (Johnsson et al., 2000) and HDG is found to be 0.2% (w/w) in flaxseed hulls (Struijs et al., 2007).

2.9 Molecular studies for seed colour

Several studies are being conducted to understand the genes or loci that are governing seed colour (especially those in the flavonoid biosynthesis pathway) in different plant species such as *Arabidopsis*, *Brassica* spp., *Zea mays* and *Glycine max* using different molecular tools. In a review by Buer et al. (2010), transparent testa (*tt*) mutations have been located in the phenylpropanoid pathway in *Arabidopsis*. The studies on *tt* mutants revealed mutations in several regulatory genes such as CHS, designated as *tt4*; CHI as *tt5*; F3H as *tt6*; F3'H as *tt7* and DFR as *tt3*. These genes are involved in controlling different points in the flavonoid pathway that are responsible for production of different compounds (reviewed by Winkel-Shirley, 2001; Buer et al., 2010). In *Arabidopsis* pale coloured seeds (either *tt* mutants or tannin deficient seeds

[tds]), genes have been characterized that regulate PA or anthocyanin accumulation such as *TT12*, which encodes for multidrug and toxic compound extrusion (MATE) secondary transporters and *TT19*, which codes for glutathione S-transferase (GST) which participates in cytoplasmic transport of PAs (Lepiniec et al., 2006). In addition to this, transcription factors and their complexes that regulate the gene expression of the flavonoid pathway has also been reviewed. For example, the transcription factor complex MYB–bHLH–WD40 (MBW) regulates the genes that encode enzymes specifically involved in the biosynthesis of anthocyanins and condensed tannins (Hichri et al., 2011). In *Arabidopsis*, most of the enzymes involved in flavonoid biosynthesis are encoded by a single copy of structural genes except flavonol synthase (FLS) (Winkel-Shirley, 2001). In a recent study, 22 flavonoid quantitative trait loci (QTLs) were identified including 16 coarsely mapped QTLs and three genes: TT7, MYB12 and TT15. Till now, 66 loci governing seed colour or the flavonoid pathway have been identified in *Arabidopsis* (Routaboul et al., 2012). In *Arabidopsis*, there are 16 mutants and 32 known genes that are involved in flavonoid biosynthesis (reviewed by Saito et al., 2013).

In some studies conducted in *Brassica* species, a random amplified polymorphic DNA (RAPD) marker is linked to the seed colour gene in a C-genome chromosome of *Brassica campestris-alboglabra* (Chen et al., 1997). Three amplified fragment length polymorphism (AFLP) markers followed by a sequence characterized amplified region (SCAR) marker have been developed in *Brassica juncea* which are tightly linked with yellow seed colour (Negi et al., 2000). In the genetic mapping study with AFLP and microsatellite markers, two microsatellite markers are found to be linked to a partially dominant allele at the *Y* locus in linkage group 9 (N9) of a *Brassica napus* responsible for yellow seed colour (Xiao et al., 2007). Two major QTLs and five minor QTLs for seed-coat colour are identified in different chromosomes of

Brassica oleracea and *Brassica napus* monosomic alien addition lines (MAALs) using microsatellite markers (Heneen et al., 2012). In *Brassica rapa*, one major QTL (SCA9-2) and one minor QTL (SCA9-1) have been identified on linkage group A9 using recombinant inbred lines (RILs) and SSR marker system (Kebede et al., 2012). These studies will help in mapping loci associated with seed colour traits in other plant systems such as flax.

2.10 Genetic and molecular marker studies in flax

Flax genetic diversity has been surveyed by using different molecular markers. These include RAPD, restriction fragment length polymorphism (RFLP), AFLP, SSR, and single nucleotide polymorphism (SNP). SSRs are also called microsatellites; they consist of tandem repeats of very short nucleotides in variable numbers and are known to be found in abundance and are co-dominant (Queller et al., 1993).

Molecular tools have been used to understand several mechanisms and traits in flax. Molecular markers have been used to understand the molecular basis of fatty acid biosynthesis especially with respect to the synthesis and various levels of alpha-linolenic acid (ALA) content in flax lines. Two *FAD3* desaturase genes (*LuFAD3A* and *LuFAD3B*) controlling the levels of ALA have been identified in flax. This was done by heterologous expression in yeast and complementing the low ALA mutants with the wild-type gene (high ALA). Later, a CAPS marker was developed for each gene (*LuFAD3A* and *LuFAD3B*) to differentiate low ALA lines from high ALA lines (Vrinten et al., 2005). Different gene expression regulators have been identified during flaxseed maturation using the cDNA-AFLP approach (Gutierrez et al., 2006). Microarray analyses was done at different stages of stem growth to identify the transcripts expressed during elongation and secondary cell wall deposition and to confirm the localization of β-galactosidase in phloem fibers (Roach and Deyholos, 2008).

Microsatellite markers are being developed for identifying polymorphism and distinguishing flax cultivars as well as genetic mapping for different traits (Roose-Amsaleg et al., 2006; Cloutier et al., 2009; Cloutier et al., 2011; Soto-Cerda et al., 2011; Soto-Cerda et al., 2012; Cloutier et al., 2012a; Cloutier et al., 2012b). One of the early and important contributions in this area is from Cloutier et al. (2009), where 248 expressed sequence tag (EST) derived SSRs were developed and polymorphism was assessed between 23 flax accessions. An extensive and invaluable resource of an expressed sequence tag (EST) database suggesting different expressed genes was generated for different developmental stages of flaxseed. In this study, higher expression of flavonoid biosynthesis genes was observed in seed coat tissues at the globular and torpedo stage of embryo formation during seed development (Venglat et al., 2011). In addition to this, several genomic resources such as ESTs, transcriptome sequences, TILLING population and the whole genome shotgun sequence of flax cultivar CDC Bethune is available through Total Utilization of Flax Genomics (TUFGEN) initiative (TUFGEN, 2012; Wang et al., 2012).

There are only a few studies in flax for genetic linkage and QTL mapping that focus on traits of agronomic value. The first comprehensive genetic linkage map developed in flax was constructed using AFLP markers, and using this map, QTLs were identified for *Fusarium* wilt resistance (Spielmeyer et al., 1998a, 1998b). A second map was developed using RAPD and RFLP markers (Oh et al., 2000). Later, an AFLP based linkage map was generated. The QTLs were identified for *Fusarium* wilt resistance in the population of a cross between fiber flax and linseed and a mapping study has been done for identifying QTLs related to agronomic traits in the cross between two fiber flax lines. In this process, a total of 60 QTLs were identified (Vromans, 2006). An SSR-based linkage map has been developed, and QTLs were identified for the traits that govern fatty acid composition in flax (Cloutier et al., 2011). In the same study, a

QTL for dominant yellow seed coat colour was identified using SP2047 as a source of dominant yellow seed colour gene (*Y1*) and UGG 5-5 as a brown seeded parent (Cloutier et al., 2011). A dominant yellow seed colour locus, *Ysc1*, is also being fine mapped to identify the genetic basis of defective flavonoid biosynthesis in a different mapping population (M. Kulkarni and G. Selvaraj, data unpublished). Utilization of SNP markers for flax genetic improvement is in its early stages. Recently, more than 55,000 SNPs were identified using a reduced genome representation (RGR) approach on eight flax genotypes through high throughput sequencing (Kumar et al., 2012). These are examples of successful linkage mapping in flax, which indicate that the molecular tools can be effectively employed to identify genomic regions associated with traits of interest.

The first physical map of flax was generated using genome wide bacterial artificial chromosome (BAC) end sequences (Ragupathy et al., 2011). The most recent integrated consensus genetic and physical map of flax was constructed using SSR markers from three populations: CDC Bethune/Macbeth, E1747/Viking and SP2047/UGG5-5. Maps from individual populations are being generated using 385 to 469 markers each. Out of a total of 770 markers, 114 were shared by all three populations to form 15 linkage groups. The total length of the genetic map is 1,551 cM with a mean marker density of 2.0 cM per marker. In that, contigs of the physical map correspond to 74% of the whole genome size of 370 Mb (Cloutier et al., 2012a). This is also a valuable resource in flax genomics where the information can be used for mapping of different traits, locating the candidate genes in the chromosomes and last but not least, it will be helpful in the flax breeding programs for desirable traits.

3. MATERIALS AND METHODS

3.1 Flax lines

Flax parental lines were selected based on their seed colour. Parental lines CDC Bethune and G1186/94 were selected for the current study. CDC Bethune is a brown-seeded variety (it is referred to in this thesis as "wild type"). G1186/94 is a yellow line with a recessive mutation at the *D* locus for seed colour and is called a European recessive yellow line (Mittapalli and Rowland, 2003). RILs in the current study; indicate reciprocal nature of the cross between parents. All flax lines used in this study were obtained from Dr. Gordon Rowland (Crop Development Centre, University of Saskatchewan).

3.2 Plant growth conditions

The $F_{7:8}$ RILs were grown in a growth chamber (GC) under controlled conditions in the Phytotron facility at the University of Saskatchewan in 2008. During the growth period, day length of 16 hours light (at 23°C) using 400 Watt high pressure sodium lamps and 8 hours of dark at 16°C was maintained on a daily basis. Appropriate moisture levels were maintained throughout the growing period. Plants were supplied with slow release fertilizer at sowing and 15-30-15 (N-P-K) fertilizer at first bud stage. Twenty seven grams of 15-30-15 (N-P-K) and 9 g of $CuSO_4$ was mixed in 9 L of water and each plant was supplied with 500 mL of that solution. They were irrigated regularly. At maturity, bolls of each RIL were harvested separately.

The $F_{8:9}$ RIL seeds were sown in six inch diameter pots in a greenhouse (GH) at Innovation Place, Saskatoon in 2011 to obtain $F_{9:10}$ seeds by single seed descent method. There, 16 hours of day length at 25°C using 400 Watt high pressure sodium lamps and 8 hours of dark at 16°C was maintained on a daily basis. Fertilizer was applied in the required amount at the

required time. Moisture levels were maintained at the required amount throughout the growing period by both manually and automated watering system, depending on the conditions. At maturity, bolls of each RIL plant were harvested, threshed and cleaned separately.

3.3 Tissue collection

Flower buds were tagged just prior to anthesis and seeds were collected at 15 days after flowering (DAF). Seeds were dissected to separate seed coat tissue from endosperm and embryo. Seed coat samples were immediately frozen in liquid nitrogen and stored at -80 ºC. Flower petals were collected from unopened flower buds and frozen in liquid nitrogen and stored at -80 ºC. Leaf samples from individual plants were collected in clean tubes on dry ice and stored at -80 ºC.

3.4 Phenotyping of RILs

3.4.1 Seed and flower colour

Eight seeds of each parent and RIL were placed in 96 well Microtest^{TM} Falcon flat bottom plates (BD Biosciences) and scanned at 400 dpi using an EPSON scanner (Expression 1680) with Silverfast (Ai V6.22r4) scanning software. The images were saved in JPEG format and modified in Adobe Photoshop® Elements 9 software to normalize the mean RGB (Red Green Blue) value to 125 with Red = 129, Green = 125, and Blue = 121; Luminosity = 126 and colour = 125. ImageJ software (ImageJ, 2013) was used to select the seed area in images to measure the average colour intensity. Finally, the mean of eight independent measurements for each of the 8 seeds from each RIL was expressed in RGB value. The seed colour values for RILs that were grown under growth chamber conditions ($F_{8:9}$) and greenhouse conditions ($F_{9:10}$) were obtained.

Flowers from plants 70-80 days after sowing were observed mid-morning (before 10:00 AM). Flower colours observed were blue and white and these were tabulated against the individual RIL.

3.5 Materials and reagents used in nucleic acid work

The following equipment was used in this procedure

Equipment	**Source**
Centrifuge	Eppendorf Centrifuge 5417C
Concentrator	Eppendorf Concentrator 5301
Incubator	Fisher Isotemp® Oven 126G
UV Spectrophotometer	Thermo Scientific NanoDrop 8000

The following reagents were used in this procedure

Reagent	**Source**
Chloroform	OmniSolv® VWR Cat. No. EM-CX1054-6
Isoamyl alcohol	Fisher Scientific Cat. No. A393-500
Isopropanol	Fisher Scientific Cat. No. A451-4
Ethanol	JT Baker Cat. No. JT92291
NaCl	EMD Cat No. SX0420-3
2-mercaptoethanol	Sigma Cat. No. M3148-100ML
Tris Base	Fisher Scientific Cat. No. BP152-1
EDTA	Fisher Scientific Cat. No. E478-1
CTAB	Sigma Cat. No. H6269-500G
Hydrochloric acid	Fisher Scientific Cat. No. SA48-1
SDS	Bio-Rad Cat. No. 161-0302
NaOH	Anachemia Cat. No. 83076-380

3.5.1 Buffers and solutions

The buffers and reagent solutions were as follows:

CTAB Extraction Buffer	**CTAB Extraction Buffer (2%) – for 1 L** • 10 g CTAB • 100 mL 1 M Tris HCl, pH 8.0 • 40 mL 0.5 M EDTA, pH 8.0 • 280 mL 5M NaCl • Add H_2O to 1 L

TE Buffer (10:1 mM)	**TE Buffer, pH 8.0 – for 250 mL** • 2.5 mL 1 M Tris HCl, pH 8.0 (10 mM) • 0.5 mL 0.5 M EDTA, pH 8.0 (1 mM) • Add H_2O to 250 mL

70 % (v/v) Ethanol	**70 % (v/v) Ethanol – for 500 mL** • 350 mL 100% ethanol • 150 mL sterile H_2O

3.6 DNA isolation protocol

Following are the steps to extract the DNA from flax leaves

Step	Action
1	Frozen leaf samples were ground with a bead beater in centrifuge tubes and for each sample, 800 µL of pre-warmed (55 ºC) CTAB extraction buffer including 2% (v/v) 2-mercaptoethanol was added.
2	Tubes were incubated for 15 min at 55 ºC and cooled at room temperature for 10 min.
3	Equal volume (800 µL) of chloroform: isoamyl alcohol (24:1) was added and mixed vigorously.
4	Tubes were centrifuged at 12000 rpm (rotations per minute) for 10 min and the upper aqueous phase was transferred into a clean tube.
5	Steps 3 and 4 were repeated.

6	DNA was precipitated in 2/3 volume isopropanol at -20 ºC for 20 min.
7	DNA was pelleted by centrifuging tubes at 13000 rpm for 15 min at 4 ºC.
8	The pellet was washed with 800 µL of 70 % ethanol and centrifuged at 12000 rpm for 5 min.
9	Ethanol was decanted and the pellet was dried in a concentrator at room temperature for 3 min to remove traces of ethanol.
10	The pellet was re-suspended in TE buffer and DNA was quantified using a UV spectrophotometer.
11	The DNA of each sample was diluted to a final concentration of 50 ng/µL for further use.

3.7 RNA isolation from seed coat tissue and flower petals

Isolation of RNA from flower petals was done using an Ambion RNAqueous kit following the manufacturer's protocol. The eluted RNA was quantified using a NanoDrop 8000 and stored at -80 ºC until it was used for cDNA synthesis. The RNA from seed coat samples were isolated by a manual method. A schematic representation of RNA isolation from seed coat tissue is shown in Figure 3.1.

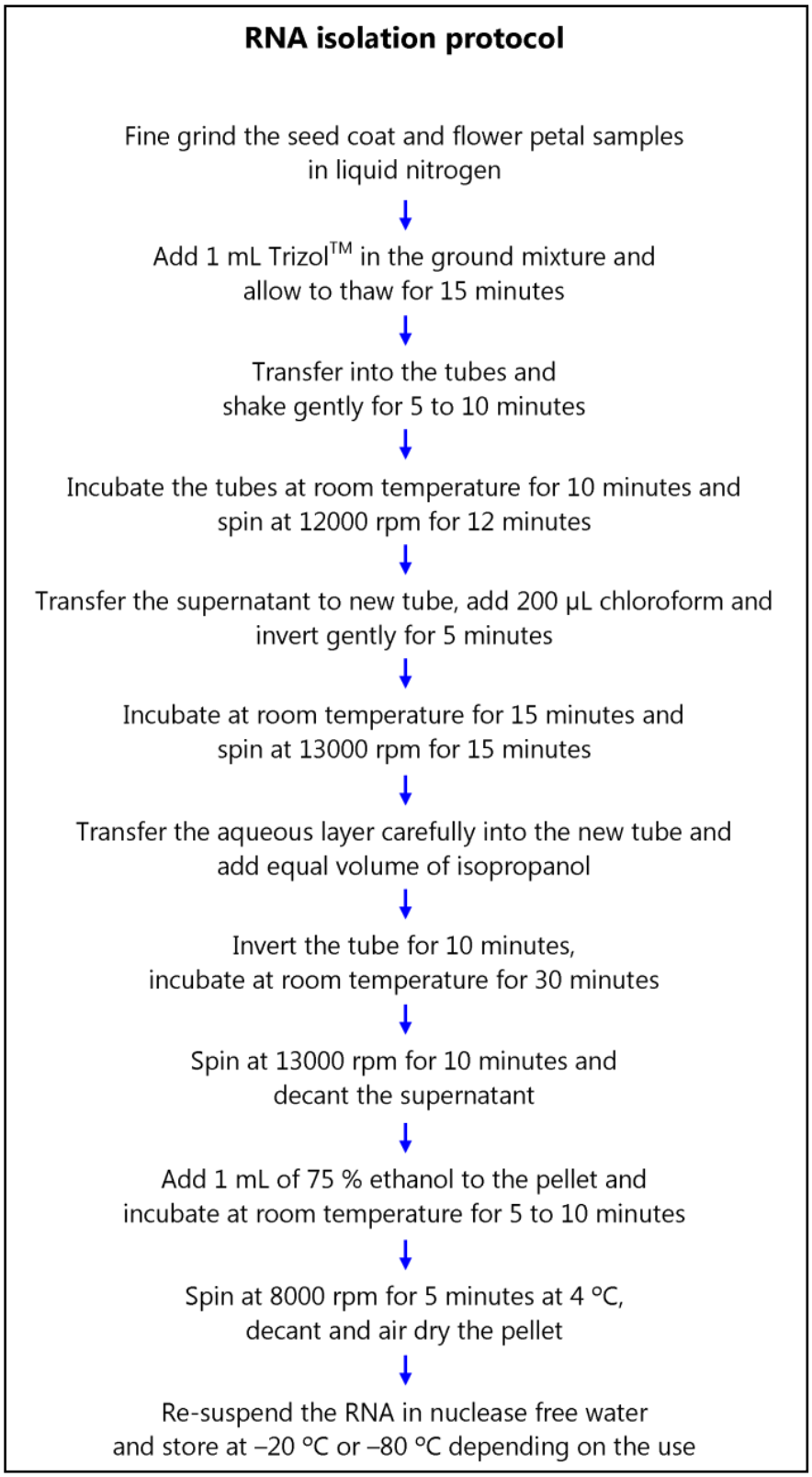


Figure 3.1: Schematic flowchart of the RNA isolation protocol

3.8 Complementary DNA (cDNA) synthesis

The seed coat and flower RNA of parental lines were used to synthesize cDNA in individual reactions. AffinityScript™ QPCR cDNA Synthesis Kit from Agilent Technologies having reverse transcriptase (RT) enzyme was used to synthesize cDNA following the manufacturer's instructions. 1 µg RNA (6 µL) was used in a total reaction volume of 20 µL. In the reaction mixture, 10 µL of 2x master mix, 2.5 µL of oligodT, 0.5 µL of random primer and 1 µL of Affinityscript RT enzyme was added and mixed gently. RT was performed in a Techne Genius thermal cycler (Model: FGEN05TP) with an initial incubation at 25 °C for 5 min, followed by 42 °C for 5 min, 50 °C for 45 min and final incubation at 95 °C for 5 min. The final product was used for gene expression analysis as outlined below.

3.9 Quantitative PCR (qPCR) of putative candidate gene

The putative F3'H gene sequence was obtained from the genome sequence database. Primers used for qPCR experiments were designed using Primer3 software (Rozen and Skaletsky, 2000) in the third exonic region of the putative gene with an expected amplicon length of 158 bp. The gene specific and reference gene specific primers are listed in Table 3.1. qPCR was performed using SYBR GreenER™ qPCR SuperMix Universal (Invitrogen, Karlsruhe, Germany; Cat no. 11762-500) following manufacturer's instructions. For qPCR, cDNA from seed coat and flower petals from three biological replicates were run twice each. Elongation factor (EF1α) was used as a reference gene (Roach and Deyholos, 2008). qPCR reactions were performed on a StepOnePlus™ Real Time PCR system and data was analyzed using StepOne v2.1 software (Applied Biosystems).

A total volume of 50 µL per reaction contained 25 µL SYBR GreenER Master Mix reagent (Applied Biosystems) (@ 2X), 2 µL cDNA, 1 µL of forward and reverse primer each (@ 10 µM), 1 µL ROX (reference dye, Applied Biosystems) and 20 µL of double distilled sterile H_2O. During the qRT-PCR reaction, the cDNA template was denatured at 95 °C for 10 min followed by a two-step protocol with 40 cycles of 95 °C for 15 s and 58 °C for 1 min. After that, melting curve analysis was started with incubation at 95 °C for 15 s then at 60 °C for 1 min with a gradual increase in temperature (0.3 °C /15 s) to 95 °C for 15 s. Time changes in fluorescence were recorded during this time. Relative changes in gene expression were calculated using the $2^{-\Delta\Delta CT}$ method (Livak and Schmittgen, 2001) with the 15 DAF seed coat samples and petal samples of G1186/94 as the calibrator and EF1α for normalization. Data were collected from the StepOne Software v.2.1 (Applied Biosystems) and analyzed in Excel.

Table 3.1: List of gene specific and reference gene primers used in qPCR analysis

	Primers	Sequence	Length (bp)	Tm (°C)
Gene specific primers	F3'H_q_F	AGCTGATGACGGCTGTTCTT	20	60
	F3'H_q_R	ATAAACATGCTCCGCCAATC	20	60
Reference gene (EF1α) specific primers	α_F1qEF	TTGGATACAACCCCGACAAAA	21	60
	α_R1qEF	GGGCCCTTGTACCAGTCAAG	20	60

3.10 Genetic markers

3.10.1 Simple sequence repeats (SSR) marker identification

A minimum of six repeats of dinucleotide motifs, or three repeats for tetra, penta and hexanucleotide motifs were sought and primers for the flanking regions were designed using 'WebSat' online software (Martins et al., 2009). All the primer pairs (forward and reverse) were

designed with a melting temperature of approximately 58 °C, which will be helpful for multiplexing. All forward primers were synthesized using the header sequence CACGACGTTGTAAAACGAC for fluorescence detection.

3.10.2 Polymerase chain reaction (PCR) for SSR markers

One-hundred ng of genomic DNA was used as a template for SSR (simple sequence repeat; also referred to as microsatellite) amplification for each sample in a total volume of 10 µL per reaction. Each reaction had 2.0 µL genomic DNA (@ 50 ng/µL), 1 µL PCR buffer (@ 10X) (provided with Paq5000 DNA polymerase kit; Agilent Technologies), 0.4 µL dNTPs (@ 10 mM), 0.3 µL forward primer (@ 1 µM), 0.2 µL reverse primer (@ 10 µM), 0.18 µL fluorescently labeled primer (FAM, HEX, NED or PET) (@ 10 µM) that will anneal to the "header" sequence and produce fluorescently labeled amplicons, 0.35 µL Paq5000 DNA polymerase and 5.57 µL double distilled sterile water. PCRs were performed in a Techne Prime thermal cycler (Model: TC 5000, Bibby Scientific Limited) with an initial denaturation at 93 °C for 2 min, followed by 30 cycles of 94 °C for 1 min, annealing at 58 °C for 50 s (with decrease of 0.5 $°Cs^{-1}$ from 94 °C to 58 °C), extension at 72 °C for 1 min (with an increase of 0.5 $°Cs^{-1}$ from 58 °C to 72 °C) followed by a final extension of 72 °C for 5 min. PCR products having VIC: FAM: PET: NED dye reactions were pooled together in proportion of 2: 2: 3: 3 parts respectively, for multiplexing before genotyping. 1.0 µL aliquots of fluorescent-labeled PCR products or multiplexed PCR products were mixed with 3.9 µL of Hi-Di formamide and 0.1 µL of Genescan Liz-500 standard, denatured 5 min at 95 °C and cooled on ice for 5 min before being resolved on an ABI 3730 xl DNA analyzer (Applied Biosystems). The run time for 96 wells was about 25 min followed by 30 min washing of capillaries. Output files (".fsa") were analyzed by Genographer 2.1.4 (Genographer, 2013) and converted into a gel-like image

(".GEL" files). Each SSR PCR product size was estimated using GeneScan Liz-500 internal size standard. Representative images of identification of polymorphic SSR markers between the parents and genotyping in RILs are shown in Figure 3.2 and Figure 3.3, respectively. The list of polymorphic SSR markers and SSR markers used for screening are provided in Appendix 1 and Appendix 2, respectively.

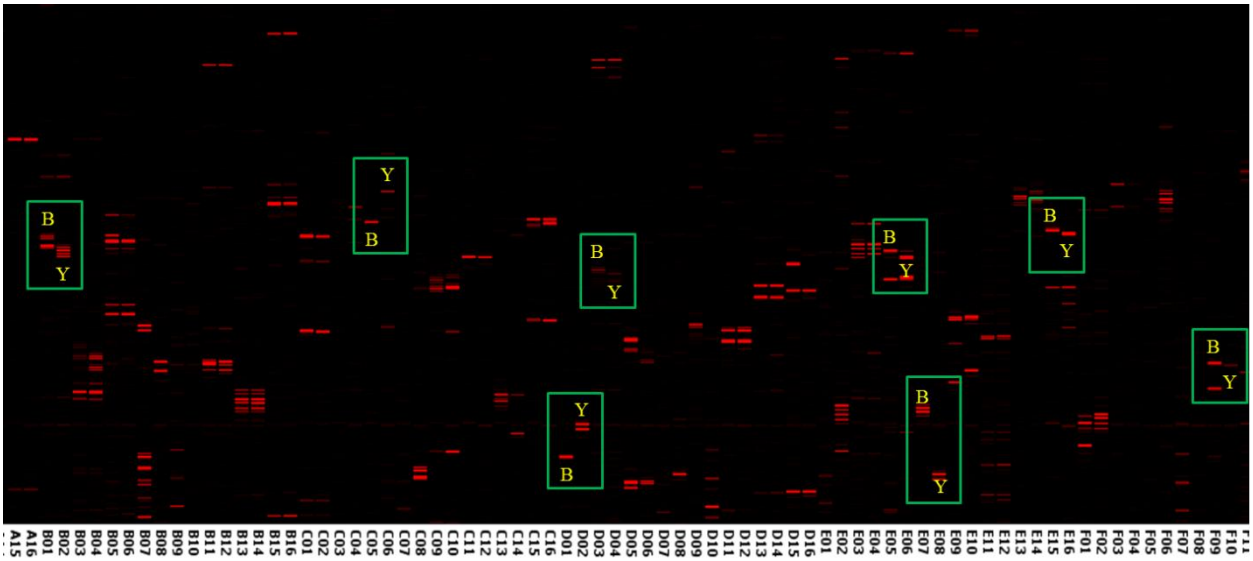


Figure 3.2: A virtual gel-like image showing polymorphic markers in the parents as a pair of lines used for a given target SSR amplification. The boxes identify polymorphic SSR that differ in size as resolved in ABI 3730 xl; the monomorphic or inconclusive pairs are not highlighted. 'B' indicates the variant allele corresponding to the brown-seeded parent CDC Bethune and 'Y' indicates the variant allele amplified from the yellow-seeded parent (G1186/94).

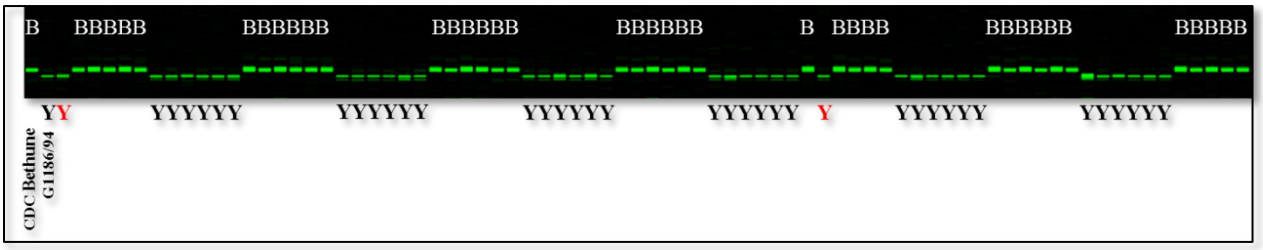


Figure 3.3: A virtual gel-like image showing the sample SSR genotyping and scoring in the mapping population. The conversion of output files from genotyping machine to gel-like image is explained in section 3.10.2. Here, VIC is used for the fluorescent detection of the alleles. Scoring of B represents brown seed and Y represents yellow seed genotype. 'Y' (in red colour) indicates the yellow type recombinants in the brown-seeded lines.

3.10.3 Finding putative single nucleotide polymorphisms (SNPs) in genomic DNA sequences

The data used were from specific scaffold sequences from the whole genome assembly of CDC Bethune (TUFGEN, 2012; Wang et al., 2012) and the reduced genome representation (RGR) sequence assembly of G1186/94 (data files were provided by Dr. Sylvie Cloutier). The former is the reference sequence. The genome sequence of G1186/94 was provided in the form of ".BAM" and ".bai" files. BAM files are the binary version of the Sequence Alignment/Map (SAM) files (Li et al., 2009) and sorted by position; indexed file of a BAM file is appended by ".bai" with the bam file name. By comparing the reference sequence and G1186/94 sequence, in CLC Genomics Workbench software (CLC bio), putative SNPs were detected *in-silico* with the help of Dr. Paula Ashe. At this point, a putative SNP was scored as such if it was present in at least two reads.

3.10.4 Development of CAPS (Cleaved Amplified Polymorphic Sequences) markers

One hundred bp of sequence from either side of the putative SNP was retrieved from the two genomic sequences (CDC Bethune and G1186/94). *In-silico* restriction digestion analysis

(using all commercially available restriction enzymes) was performed in CLC Genomics Workbench software using the restriction digestion analysis tool. The SNPs (in both parents) that affect restriction enzyme recognition sites were identified virtually. The CAPS with high read-counts (in the assembly file) were further selected as putative CAPS. A snapshot of the SNP marker is provided in Figure 3.4 using TABLET software (Milne et al., 2010). Then, 300 bp of either side of the putative SNP was extracted from the genomic database. PCR primers for these sequences were designed such that the product would be 500-550 bp. If the SNP did occur in the genomic DNA, the amplicons from the two lines would have different restriction enzyme digestion patterns (cleavage-positive and cleavage-negative).

The restriction enzyme digestion patterns were tested as follows. Fifty ng of the genomic DNA was used as a template for CAPS amplification for each sample in a total volume of 25 µL per reaction. Each reaction had 1.0 µL genomic DNA (@ 50 ng/µL), 2.5 µL PCR buffer (@ 10X), 0.75 µL $MgCl_2$ (@ 50 mM), 0.6 µL dNTPs (@ 10 mM), 0.5 µL forward primer (@ 10 µM), 0.5 µL reverse primer (@ 10 µM), 0.25 µL Taq polymerase and 18.9 µL double distilled sterile water. PCRs were performed on a Techne Prime thermal cycler (Model: TC 5000, Bibby Scientific Limited) with initial denaturation at 94 °C for 5 min, followed by 35 cycles of 94 °C for 30 s, annealing at 60 °C for 30 s, extension at 72 °C for 30 s, and a final extension of 72 °C for 8 min at the end of the 35 cycles. The PCR product was treated with a diagnostic restriction enzyme in a total volume of 15 µL per reaction under the conditions specified by the supplier of the enzyme. Five µL of the reaction solution, after completing and terminating the reaction, was mixed with 2 µL of loading dye (@ 1X) and resolved on an agarose (2 % w/v) gel including GelRed nucleic acid stain, 10,000X in water (Biotium, Inc.) (0.005 % v/v). Gel electrophoresis was performed using Tris acetate EDTA (TAE) buffer (@ 1 X) at 50 volts. The gel was

visualized under ultra violet (UV) light in a gel documentation unit (BIO-RAD Gel Doc 2000) and the image was captured.

The stock buffer solution used for agarose gel electrophoresis is as follows:

50X TAE Buffer	**50X TAE Buffer – for 1 L** • 242 g Tris Base • 100 mL 0.5 M EDTA, pH 8.0 • 57.2 mL Glacial acetic acid • Add H_2O to 1 L

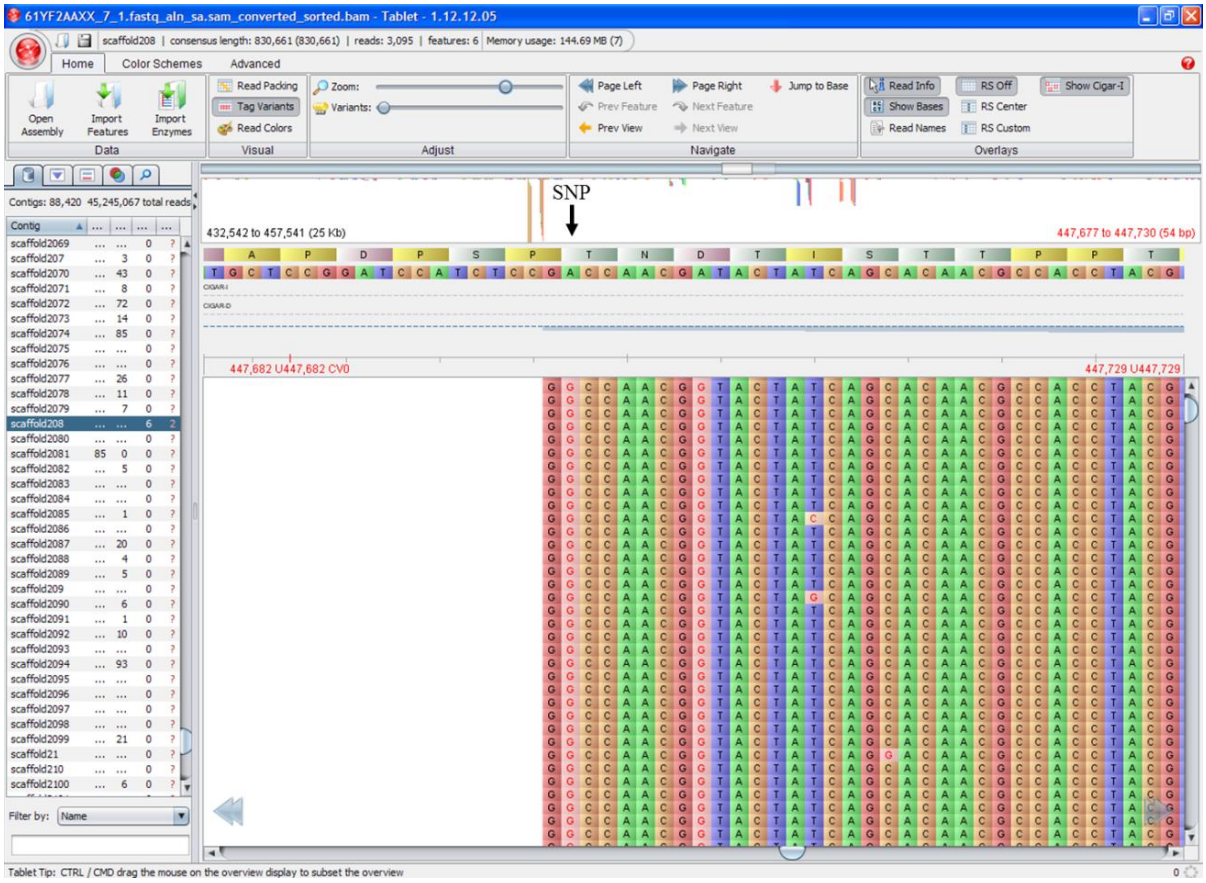


Figure 3.4: A screen shot of the SNP marker in TABLET software. Here, scaffold208 is used as a reference sequence (highlighted in green colour in left pane) to align G1186/94 sequences. The arrow indicates the SNP, where CDC Bethune has 'A' nucleotide and G1186/94 has 'G' in multiple reads.

3.11 Linkage mapping and seed colour locus detection

The expected ratio (1:1) of brown: yellow seed colour in the RIL population [based on proposed one gene model for the European recessive yellow mutation (Mittapalli and Rowland, 2003)] was tested by chi-square analysis using Yates' correction. The method (described by Cloutier et al., 2011) for linkage analysis and locus detection is as follows. Linkage analysis was performed for the marker data using MAPMAKER/EXP 3.0 (Lander et al., 1987). First, data was prepared by 'prepare data' command using the '.raw' marker data file. Markers were grouped by using the "group" command with a minimum logarithm of odds (LOD) score of 3.0 and maximum recombination factor of 40. When the linkage group was formed, the markers in the group were arranged in order using the "order" command. Then the "ripple" command was used to verify the order of the markers. Later, Kosambi (Kosambi, 1943) mapping function was used to get the genetic distances expressed in cM between the markers in a linkage group.

After this, the phenotypic data of seed colour and genotypic data were arranged and subjected to locus identification using QTL Cartographer v2.5. Then, Composite Interval Mapping (CIM) was performed using a Kosambi function of genetic distances with a walk speed of 1 cM and cofactors were chosen using the 'forward and backward regression' method. The locus was identified at the highest LOD score which was higher than the threshold value (LOD of 2.5) in the data set. The CIM analysis gave estimates of the percentage of phenotypic variance controlled by the identified locus in the form of coefficient of variation (R^2).

3.12 Genomics-assisted gene search

Gene prediction analysis was performed on a given genomic sequence using FGENESH online software (FGENESH, 2012). For the analysis, *Hevea* was used as a model plant for gene

prediction as it was found to be closely related to *Linum usitatissimum* (Lee et al., 2011). Then, predicted genes were annotated using TAIR10 database. The possible putative candidate gene was identified based on the enzyme function description that shows an involvement in the flavonoid biosynthesis pathway. The list of predicted genes in the region from start to close to the SSR marker (on scaffold208) in the *D* locus region is given in Table 4.5 of Chapter 4.

4. RESULTS

4.1 Phenotyping of Recombinant inbred lines (RILs) and their parents

4.1.1 Production of single seed descent lines of a recombinant inbred line population for further analysis

The seed colour was a criterion in selecting flaxseed lines. Parental lines CDC Bethune and G1186/94 were selected for the current study. CDC Bethune is a brown seeded variety (Rowland et al., 2002) which is the most extensively cultivated variety in Western Canada (SaskFlax, 2007). G1186/94 is a yellow-seeded line (hereafter "yellow line") with a mutation at the *D* locus for seed colour and is called a European recessive yellow line (Mittapalli and Rowland, 2003). The parental lines with a difference in seed and flower colour are shown in Figure 4.1. A population of recombinant inbred lines (RILs) from selfing of F_1 plants from a cross between G1186/94 X CDC Bethune and its reciprocal cross followed by single-seed descent method was developed by Dr. Gordon Rowland at the Crop Development Centre, University of Saskatchewan, Canada. This RIL population contains 479 lines of $F_{8:9}$ generation. To initiate the current study, the seeds of the RIL population were obtained from Dr. G. Rowland. For matching the genotype to phenotype, it is essential to have single plants whose seeds are phenotyped. It was essential to grow a single plant per line from the $F_{8:9}$ seed lot [growth chamber (GC), 2008] and use the plant for genotyping and the resulting $F_{9:10}$ seeds (GH, 2011) for phenotyping. Thus, there will be two phenotyping experiments that will be presented here. The segregation pattern of seed colour will also be presented. For this work, it was necessary to have a reproducible method to assign colour value to seeds. Visual rating was considered unreliable and operator-biased.

Figure 4.1: Seeds and flowers of parental lines. Left side, brown-seeded flax cultivar CDC Bethune produces blue petals with dark blue veins; and right side, yellow-seeded G1186/94 flax line produces white petalled flowers with blue veins.

4.1.2 Re-evaluation of a scanner-based method for seed colour assignment shows that it is well-suited for seed colour analysis.

Previous work in this laboratory had shown that single seeds in 96-well plates could be scanned in a document scanner in colour mode and the images could then be analyzed for red, green and blue (RGB) values that provided a reproducible digital value for seed colour (S. Nandy, P. Ashe and G. Selvaraj, unpublished). This method had also been subsequently used by Dr. M. Kulkarni in the laboratory to distinguish seed colour segregation in another flax

population. Note that a similar method was also published for assigning colour values to *Brassica napus* seeds where they transformed RGB values to HSB (hue, saturation and brightness) values; then YSD (yellow-seeded degree) was calculated by setting the YSD value of black-seeded parent to zero (Fu et al., 2007). Using the method developed in the laboratory (described in Materials and Methods) where RGB values were considered directly by normalizing all the scanned images with mean RGB and luminosity, the $F_{8:9}$ and $F_{9:10}$ RIL populations were screened for seed colour. A representative figure with 12 lines is shown here (Figure 4.2).

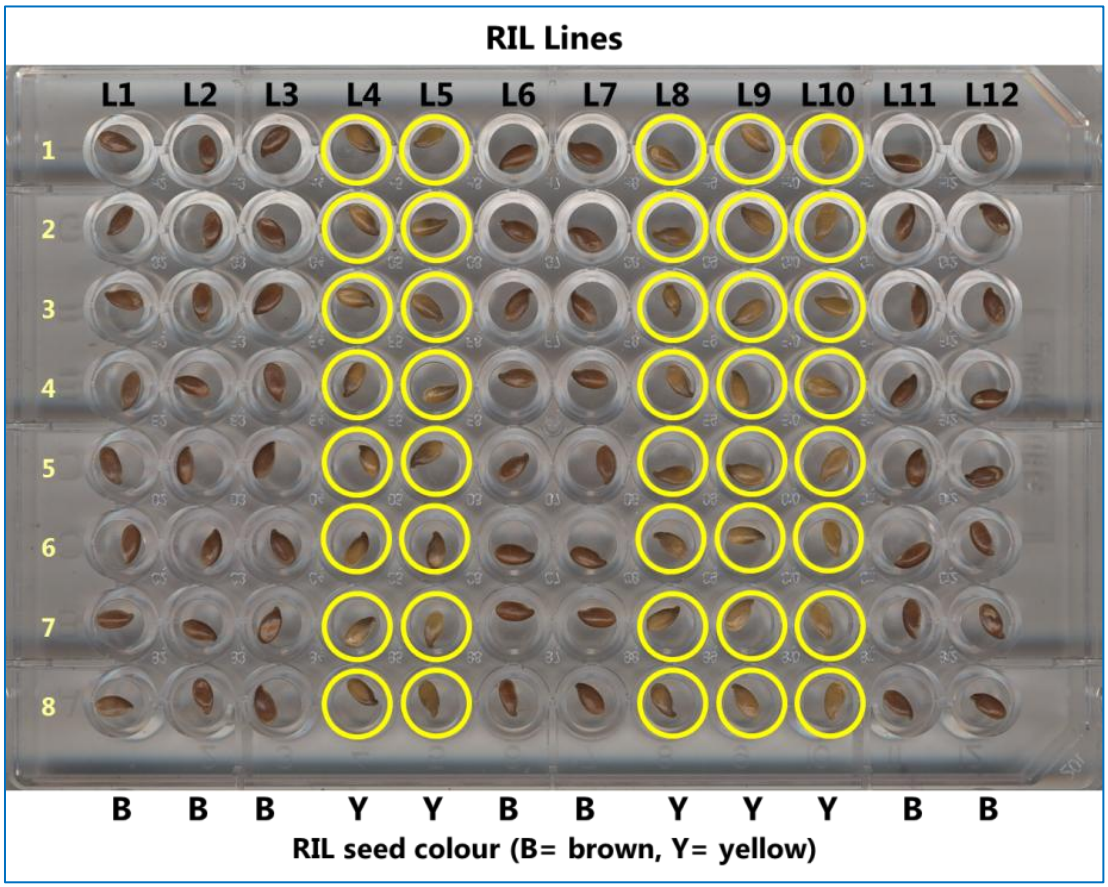


Figure 4.2: A scanned photograph of a representative 96 well plate having 8 seeds (rows) for each of 12 flax lines (L1 to L12 in columns). The plate contains both brown and yellow-seeded lines (GC, 2008). The wells with yellow circles have yellow seeds in them and others have brown.

The examples of seed colour and their corresponding RGB values for the parental lines (GC, 2008) are shown in Table 4.1.

Table 4.1: Seed colour values (RGB) of individual seeds and mean RGB values corresponding to their seed colour in parental lines CDC Bethune and G1186/94 (GC, 2008).

Parental lines	Seed colour	RGB values of individual seeds								Mean RGB value
		Seed 1	Seed 2	Seed 3	Seed 4	Seed 5	Seed 6	Seed 7	Seed 8	
CDC Bethune	Brown	57.3	56.7	56.2	56.9	56.1	56.8	57.1	56.5	56.7 ± 0.4
G1186/94	Yellow	99.6	98.5	99.8	98.6	100.5	104.8	98.0	94.2	99.3 ± 2.9

Since the parental lines differed in flower colour drastically as CDC Bethune flowers were blue-petalled and G1186/94 flowers were white-petalled, phenotypic linkage of seed colour and flower petal colour was also examined with petal color being scored as blue or white (details presented later in this chapter). This examination was based on a previous study in an F_3 population of a cross between CPI84495 (dominant yellow-seeded female parent, producing blue flower petals) and G1186/94 (male parent) where white or pink flower petals were observed in yellow-seeded homozygous rows having the *d* locus, indicating its role in conditioning both seed and flower colour (Mittapalli, 2002).

4.1.3 Segregation of seed colour in mapping populations

The $F_{8:9}$ RIL seeds of G1186/94 X CDC Bethune and its reciprocal cross were measured for seed colour values. The mean seed colour values of all RILs are listed in Appendix 3.

The frequency distribution of the 479 $F_{8:9}$ RIL seeds grown in a fully controlled growth chamber showed a bimodal distribution suggesting that the seed colour trait is controlled by a single major gene. The frequency distribution of the seed colour values is shown in Figure 4.3. The $F_{9:10}$ RIL seeds that were grown in a greenhouse also showed a bimodal distribution,

separating into clear brown and yellow seed colour. In the case of the $F_{9:10}$ RILs, 16 of the 479 RILs from $F_{8:9}$ did not germinate or did not set seeds and therefore the remaining 463 lines were evaluated. All yellow lines from $F_{8:9}$ were also yellow in $F_{9:10}$ and likewise all brown lines were brown in the second evaluation even though the growth environments were different. The frequency distribution is shown in Figure 4.4.

The ratio between brown seed to yellow lines in the G1186/94 X CDC Bethune cross was 1.46:1 (142 brown and 97 yellow RILs) and CDC Bethune X G1186/94 was 1.35:1 (138 brown and 102 yellow RILs) in the $F_{8:9}$ RILs. Mittapalli and Rowland (2003) proposed that seed colour is governed by a single genetic locus in G1186/94. Considering the expected segregation ratio of 1:1 for a single gene model, Chi-Square analysis (with Yates' correction) for the observed ratio showed a significant difference in both of the populations ($P < 0.05$ i.e, $P= 0.0044$ and 0.024, respectively). The average ratio for brown to yellow seeded lines was 1.41: 1 (280 brown and 199 yellow RILs). In the $F_{9:10}$ RILs, brown as to yellow seed line ratio in the G1186/94 X CDC Bethune cross was 1.43:1 (134 brown and 94 yellow) and in the CDC Bethune X G1186/94 cross was 1.33:1 (134 brown and 101 yellow). In this trial also, Chi-Square analysis gave a significant difference in both populations from the expected 1:1 ratio. The Chi-Square statistics for the two growth trials is summarized in Table 4.2.

Table 4.2: Chi-Square test for seed colour segregation in both growth chamber (GC) and greenhouse (GH) grown populations.

Growth trials	Parental cross	Brown RILs	Yellow RILs	Total RILs	χ^2	p-value	Brown to Yellow ratio
$F_{8:9}$ RILs (GC, 2008)	G1186/94 X CDC Bethune	142	97	239	8.1	0.004	1.46:1
	CDC Bethune X G1186/94	138	102	240	5.1	0.024	1.35:1
	Combined	280	199	479	13.4	0.0003	1.41:1
$F_{9:10}$ RILs (GH, 2011)	G1186/94 X CDC Bethune	134	94	228	6.7	0.009	1.43:1
	CDC Bethune X G1186/94	134	101	235	4.4	0.04	1.33:1
	Combined	268	195	463	11.2	0.0008	1.38:1

The range of RGB colour values within the bimodal distribution of the $F_{8:9}$ RILs was from 46.8 to 108.1, where the mean RGB colour value for brown lines was 58.3 ± 4.4 and for yellow lines was 89.3 ± 7.0. The brown parental line had a seed colour value of 56.7 ± 0.4 and the recessive yellow parental line G1186/94 had a value of 99.3 ± 3.0. See Appendix 3 for the data points of mean seed colour values of RILs in both growth trials (GC, 2008 and GH, 2011). The range of RGB colour values within the bimodal distribution of RILs in the $F_{9:10}$ RILs was from 44.0 to 106.2, where mean RGB colour value for brown lines was 59.6 ± 4.5 and for yellow lines was 90.2 ± 8.2. The brown parental line had a seed colour value of 63.5 ± 1.7 and the recessive yellow parental line G1186/94 had a value of 99.7 ± 3.1. There was 91.7 % correlation for RGB values between two growth trials analyzed for 463 RILs.

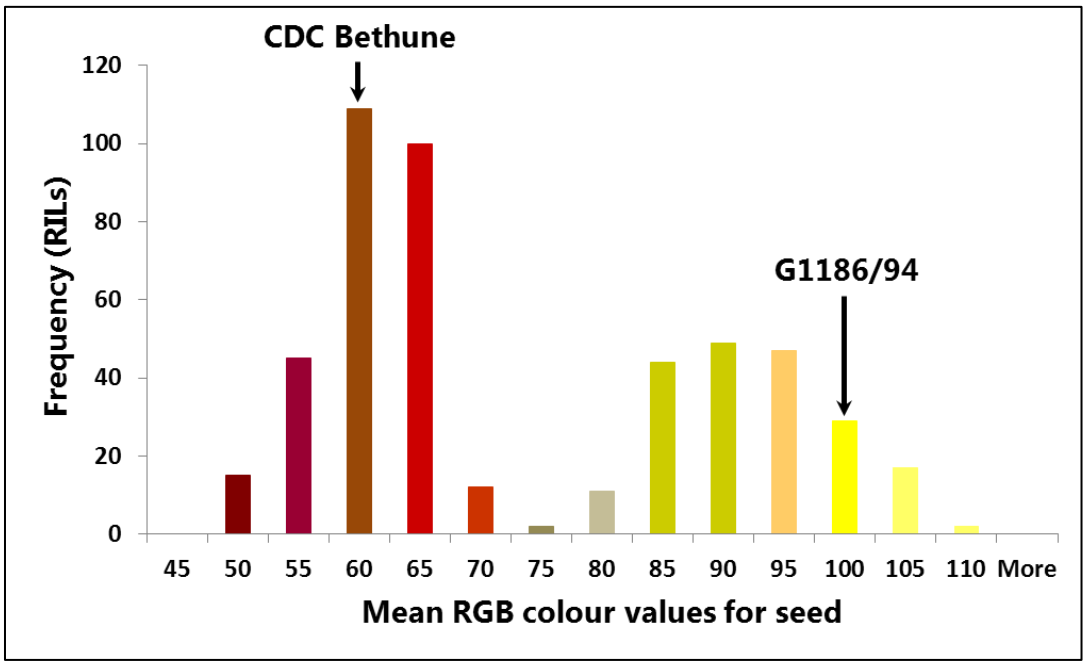


Figure 4.3: Phenotypic distribution of seed colour (mean RGB value) in flax G1186/94 X CDC Bethune RIL population grown in a growth chamber in 2008.

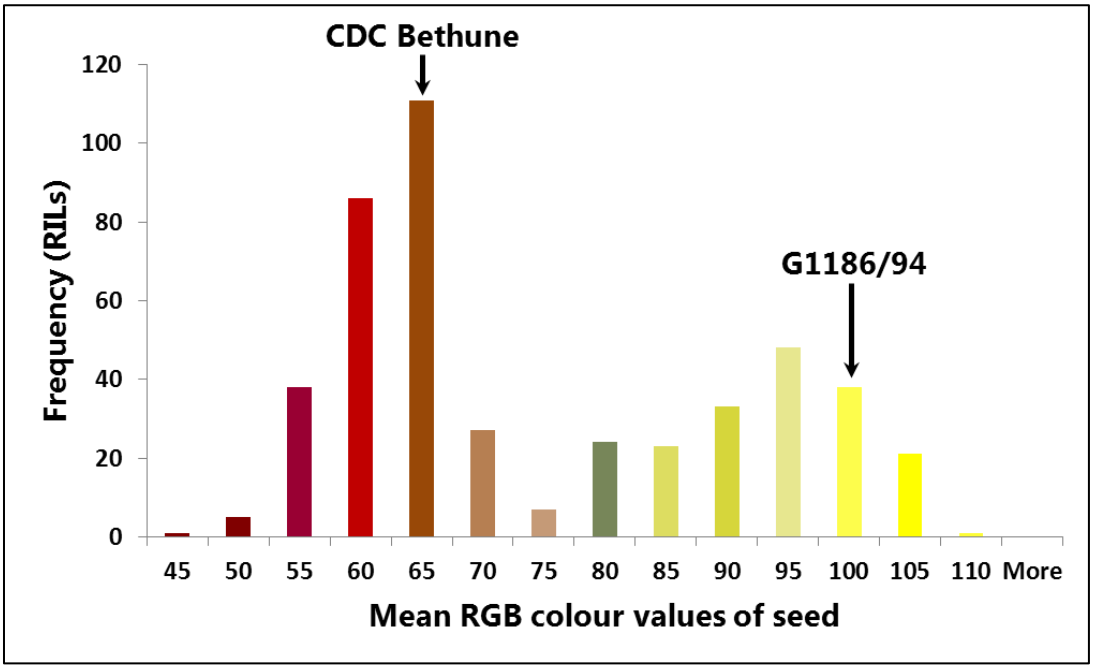


Figure 4.4: Phenotypic distribution of seed colour (mean RGB value) in flax G1186/94 X CDC Bethune $F_{9:10}$ RIL populations grown in the greenhouse (GH) at Innovation Place, Saskatoon in 2011.

There is a relationship between the *d* locus and flower colour. Mittapalli (2002) reviewed as well as observed that the mutation at the *D* locus is also linked to change in flower colour i.e., causes white or pink coloured flower petals in G1186/94. In accordance, during this investigation it was observed that there was a tight linkage between yellow seed colour and white flower colour. During the growth trial in 2011, 9 of the 479 plants either did not flower or died pre-maturely. So, observations were taken for a total of 470 RILs. Therefore, in a mapping population of total 470, there was no recombinant which had yellow seed and blue flower nor brown seed and white flower. Phenotypic data is given in Appendix 3. As the seed colour was mapped to the *D* locus, it was also evident that the same genetic locus also governed flower colour mutation.

4.2 Screening of SSR markers for polymorphism in parental lines and development of framework genetic map

Initially, a total of 477 genomic SSR markers (identified by Dr. M. Kulkarni) were screened for polymorphism between CDC Bethune and G1186/94. I found 74 polymorphic SSRs (15.5 %). A list of polymorphic markers is provided in Appendix 1. Just prior to using these polymorphic markers, I became aware of the construction of linkage groups (LGs) from the personal communication about the three bi-parental mapping populations (CDC Bethune/Macbeth, E1747/Viking and SP2047/UGG5-5) in Dr. S. Cloutier's laboratory. This laboratory had also generated a list of polymorphic markers in G1186/94 that would correspond to CDC Bethune and thus positioned on potential LGs. Note that these markers had not been tested for genetic linkage in the CDC Bethune X G1186/94 population. We obtained 148 such markers spanning a collection of preliminary linkage groups (shown in Table 4.3). Of this, I chose 73 that would span all these preliminary LGs (highlighted with yellow colour in Table 4.3). Of these, 4 did not amplify (marked with an asterisk in Table 4.3) but the remainder were confirmed to be polymorphic between CDC Bethune and G1186/94.

Table 4.3: Polymorphic SSR markers (148) spanning a collection of preliminary linkage groups obtained from Dr. S. Cloutier.

Linkage groups having LG specific markers														
LG1	LG2	LG3	LG4	LG5	LG6	LG7	LG8	LG9	LG10	LG11	LG12	LG13	LG14	LG15
Lu2268	Lu390	Lu587	Lu943	Lu318	Lu69	Lu125	Lu146	Lu682	Lu197*	Lu2746	Lu3033	Lu91	Lu291	Lu462a
Lu2428	Lu2913		Lu2010b	Lu1039	Lu442a	Lu209	Lu138	Lu330	Lu650	Lu371b	Lu514	Lu283	Lu575	
Lu2921	Lu1052		Lu3053	Lu1161	Lu699	Lu824	Lu151	Lu652	Lu805	Lu458	Lu3219a	Lu757	Lu512	
Lu447	Lu3064		Lu2183a	Lu2628	Lu502	Lu532	Lu235	Lu176		Lu2728	Lu813	Lu798	Lu785	
Lu625*	Lu381		Lu47	Lu774	Lu2917b	Lu257	Lu296	Lu2014		Lu3007	Lu3038	Lu801	Lu325	
Lu2587	Lu787b		Lu46*	Lu2767		Lu910	Lu672	Lu223		Lu273	Lu476		Lu292	
Lu2618	Lu2779		Lu2687	Lu2764		Lu344	Lu585b	Lu227		Lu2052	Lu461			
Lu2317	Lu2774		Lu2698	Lu64			Lu566	Lu643		Lu2162	Lu613			
Lu2316	Lu2775a		Lu868	Lu787a				Lu361*		Lu685	Lu959			
Lu2312	Lu2778		Lu2055	Lu2194										
Lu2105	Lu2090		Lu796	Lu2161										
Lu857	Lu2081		Lu299	Lu2047										
Lu2203	Lu2773		Lu56	Lu2044										
Lu2923	Lu728a		Lu49b	Lu2040										
Lu2056	Lu263			Lu2049										
Lu2957	Lu1135			Lu2164										
Lu3157	Lu896			Lu3111										
Lu3059	Lu439			Lu639										
Lu1171	Lu2612			Lu638										
Lu339	Lu2600			Lu3153										
Lu2840	Lu2996			Lu933										
	Lu220			Lu2706										
	Lu140			Lu658										
	Lu628			Lu558										
				Lu3144										
				Lu422a										
LG8	LG12	LG4	LG1	LG3	LG6	LG2	LG7	LG5	LG13	LG10	LG14	LG9	LG11	LG15
Corresponding linkage groups published in Cloutier et al (2012a)														

MAPMAKER/EXP 3.0 (Lander et al., 1987) clusters co-segregating markers in a mapping population at estimated frequencies based on the genetic recombination and stringency of the method can be set for various LOD values and recombination factors. For example, if LOD is 3 and recombination factor is 30, it means that the two markers (or loci) are considered to be linked whenever the recombination fraction between them is less than 30 % and the maximum LOD exceeds 3.0. In this process, an order of the markers can also be obtained. Generally, such genetic scaffolding would be done at the outset to derive a working group of markers arranged in LGs. I used a subset of 94 lines from the population of 463 lines in my

further work. These 94 lines were sampled as follows: 47 yellow lines and 47 brown lines were picked at random from a collection of 195 yellow and 268 brown lines, respectively. The RGB colour index for the chosen lines is provided in Appendix 4. For the remainder of the work, these 94 lines constituted the mapping population. The coarse genetic linkage map based on 69 markers tested on this population is shown in Figure 4.5. At this point, Cloutier et al. (2012a) published their linkage group analysis consisting of 15 LGs developed by combining the data of three mapping populations (CDC Bethune/Macbeth, E1747/Viking and SP2047/UGG5-5). The correspondence of the two LGs (published LGs highlighted in blue colour) is shown in Table 4.3.

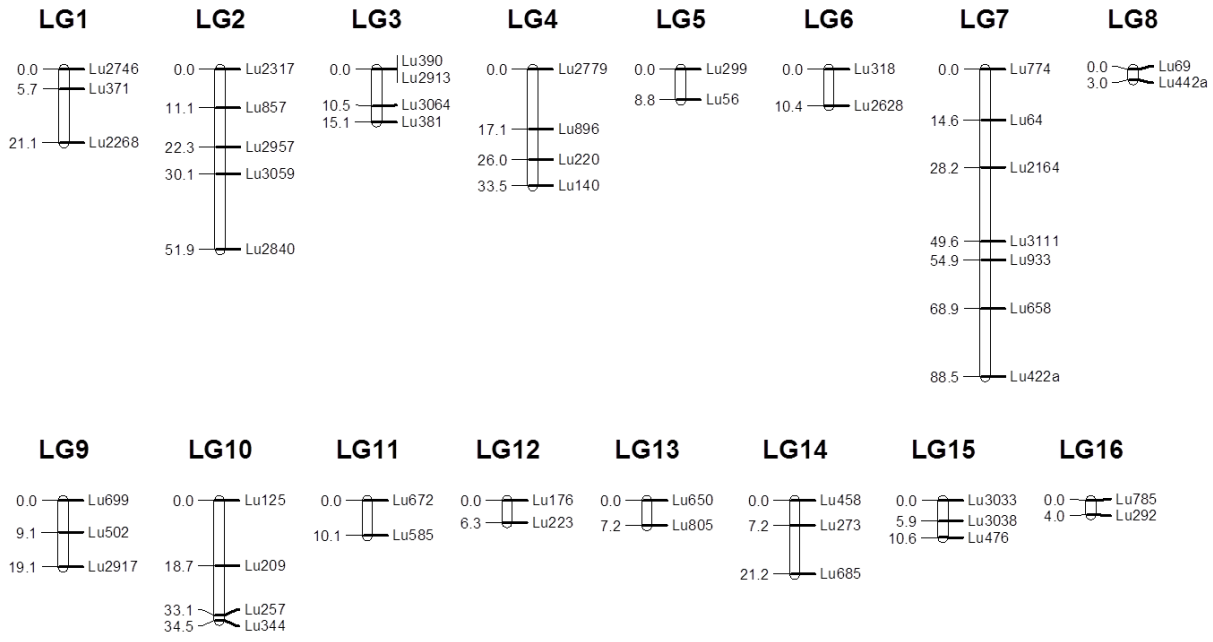


Figure 4.5: First framework genetic linkage map of CDC Bethune X G1186/94 consisting of 50 SSR markers forming 16 linkage groups. Nineteen markers were found to be unlinked.

Because there were fewer markers for LG 4, LG 6 and LG 15 of Cloutier et al. (2012a) in the marker collection I had used up to this point, 109 additional primer pairs were re-synthesized

for those SSRs that are not included in Figure 4.5. These primers were based on Cloutier et al. (2012a) and were re-synthesized from the previously published sources (Roose-Amsaleg et al., 2006; Cloutier et al., 2009; Deng et al., 2010; Cloutier et al., 2011; Deng et al., 2011; Cloutier et al., 2012a, 2012b). Of these 109 primer pairs, 51 (46.7 %) were found to be polymorphic between CDC Bethune and G1186/94. From these 51, 16 markers that spanned an interval of no more than 25 cM within LG 4, LG 6 and LG 15 of Cloutier et al. (2012a) were added to the LG analysis described in Figure 4.5. The updated LG is shown in Figure 4.6.

4.3 Linkage mapping shows association of the *D* seed colour locus to LG 1

Linkage mapping with the additional markers gave LGs. During the analysis a hypothetical marker "d" representing the *D* locus was considered to identify the statistically significant association with the markers in LGs. Out of 86 markers (includes "d"), 69 formed 19 LGs and 17 were unlinked. The framework genetic map spanned a total genetic distance of 469.3 cM with 68 markers and a hypothetical marker (d). The average marker density in this genetic map was 6.8 cM per marker. With LOD 3.0 and recombination factor of 40 % settings, 4 markers of LG 1 (of my study) were found to be associated with the "d" locus (for seed colour *D* locus). The closest marker (Lu209) on one side was at 2.3 cM from the *D* locus and Lu125 was on the other side of the D locus at a distance of 14.3 cM. Hence, markers associated with the seed colour locus in the mapping population were identified. LG 1 with the "d" locus (pointed with brown arrow) and the other 18 LGs are shown in Figure 4.6. The list of markers used to create the frame work genetic map and their corresponding LGs in Cloutier et al. (2012a) is given in Appendix 5.

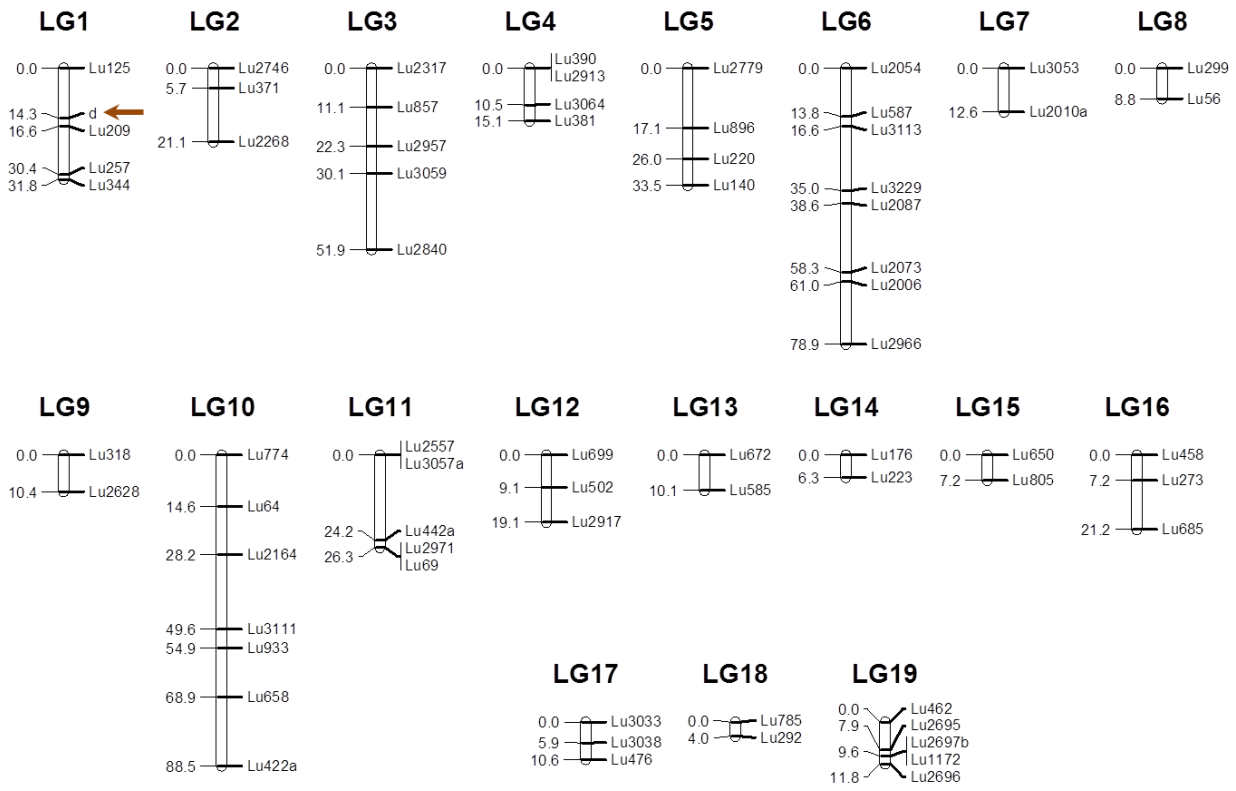


Figure 4.6: Second framework genetic linkage map of CDC Bethune X G1186/94 consisting of 68 SSR markers and a seed colour marker (d) forming 19 linkage groups. Seventeen markers were found to be unlinked. LG 1 is showing 4 markers associated with seed colour locus "d" (indicated with a brown arrow) and the closest one is Lu209.

4.4 Fine-mapping of the *D* locus

LG 1 was focused on further for fine mapping based on the linkage mapping (Figure 4.6). Both Lu209 and Lu125 are EST-derived SSR markers (Cloutier et al., 2009; Venglat et al., 2011). ESTs containing primer sequences for Lu209, the closer marker to the *D* locus, were identified from the *Linum* EST database in GenBank (NCBI, 2013). The sequences of forward and reverse primers of Lu209 used for BLAST against the NCBI EST database were 5' AACAAAGGGGGAAACAGCTT 3' and 5'GCCAAAAGAATTGGCAAGAC 3' respectively. Two ESTs (NCBI accession numbers JG040203.1 and JG047082.1) were found in the search.

Alignment of sequences of both ESTs showed 100 % identity although JG047082.1 is 5 bases longer than the other one. Names and aligned sequences of both ESTs are shown here (Figure 4.7).

Figure 4.7: EST sequences harboring the Lu209 marker. A) A selected portion of a screenshot showing two ESTs obtained from NCBI *Linum* EST database BLAST search of Lu209 primers as query; B) Aligned sequences of both of the ESTs showing 100 % match of forward and reverse primers of Lu209.

After obtaining the EST sequence, the entire EST sequence (combination of both) was used as the query sequence in BLAST tool (e-value: 10, maximum target sequences: 50) to search the nucleotide sequences in the flax genome database. The genomic sequence is available as computationally assembled scaffolds from individual sequence reads; authors first obtained individual reads of approximately 44-100 bp that were assembled into contigs with N_{50} = 20.1kb which produced 88,384 scaffolds (N_{50} = 693.5 kb) (TUFGEN, 2012; Wang et al., 2012). Both EST sequences containing the SSR marker Lu209 were located on scaffold 208.

Markers present on LG 1 were tested for their presence or absence in scaffold 208 which is 830,661 bp. Figure 4.8 shows the inter-marker genetic distance and the physical distance. At this point, only Lu209 was found on scaffold 208 at the physical position between 184,090 and 184,515 bp from the start of the scaffold. Because single nucleotide polymorphisms (SNPs) are more prevalent than SSR polymorphism, potential SNPs with G1186/94 were sought. As explained in the Materials and Methods chapter, CDC Bethune scaffold 208 of the whole genome assembly (TUFGEN, 2012; Wang et al., 2012) was used as the reference sequence and the BAM file of reduced genome representation (RGR) sequence assembly of G1186/94 (provided by Dr. Sylvie Cloutier) was used to find the putative SNPs using CLC Genomics Workbench software (CLC bio). A total of 262 SNPs (from both CDC Bethune and G1186/94) from scaffold 208 genomic sequence region were identified and *in-silico* analysis for change in restriction enzyme recognition sites at the SNPs was performed. A total of 43 SNPs with altered restriction sites were found and these were considered putative candidates for CAPS (Cleave Amplified Polymorphic Sequences). CAPS from SNPs would be easier to test in the restriction enzyme analysis following amplification of the regions flanking the CAPS. Out of these, six

putative CAPS were selected as they had at least 70 reads showing nucleotide variation between CDC Bethune and G1186/94. Their positions are indicated in Figure 4.8.

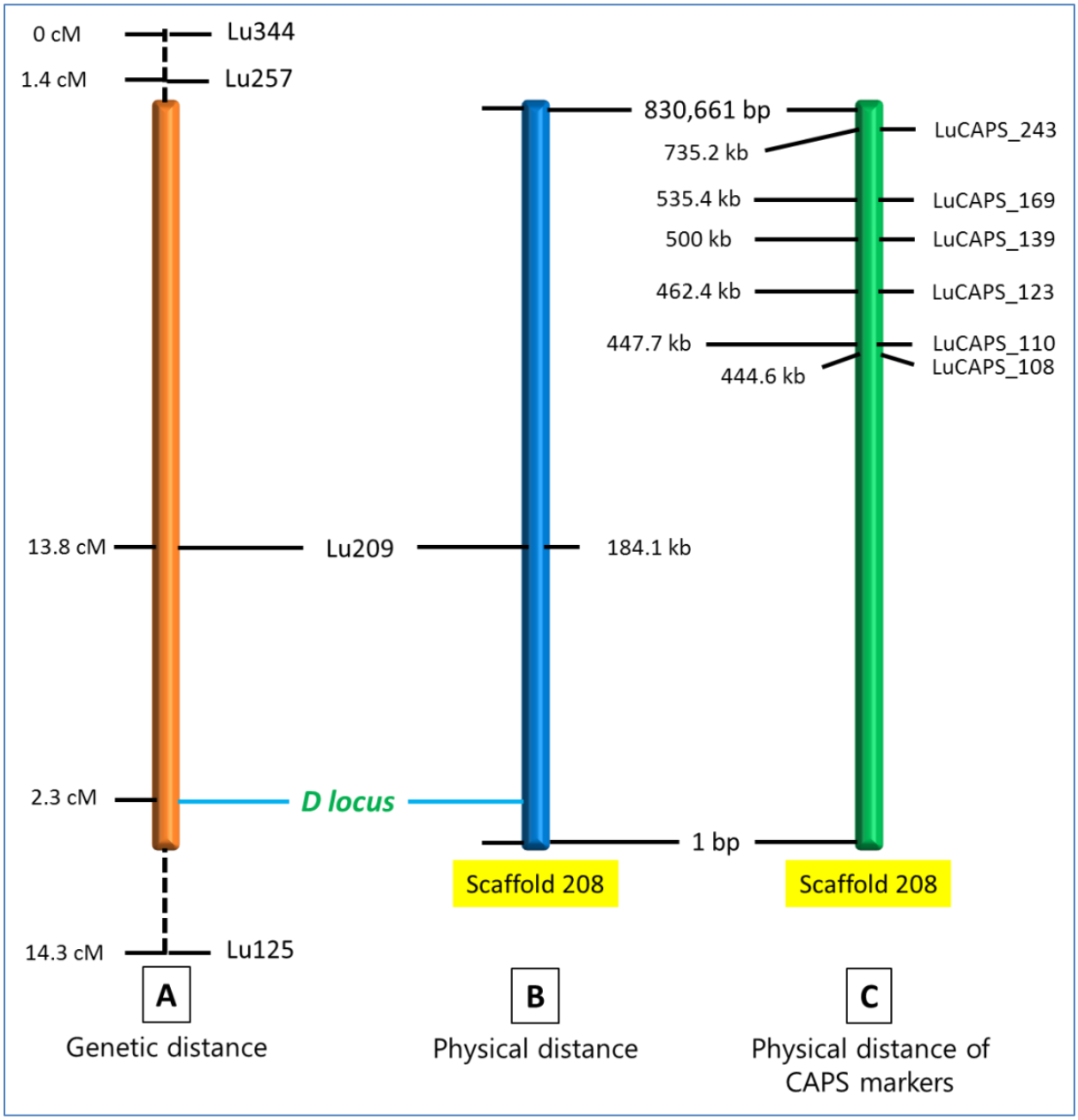


Figure 4.8: Genetic and physical map of the *D* locus showing co-linearity for marker Lu209 pertaining to scaffold 208. The length of the scaffold 208 region is the same in A, B and C; however, the distances shown are not to the scale. Here, A) showing genetic distance of markers in LG1; B) showing approximate physical positions of the markers in scaffold 208; and C) showing presence of putative CAPS markers with their positions on scaffold 208.

The CAPS markers were tested in parental lines first. The successful marker found that was near to Lu209 is LuCAPS_110. This marker was selected to check the orientation of scaffold with respect to the *D* locus. CAPS marker analysis with LuCAPS_110 on the mapping population was performed. There were only 5 recombinants out of the 94 lines in the population (Figure 4.9). Thus, this marker is tightly linked with the yellow seed colour locus. This marker was also added to the genetic linkage map through linkage analysis using MAPMAKER/EXP 3.0 (Lander et al., 1987). This marker, however, was found to be positioned further away from Lu209 and distal to the *D* locus.

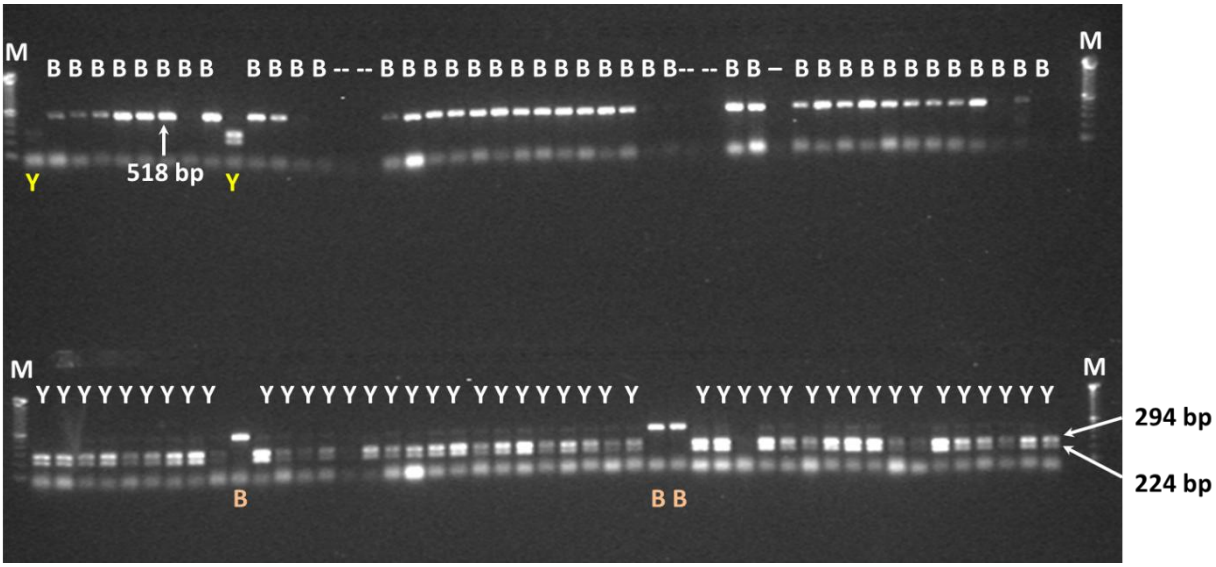


Figure 4.9: Agarose gel showing genotyping of 94 RILs using LuCAPS_110; 47 each of brown (upper lanes) and yellow seeded genotypes (lower lanes). Genotypes were scored as B (brown type with ~500 bp and no restriction digestion), and Y (yellow type, with a restriction digested product giving two bands at ~ 290 bp and ~ 220 bp). Yellow coloured 'Y' and light brown coloured 'B' showing recombinants (phenotypically parental type but genotypically non-parental type) in the population for the marker.

There were no other potential CAPS on the other side of Lu209 i.e. towards the *D* locus. As the number of reads for putative CAPS in that region was less, there was no confidence in those markers to test. We considered KASP^{TM} (Kompetitive Allele Specific PCR) genotyping assay (www.lgcgenomics.com) for other SNPs but technical issues precluded its use. Apart from this, there was discontinuity in the assembled reads in G1186/94 bam files in the region of interest (first 200 kb of scaffold) with reference to CDC Bethune scaffold 208 sequence. Therefore, neither potential SNPs nor CAPS were found in the first 200 kb of scaffold 208.

In addition, other SSR markers (Lu125, Lu257 and Lu344) were searched for their physical location on scaffolds in the CDC Bethune genome sequence assembly (TUFGEN, 2012; Wang et al., 2012). Marker Lu125 was found on scaffold 43 (1.64 Mb) at approximately 1.11 Mb region; Lu257 was located on scaffold 89 (at 79.1 kb) and Lu344 was positioned on scaffold 37 (at 1.29 Mb). None of the three markers tested was found on scaffold 208. Those scaffolds were unconnected and orientation was not known. Therefore, those scaffolds were not used for further studies.

Because scaffold 208 was the only usable scaffold at this point, additional potential SSR marker candidates were sought by analyzing the sequence in WebSat online software (Martins et al., 2009). In total, 229 SSR markers were located on this scaffold at a setting of a minimum of six repeats of dinucleotide motifs, or three repeats for tetra, penta and hexanucleotide motifs. As stated above, the *D* locus was placed between Lu125 and Lu209 in the coarse mapping work, but scaffold 208 does not include Lu125 sequence. Therefore, the sequence of scaffold 208 from nt 1 to nt 200,000 that covers the Lu209 primer sequence was screened for SSR polymorphism.

As described earlier, LuCAPS_110 was located at position 447,697 bpin the genomic sequence and Lu209 was positioned between 184,090 and 184,515 bp of scaffold 208. Therefore, the genomic sequence of the first 200 Kb was considered to search for markers for fine mapping. This search identified 52 SSR markers, out of which 8 mapped to the *D* locus. Other than these, 4 out of 9 markers in this region known from the published literature (Cloutier et al., 2012a) also mapped to the *D* locus. The list of markers used in fine mapping is given in Table 4.4. The genetic and physical map of the *D* locus fine-mapped region with co-linearity of markers in scaffold 208 is shown in Figure 4.10.

Table 4.4: List of polymorphic markers and their primers, used in fine-mapping of *D* locus.

Marker	Primer name	Sequence (5' to 3')	Length (bp)	Tm (°C)	Source
LuCAPS_110	LuCAPS_110_F	CCTTTATCTCTGCCTCTTCTCC	22	58	
	LuCAPS_110_R	ACAACCCCAACACAATCTCG	20	60	
LuM71	LuMF71	AGAAAATCGAAAGATGAAGGGG	22	60	
	LuMR71	CTACTCTTCCCACGTTGACCA	21	59	
LuM597	LuM597_F	TCCGACTATTTCGGGGTTAT	20	59	
	LuM597_R	TTTTCCTGTGGAGTGTCGAT	20	59	
LuM595	LuM595_F	AATCCTCTGTTTCTCCCCTG	20	59	
	LuM595_R	AGAGTGCCATCAGTTTGAGC	20	59	
LuM593	LuM593_F	TTGCTCCTTCCTATGCTCTC	20	58	
	LuM593_R	ACTGGGAAATTAAAGGGGAA	20	58	
LuM592	LuM592_F	TTAATTGGTGGGATGGAGAA	20	58	
	LuM592_R	CTTATCACAAAGCGAGTAGTTACG	24	58	
LuM588	LuM588_F	CACACACAAAGATGCCGTTA	20	59	
	LuM588_R	ATGTGAGATGGGAATGATGG	20	59	
LuM569	LuM569_F	ATCCTCCCTCCGTAGCATAG	20	59	
	LuM569_R	AGTGTTGAAGAATTGCAGGC	20	58	
LuM568	LuM568_F	ATCCTCCCTCCGTAGCATAG	20	59	
	LuM568_R	AGTGTTGAAGAATTGCAGGC	20	58	
LuM566	LuM566_F	TTTTCCTTCACCTAGCTCACTT	22	58	
	LuM566_R	AAGTTTAGGGGTCCAATCGT	20	58	
Lu2351	Lu2351_F	GGAAGCGAGTCATTCAATACG	21	60	Cloutier et al. (2012b)
	Lu2351_R	GCTGCGTAGCTACAATTTGATTAC	24	60	
Lu2341	Lu2341_F	CTGGTGGTTGATGCTAGTGC	20	59	Cloutier et al. (2012b)
	Lu2341_R	AAATGGGGGACTTGATTTAGC	21	59	
Lu2347	Lu2347_F	ACGTCATGTCCTCCACGTC	19	60	Cloutier et al. (2012b)
	Lu2347_R	GACGAGGGAAAGTTGTGCTC	20	60	
Lu2346	Lu2346_F	GAAAAGCAAAGAAGCTGAAAGG	22	60	Cloutier et al. (2012b)
	Lu2346_R	TTGGCCAAAATCACTCACC	19	60	

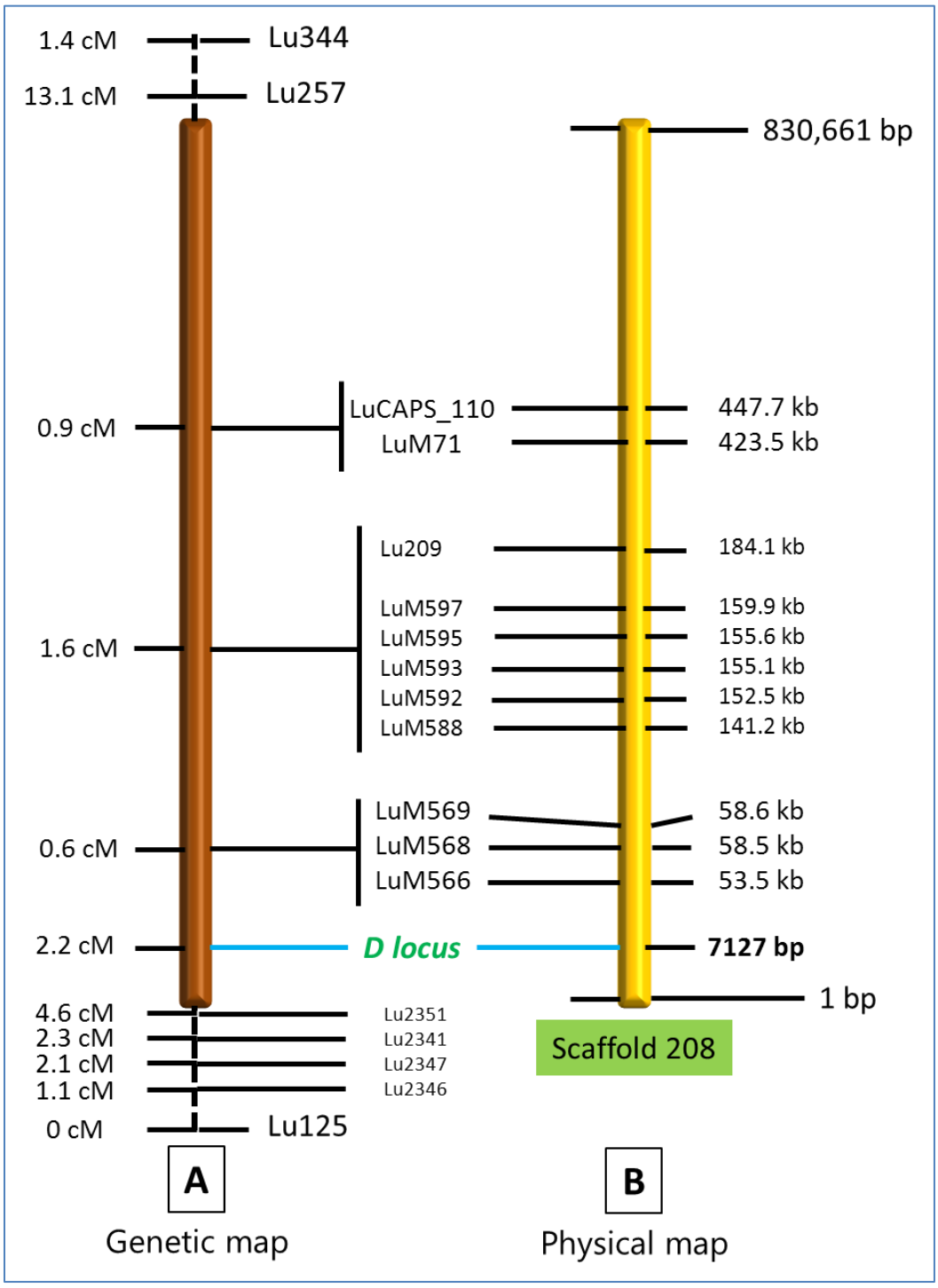
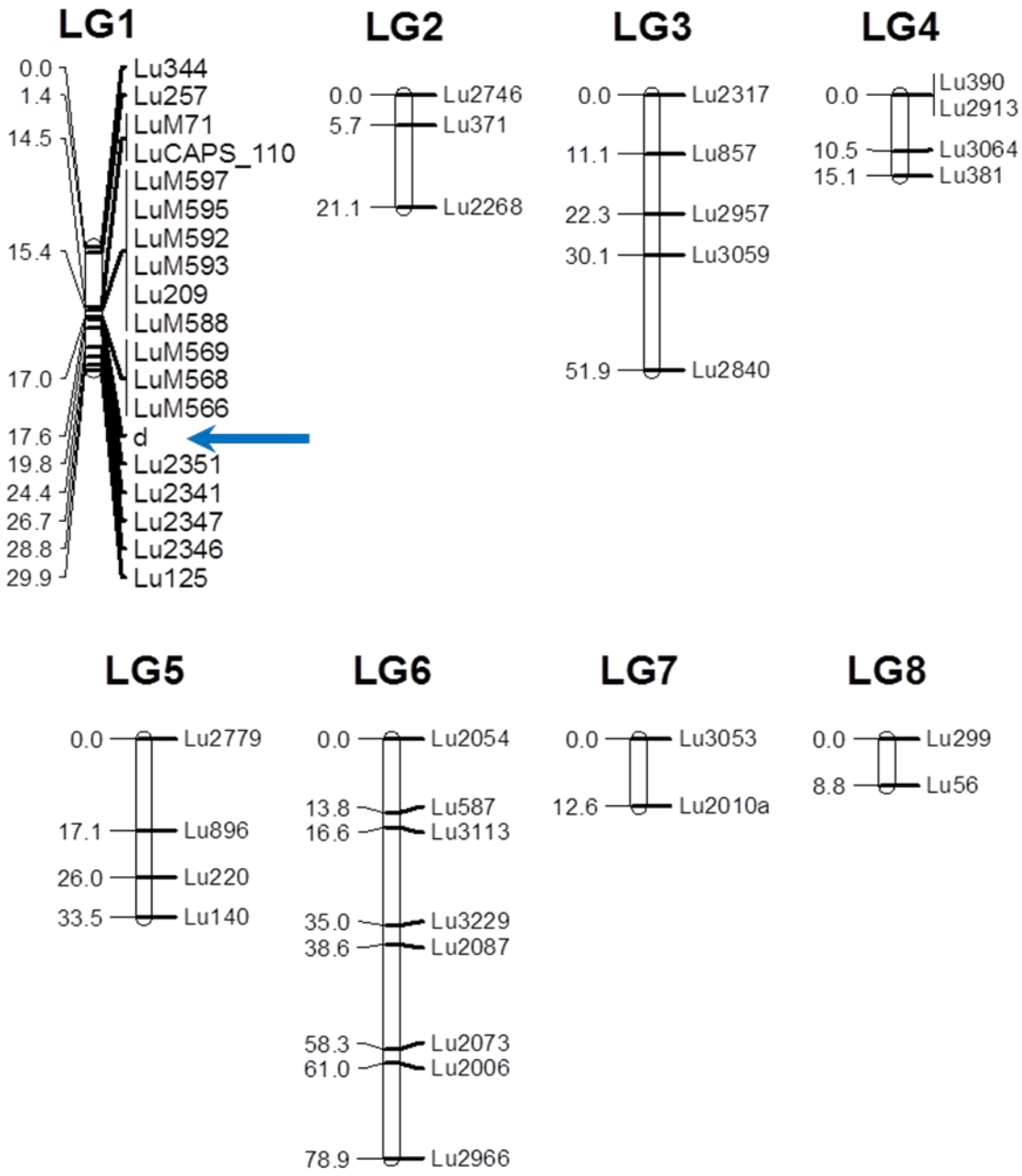


Figure 4.10: Genetic and physical map of fine-mapped region of the *D* seed colour and flower petal colour locus showing co-linearity of markers of scaffold 208 at their physical positions (not to scale). Here, A) is a genetic map with inter-marker genetic distances and B) is physical map of scaffold 208 with genomic positions between the co-linear markers.

4.5 Final genetic linkage map after fine-mapping

In the final genetic linkage analysis, a total of 100 markers [includes all the markers (85) from the second frame work linkage map, 13 SSR markers identified during fine-mapping, one CAPS and a hypothetical "d" locus marker] were used after genotyping in the flax mapping population. Linkage analysis was performed using MAPMAKER/EXP 3.0 (Lander et al., 1987) with a setting of LOD value of 3 and recombination factor of 40 %. Out of 100 markers screened, 83 formed 19 LGs in total similar to the intermediate LGs. Seventeen of the 100 markers did not map to any of the LGs as they remained unlinked. These markers in the LGs corresponded to the LGs published in an integrated genetic and physical map (Cloutier et al., 2012a). The genetic map spanned a total genetic distance of 467.8 cM. Average marker density in this genetic map was 5.64 cM per marker. The list of markers used to create the genetic map and comparison of present LGs with that of Cloutier et al. (2012a) are given in Appendix 5. Fourteen out of 99 markers exhibited significant segregation distortion ($P < 0.05$) as tested for goodness of fit using Chi-Square test for 1:1 segregation ratio (brown as to yellow seeded phenotypes). The expected segregation ratio of 1:1 in RILs is based on a single gene model proposed for seed colour in G1186/94 (Mittapalli and Rowland, 2003). Only eight of the markers are present on LGs; however, none of the markers on LG 1 had significant segregation distortion. Details of the Chi-Square analysis are provided in Appendix 6.

In this analysis, the yellow seed colour and flower petal colour locus (*D*) was found to be on LG 1 spanned by SSR markers LuM566 and Lu2351 with genetic distance of 2.8 cM. The rest of the LGs as well as unlinked single SSR markers did not show association with the yellow seed colour locus. Graphical representation of LGs is shown in Figure 4.11.


LG1
0.0 Lu344
1.4 Lu257
LuM71
14.5 LuCAPS_110
LuM597
LuM595
LuM592
15.4 LuM593
Lu209
LuM588
LuM569
17.0 LuM568
LuM566
17.6 d
19.8 Lu2351
24.4 Lu2341
26.7 Lu2347
28.8 Lu2346
29.9 Lu125
LG2
0.0 Lu2746
5.7 Lu371
21.1 Lu2268
LG3
0.0 Lu2317
11.1 Lu857
22.3 Lu2957
30.1 Lu3059
51.9 Lu2840
LG4
0.0 Lu390
Lu2913
10.5 Lu3064
15.1 Lu381
LG5
0.0 Lu2779
17.1 Lu896
26.0 Lu220
33.5 Lu140
LG6
0.0 Lu2054
13.8 Lu587
16.6 Lu3113
35.0 Lu3229
38.6 Lu2087
58.3 Lu2073
61.0 Lu2006
78.9 Lu2966
LG7
0.0 Lu3053
12.6 Lu2010a
LG8
0.0 Lu299
8.8 Lu56

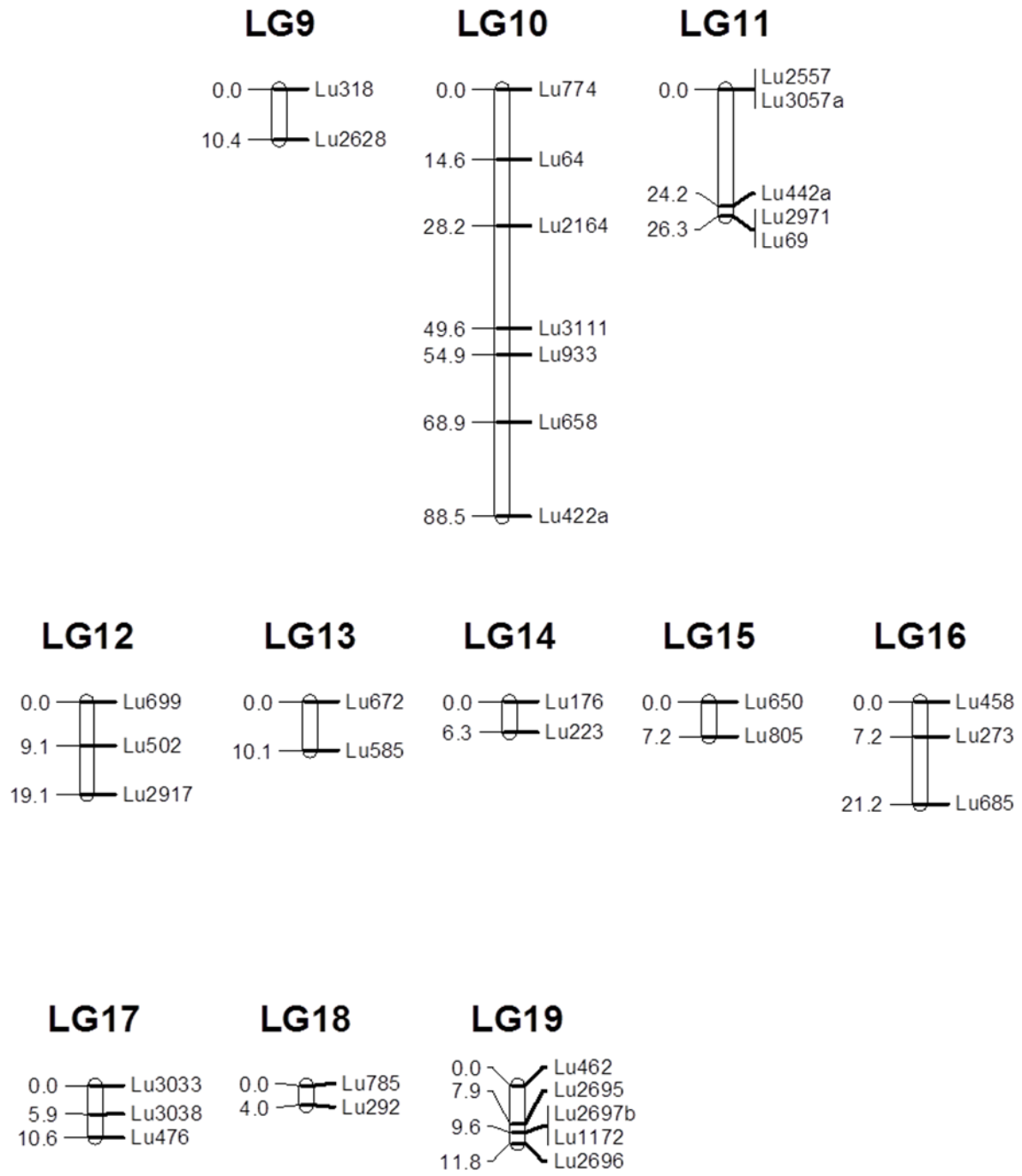


Figure 4.11: A frame work genetic linkage map of flax (*Linum usitatissimum* L.) using $F_{9:10}$ RIL populations of CDC Bethune X G1186/94 and its reciprocal cross. The map consists of 19 LGs comprising 81 SSRs, one CAPS marker and a hypothetical marker for the seed colour and flower petal colour locus (d). The arrow in LG 1 indicates the position of detected the seed colour and flower petal colour locus *D*.

4.6 High density linkage map at the *D* locus and co-linearity of markers of LG 1 with published flax LG 2

The final map after fine-mapping for the *D* locus had 19 markers [include 17 SSR markers, one CAPS marker and a hypothetical marker (d) for the seed colour locus] on LG 1. The total genetic distance of the LG 1 was 30.2 cM with an average distance per marker of 1.59 cM. The *D* locus was mapped between LuM566 and Lu2351 SSR markers as mentioned earlier. The closest marker was LuM566 with a genetic distance of 0.6 cM and Lu2351 was 2.2 cM away from the *D* locus. There was co-linearity of marker order in my LG 1 to the LG 2 of Cloutier et al. (2012a). The graphical representation of the markers spanning LG 1, the close view of the fine mapped *D* locus and co-linearity between the markers of LG 1 of the present study and LG 2 of Cloutier et al. (2012a) is shown in Figure 4.12.

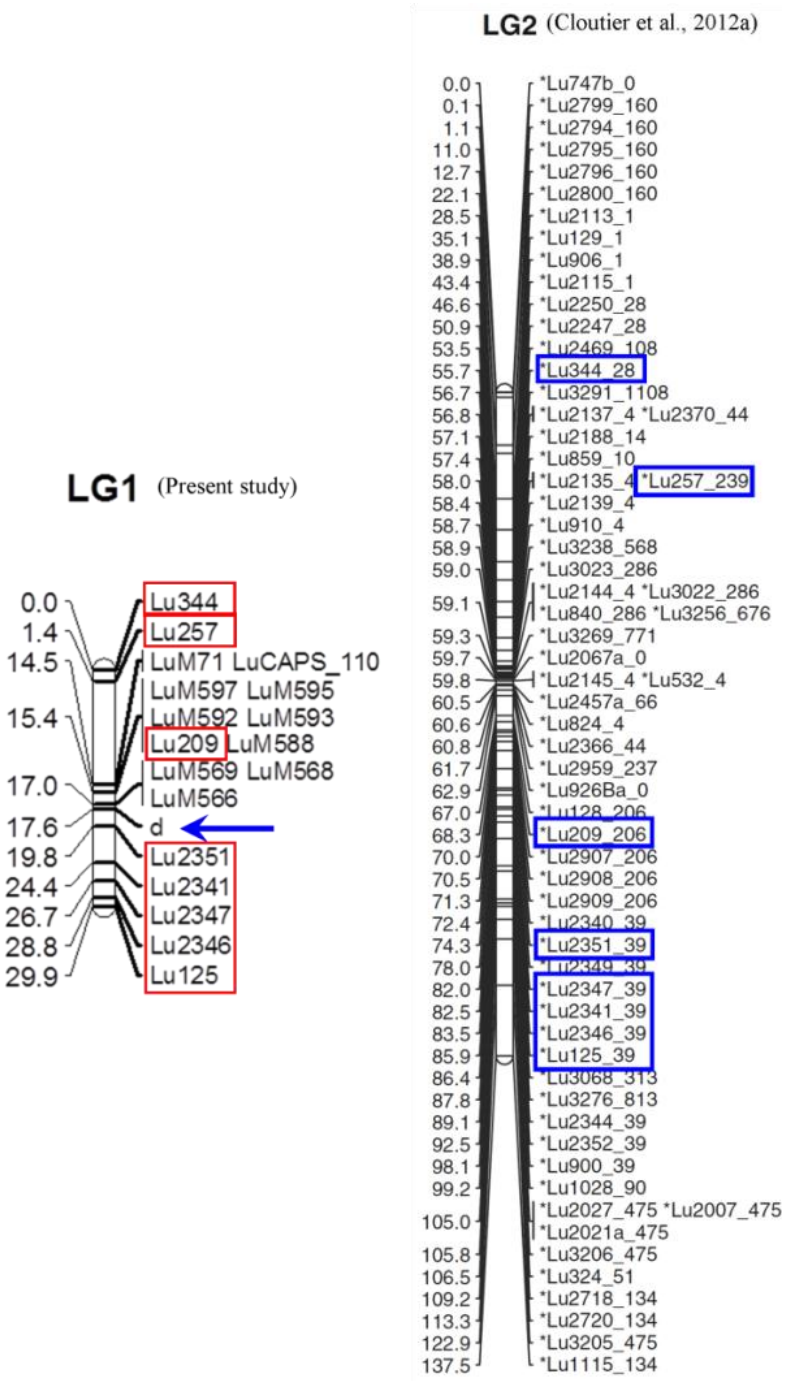


Figure 4.12: The *D* locus (blue arrow) in the fine-mapped region of LG 1 and co-linearity between markers of LGs from two studies. Markers of LG 1 of the present study (in red box) are highlighted on LG 2 of Cloutier et al. (2012a) in blue boxes. Here, genetic distance between markers and marker order is the same except Lu2341 and Lu2347; however, all the markers are linked. The blue arrow in the LG 1 indicates the '*D'* seed colour and flower petal colour locus.

4.7 Composite Interval Mapping (CIM) analysis for the seed colour locus

The marker-trait association was investigated using QTL Cartographer v2.5 software. The CIM analyses were performed on two phenotypic data sets of the RILs. In both cases, highly significant marker-trait association was observed indicating the presence of the *D* locus between markers LuM566 and Lu2351. In the first phenotypic data set (GC, 2008) the LOD value was 47.1 and 89% [coefficient of variation (R^2 = 0.89)] of the phenotypic variation was represented and in the second data set (GH, 2011), LOD value was observed to be 39.3 with R^2=0.84. Marker-trait association analysis indicated stability of the mapped locus for seed colour for two independent growing conditions. The loci from both trials are graphically shown in Figure 4.13.

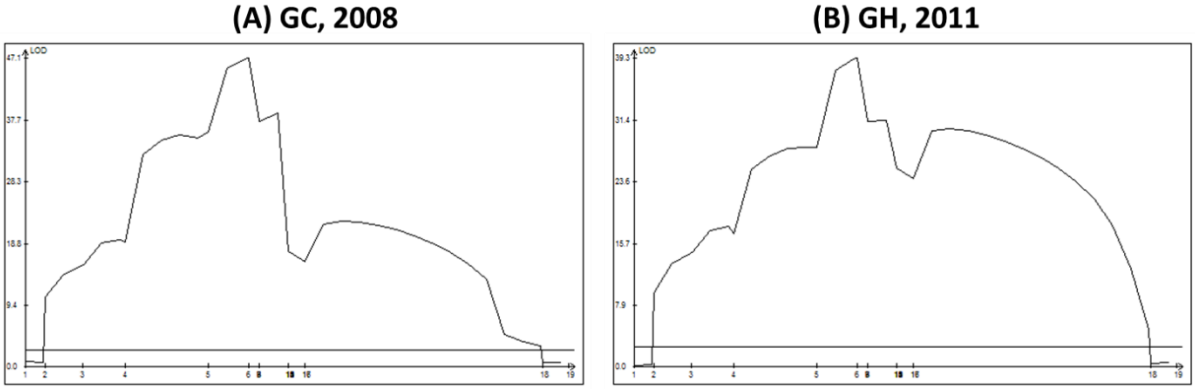


Figure 4.13: Logarithm of odds (LOD) graph of the *D* locus in two phenotypic data sets. The threshold LOD score in both cases was 2.5. A) The graph represents CIM analysis with four SSR markers of RILs grown in a growth chamber (GC, 2008) where LOD was 47.1 for the *D* locus spanning a genetic distance of 30.2 cM and B) where RILs grown in a greenhouse (GH, 2011) showed an LOD of 39.3 and 30.2 cM of genetic distance for the same locus with same markers.

4.8 Putative candidate gene for yellow seed colour mutation

In the mapped region, LuM566 was located between 53427 and 53720 bp. Therefore, the first 54Kb region of scaffold 208 was subjected to gene prediction using FGENESH software (FGENESH, 2012) using *Hevea* as model plant species as it is phylogenetically close to flax (Lee et al., 2011). This search resulted in a total of 14 predicted genes. Those genes were annotated using the TAIR database. Predicted genes are listed in Table 4.5. Out of these 14 genes, there was only one gene which could be directly related to the flavonoid biosynthesis pathway and that was one of the key structural genes. That was found to be a putative flavonoid 3' hydroxylase (F3'H) gene. The gene was located between 7127 to 9095 bp (negative strand) on scaffold 208. The predicted gene had three exonic regions that code for 521 amino acids and showed 61.69% identity with the F3'H of the *Arabidopsis* TT7 gene (AT5G07990.1) over the entire length of the polypeptide. The predicted gene in the flax genome database (TUFGEN, 2012) is Lus10021620.

Table 4.5: The list of TAIR predicted genes in the *D* locus region using scaffold 208 genomic sequence (1 to 50,000 bp).

	TAIR ID	%ID match	Length	e-value	Bit score	Gene description	Region in scaffold 208 (bp)	Amino acids	Strand
1	AT1G18650.1	68.89	90	4e-38	138	Plasmodesmata callose-binding protein 3 (PDCB3)	654 - 4897	436 aa	chain -
2	AT4G19710.2	50	30	0.59	29.3	Aspartate kinase-homoserine dehydrogenase II (AK-HSDH II)	5224 - 6099	124 aa	chain +
3	AT5G07990.1	61.69	509	0	656	Required for flavonoid 3' hydroxylase activity. protein_coding TRANSPARENT TESTA 7 (TT7) TRANSPARENT TESTA 7 (TT7); (D501);CYTOCHROME P450 75B1 (CYP75B1)	7127 - 9095	521 aa	chain -
4	AT5G11040.1	26.47	68	1.3	29.3	Encodes a tethering factor (TRS120) (TRS120); (AtTRS120)	10278 - 13980	178 aa	chain +
5	AT1G11100.2	26.95	141	3.1	27.3	SNF2 domain-containing protein / helicase domain-containing protein / zinc finger protein-related. (TAIR:AT1G61140.1)	15246 - 15611	121 aa	chain -
6	AT5G03040.1	30.77	39	0.76	29.3	IQ-domain 2 (iqd2); FUNCTIONS IN: calmodulin binding	17050 - 17603	138 aa	chain -
7	AT2G21590.1	29.33	75	0.89	26.9	Encodes the large subunit of ADP-glucose pyrophosphorylase. Protein_coding (APL4) (APL4)	20388 - 20612	74 aa	chain -
8	AT5G07980.1	37.78	540	7e-72	265	Dentin sialophosphoprotein-related; (TAIR:AT5G07940.2)	22686 - 27964	1366 aa	chain +
9	AT5G07950.1	41	239	1e-47	164	Unknown protein	28131 - 30004	314 aa	chain -
10	AT5G07960.1	55.71	140	6e-47	150	unknown protein; contains interpro domain/s: Uncharacterised protein family UPF0139 (InterPro:IPR005351)	30347 - 32172	145 aa	chain +
11	AT5G39510.1	74.87	199	1e-104	319	Encodes a member of SNARE gene family. Homologous with yeast VTI1. Protein_coding shoot gravitropsim 4 (SGR4) (ZIG1); (VTI1A); Vesicle transport v-snare 11 (VTI11)	33577 - 37672	637 aa	chain -
12	ATMG00810.1	43.14	51	5e-10	55.8	Hypothetical protein protein_coding (ORF240B) (ORF240B)	43051 - 44019	161 aa	chain -
13	AT1G77470.1	29.17	72	0.24	29.6	Encodes a protein with high homology to the Replication Factor C, Subunit 3 (RFC3) of yeast and other eukaryotes.	46323 - 47364	108 aa	chain +
14	AT5G07920.1	71.64	744	0	1096	Diacylglycerol kinase protein_coding diacylglycerol kinase1 (DGK1)	48975 - 52821	726 aa	chain +

A BLAST search using this gene as a query with default parameters at NCBI *Linum* EST database (Venglat et al., 2011) identified six matching ESTs. A graphical representation of six ESTs and their alignment with the reference gene is shown in Figure 4.14, and sequences of these six ESTs are given in Appendix 7. Five of the ESTs matched at the 3' end of the gene sequence and one at the 5' end of the gene. Out of those, one was from torpedo stage seed coat tissue; two were from flower tissues and three from endosperm tissue of CDC Bethune. Those results indicate that the putative gene was expressed in both seed coat as well as flower tissue. All the six EST clones were obtained from authors (Venglat et al., 2011) and re-sequenced to get full length or additional sequence of the gene. This exercise did not add any additional sequence information than that published. The sequence length was the same as given in the NCBI *Linum* EST database (Venglat et al., 2011) and the clones contain only short inserts. The hypothesis that the predicted gene is expressed in both seed coat and flower tissues based on ESTs, was tested by gene expression analysis of the putative candidate gene (F3'H) in seed coat and flower petal tissues.

Note there is no reliable sequence data available for the region between Lu2351 and the start of scaffold 208. Thus, this gene search could not be extended to the missing region which is up to 2.2 cM in genetic distance.

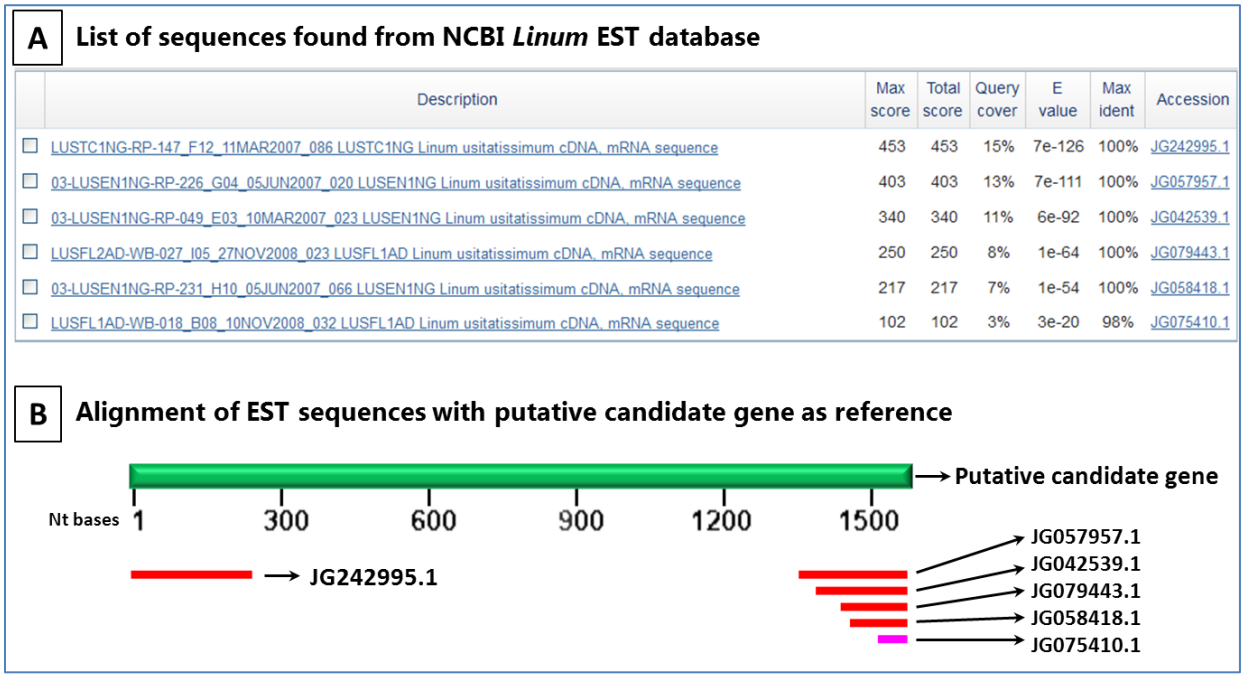


Figure 4.14: EST sequences found in the NCBI *Linum* EST database using the putative candidate gene F3'H as a query. Here, A) is a screen shot of the list of EST sequences with the BLAST result from the NCBI site and B) is a modified figure of the alignment of six ESTs with their matching region of F3'H.

Because F3'H is involved in flavonoid biosynthesis, its expression in G1186/94 flower petals and in developing seed coat tissue was tested. The primers were designed in the third exon of the putative candidate gene amplifying 158 bp (primers sequences are in Chapter 3). This 158 bp region was used for searching similar mRNA and *Linum* EST sequences in NCBI database. The maximum identity was 99 % with *Glycine max* flavonoid 3'-hydroxylase (SF3'H1) mRNA and 92 out of top 100 hits were showing F3'H gene. The same five ESTs (shown in Figure 4.14) other than JG242995.1 were matched with 158 bp, giving confidence for the presence of one gene. Relative to the expression levels in CDC Bethune, the expression level in G1186/94 was 25-fold lower in 15 DAF seed coat tissues and 20-fold lower in flower petal tissue (Figure 4.15

A and B, respectively). The analysis indicated that expression of the putative candidate gene (F3'H) was very low in both seed coat as well as flower petal tissues of G1186/94.

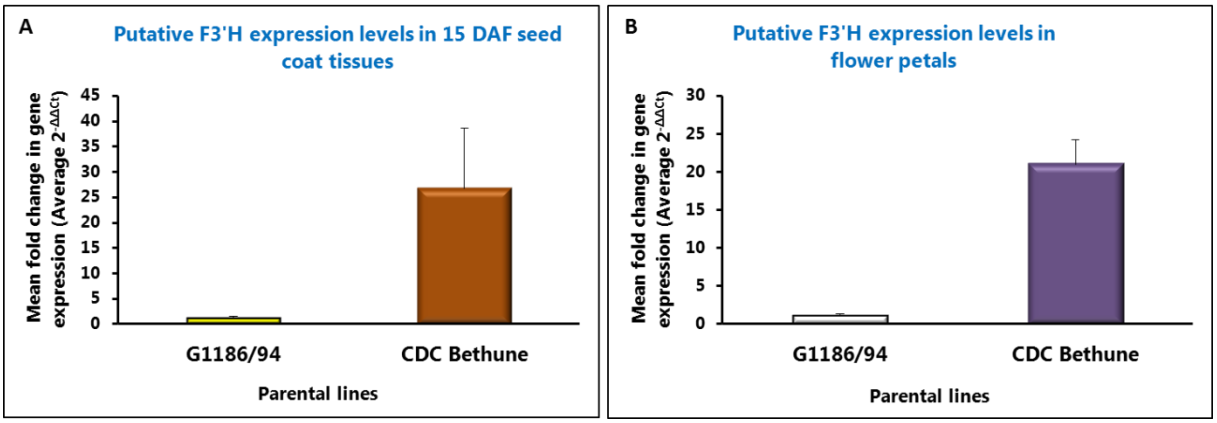


Figure 4.15: A comparative gene expression analysis of putative F3'H in seed coat (15 DAF) and young flower petal tissues of CDC Bethune and G1186/94. First, Ct values from seed coat and young flower petal tissues were normalized with Ct values of reference gene (EF1α) of respective samples; then, interpreted as fold change in gene expression compared to G1186/94 (1 part). A) CDC Bethune had 25-fold higher expression of the putative F3'H in seed coat; and B) 20-fold higher expression in young flower petals compared to G1186/94.

Thus, this work has identified putative F3'H as a candidate gene underlying the *D* locus. I hypothesize that impairment of this gene's expression in G1186/94 causes white petal colour and yellow seed phenotype.

5. DISCUSSION

The main objective of my research was to identify the locus governing yellow seed colour in the G1186/94 flax line through genetic linkage mapping. The trait was reported to be governed by a recessive allele at the *D* locus through classic genetic studies. Another objective was to map the locus responsible for the white petal colour in the same parental line. The RIL mapping population was used to study the proposed objectives. Successful research findings in genetic mapping of the *D* locus, issues and challenges addressed during the study and future perspectives of this research are discussed in this chapter.

5.1 Importance of seed colour genetics in flax

Flax is one of the important crops in Saskatchewan. It is mainly grown for oil purposes. A major aim of flax breeders in Canada is to improve crop yield. Oil yield per unit area can be increased either via increasing productivity (seed yield per unit area) or increasing oil content of the seed even if the seed yield is maintained. Yellow-seed colour in flax is one of the important traits in that context as yellow-seeded flax is reported to have slightly higher 1000-seed weight (6.42 ± 1.29 g) as compared to brown seed (5.98 ± 1.15 g) and to also have 1% higher oil content based on a study using the flax Plant Gene Resources of Canada (PGRC) world collection (Diederichsen et al., 2006). My study has found that the 1000-seed weight of brown-seeded parental line (CDC Bethune) is 4.2 ± 0.2 g and that of the yellow-seeded line G1186/94 is 4.4 ± 0.1 g under the experimental conditions of a greenhouse (GH) at Innovation Place, Saskatoon in 2011.

It has been proposed that the contribution of seed coat to total seed weight in yellow-seeded flax is low in comparison with brown-seeded flax (Mittapalli, 2002). This in turn results

in a higher embryo to seed coat weight ratio making the seed size larger. A similar phenomenon is known in yellow-seeded *Brassica* sp. (Chen and Heneen, 1992). Apart from higher seed weight, yellow-seeded flax has other advantages. They have significantly lower accumulation of phenolic acids (pigments) in seeds such as in Linola 947 (Oomah et al., 1995). The co-extraction of pigments with oil negatively impacts the oil quality as in the case of brown seeds. Yellow seeds have significantly higher protein content (3-4%) in comparison with brown-seeded flax (M. Kulkarni and G. Selvaraj, data unpublished). After oil extraction, flaxseed meal is used as animal feed and yellow flaxseed meal would have higher protein content. Therefore, yellow-seeded flax makes excellent animal feed after oil extraction and could increase overall value or profit from a crop. Because of all of the above mentioned advantages, it is interesting to study the genetic basis of the trait and its associations at molecular level.

As we know, yellow seed colour is one of the important agronomic traits in flax but the hereditary pattern of seed colour loci in flax is not yet very clear. Brown seed colour is predominant in the world collection and yellow-seeded flax lines are less than 8% in the total PGRC collection (Diederichsen et al., 2006). Three basic seed colour loci (G, D and B) are known and reported since 1922 (Tammes, 1922, 1928). In addition to the basic seed colour loci, another seed colour locus which is a dominant negative regulator of seed colour was reported in CPI84495 (Green and Dribnenki, 1995) and Bionda (Popescu and Marinescu, 1996) in two independent studies. Three recessive seed colour loci (*g*, *d* and *b1*) and one dominant yellow locus (*Y1*) are reviewed and reconfirmed by the extensive genetic studies conducted by Mittapalli and Rowland (2003).

In my study, a brown-seeded cultivar (CDC Bethune) is used as the “wild-type” and a European recessive yellow line (G1186/94) that is believed to have a mutation at the *D* locus is

used as the yellow-seeded line to perform genetic linkage mapping. The pedigree of CDC Bethune is Norman/FP857 and G1186/94 was derived from Atlanta/Avantgard (European variety). Genetic linkage mapping to find the recessive yellow seed mutations have not been reported yet. Neither tightly linked molecular markers are available nor are the genes or genetic mechanisms governing yellow seed colour of the three basic recessive loci in flax known. The only seed colour mapping has been done for the dominant yellow seed colour locus Y1 (Cloutier et al., 2011, 2012a); however, the molecular mechanism governing the dominant yellow seed colour locus has not been deciphered.

Elucidating genetic regulation of seed colour traits in flax will promote its breeding efficiency for developing value added products. In nature, flax seed shows different colours such as brown, yellow, green and variegated; however, brown and yellow are the most prevalent ones. Intensity of seed colour could vary in each category depending on the genotypes. Based on visual appearance, brown or yellow seed colour in flax ranges from darker to lighter shades. Even though phenotypically seeds look yellow, genotypically mutations in these seeds could be different. Therefore, genetic markers that are tightly linked to mutations will help to provide molecular identity to each yellow seed phenotype. Consequently, this achievement will help to categorize all yellow-seeded flax accessions in world collections into respective mutations and identification of new loci governing yellow seed colour could be possible.

In general, the undesirable effect that yellow seed colour exhibited in different studies was lower germination percentage or less seedling vigour (Culbertson et al., 1960). This could be because of thinner seed coat, lower permeability and less mucilage content in the seed coat compared to brown seeds. All of these can affect seed water imbibition during germination. Although that is true for other yellow-seeded accessions (having *g* and *b1* mutations), it has been

reported that yellow seed colour mutation with the *d* locus had no association with seed yield i.e., there are no major genes affecting seed yield that are linked to the *d* locus (derived from Bolley Golden) as reviewed by Mittapalli (2002). In this context, *d* seed colour mutation would be an ideal trait for breeding yellow-seeded flax varieties. Effective selection for yellow seed colour aided by molecular markers will speed up breeding this trait more efficiently.

The F_1 cross studies were conducted by Mittapalli and Rowland (2003) using G1186/94 as a source of European yellow mutation. The study confirmed that the yellow seed is formed in the presence of homozygous recessive allele *d* and the trait is governed by a single locus. This particular mutation was called the European recessive yellow mutation (Mittapalli and Rowland, 2003). The flower colour was observed in the F_3 populations and the only prominent association of flower colour was observed in the rows wherever G1186/94 (producing white or pink flower petals) was involved as a yellow parent. White or pink flower petals were observed in most of the yellow seeded homozygous rows indicating the involvement of the *d* locus in conditioning both seed and flower colour. These observations were done because of the previous reports on a pleiotropic effect of the *d* locus (from Bolley Golden and Viking) as reviewed by Mittapalli (2002). Interestingly, the flower colour data observed during the trial of my study indicated that brown-seeded lines were producing blue flower petals whereas yellow-seeded lines produced white flower petals. There was no recombination for this trait in 470 RILs. The qualitative data was observed in $F_{9:10}$ RILs of the mapping population. This indicates the tight linkage between the seed colour and petal colour trait and also could have the same locus in genetic linkage analysis. The QTL and percent contribution of the petal colour trait cannot be explained because quantitative data for petal colour was not observed as my main focus was on seed colour trait. However, the *D* locus in LG1 holds good for petal colour trait..

5.2 Genetic segregation of seed colour in the recombinant inbred line population

In the present study, $F_{9:10}$ generation RIL populations of CDC Bethune X G1186/94 and its reciprocal cross were used for genetic linkage mapping. Mittapalli and Rowland (2003) have shown that F_2 plants segregated in a 3:1 ratio for brown as to yellow in crosses between G1186/94 (female) and brown lines (male) and proposed that the inheritance of seed colour in the yellow parent is governed by a single recessive gene.

This population of 479 lines ($F_{8:9}$) showed a phenotypic segregation ratio of 1.41:1 for, brown and yellow lines, respectively. (Note that the terms "yellow-seeded", "yellow lines", and "yellow seed" are used interchangeably). This significantly deviated from a classic Mendelian segregation ratio (1:1) assuming that the production of single seed descent population had no bias toward brown-seed lines at an early stage. Additionally, some individuals in this population might be heterozygotes for a gene that controls or conditions seed colour. Genotyping of the individuals with respect to the locus of interest will resolve this.

Cloutier et al. (2011) used SP2047 x UGG-5-5 double haploid (DH) mapping population of 78 individuals generated from anther culture to map the dominant yellow seed colour locus. That population segregated approximately 1 brown: 2 yellow-seeded types (personal communication). Even though the segregation ratio of phenotypes does not accord with a single gene model, the genetic linkage analysis in a high density linkage map shows that there is only one genomic region associated with dominant yellow seed colour (Cloutier et al., 2012a). In a study by Lambrides et al. (2004), for understanding the complex inheritance pattern of seed colour trait in mungbean (*Vigna radiata* L. Wilcek), a di-genic model was proposed using a classical genetic analysis method. They got 5.1: 1 F_2 and 2.7: 1 F_7 frequencies for green-

speckled-black testa color as to green testa color, that were significantly deviating from expected 3:1 and 1:1 ratios respectively for a single gene model. However, it was concluded through genetic linkage mapping with a population of 207 individuals that there is only one genomic region associated with seed colour. Three markers (one RFLP and two RAPD) that were linked with the testa colour trait showed segregation distortion. The study highlighted the importance of molecular markers to resolve issues related to inheritance of traits of economic importance influenced by segregation distortion (Lambrides et al., 2004). Even though in my case, phenotypic data shows significant deviation from a one gene model, this issue was resolved through genetic linkage mapping and only one major genetic locus was identified associated with yellow seed colour trait.

5.3 Phenotyping: Seed colour values and their bimodal frequency distribution in the mapping population

In order to quantify the intensity of seed colour, various methods have been used. In earlier reports, seed colour phenotyping was done using a visual method (Rahman, 2001); a colorimetric method (Zhi-wen et al., 2005) and dissecting microscopy (Liu et al., 2006) in *Brassica* spp. In the current study for phenotyping, a seed scanning method using optical well plates was adopted. The scanned digital images were used to measure seed colour in RGB values using ImageJ software (ImageJ, 2013). A similar method was reported to be efficient for seed colour measurement in *Brassica napus* (Fu et al., 2007). Cloutier et al. (2011) used a method based on CIE (Commission Internationale de l'Eclairage) colour co-ordinates (L*, a* and b*) using a Minolta spectrophotometer (Note "*" is colour space: method for stating the color of an object or a light source using some kind of symbolization, such as numbers). Here, L* indicates lightness, and a* and b* are chromaticity coordinates on the circular chromaticity diagram which

is divided into 4 parts (+a*, -a*, +b* and –b*). Plus and minus signs gives the direction of the coordinate where +a* is the red direction, -a* is the green direction, +b* is the yellow direction and –b* is the blue direction (as described in www.konicaminolta.eu). The colour is measured based on the value of coordinates and this method is satisfactory to present colour values but requires specific instrumentation. However, the method used in the present study is easy to use and does not require special equipment other than a colour scanner which is widely used. This was used first with the parental lines and then applied to the RIL populations of 479 ($F_{8:9}$) and 463 ($F_{9:10}$) lines, and it proved to be efficient and consistent in the numerical values.

The bimodal distribution of phenotypic trait in both $F_{8:9}$ and $F_{9:10}$ RILs indicated single major gene governing the seed colour trait in the current study. Plants of the same population at $F_{2:3}$ generation showed 3:1 brown as to yellow ratio (Mittapalli and Rowland, 2003) indicating the single gene inheritance of seed colour. Dominant yellow seed colour phenotypic segregation is also bimodal in nature (Cloutier et al., 2011) and one major locus governing this trait is identified through genome wide linkage analysis (Cloutier et al., 2012a). During the measurement of seed colour, manual and technical errors were removed by normalizing all the scanned images to similar over all red, green, blue and luminosity values. The variability in seed colour in the population may be caused by the influence of environmental factors such as light, temperature and fertilization. The effects of temperature on *Brassica napus* seed color were observed by growing yellow-seeded genotypes under 20ºC, 28ºC and above 30ºC. Seed colour becomes increasingly yellow at higher temperatures whereas cooler temperatures tend to produce dark coloured seeds (Burbulis and Kott, 2005). Therefore, genetic inheritance of seed colour trait was confirmed by phenotyping in two subsequent generations as well as two growth conditions i.e. $F_{8:9}$ and $F_{9:10}$ RILs grown in GC, 2008 and GH, 2011 respectively.

5.4 Genotyping: genetic linkage map and association of markers with seed colour locus in the mapping population

Amongst the molecular markers, SSR markers were chosen because of their user friendly nature. They are co-dominant, highly polymorphic, abundant and distributed evenly across the genome. This is a PCR based marker technique and requires a small amount of DNA as compared to non-PCR based techniques such as RFLP. This type of marker (SSRs) has been shown to be applicable within *Linum* species (Cloutier et al., 2009; Bickel et al., 2011; Cloutier et al., 2011; Fu, 2011). These types of markers are preferred by breeders for use in marker-assisted breeding. For example, in bread wheat for selection of scab resistant near-isogenic lines, two flanking SSR markers to quantitative trait loci (QTL) were used from $F_{2:3}$ until F_6 generation (Zhou et al., 2003) and during selection of submergence tolerant rice genotypes, SSR markers linked to the *Sub1* locus were used (Biswas et al., 2013). Therefore, identification of SSR markers linked to seed colour was one of the tasks in the study.

In the current study, initially, I found 74 (15.5 %) polymorphic out of 477 genomic SSR markers (identified by Dr. M. Kulkarni) tested. Then, in the later part of my study, SSR markers identified by myself were tested. From that group, 8 out of 52 SSR primers tested were polymorphic (15.4 %) between parents CDC Bethune and G1186/94. In a study comprising 23 flax accessions, 40.7% of the primers were found to be polymorphic (Cloutier et al., 2009). The level of SSR polymorphism in other crop species varies depending on the mode of pollination; examples of such crops with polymorphic markers are wheat (25%) (Eujayl et al., 2002); barley (33 %) (Varshney et al., 2006); and *Brassica napus* (33.5 %) (Piquemal et al., 2005). Two of the many factors which affect the level of polymorphism are relatedness of varieties and ploidy

levels. As the two parental lines used in the present study might not be diverse in origin (as flax is a self-pollinated crop), the level of polymorphism between these lines was lower.

Earlier published linkage maps used marker systems such as RAPD and RFLP markers (Oh et al., 2000) and AFLP markers (Spielmeyer et al., 1998a) that have a limitation of reproducibility and unreliability of markers. Using RAPD, AFLP, RFLP or SRAP markers is not as beneficial as SSR markers because flax is an autopolyploid plant species and it is difficult to construct a genetic linkage map using dominant marker system such as RAPD. Recently, high density genetic linkage maps in flax (using three mapping populations) were generated by using SSR markers from EST and BES (BAC end sequencing) (Cloutier et al., 2011, 2012a) which is of immense help for fine mapping and marker assisted selection of many traits of economic importance after linked markers for the traits are found. Availability of SNP markers is just in its early stages and there are very few published studies indicating the utilization of SNP markers for flax genetic improvement (Kumar et al., 2012). Molecular markers have been employed in other crop species such as *Brassica* (Negi et al., 2000; Liu et al., 2006; Rahman et al., 2007; Heneen et al., 2012; Kebede et al., 2012) to map genetic loci governing seed colour.

In this research project, we successfully constructed a genetic linkage map using $F_{9:10}$ RIL populations. For this goal, single plants from $F_{8:9}$ were grown and seeds were phenotyped. In the genetic linkage map, the average genetic distance between markers is 5.64 cM per marker and the total genetic distance covered is 467.8 cM in 19 LGs. The genetic linkage map published by Cloutier et al. (2011) using EST SSR markers has a total genetic distance of 833.8 cM in 24 LGs. This map has average distance of 7.3 cM per marker. Although the total genetic distance in the present investigation is less than the previous reports, this served the main objective of this study i.e. for locating the genomic region controlling yellow seed colour in the mapping

population. The other reason for lower coverage of genetic distance was because we followed a strategy for mapping, where we chose minimum number of markers spanning 25 cM each in a LG from the most recent high density linkage map of flax covering a total genetic distance of 1,551 cM (Cloutier et al., 2012a). Flax (2n = 30), a diploid species with 15 chromosome pairs, should have got 15 LGs as in the case of Cloutier et al. (2012a). However, in the present study, 19 LGs were formed because of a lower number of markers. The experiment resulted in the association of the seed colour locus (*D*) with LG 1. Even though the mapping population used in the present study was only 94 RILs, the expected recombination events were enough to map the seed colour locus. In addition to this result, a fast track mapping approach was also implemented where markers were used from previously published literature to create a coarse genetic linkage map. Once the yellow seed colour linked marker was located on LG 1, we focused on only that linkage group for further marker development as well as fine mapping of the locus. Note that this LG 1 that I named during the course of my work corresponds to LG 2 in Cloutier et al. (2012a).

5.5 Segregation distortion issues in linkage mapping

In this study, χ^2 test was used to identify markers with segregation distortion or significant deviation from the expected 1:1 ratio assuming the trait fits into a single gene model. None of the markers on LG 1 harbouring the recessive yellow seed colour locus showed any segregation distortion which indicates authenticity of linkage map generated. The genetic linkage map published by Cloutier et al. (2009) had 113 markers in total. Out of those, 27 markers showed significant deviation from the expected 1:1 ratio in genotyping. It was lower in the present study (14 out of 99 and $p < 0.05$) possibly because of using RIL populations while in the earlier reported study, a doubled haploid (DH) population was used, and that could show higher segregation distortion as compared to RIL population. The marker order in our genetic map is

comparable with the published high density linkage map (Cloutier et al., 2012a) with some exceptions. This expectation would be because of small LGs and a different mapping population in my study.

Segregation distortion in genetic markers may result in false interpretations. Doubled haploid populations of flax used in previous studies showed significant segregation distortion for markers from the expected Mendelian segregation ratio of 1:1 [Cloutier et al., (2009) (~27%); Spielmyer et al., (1998) (28 %)]. Apart from those, other crop species like *Brassica* (Ferreira et al., 1994; Cloutier et al., 1995) and maize (Bentolila et al., 1992) also showed segregation distortion in DH populations. Segregation distortion could be the result of various factors including selective abortion of specific gametes or selection pressures at various developmental stages like seed germination and efficient seedling establishment (Bajaj, 1983; Cloutier et al., 1995; Zhu et al., 2007). Distortion could lead to spurious linkage groups (Cloutier et al., 1997). However, comparing the newly developed linkage map, LGs and marker order with existing high density linkage maps (developed with three populations and then combined together) (Cloutier et al., 2012a) would help in resolving potential issues.

5.6 Fine-mapping of the *D* locus and genome assembly issues

Fine mapping was done by identifying scaffolds (genomic assembly sequences) (TUFGEN, 2012) harbouring the closest SSR markers to the *D* locus. Scaffold 208 harboured markers on one side of the *D* locus. There was consistency in marker order of the genetic linkage map and order of markers that are physically present on scaffold 208. By identifying SSRs in the scaffold and locus region, I could narrow down the genetic distance between markers and eventually a close marker was identified. The *D* locus is present between LuM566 of my study

and Lu2351 (from Cloutier et al., 2012b) within a genetic distance of 2.8 cM. Following are the important markers and their physical location identified on scaffold 208 that are linked with the *D* locus and follow the order on LG 1; LuM566 (at 53.4 kb); Lu209 (at 184.1 kb) and LuCAPS_110 (at 447.6 kb). Close to Lu209 there are other SSR markers: LuM588 (at 141.2 kb), LuM592 (at 152.5 kb), LuM593 (at 155.1 kb), LuM595 (at 155.6 kb) and LuM597 (at 159.9 kb) which map very closely with no recombination in that region. Near LuM566, there are two markers, LuM569 (at 58.5 kb) and LuM568 (at 58.6 kb), showing no recombination in LG 1.

When BES of markers on other side of the *D* locus are used to identify physical location using BLAST analysis in the TUFGEN flax genome database, it is found that Lu2351 is on scaffold 35 (at 545.9 kb) whereas Lu125 is on scaffold 43 (at 1.11 Mb). As scaffold assemblies do not have overlapping regions, *in-silico* joining of scaffolds to create pseudomolecules for that region is not possible. The identified scaffolds are very large in size (scaffold 35 is 679.5 kb and scaffold 43 is 1640 kb) in proportion to genetic distance that is covered by the markers. This gives low confidence to proceed further in terms of fine mapping. From this physical location analysis it can be concluded that either these markers are duplicated in the flax genome or there are mis-assembly issues in the flax genome database. This mis-assembly issue can be resolved when whole BAC sequence based physical maps are available.

At present, the published physical map of flax is based on short read sequences obtained from reduced genome representation and assembled together in fingerprint contigs (FPC) with supportive data combined from the EST database and BAC end sequencing. The first physical map of flax genome covering ~368 Mb with 416 FPC contigs is published (Ragupathy et al., 2011). The physical map covering approximately 74% of the flax genome (~370 Mb in case of variety CDC Bethune) is the only high density physical map available in flax. In the study by

Cloutier et al. (2012a), they were of opinion that there is a requirement for SNP markers to be added to get high accuracy ordering and orientation of scaffolds of genomic sequences. Because of these issues, we do not have enough confidence for further marker development using scaffolds 43 and 35. Scaffold 208 was used for new SSR as well as CAPS marker development and was successfully used to fine map the recessive yellow locus in this region from one side. The physical location of each marker follows marker order in the genetic linkage map which indicates the accuracy of the mapping results.

5.7 Marker-trait association and stability of the locus

There are few reports of marker-trait association studies in the flax system. Traits of economic importance have been mapped using linkage analysis; these include *Fusarium* wilt resistance (Spielmeyer et al., 1998a); 1000-seed weight; fiber content (Vromans, 2006); fatty acid composition and dominant yellow seed colour loci (Cloutier et al., 2011).

The stability of a QTL can be detected by doing marker-trait association of a mapping population grown in different conditions or time points. Phenotype is the result of genotype and its interaction with environment. The environment (growing conditions) can influence gene expression and eventually the phenotype of interest; seed colour trait is one such example. In the current study, an advanced RIL population ($F_{9:10}$) was used for locus detection and QTL stability was detected with phenotypic data from two independent data sets of seed colour (RGB) value. A major locus was detected by CIM (with LOD threshold of 2.5) underlying the same tightly linked marker within the same linkage group (LG 1) in both datasets. These results strongly indicate that this is a major and stable locus for the yellow seed colour locus (*D*) in the mapping population. Analysis of variance (ANOVA) between RIL populations of two growth conditions

shows a significant difference in colour values at the 5 % level. Even though there is a significant difference in the seed colour values between individual RILs as well as the whole population ($p < 0.05$), the LOD scores for the seed colour locus were high (47.1 in GC, 2008 and 39.3 in GH, 2011 RILs). Hence, the seed colour trait is stable in two growth conditions. Yellow seed colour loci in *Brassica* species are significantly influenced by environmental factors and due to which, the phenotype was unstable (Burbulis and Kott, 2005). This kind of interaction makes it difficult to predict a specific genetic effect of a QTL across different environments as reproducibility is a major issue. To increase the breeding value of QTLs, checking their stability under different environmental conditions is a better approach. The locus detected in the present study explains a high marker-trait association with 89 % of phenotypic variation. This is consistently observed across two different growth conditions (89 % in GC, 2008 RILs and 84 % in GH, 2011 RILs) indicating involvement of a major locus. The identified locus mainly governs the trait. Cloutier et al. (2011; 2012a) reported a major QTL for the dominant yellow seed colour locus. In that, the QTL for the dominant yellow seed colour locus had peak LOD value of 46.6 and accounted for 72% of phenotypic variation.

5.8 Flavonoid biosynthesis and putative candidate gene

The pigments or colours in nature are derived from the phenylpropanoid pathway followed by the flavonoid pathway and condensation of PAs; betalaines (crimson red compounds) are derived from tyrosine and carotenoids (yellow-red compounds) are formed from isopentenyl pyrophosphate (IPP) (reviewed by Tanaka et al., 2008). Yellow seed colour in flax is because of very low levels of condensed tannins in seed (Oomah et al., 1995). Studies in *Brassica napus* show total oil and protein content are 2.6% and 5.0% higher respectively (Simbaya et al., 1995) in yellow-seeded as compared to black-seeded genotypes. High fiber

content in rapeseed meal reduces its digestibility and decrease the value of its meal (Lee et al., 1984). Yellow-seeded turnip rape (5%) and rape seed genotypes (3-4%) have lower fiber contents as compared to black seeded genotypes (Stringam, 1980). Seed meal of yellow *B. napus* has up to 55 % lower fiber content than black seeds (Rahman et al., 2001). Therefore, seed colour is directly or indirectly associated with other traits such as fiber, oil and protein content.

Yellow seed colour and lower fiber content are interconnected as the biochemical pathway leading to lignin and condensed tannin accumulation have common precursors from the phenylpropanoid pathway (Grisebach, 1981). Studies indicate that yellow-seeded flax accumulates lower levels of condensed tannins in the seed-coat pigment layer as compared to brown-seeded flax genotypes (M. Kulkarni and G. Selvaraj, data unpublished). If condensed tannin accumulation is defective in the seed coat pigment layer it might have a direct or indirect effect on the lignan biosynthesis pathway. Flavonols are required for production of anthocyanins and PAs. The production of flavonols is catalyzed by several flavonoid hydroxylases and cytochrome P450 based enzymes such as flavanone 3 hydroxylase (F3H), flavonoid 3' hydroxylase (F3'H), flavonoid 3' 5' hydroxylase (F3'5'H), flavonol synthase (FLS), dihydroflavonol reductase (DFR) (Winkel-Shirley, 2001; Tanaka et al., 2008).

Based on fine mapping, the SSR marker, LuM566, tightly linked to the *D* locus is found at the 53.4 kb region from the start of scaffold 208. The physical sequence of the scaffold was obtained from the flax genome database (TUFGEN, 2012) to search for a putative candidate gene. Analyses for gene prediction were performed on scaffold 208 sequence using FGENESH online software on the first 50 kb region based on location of the closest marker, LuM566. A putative F3'H gene among 14 predicted genes was identified as the most probable candidate gene. This is a key structural gene involved in the flavonoid biosynthesis pathway prior to

condensed tannin biosynthesis as reported from its functional analysis in Asteraceae family (Seitz et al., 2006) and recently in *Pericallis* × *hybrid* where a homolog of the F3'5'H gene was tested for its function through heterologous expression (Sun et al., 2013). qPCR analysis of the putative candidate gene in seed-coat (15 DAF) and flower petal tissues shows very low expression in both tissues of G1186/94. This F3'H or F3'5'H could be the potential candidate gene responsible for low pigments in both of the tissues in the recessive yellow line (G1186/94) as it was found in the mapped region of the *D* locus.

6. FUTURE RESEARCH AND CONCLUSION

6.1 Future research

The candidate gene identified through fine mapping in the current study is interesting as it shows potential to be the gene regulating seed as well as flower colour. It is possible that this gene is defective in G1186/94 resulting in yellow seed as well as white petals. The petal colour of CDC Bethune is blue, possibly because of presence of delphinidin as most of the blue or purple colour in nature is related to presence of delphinidin (Winkel-Shirley, 2001). Flower petals of G1186/94 are white because they may not synthesize delphinidin. It is also a substrate for condensed tannin accumulation in seed-coats.

In the present prediction and annotation, the putative candidate gene is showing as F3'H. Cyanidin, which is derived from F3'H activity, gives a reddish colour whereas delphinidin and its derivatives, derived from F3'5'H activity, impart blue colour to flower petals. Because CDC Bethune has blue flower petals, it is possible that the putative candidate gene is actually playing a role in 3'5' hydroxylation. To substantiate this hypothesis, positional cloning of the predicted candidate gene has to be performed and sequenced in both the brown and the yellow-seeded parent. The gene or genes in the locus region might be a structural or regulatory gene involved in regulation of seed-coat pigment biosynthesis and deposition, so it is important to clone and functionally characterize it in future. If the hypothesis is true, then it is possible that a mutation in a single copy of the gene is responsible for both yellow seed colour and white petals. It should also be noted that these two traits are tightly linked to each other and in a population of 470 RILs, no recombinant was found. Another possibility is that there might be two genes (each controlling seed colour and petal colour) which are close to each other such that they are tightly linked and always go together in mapping.

Preliminary gene expression analysis of the predicted candidate gene showed low expression both in seed coat as well as flower petal tissues of G1186/94 indicating that this is an important candidate gene that should be characterized. Currently, available resources are insufficient to resolve this issue and further fine mapping and gene searches in that region are required. To resolve the issue of physical location of markers and finding scaffolds which cover the entire mapped region, developing and screening a BAC library of G1186/94 and CDC Bethune has to be done. This approach will help in finding other possible candidate gene/s.

6.2 Conclusion

The yellow seed colour locus has been a mystery in flax for the last 90 years since Tammes (1922) proposed *G*, *D* and *B* as three basic seed colour loci. In the present study, for the first time, a locus for the European recessive yellow seed colour mutation *d* was mapped and a stable QTL was identified governing seed colour. A tightly linked SSR marker (LuM566) for yellow seed colour was made available through this study that can be used by breeders for marker assisted selection of the European recessive yellow seed colour locus. The *D* locus has been confined in a 2.8 cM region and a potential and interesting putative candidate gene was provided. This gene can be further cloned and functionally characterized to understand the molecular basis of the yellow seed colour phenotype. To summarize, this study marks the first successful report on genetic mapping based putative candidate gene identification for the recessive yellow seed colour mutation in the *D* locus in flax.

7. REFERENCES

Abeynayake SW, Panter S, Chapman R, Webster T, Rochfort S, Mouradov A, Spangenberg G (2012) Biosynthesis of proanthocyanidins in white clover flowers: cross talk with in the flavonoid pathway. Plant Physiology **158:** 666-678

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

American Cancer Society (2011). *In* Flaxseed: Find Support & Treatment http://www.cancer.org/treatment/treatmentsandsideeffects/complementaryandalternativemedicine/herbsvitaminsandminerals/flaxseed, Atlanta, Georgia

American Diabetes Services (2012). *In* Flax seed: nature's little powerhouse. http://www.americandiabetes.com/living-diabetes/diabetes-nutrition-articles/flax-seed-natures-little-powerhouse

Attoumbre J, Bienaime C, Dubois F, Fliniaux M-A, Chabbert B, Baltora-Rosset S (2010) Development of antibodies against secoisolariciresinol - Application to the immunolocalization of lignans in *Linum usitatissimum* seeds. Phytochemistry **71:** 1979-1987

Bajaj YPS (1983) Regeneration of plants from pollen-embryos of *Arachis, Brassica* and *Triticum* sps. cryopreserved for one year. Current Science **52:** 484-486

Barnes DK, Culbertson JD, Lambert JW (1960) Inheritance of seed and flower colors in flax. Agronomy Journal **52:** 456–459

Beard BH, Comstock VE (1980) Flax. *In* WR Fehr, HH Hadley, eds, Hybridization of crop plants. American Society of Agronomy - Crop Science Society of America, Madison, WI, pp 357-366

Bentolila S, Hardy T, Guitton C, Freyssinet G (1992) Comparative genetic analyses of F_2 plants and anther culture derived plants of maize. Genome **35:** 575-582

Bickel CL, Gadani S, Lukacs M, Cullis CA (2011) SSR markers developed for genetic mapping in flax (*Linum usitatissimum* L.). Research and Reports in Biology **2:** 23-29

Biswas T, Das A, Bhattacharyya S (2013) Marker assisted selection for developing high yielding submergence tolerant rice (*Oryza sativa* L.) genotypes with slender grain. Cereal Research Communications **41:** 35-44

Boesewinkel FD (1980) Development of ovule and testa of *Linum usitatissimum L.* Acta Botanica Neerlandica **29:** 17-32

Buer CS, Imin N, Djordjevic MA (2010) Flavonoids: new roles for old molecules. Journal of Integrative Plant Biology **52:** 98-111

Burbulis N, Kott LS (2005) A new yellow-seeded canola genotype originating from double low black-seeded *Brassica napus* cultivars. Canadian Journal of Plant Science **85:** 109-114

Charles DJ (2013) Natural Antioxidants. *In* Antioxidant properties of spices, herbs and other sources. Springer New York, pp 39-64

Chen BY, Heneen WK (1992) Inheritance of seed colour in *Brassica campestris* L. and breeding for yellow-seeded *B. napus* L. Euphytica **59:** 157-163

Chen BY, Jørgensen RB, Cheng BF, Heneen WK (1997) Identification and chromosomal assignment of RAPD markers linked with a gene for seed colour in a *Brassica campestris*-Alboglabra addition line. Hereditas **126:** 133-138

Cloutier S, Cappadocia M, Landry BS (1995) Study of microspore-culture responsiveness in oilseed rape (*Brassica napus* L.) by comparative mapping of a F_2 population and two microspore-derived populations. Theoretical and Applied Genetics **91:** 841-847

Cloutier S, Cappadocia M, Landry BS (1997) Analysis of RFLP mapping inaccuracy in *Brassica napus* L. Theoretical and Applied Genetics **95:** 83-91

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

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

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

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

Culbertson JO, Comstock VE, Frederiksen RA (1960) Further studies on the effect of seed coat color on agronomic and chemical characters and seed injury in flax. Agronomy Journal **52:** 210-212

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

Cullis CA, Kole C (2007) Flax. *In* C Kole, ed, Oilseeds: Genome mapping and molecular breeding in plants, Vol 2. Springer Berlin Heidelberg, pp 275-295

Davies KM, Schwinn KE (2006) Molecular biology and biotechnology of flavonoid biosynthesis. *In* Flavonoids: chemistry, biochemistry and applications. CRC Press LLC, pp 143-218

Deng X, Long S, He D, Li X, Wang Y, Liu J, Chen X (2010) Development and characterization of polymorphic microsatellite markers in *Linum usitatissimum*. Journal of Plant Research **123:** 119-123

Deng X, Long S, He D, Li X, Wang Y, Hao D, Qiu C, Chen1 X (2011) Isolation and characterization of polymorphic microsatellite markers from flax (*Linum usitatissimum* L.). African Journal of Biotechnology **10:** 734-739

Dexter JE (2009) Quantification and risk assessment of seed-mediated gene flow with flax as a platform crop for bioproducts (Doctoral dissertation). University of Alberta, Edmonton

Diederichsen A, Raney JP, Duguid SD (2006) Variation of mucilage in flax seed and its relationship with other seed characters. Crop Science **46:** 365-371

Diederichsen A, Richards K (2003) Cultivated flax and the genus *Linum* L.,Taxonomy and germplasm conservation. *In* AD Muir, ND Westcott, eds, Flax: The genus *Linum*. Taylor & Francis, London

Dillman AC (1938) Natural crossing in flax. Agronomy Journal **30:** 279-286

Dillman AC, Brinsmade JC (1938) Effect of spacing on the development of the flax plant. Agronomy Journal **30:** 267-278

Dixon RA, Liu C, Jun JH (2012) Metabolic engineering of anthocyanins and condensed tannins in plants. Current Opinion in Biotechnology **24:** 1-7

Eujayl I, Sorrells ME, Baum M, Wolters P, Powell W (2002) Isolation of EST-derived microsatellite markers for genotyping the A and B genomes of wheat. Theoretical and Applied Genetics **104:** 399-407

FAOSTAT (2013) Countries by commodity. *In*. http://faostat3.fao.org/home/index.html#VISUALIZE_TOP_20, Rome, Italy

Ferreira ME, Williams PH, Osborn TC (1994) RFLP mapping of *Brassica napus* using doubled haploid lines. Theoretical and Applied Genetics **89:** 615-621

FGENESH (2012) http://linux1.softberry.com/berry.phtml?topic=fgenesh&group=programs&subgroup=gfind. *In* HMM-based gene structure prediction (multiple genes, both chains). www.softberry.com

Flax Council of Canada (2012). *In*. www.flaxcouncil.ca

Fu F-Y, Liu L-Z, Chai Y-R, Chen L, Yang T, Jin M-Y, Ma A-F, Yan X-Y, Zhang Z-S, Li J-N (2007) Localization of QTLs for seed color using recombinant inbred lines of *Brassica napus* in different environments. Genome **50:** 840-854

Fu Y-B (2005) Geographic patterns of RAPD variation in cultivated flax. Crop Science **45:** 1084-1091

Fu Y-B (2011) Genetic evidence for early flax domestication with capsular dehiscence. Genetic Resources and Crop Evolution **58:** 1119-1128

Genographer (2013) Genographer-2.1.4. *In* http://sourceforge.net/projects/genographer/files/. Dice Holdings, Inc

Green AG (1986) A mutant genotype of flax (*Linum usitatissimum* L.) containing very low levels of linolenic acid in its seed oil. Canadian Journal of Plant Science **66:** 499-503

Green AG, Dribnenki JCP (1995) Breeding and development of LINOLA (low-linolenic flax). *In* Proceedings of the third meeting of the International Flax Breeding Group, St.Valery en Caux, France, pp 145-150

Green AG, Marshall DR (1984) Isolation of induced mutants in linseed (*Linum usitatissimum*) having reduced linolenic acid content. Euphytica **33:** 321-328

Grisebach H (1981) Lignins. *In* EE Conn, ed, The biochemistry of plants, Vol 7. Academic, New York, pp 457-478

Gutierrez L, Conejero G, Castelain M, Guenin S, Verdeil J-L, Thomasset B, Van Wuytswinkel O (2006) Identification of new gene expression regulators specifically expressed during plant seed maturation. Journal of Experimental Botany **57:** 1919-1932

Hayward HE (1938) *Linum usitatissimum*. *In* HE Hayward, ed, The structure of economic plants. Macmillan, New York, pp 371-410

Heneen WK, Geleta M, Brismar K, Xiong Z, Pires JC, Hasterok R, Stoute AI, Scott RJ, King GJ, Kurup S (2012) Seed colour loci, homoeology and linkage groups of the C genome chromosomes revealed in *Brassica rapa* and *B. oleracea* monosomic alien addition lines. Annals of Botany **109:** 1227-1242

Hichri I, Barrieu F, Bogs J, Kappel C, Delrot S, Lauvergeat V (2011) Recent advances in the transcriptional regulation of the flavonoid biosynthetic pathway. Journal of Experimental Botany **62:** 2465-2483

ImageJ (2013) Image processing and analysis in Java. *In*. http://rsbweb.nih.gov/ij/

Johnsson P, Kamal-Eldin A, Lundgren LN, Aman P (2000) HPLC method for analysis of secoisolariciresinol diglucoside in flaxseeds. Journal of Agricultural and Food Chemistry **48:** 5216-5219

Kebede B, Cheema K, Greenshields DL, Li C, Selvaraj G, Rahman H (2012) Construction of genetic linkage map and mapping of QTL for seed color in *Brassica rapa*. Genome **55:** 813-823

Kosambi DD (1943) The estimation of map distances from recombination values. Annals of Eugenics **12:** 172-175

Kumar S, You F, Cloutier S (2012) Genome wide SNP discovery in flax through next generation sequencing of reduced representation libraries. BMC Genomics **13:** 684

Kvavadze E, Bar-Yosef O, Belfer-Cohen A, Boaretto E, Jakeli N, Matskevich Z, Meshveliani T (2009) 30,000-year-old wild flax fibers. Science **325:** 1359

Lambrides CJ, Godwin ID, Lawn RJ, Imrie BC (2004) Segregation distortion for seed testa color in mungbean (*Vigna radiata* L. Wilcek). Journal of Heredity **95:** 532-535

Lander ES, Green P, Abrahamson J, Barlow A, Daly MJ, Lincoln SE, Newberg LA (1987) MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. Genomics **1:** 174-181

Lee EK, Cibrian-Jaramillo A, Kolokotronis S-O, Katari MS, Stamatakis A, Ott M, Chiu JC, Little DP, Stevenson DW, McCombie WR, Martienssen RA, Coruzzi G, DeSalle R (2011) A functional phylogenomic view of the seed plants. PLoS Genetics **7:** e1002411

Lee PA, Pittam S, Hill R (1984) The voluntary food intake by growing pigs of diets containing 'treated' rapeseed meals or extracts of rapeseed meal. British Journal of Nutrition **52:** 159-164

Lepiniec L, Debeaujon I, Routaboul J-M, Baudry A, Pourcel L, Nesi N, Caboche M (2006) Genetics and biochemistry of seed flavonoids. Annual Review of Plant Biology **57:** 405-430

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R (2009) The sequence alignment/map format and SAMtools. Bioinformatics **25:** 2078 - 2079

Lin JK, Weng MS, Grotewold E (2006) Flavonoids as nutraceuticals. *In* The science of flavonoids. Springer New York, pp 213-238

Liu L-Z, Meng J-L, Lin N, Chen L, Tang Z-L, Zhang X-K, Li J-N (2006) QTL mapping of seed coat color for yellow seeded *Brassica napus*. Acta Genetica Sinica **33:** 181-187

Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using Real-Time quantitative PCR and the 2-[delta][delta]CT method. Methods **25:** 402-408

Martins WS, Lucas DCS, Neves KFS, Bertioli DJ (2009) WebSat - A web software for microsatellite marker development. Bioinformation **3:** 282-283

Mazur W, Adlercreutz H (1998) Naturally occuring oestrogens in food. Pure and Applied Chemistry **70:** 1759–1776

Mazur W, Fotsis T, Wähälä K, Ojala S, Salakka A, Adlercreutz H (1996) Isotope dilution gas chromatographic-mass spectrometric method for the determination of isoflavonoids, coumestrol, and lignans in food samples. Analytical Biochemistry **233:** 169-180

Milne I, Bayer M, Cardle L, Shaw P, Stephen G, Wright F, Marshall D (2010) Tablet - next generation sequence assembly visualization. Bioinformatics **26:** 401-402

Mittapalli O (2002) Inheritence of seed coat colour in flax (*Linum usitatissimum L.*) (Masters dissertation). University of Saskatchewan, Saskatoon

Mittapalli O, Rowland G (2003) Inheritance of seed color in flax. Crop Science **43:** 1945-1951

Muir AD (2009) Flax lignans: new opportunities for functional foods. Food Science and Technology Bulletin: Functional Foods **6:** 61-79

Negi MS, Devic M, Delseny M, Lakshmikumaran M (2000) Identification of AFLP fragments linked to seed coat colour in *Brassica juncea* and conversion to a SCAR marker for rapid selection. Theoretical and Applied Genetics **101:** 146-152

NCBI (2013) http://www.ncbi.nlm.nih.gov/nucest

Oh TJ, Gorman M, Cullis CA (2000) RFLP and RAPD mapping in flax (*Linum usitatissimum*). Theoretical and Applied Genetics **101:** 590-593

Oomah BD, Kenaschuk EO, Mazza G (1995) Phenolic acids in flaxseed. Journal of Agricultural and Food Chemistry **43:** 2016-2019

Oomah BD, Mazza G (1993) Flaxseed proteins- a review. Food Chemistry **48:** 109-114

Pang Y, Peel GJ, Wright E, Wang Z, Dixon RA (2007) Early steps in proanthocyanidin biosynthesis in the model legume *Medicago truncatula*. Plant Physiology **145:** 601-615

Peterson J, Dwyer J, Adlercreutz H, Scalbert A, Jacques P, McCullough ML (2010) Dietary lignans: physiology and potential for cardiovascular disease risk reduction. Nutrition Reviews **68:** 571-603

Piquemal J, Cinquin E, Couton F, Rondeau C, Seignoret E, doucet I, Perret D, Villeger MJ, Vincourt P, Blanchard P (2005) Construction of an oilseed rape (*Brassica napus* L.) genetic map with SSR markers. Theoretical and Applied Genetics **111:** 1514-1523

Popescu F, Marinescu I (1996) Y1-a dominant gene determining yellow seed color in linseed. Problem. Genetic. Theory **28:** 99-106

Qiu S-X, Lu Z-Z, Luyengi L, Lee SK, Pezzuto JM, Farnsworth NR, Thompson LU, Fong HHS (1999) Isolation and characterization of flaxseed (*Linum usitatissimum*) constituents. Pharmaceutical Biology **37:** 1-7

Queller DC, Strassmann JE, Hughes CR (1993) Microsatellites and kinship. Trends in Ecology and Evolution **8:** 285-288

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

Rahman M, McVetty PE, Li G (2007) Development of SRAP, SNP and Multiplexed SCAR molecular markers for the major seed coat color gene in *Brassica rapa* L. Theoretical and Applied Genetics **115:** 1101-1107

Rahman MH (2001) Production of yellow-seeded *Brassica napus* through interspecific crosses. Plant Breeding **120:** 463-472

Rahman MH, Joersbo M, Poulsen MH (2001) Development of yellow-seeded *Brassica napus* of double low quality. Plant Breeding **120:** 473-478

Roach MJ, Deyholos MK (2008) Microarray analysis of developing flax hypocotyls identifies novel transcripts correlated with specific stages of phloem fibre differentiation. Annals of Botany **102:** 317-330

Roose-Amsaleg C, Cariou-Pham E, Vautrin D, Tavernier R, Solignac M (2006) Polymorphic microsatellite loci in *Linum usitatissimum*. Molecular Ecology Notes **6:** 796-799

Routaboul J-M, Dubos C, Beck G, Marquis C, Bidzinski P, Loudet O, Lepiniec L (2012) Metabolite profiling and quantitative genetics of natural variation for flavonoids in *Arabidopsis*. Journal of Experimental Botany **63:** 3749-3764

Routaboul J-M, Kerhoas L, Debeaujon I, Pourcel L, Caboche M, Einhorn J, Lepiniec L (2006) Flavonoid diversity and biosynthesis in seed of *Arabidopsis thaliana*. Planta **224:** 96-107

Rowland G, McHughen A, Gusta L, Bhatty R, MacKenzie S, Taylor D (1995) The application of chemical mutagenesis and biotechnology to the modification of linseed (*Linum usitatissimum* L.). Euphytica **85:** 317-321

Rowland GG (1991) An EMS-induced low-linolenic-acid mutant in McGregor flax (*Linum usitatissimum* L.). Canadian Journal of Plant Science **71:** 393-396

Rowland GG, Hormis YA, Rashid KY (2002) CDC Bethune flax. Can J Plant Sci **82:** 101 - 102

Rozen S, Skaletsky HJ (2000) Primer3 on the WWW for general users and for biologist programmers. *In* S Krawetz, S Misener, eds, Bioinformatics Methods and Protocols: Methods in Molecular Biology. Humana Press, Totowa, NJ, pp 365-386

Saito K, Yonekura-Sakakibara K, Nakabayashi R, Higashi Y, Yamazaki M, Tohge T, Fernie AR (2013) The flavonoid biosynthetic pathway in Arabidopsis: Structural and genetic diversity. Plant Physiology and Biochemistry **(online only):** 1-14

SaskFlax (2007) Flax grower survey – Seeding details. *In* The Saskatchewan flax grower. http://www.saskflax.com/article_seeding_details.html, Saskatoon, SK

Schewe LC, Sawhney VK, Davis AR (2011) Ontogeny of floral organs in flax (*Linum usitatissimum*; Linaceae). American Journal of Botany **98:** 1077-1085

Seitz C, Eder C, Deiml B, Kellner S, Martens S, Forkmann G (2006) Cloning, functional identification and sequence analysis of flavonoid 3′-hydroxylase and flavonoid 3′,5′-hydroxylase cDNAs reveals independent evolution of flavonoid 3′,5′-hydroxylase in the Asteraceae family. Plant Molecular Biology **61:** 365-381

Sharma HSS, Sumere CFV (1992) Enzyme treatment of flax. Genetic Engineer and Biotechnologist **12:** 19-23

Shaw FJ, Khan AR, Alam M (1931) Studies in Indian oilseeds. V. The inheritance of characters in Indian linseed. Indian Journal of Agricultural Science **1:** 1-57

Simbaya J, Slominski BA, Rakow G, Campbell LD, Downey RK, Bell JM (1995) Quality characteristics of yellow-seeded *Brassica* seed meals: Protein, carbohydrate, and dietary fiber components. Journal of Agricultural and Food Chemistry **43:** 2062-2066

Sood SP, Bhateria S, Sood S (2012) Inheritance of flower and seed colour in flax (*Linum usitatissimum* L.) over environments in north western Himalaya. National Academy Science Letters **35:** 127-130

Soto-Cerda JB, Carrasco AR, Aravena AG, Urbina AH, Navarro SC (2011) Identifying novel polymorphic microsatellites from cultivated flax (*Linum usitatissimum* L.) following data mining. Plant Molecular Biology Reporter **29:** 753-759

Soto-Cerda JB, Maureira-Butler I, Muñoz G, Rupayan A, Cloutier S (2012) SSR-based population structure, molecular diversity and linkage disequilibrium analysis of a collection of flax (*Linum usitatissimum* L.) varying for mucilage seed-coat content. Molecular Breeding **30:** 875-888

Spielmeyer W, Green AG, Bittisnich D, Mendham N, Lagudah ES (1998a) Identification of quantitative trait loci contributing to *Fusarium* wilt resistance on an AFLP linkage map of flax (*Linum usitatissimum*). Theoretical and Applied Genetics **97:** 633-641

Spielmeyer W, Lagudah ES, Mendham N, Green AG (1998b) Inheritance of resistance to flax wilt (*Fusarium oxysporum* f.sp. lini Schlecht) in a doubled haploid population of *Linum usitatissimum* L. Euphytica **101:** 287-291

Statistics Canada (2012). *In* Cereals and Oilseeds Review - June 2012. http://www.statcan.gc.ca/pub/22-007-x/22-007-x2012006-eng.pdf

Stringam GR (1980) Inheritance of seed color in turnip rape. Canadian Journal of Plant Science **60:** 331-335

Struijs K, Vincken J-P, Verhoef R, van Oostveen-van Casteren WHM, Voragen AGJ, Gruppen H (2007) The flavonoid herbacetin diglucoside as a constituent of the lignan macromolecule from flaxseed hulls. Phytochemistry **68:** 1227-1235

Sun Y, Huang H, Meng L, Hu K, Dai S-L (2013) Isolation and functional analysis of a homolog of flavonoid 3′,5′-hydroxylase gene from *Pericallis×hybrida*. Physiologia Plantarum (online)

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

Tammes T (1928) The genetics of genus *Linum*. Bibliographia Genetica **IV:** 1-36

Tanaka Y, Sasaki N, Ohmiya A (2008) Biosynthesis of plant pigments: anthocyanins, betalains and carotenoids. The Plant Journal **54:** 733-749

Tanner GJ, Francki KT, Abrahams S, Watson JM, Larkin PJ, Ashton AR (2003) Proanthocyanidin biosynthesis in plants: Purification of legume leucoanthocyanidin reductase and molecular cloning of its cDNA. Journal of Biological Chemistry **278:** 31647-31656

Thompson LU, Seidl MM, Rickard SE, Orcheson LJ, Fong HHS (1996) Antitumorigenic effect of a mammalian lignan precursor from flaxseed. Nutrition and Cancer **26:** 159-165

Tian L, Pang Y, Dixon R (2008) Biosynthesis and genetic engineering of proanthocyanidins and (iso)flavonoids. Phytochemistry Reviews **7:** 445-465

TUFGEN (2012) www.linum.ca.

Vaisey-Genser M, Morris DH (2003) Introduction: History of the cultivation and uses of flaxseed. *In* AD Muir, ND Westcott, eds, Flax: The genus *Linum*. Taylor & Francis, London, pp 1-21

Varshney RK, Grosse I, Hahnel U, Siefken R, Prasad M, Stein N, Langridge P, Altschmied L, Graner A (2006) Genetic mapping and BAC assignment of EST-derived SSR markers shows non-uniform distribution of genes in the barley genome. Theoretical and Applied Genetics **113:** 239-250

Venglat P, Xiang D, Qiu S, Stone S, Tibiche C, Cram D, Alting-Mees M, Nowak J, Cloutier S, Deyholos M, Bekkaoui F, Sharpe A, Wang E, Rowland G, Selvaraj G, Datla R (2011) Gene expression analysis of flax seed development. BMC Plant Biology **11:** 74

Vrinten P, Hu Z, Munchinsky M-A, Rowland G, Qiu X (2005) Two FAD3 desaturase genes control the level of linolenic acid in flax seed. Plant Physiology **139:** 79-87

Vromans J (2006) Molecular genetic studies in flax (*Linum usitatissimum* L.) (Doctoral dissertation). Wageningen University, The Netherlands

Wang Z, Hobson N, Galindo L, Zhu S, Shi D, McDill J, Yang L, Hawkins S, Neutelings G, Datla R, Lambert G, Galbraith DW, Grassa CJ, Geraldes A, Cronk QC, Cullis C, Dash PK, Kumar PA, Cloutier S, Sharpe AG, Wong GKS, Wang J, Deyholos MK (2012) The genome of flax (*Linum usitatissimum*) assembled de novo from short shotgun sequence reads. The Plant Journal **72:** 461-473

Winkel-Shirley B (2001) Flavonoid biosynthesis. A colorful model for genetics, biochemistry, cell biology, and biotechnology. Plant Physiology **126:** 485-493

www.konicaminolta.eu Precise colour communication. *In*. Konica Minolta

www.lgcgenomics.com http://www.lgcgenomics.com/kasp-genotyping-reagents. *In* KASP genotyping reagents. LGC Ltd., Teddington, UK

Xiao S, Xu J, Li Y, Zhang L, Shi S, Shi S, Wu J, Liu K (2007) Generation and mapping of SCAR and CAPS markers linked to the seed coat color gene in *Brassica napus* using a genome-walking technique. Genome **50:** 611-618

Xie DY, Sharma SB, Paiva NL, Ferreira D, Dixon RA (2003) Role of anthocyanidin reductase, encoded by BANYULS in plant flavonoid biosynthesis. Science **299:** 396-399

Zhi-wen L, Ting-dong F, Jin-xing T, Bao-yuan C (2005) Inheritance of seed colour and identification of RAPD and AFLP markers linked to the seed colour gene in rapeseed (*Brassica napus* L.). Theoretical and Applied Genetics **110:** 303-310

Zhou WC, Kolb FL, Bai GH, Domier LL, Boze LK, Smith NJ (2003) Validation of a major QTL for scab resistance with SSR markers and use of marker-assisted selection in wheat. Plant Breeding **122:** 40-46

Zhu C, Wang C, Zhang Y-M (2007) Modeling segregation distortion for viability selection I. Reconstruction of linkage maps with distorted markers. Theoretical and Applied Genetics **114:** 295-305

8. APPENDICES

Appendix 1: The list of polymorphic SSR markers between G1186/94 and CDC Bethune developed by genomic sequences

	Marker name	Fragment length (bp)		Primer name	Sequence (5' to 3')	Tm (°C)	primer length (mer)
		CDC Bethune	G1186/94				
1	LuM3	204	207	LuMF3	ACGGAAGGTTTAGTGTCGAGAA	58	22
				LuMR3	ATGGGAAGAAGGTAAAGCCAAT	58	22
2	LuM18	211	229	LuMF18	TTGCATAGACAAATGGCTCATC	58	22
				LuMR18	ACCAACAACGCAAAAGCATAG	58	21
3	LuM47	233	231	LuMF47	CAAATCAGAATGTGCGTGTGTA	58	22
				LuMR47	GGACGTTATGGTCTCTGCTCTC	58	22
4	LuM56	247	244	LuMF56	AATAGATGCTAGAGTGCTCGCC	58	22
				LuMR56	TCGTCCATAAGCTGGAAATCTT	58	22
5	LuM63	239	245	LuMF63	ACAGCTAGGGGTAGGCCAGT	58	20
				LuMR63	CCCACAACACACACAAATAACC	58	22
6	LuM67	132	133	LuMF67	ATAGGATAGGACATGACGAGCC	58	22
				LuMR67	TGCTGAGAAGGTGAAGACTGAA	58	22
7	LuM68	298	300	LuMF68	TAAATCAGTCAGGTTCGGTTTG	57	22
				LuMR68	CGCATCAGAATCCATCGTATAA	58	22
8	LuM71	369	367	LuMF71	AGAAAATCGAAAGATGAAGGGG	59	22
				LuMR71	CTACTCTTCCCACGTTGACCA	58	21
9	LuM73	229	224	LuMF73	GGTCTAACGGAATGCCTATCAG	58	22
				LuMR73	CATGTCTTCCCCTTCTCACTTC	58	22
10	LuM82	295	298	LuMF82	CGCATCAGAATCCATCGTATAA	58	22
				LuMR82	TAAATCAGTCAGGTTCGGTTTG	57	22
11	LuM84	231	250	LuMF84	CGACAGTTCGTAGGAGAGAAATAA	57	24
				LuMR84	TTGAGTTGATATGTTCCGCTGT	58	22
12	LuM88	146	140	LuMF88	ACTGCTTTGAATCGGAGAGAAG	58	22
				LuMR88	GGGGATGAAGGTCTACCATGT	58	21
13	LuM93	341	313	LuMF93	ATTTACTAGCAGCAGGTTGGGA	58	22
				LuMR93	GGAAAACATACGATAGCCAAGC	58	22
14	LuM101	196	184	LuMF101	ATAGGTAGTGCCGTTGGGTTT	58	21
				LuMR101	AACAAAGGGTAGCCATGAAGAA	58	22
15	LuM106	174	162	LuMF106	AGAAGAGGGCATCCAACAAAG	59	21
				LuMR106	AACAAAGGGTAGCCATGAAGAA	58	22
16	LuM109	255	313	LuMF109	CATTTTGTGGAAGGACAACAAG	58	22
				LuMR109	ACAATGATAGGAACAGCACGTC	57	22
17	LuM126	346	343	LuMF126	GCTTCTTCTTCTTCACAACATCAC	58	24

				LuMR126	CCGGAACCCCTTACAAAGAT	58	20
18	LuM132	231	240, 241	LuMF132	AGAAGAGAGGATGGGGTTGAAG	59	22
				LuMR132	ATGTTTATTAGTGGGTGGTCGG	58	22
19	LuM134	174	162	LuMF134	AGAAGAGGGCATCCAACAAAG	59	21
				LuMR134	AACAAAGGGTAGCCATGAAGAA	58	22
20	LuM137	212	211	LuMF137	AACCAGAAACAGAAACAGAGGC	58	22
				LuMR137	ATACCTGGATTGGAGTTGGAGA	58	22
21	LuM140	255	252	LuMF140	CAAACAAGAACCCTAAAACCCA	58	22
				LuMR140	TTGGATCACTAACGCTGCAAT	59	21
22	LuM147	225	226	LuMF147	TGGAGGAGGAGTGGTAAGAAAA	58	22
				LuMR147	ATAGGATAGGACATGACGAGCC	58	22
23	LuM149	248	304	LuMF149	ACAATGATAGGAACAGCACGTC	57	22
				LuMR149	CATTTTGTGGAAGGACAACAAG	58	22
24	LuM152	276	282	LuMF152	GACTGTAGCAAACACCATGCAG	59	22
				LuMR152	CTCCTAAACCCACTCACCAAAG	58	22
25	LuM159	321	251	LuMF159	TCAACATCCTTTTGTGTCCAAC	58	22
				LuMR159	TTCTGCTGTTTCGTTAGCAAAG	58	22
26	LuM162	127	151	LuMF162	CCATCTCCTTCATTCTTACCTCC	58	23
				LuMR162	TCAATAAACAACAGCGGAACTG	58	22
27	LuM167	162	138	LuMF167	TGCGTTTCAGTCTCTTTTGTGT	58	22
				LuMR167	CTTTGTCAGGCTCCTTCTTTTG	58	22
28	LuM170	255	313	LuMF170	CATTTTGTGGAAGGACAACAAG	58	22
				LuMR170	ACAATGATAGGAACAGCACGTC	57	22
29	LuM171	191, 194	179, 182	LuMF171	AAGAGGAACAAAGGGTAGCCA	58	21
				LuMR171	GTAGTGCCGTTGGGTTTGAG	59	20
30	LuM172	288	279	LuMF172	AAAGCGATGGAGAAATTAGGTG	58	22
				LuMR172	ACAGTGCGTAGGGGAGAAATAA	58	22
31	LuM173	288	278	LuMF173	AAAGCGATGGAGAAATTAGGTG	58	22
				LuMR173	ACAGTGCGTAGGGGAGAAATAA	58	22
32	LuM177	282	273	LuMF177	ACAGTGCGTAGGGGAGAAATAA	58	22
				LuMR177	AAAGCGATGGAGAAATTAGGTG	58	22
33	LuM185	287	278, 279	LuMF185	AAAGCGATGGAGAAATTAGGTG	58	22
				LuMR185	ACAGTGCGTAGGGGAGAAATAA	58	22
34	LuM186	238	232	LuMF186	ATGGAGTGTATGACAGCAGACG	58	22
				LuMR186	CAAAATGTTCCTTCCTTCCTTG	58	22
35	LuM188	407	401	LuMF188	GCTTTATGGCAAGCTCTATCGT	58	22
				LuMR188	GCCAGATTTATGCTCGTGATCT	59	22
36	LuM193	129	127	LuMF193	TTATGTGTGGGAATTGGACACT	57	22
				LuMR193	GGCCAACTAACTCCTGAAACTC	57	22
37	LuM221	146	148	LuMF221	CAGTGCGATCAATAGAGTTGCT	58	22
				LuMR221	AAAGCATGGAGATAGGGTGAGA	58	22

38	LuM231	145	148	LuMF231	CAGTGCGATCAATAGAGTTGCT	58	22
				LuMR231	AAAGCATGGAGATAGGGTGAGA	58	22
39	LuM241	145	148	LuMF241	CAGTGCGATCAATAGAGTTGCT	58	22
				LuMR241	AAAGCATGGAGATAGGGTGAGA	58	22
40	LuM246	129	104	LuMF246	TTATGTGTGGGAATTGGACACT	57	22
				LuMR246	GGCCAACTAACTCCTGAAACTC	57	22
41	LuM248	133	130, 131	LuMF248	ATCGAGGAGATTTGTTGCC	56	19
				LuMR248	CAAGCCTAGCATCTGAAGTTTT	56	22
42	LuM249	113	105	LuMF249	TTATGTGTGGGAATTGGACACT	57	22
				LuMR249	GGCCAACTAACTCCTGAAACTC	57	22
43	LuM253	129	127	LuMF253	TTATGTGTGGGAATTGGACACT	57	22
				LuMR253	GGCCAACTAACTCCTGAAACTC	57	22
44	LuM262	130	127	LuMF262	TTATGTGTGGGAATTGGACACT	57	22
				LuMR262	GGCCAACTAACTCCTGAAACTC	57	22
45	LuM263	145	149	LuMF263	CAGTGCGATCAATAGAGTTGCT	58	22
				LuMR263	AAAGCATGGAGATAGGGTGAGA	58	22
46	LuM266	216	222	LuMF266	ATACTTGCTGAGTGCGAAAGC	58	21
				LuMR266	GTCCTAATACTGCCCTCTTCCA	58	22
47	LuM272	230	241	LuMF272	TTCCAGTTCTCACCATTCTCAC	57	22
				LuMR272	TAGAGCCCGAAATCAAAAGAAG	58	22
48	LuM280	215	216	LuMF280	TCTCTCTCTCTGTTTTCTGGGAGT	58	24
				LuMR280	AGATTGAGGAGTTTGGTTGGTG	58	22
49	LuM286	174	170	LuMF286	AGTGGAAAGTGCCATTCTGTTT	58	22
				LuMR286	CGGTGTTAGTAGATGCTTCGGT	59	22
50	LuM322	266	254	LuMF322	CTTTGCTTCCTACTCACCCCTT	59	22
				LuMR322	ACAGAGACAGAACCGCAGTCAT	59	22
51	LuM327	266	254	LuMF327	CTTTGCTTCCTACTCACCCCTT	59	22
				LuMR327	ACAGAGACAGAACCGCAGTCAT	59	22
52	LuM331	266	254	LuMF331	CTTTGCTTCCTACTCACCCCTT	59	22
				LuMR331	ACAGAGACAGAACCGCAGTCAT	59	22
53	LuM333	266	254	LuMF333	CTTTGCTTCCTACTCACCCCTT	59	22
				LuMR333	ACAGAGACAGAACCGCAGTCAT	59	22
54	LuM337	274	262	LuMF337	TTATTCCCCTTTGCTTCCAACT	59	22
				LuMR337	ACAGAGACAGAACCGCAGTCAT	59	22
55	LuM339	271	259	LuMF339	TATTCCCCTTTGCTTCCAACT	58	21
				LuMR339	CAGAGACAGAATCGCAGTCATC	58	22
56	LuM347	272	260	LuMF347	TTATTCCCCTTTGCTTCCTACTC	58	23
				LuMR347	TAGAGACAGAACCGCAGTCATC	58	22
57	LuM358	272	262	LuMF358	TTATTCCCCTTTGCTTCCAACT	59	22
				LuMR358	ACAGAGACAGAACCGCAGTCAT	59	22
58	LuM361	271	259	LuMF361	TATTCCCCTTTGCTTCCAACT	58	21

				LuMR361	CAGAGACAGAATCGCAGTCATC	58	22
59	LuM365	266	254	LuMF365	CTTTGCTTCCTACTCACCCCTT	59	22
				LuMR365	ACAGAGACAGAACCGCAGTCAT	59	22
60	LuM370	275	263	LuMF370	CTTATTCCCCTTTGCTTCCAAC	59	22
				LuMR370	ACAGAGACAGAACCGCAGTCAT	59	22
61	LuM371	219	207	LuMF371	ACCCCTCCACTCCCTTTATTC	59	21
				LuMR371	GGATGACGAGGAAATTGGGTAT	59	22
62	LuM372	274	262	LuMF372	TTATTCCCCTTTGCTTCCAACT	59	22
				LuMR372	ACAGAGACAGAACCGCAGTCAT	59	22
63	LuM374	201	188	LuMF374	AATCCCTCCACTCCCTTTATTC	58	22
				LuMR374	TATACAGCCAAACGCCATTGTA	58	22
64	LuM377	266	254	LuMF377	CTTTGCTTCCTACTCACCCCTT	59	22
				LuMR377	ACAGAGACAGAACCGCAGTCAT	59	22
65	LuM391	264	266	LuMF391	GTTGGCCTGTTTGGTTAGGTT	58	21
				LuMR391	CGGGGAGGTATAGATTGTTCTG	58	22
66	LuM404	184	186	LuMF404	GTGCGATCAATAGAGTTGCTTG	58	22
				LuMR404	ACACAGAATCGAAACACAAACG	58	22
67	LuM408	239	234	LuMF408	GGCCAACTAACTCCTGAAACAC	58	22
				LuMR408	GGGAGGCAACCCATGTCTA	59	19
68	LuM413	400	403	LuMF413	CCAAATCAACCCATTAGATGCT	58	22
				LuMR413	ACCGTTGTGTCCTTGTCTTCTT	58	22
69	LuM426	383	385	LuMF426	GGTAGAGTGACCGATGAGTTCC	58	22
				LuMR426	ACAACAACAACCACAACAGGTC	58	22
70	LuM427	332	323	LuMF427	GGGGAGATACGAGTTGATGATT	57	22
				LuMR427	ATGCTTGCAGTCACAGACATTT	58	22
71	LuM429	384	363	LuMF429	AAGGATTTGGTAAAGAGGGGAG	58	22
				LuMR429	AGTCCAAGGGAAGCACAAGTAG	58	22
72	LuM448	400	403	LuMF448	CCAAATCAACCCATTAGATGCT	58	22
				LuMR448	ACCGTTGTGTCCTTGTCTTCTT	58	22
73	LuM457	109	111	LuMF457	GTAAAGCAGTCAAACCCTGGTC	58	22
				LuMR457	TCTTGAAACAGCCAACCCTTAT	58	22
74	LuM468	378	381	LuMF468	TTTGCTCTACCTCTTTGGGTTC	58	22
				LuMR468	TGTGGTCGATACAAGTGAAAGG	58	22

Appendix 2: The list of SSR primers resynthesized from previous published sources for screening for polymorphism between CDC Bethune and G1186/94.

	SSR primers	**Linkage group**	**Source**
1	Lu3281	4	Cloutier et al. (2012b)
2	Lu2966	4	Cloutier et al. (2012b)
3	Lu996	4	Cloutier et al. (2012b)
4	Lu2006	4	Deng et al. (2011)
5	Lu2004	4	Deng et al. (2011)
6	Lu2002	4	Deng et al. (2011)
7	Lu2968	4	Cloutier et al. (2012)
8	Lu2073	4	Deng et al. (2010)
9	Lu722B	4	Cloutier et al. (2009)
10	Lu2025	4	Deng et al. (2011)
11	Lu2008	4	Deng et al. (2011)
12	Lu2059	4	Deng et al. (2010)
13	Lu2207	4	Cloutier et al. (2012b)
14	Lu3228	4	Cloutier et al. (2012b)
15	Lu3229	4	Cloutier et al. (2012b)
16	Lu2399	4	Cloutier et al. (2012b)
17	Lu2396	4	Cloutier et al. (2012b)
18	Lu3252	4	Cloutier et al. (2012b)
19	Lu3213	4	Cloutier et al. (2012b)
20	Lu2087	4	Roose-Amsaleg et al. (2006)
21	Lu717	4	Cloutier et al. (2009)
22	Lu2397	4	Cloutier et al. (2012b)
23	Lu2942	4	Cloutier et al. (2012b)
24	Lu2944	4	Cloutier et al. (2012b)
25	Lu2940	4	Cloutier et al. (2012b)
26	Lu2943	4	Cloutier et al. (2012b)
27	Lu989	4	Cloutier et al. (2012b)
28	Lu207	4	Cloutier et al. (2009)
29	Lu3113	4	Cloutier et al. (2012b)
30	Lu3116	4	Cloutier et al. (2012b)
31	Lu2983	4	Cloutier et al. (2012b)
32	Lu2980	4	Cloutier et al. (2012b)
33	Lu587	4	Cloutier et al. (2009)
34	Lu833	4	Cloutier et al. (2009)
35	Lu1049	4	Cloutier et al. (2012b)
36	Lu2984	4	Cloutier et al. (2012b)
37	Lu851	4	Cloutier et al. (2009)

38	Lu2981	4	Cloutier et al. (2012b)
39	Lu2239	4	Cloutier et al. (2012b)
40	Lu2043	4	Deng et al. (2010)
41	Lu2054	4	Deng et al. (2010)
42	Lu2237	4	Cloutier et al. (2012b)
43	Lu2031	4	Deng et al. (2011)
44	Lu2233	4	Cloutier et al. (2012b)
45	Lu2230	4	Cloutier et al. (2012b)
46	Lu919	4	Cloutier et al. (2012b)
47	Lu2235	4	Cloutier et al. (2012b)
48	Lu2076	4	Roose-Amsaleg et al. (2006)
49	Lu2286	4	Cloutier et al. (2012b)
50	Lu2287	4	Cloutier et al. (2012b)
51	Lu2009	4	Deng et al. (2011)
52	Lu2011	4	Deng et al. (2011)
53	Lu2024	4	Deng et al. (2011)
54	Lu2542	6	Cloutier et al. (2012b)
55	Lu2549	6	Cloutier et al. (2012b)
56	Lu2543	6	Cloutier et al. (2012b)
57	Lu2071	6	Deng et al. (2010)
58	Lu2064	6	Deng et al. (2010)
59	Lu836	6	Cloutier et al. (2009)
60	Lu2975	6	Cloutier et al. (2012b)
61	Lu2974	6	Cloutier et al. (2012b)
62	Lu2971	6	Cloutier et al. (2012b)
63	Lu2556	6	Cloutier et al. (2012b)
64	Lu1002B	6	Cloutier et al. (2012b)
65	Lu2561b	6	Cloutier et al. (2012b)
66	Lu2560	6	Cloutier et al. (2012b)
67	Lu2564	6	Cloutier et al. (2012b)
68	Lu2565	6	Cloutier et al. (2012b)
69	Lu2553	6	Cloutier et al. (2012b)
70	Lu2554	6	Cloutier et al. (2012b)
71	Lu2555	6	Cloutier et al. (2012b)
72	Lu2557	6	Cloutier et al. (2012b)
73	Lu3057a	6	Cloutier et al. (2012b)
74	Lu60	6	Cloutier et al. (2009)
75	Lu861	6	Cloutier et al. (2012b)
76	Lu1112	6	Cloutier et al. (2012b)
77	Lu3091	6	Cloutier et al. (2012b)
78	Lu2078	6	Roose-Amsaleg et al. (2006)

79	Lu2497	15	Cloutier et al. (2012b)
80	Lu451	15	Cloutier et al. (2009)
81	Lu1001	15	Cloutier et al. (2012b)
82	Lu637	15	Cloutier et al. (2009)
83	Lu271	15	Cloutier et al. (2009)
84	Lu2931	15	Cloutier et al. (2012b)
85	Lu510	15	Cloutier et al. (2009)
86	Lu3186	15	Cloutier et al. (2012b)
87	Lu3026	15	Cloutier et al. (2012b)
88	Lu3185	15	Cloutier et al. (2012b)
89	Lu3028	15	Cloutier et al. (2012b)
90	Lu2707	15	Cloutier et al. (2012b)
91	Lu2383	15	Cloutier et al. (2012b)
92	Lu1163	15	Cloutier et al. (2012b)
93	Lu2057	15	Deng et al. (2010)
94	Lu357	15	Cloutier et al. (2009)
95	Lu2382	15	Cloutier et al. (2012b)
96	Lu2965	15	Cloutier et al. (2012b)
97	Lu1127	15	Cloutier et al. (2012b)
98	Lu2001	15	Deng et al. (2011)
99	Lu2010a	15	Deng et al. (2011)
100	Lu359	15	Cloutier et al. (2009)
101	Lu2354	15	Cloutier et al. (2012b)
102	Lu1007	15	Cloutier et al. (2012b)
103	Lu1172	15	Cloutier et al. (2012b)
104	Lu838	15	Cloutier et al. (2009)
105	Lu462a	15	Cloutier et al. (2009)
106	Lu2696	15	Cloutier et al. (2012b)
107	Lu2695	15	Cloutier et al. (2012b)
108	Lu113	15	Cloutier et al. (2009)
109	Lu2697b	15	Cloutier et al. (2012b)

Appendix 3: The list of all of the RILs with their mean seed colour values in both GC, 2008 (growth chamber, 2008) and GH, 2011 (greenhouse, 2011) growth trials and flower colour in $F_{8:9}$ plants (GH, 2011). Here, B is brown-seeded line and Y is yellow-seeded line. All brown-seeded lines produced blue petalled flowers and all yellow-seeded produced white petalled flowers in the GH trial. Therefore, B = brown seed + blue petals and Y = yellow seed + white petals. In addition to tested RILs in GH, a few also flowered (recorded as "blue" and "white" for their petal colour) but did not set seeds so, only petal colour is mentioned. Here, "No flowers" means plants neither flowered nor set seeds; therefore, no observations available for those RILs.

$F_{8:9}$ RIL seeds (GC, 2008)				**$F_{9:10}$ RIL seeds (GH, 2011)**			
	RIL name	**Seed colour**	**Mean RGB values**		**RIL name**	**Seed/ Flower colour**	**Mean RGB values**
1	GB1	Y	108.2 ± 5.6	1	GB1	Y	103.1 ± 1.6
2	GB2	B	63.4 ± 4.7	2	GB2	B	63.2 ± 1.1
3	GB3	Y	92.4 ± 6.2	3	GB3	Y	81.8 ± 1.8
4	GB4	B	56 ± 4.8	4	GB4	B	63.6 ± 2
5	GB5	B	58.2 ± 3.7	5	GB5	B	57.7 ± 4.5
6	GB6	Y	84.3 ± 9.2	6	GB6	Y	90.1 ± 2.8
7	GB7	B	59.6 ± 3	7	GB7	B	60.7 ± 5.3
8	GB8	B	64.9 ± 6.2	8	GB8	B	65.4 ± 1.3
9	GB9	B	65.2 ± 3.7	9	GB9	B	70.2 ± 1.4
10	GB10	B	67.4 ± 5.4	10	GB10	B	67.8 ± 7
11	GB11	Y	96.4 ± 6.9	11	GB11	Y	95.3 ± 2.2
12	GB12	B	66.9 ± 2.8	12	GB12	B	67.1 ± 4.2
13	GB13	Y	94.1 ± 4.7	13	GB13	Y	98.4 ± 3
14	GB14	Y	94 ± 3.3	14	GB14	Y	72.8 ± 1.5
15	GB15	Y	103.2 ± 3	15	GB15	Y	92.4 ± 7.3
16	GB16	B	65.1 ± 2.2	16	GB16	Blue	
17	GB17	Y	87.5 ± 4	17	GB17	Y	91.2 ± 4.2
18	GB18	B	61 ± 3.6	18	GB18	B	65.3 ± 2.1
19	GB19	Y	82.2 ± 1.7	19	GB19	Y	81.5 ± 2.3
20	GB20	Y	80.8 ± 1.9	20	GB20	Y	76.8 ± 1.6
21	GB21	Y	91.3 ± 3.3	21	GB21	Y	85.8 ± 9.1
22	GB22	B	62.5 ± 3	22	GB22	B	57 ± 3.1
23	GB23	B	65.9 ± 2.6	23	GB23	B	62.1 ± 1.6
24	GB24	Y	83.2 ± 3	24	GB24	Y	87.8 ± 3.5
25	GB25	B	64.5 ± 2.1	25	GB25	B	52.8 ± 0.6
26	GB26	B	51.4 ± 2	26	GB26	B	59.6 ± 1.4
27	GB27	Y	94 ± 2.2	27	GB27	Y	94.2 ± 3.1
28	GB28	B	46.8 ± 2.4	28	GB28	B	57.5 ± 1.2
29	GB29	B	49.4 ± 1.3	29	GB29	B	61.2 ± 2.5
30	GB30	B	57.6 ± 0.8	30	GB30	B	60.6 ± 1.6
31	GB31	B	63.9 ± 3.9	31	GB31	B	54.1 ± 2.2
32	GB32	Y	96.8 ± 2.6	32	GB32	Y	76.7 ± 2.2
33	GB33	B	62.8 ± 2.1	33	GB33	B	64.3 ± 3.5
34	GB34	Y	85.5 ± 2.6	34	GB34	Y	87.1 ± 1.6
35	GB35	Y	94.4 ± 3.6	35	GB35	Y	99.9 ± 2
36	GB36	B	63.1 ± 2.2	36	GB36	B	61.8 ± 2.3
37	GB37	Y	80.8 ± 2.7	37	GB37	Y	82.7 ± 0.3
38	GB38	B	61.1 ± 1.5	38	GB38	B	56.9 ± 2.4

39	GB39	B	60.5 ± 3.1
40	GB40	B	61.9 ± 1.9
41	GB41	Y	83.7 ± 1.3
42	GB42	B	55.9 ± 1.5
43	GB43	Y	82.5 ± 2.4
44	GB44	Y	95.6 ± 2.1
45	GB45	B	56.6 ± 3.1
46	GB46	B	64.8 ± 2.8
47	GB47	Y	92 ± 4
48	GB48	B	60.2 ± 3.5
49	GB49	Y	82.4 ± 2.7
50	GB50	B	62.2 ± 1.7
51	GB51	B	60.8 ± 2.1
52	GB52	Y	88.5 ± 3.8
53	GB53	B	62.9 ± 1.9
54	GB54	Y	91 ± 2.9
55	GB55	B	58.7 ± 2.8
56	GB56	Y	94.2 ± 2.3
57	GB57	Y	85 ± 5.4
58	GB58	B	60.3 ± 2.6
59	GB59	B	63 ± 1.5
60	GB60	B	60.9 ± 3.7
61	GB61	B	62.2 ± 3.4
62	GB62	B	63 ± 1.3
63	GB63	B	62.4 ± 3.7
64	GB64	Y	98.7 ± 4.3
65	GB65	Y	80.2 ± 2.1
66	GB66	B	55.2 ± 1.9
67	GB67	B	61.1 ± 2.8
68	GB68	B	60.5 ± 3.5
69	GB69	B	59.5 ± 2.9
70	GB70	Y	81 ± 3.3
71	GB71	B	54.5 ± 4.1
72	GB72	B	63.7 ± 2.5
73	GB73	Y	87.3 ± 2.4
74	GB74	Y	99.4 ± 6.3
75	GB75	Y	86.5 ± 2.2
76	GB76	B	57.8 ± 2.1
77	GB77	B	60.4 ± 3.7
78	GB78	B	60.9 ± 4.7
79	GB79	Y	89.2 ± 2.9
80	GB80	B	54.7 ± 2.7
81	GB81	Y	88.5 ± 5.4
82	GB82	B	49.8 ± 2.4
83	GB83	B	58.8 ± 3.2
84	GB84	B	58.1 ± 2.9
85	GB85	Y	93 ± 3.4
86	GB86	Y	96.8 ± 5.1
87	GB87	Y	95.1 ± 2.3
88	GB88	Y	83.1 ± 1.9
89	GB89	Y	84.8 ± 5.6
90	GB90	B	61.3 ± 2.3
91	GB91	B	61.7 ± 3.8

39	GB39	B	58.6 ± 2.6
40	GB40	B	61.6 ± 3.9
41	GB41	Y	85.7 ± 2.1
42	GB42	B	58.6 ± 1.3
43	GB43	Y	94.2 ± 2.3
44	GB44	Y	95.9 ± 1.8
45	GB45	B	62.1 ± 1.2
46	GB46	B	52.5 ± 1.2
47	GB47	Y	88.8 ± 1.6
48	GB48	B	53 ± 3
49	GB49	Y	78.9 ± 6.7
50	GB50	B	61.6 ± 2.9
51	GB51	Blue	
52	GB52	Y	100.4 ± 1.8
53	GB53	B	58.5 ± 1.9
54	GB54	Y	92.1 ± 1.5
55	GB55	B	53.8 ± 2.6
56	GB56	Y	92.5 ± 0.6
57	GB57	Y	81.1 ± 1.7
58	GB58	B	62.9 ± 2.5
59	GB59	B	59.7 ± 0.8
60	GB60	B	49.6 ± 1.3
61	GB61	B	54.6 ± 1.1
62	GB62	B	54 ± 2.1
63	GB63	B	63.6 ± 1.8
64	GB64	Y	94.4 ± 3.3
65	GB65	Y	86 ± 3.7
66	GB66	B	53.7 ± 0.5
67	GB67	B	53.1 ± 2.7
68	GB68	B	56.9 ± 1.8
69	GB69	B	59.7 ± 2.6
70	GB70	White	
71	GB71	B	52.9 ± 1.6
72	GB72	B	60 ± 3.5
73	GB73	Y	79.7 ± 6.6
74	GB74	Y	90.2 ± 2.5
75	GB75	Y	86.2 ± 2.7
76	GB76	B	51.1 ± 1.4
77	GB77	B	48.8 ± 1.8
78	GB78	B	59.6 ± 1.4
79	GB79	Y	106.2 ± 0.9
80	GB80	B	63.4 ± 1.8
81	GB81	Y	78.7 ± 1.8
82	GB82	B	52.7 ± 1.2
83	GB83	B	44 ± 0.1
84	GB84	B	46.6 ± 1
85	GB85	Y	89.1 ± 2.6
86	GB86	Y	96.2 ± 0.9
87	GB87	Y	82.1 ± 5.8
88	GB88	Y	79.1 ± 0.9
89	GB89	Y	89.6 ± 6.9
90	GB90	B	62.5 ± 0.9
91	GB91	B	61.3 ± 1.9

92	GB92	B	62.1 ± 4.8
93	GB93	Y	101.8 ± 5.4
94	GB94	B	50.2 ± 4.9
95	GB95	B	60.5 ± 2.5
96	GB96	B	57.1 ± 0.7
97	GB97	B	59.7 ± 4.6
98	GB98	Y	92.5 ± 2.9
99	GB99	B	62.4 ± 2.8
100	GB100	B	61.9 ± 2.7
101	GB101	B	53.1 ± 1.3
102	GB102	Y	85.5 ± 1.2
103	GB103	B	56.4 ± 1.2
104	GB104	Y	103.3 ± 1.3
105	GB105	B	62.2 ± 1
106	GB106	B	56.9 ± 3
107	GB107	B	56 ± 1.8
108	GB108	Y	98.4 ± 7
109	GB109	Y	84.3 ± 4
110	GB110	B	56.4 ± 5.5
111	GB111	Y	85 ± 1.9
112	GB112	Y	101.3 ± 1.2
113	GB113	Y	83.5 ± 5
114	GB114	B	58.8 ± 2.4
115	GB115	Y	78.2 ± 3.2
116	GB116	B	55.1 ± 2.3
117	GB117	Y	96.6 ± 1.3
118	GB118	Y	97.5 ± 6.1
119	GB119	Y	93.9 ± 5.1
120	GB120	B	57.9 ± 1.3
121	GB121	B	55.7 ± 1.8
122	GB122	B	57.4 ± 1.8
123	GB123	Y	90.5 ± 1.9
124	GB124	Y	89.3 ± 2.3
125	GB125	Y	88.1 ± 1.7
126	GB126	B	54.1 ± 0.9
127	GB127	B	62.4 ± 1.7
128	GB128	Y	96.4 ± 3.1
129	GB129	B	65 ± 2.3
130	GB130	B	59.8 ± 3.3
131	GB131	Y	85.6 ± 1.2
132	GB132	B	56.5 ± 3.4
133	GB133	B	59 ± 3.5
134	GB134	B	58.6 ± 4.5
135	GB135	Y	82.3 ± 4.1
136	GB136	Y	86.8 ± 6.4
137	GB137	Y	89.1 ± 5.5
138	GB138	Y	94.7 ± 2
139	GB139	B	49.8 ± 1.1
140	GB140	Y	101.4 ± 1.6
141	GB141	Y	87.5 ± 7
142	GB142	B	53.3 ± 3.7
143	GB143	B	50.1 ± 0.5
144	GB144	B	57.5 ± 2.2

92	GB92	B	61.8 ± 2.2
93	GB93	Y	92.4 ± 4.3
94	GB94	B	54 ± 1.7
95	GB95	B	60.5 ± 1.8
96	GB96	B	62.2 ± 0.8
97	GB97	B	55.5 ± 0.9
98	GB98	Y	83.8 ± 0.9
99	GB99	B	64.9 ± 0.7
100	GB100	B	59.4 ± 1
101	GB101	B	57.6 ± 1.2
102	GB102	Y	97.3 ± 5.4
103	GB103	B	61.6 ± 0.6
104	GB104	Y	98.6 ± 2.4
105	GB105	B	63.1 ± 2.8
106	GB106	B	59.4 ± 1.1
107	GB107	B	66.8 ± 1.8
108	GB108	Y	101.9 ± 1.4
109	GB109	Y	92.5 ± 5.7
110	GB110	No flowers	
111	GB111	No flowers	
112	GB112	Y	97.6 ± 0.9
113	GB113	Y	79.9 ± 1.3
114	GB114	B	64.7 ± 0.7
115	GB115	Y	77.5 ± 1.2
116	GB116	B	67.1 ± 0.4
117	GB117	Y	102 ± 0.5
118	GB118	Y	100.6 ± 2.4
119	GB119	Y	83.9 ± 2.4
120	GB120	B	64 ± 2.3
121	GB121	B	56.4 ± 1.7
122	GB122	B	65.4 ± 1.4
123	GB123	Y	81.2 ± 3
124	GB124	Y	91.3 ± 1.1
125	GB125	Y	93.3 ± 2.2
126	GB126	No flowers	
127	GB127	B	63.7 ± 4.3
128	GB128	Y	96.6 ± 1.5
129	GB129	B	62.7 ± 2
130	GB130	B	58.8 ± 1.8
131	GB131	Y	94.4 ± 3.3
132	GB132	B	57.2 ± 1.3
133	GB133	B	61.4 ± 0.8
134	GB134	B	60.5 ± 3.7
135	GB135	Y	76.2 ± 6.1
136	GB136	Y	76.9 ± 5.1
137	GB137	Y	83.1 ± 3.1
138	GB138	Y	98.7 ± 2.9
139	GB139	B	60.2 ± 1.5
140	GB140	Y	91.6 ± 3.3
141	GB141	Y	90.8 ± 5.7
142	GB142	B	59.1 ± 0.5
143	GB143	B	51.3 ± 1.3
144	GB144	B	65.4 ± 0.9

145	GB145	Y	79.2 ± 6.1
146	GB146	B	59.5 ± 4.4
147	GB147	Y	90.6 ± 4.2
148	GB148	B	54.8 ± 4.6
149	GB149	Y	81.2 ± 4
150	GB150	B	55.6 ± 2.6
151	GB151	Y	84.9 ± 2.2
152	GB152	B	58 ± 3.6
153	GB153	B	57.2 ± 2.3
154	GB154	B	57.5 ± 3.4
155	GB155	B	55.9 ± 1.4
156	GB156	B	63.6 ± 1.2
157	GB157	B	56.9 ± 1.8
158	GB158	Y	91 ± 2.4
159	GB159	B	52.5 ± 2.3
160	GB160	B	61.4 ± 2.6
161	GB161	Y	88.2 ± 3.6
162	GB162	B	59.6 ± 1.7
163	GB163	B	57.4 ± 2.1
164	GB164	Y	89.4 ± 1.8
165	GB166	B	49 ± 0.6
166	GB167	B	52 ± 1.5
167	GB168	B	56.8 ± 3.1
168	GB169	B	64.6 ± 1.7
169	GB170	Y	92 ± 2.4
170	GB171	Y	102.1 ± 5.6
171	GB172	Y	80.2 ± 3.1
172	GB173	B	52.9 ± 1.9
173	GB174	B	53.5 ± 2.8
174	GB175	B	58.5 ± 0.8
175	GB176	B	53.9 ± 2.4
176	GB177	Y	95.7 ± 3.4
177	GB178	Y	82 ± 1.5
178	GB179	Y	94.3 ± 2.3
179	GB180	Y	93.8 ± 2
180	GB181	B	62 ± 1.9
181	GB182	B	64.1 ± 1.9
182	GB183	B	56.7 ± 2.6
183	GB184	B	62.3 ± 4.9
184	GB185	Y	90.1 ± 4.7
185	GB186	Y	91.9 ± 3.9
186	GB187	B	61.8 ± 12.7
187	GB188	Y	85.2 ± 3.1
188	GB189	B	58.3 ± 3
189	GB190	B	58.6 ± 2.6
190	GB191	B	57.8 ± 2.1
191	GB192	B	65.3 ± 0.7
192	GB193	B	61.2 ± 2.2
193	GB194	Y	101 ± 2.6
194	GB195	Y	86.5 ± 6.7
195	GB196	B	59 ± 2
196	GB197	Y	90 ± 4.5
197	GB198	Y	86.5 ± 3.4

145	GB145	Y	80.5 ± 5.7
146	GB146	B	56.7 ± 1.3
147	GB147	Y	92.8 ± 2.3
148	GB148	B	55.1 ± 1.2
149	GB149	Y	87.2 ± 1.8
150	GB150	B	63.1 ± 4.3
151	GB151	Y	99.4 ± 2.8
152	GB152	B	58.2 ± 1.1
153	GB153	B	61.4 ± 2.5
154	GB154	B	69.9 ± 0.5
155	GB155	B	58.7 ± 2
156	GB156	B	63.5 ± 1.3
157	GB157	B	56.6 ± 0.7
158	GB158	Y	91.4 ± 3.5
159	GB159	B	58.4 ± 1.6
160	GB160	B	63.5 ± 1.4
161	GB161	Y	92.1 ± 1.5
162	GB162	Blue	
163	GB163	B	64.5 ± 1.7
164	GB164	Y	88.7 ± 4.3
165	GB166	B	61.2 ± 0.8
166	GB167	B	58.2 ± 0.6
167	GB168	B	60.8 ± 1.1
168	GB169	B	60.9 ± 1.2
169	GB170	Y	100.1 ± 2.1
170	GB171	Y	91.7 ± 2.7
171	GB172	Y	101.5 ± 2.7
172	GB173	No flowers	
173	GB174	B	62.3 ± 2.7
174	GB175	No flowers	
175	GB176	B	56.3 ± 0.7
176	GB177	Y	99.7 ± 0.9
177	GB178	Y	82.5 ± 6.3
178	GB179	Y	94 ± 2.6
179	GB180	Y	88.5 ± 0.8
180	GB181	B	63.6 ± 1.9
181	GB182	B	62.2 ± 3.5
182	GB183	B	59.1 ± 1.5
183	GB184	B	58 ± 3.1
184	GB185	Y	84.9 ± 4.8
185	GB186	Y	94.4 ± 2.5
186	GB187	B	59.6 ± 2
187	GB188	Y	73.6 ± 2.7
188	GB189	B	64.4 ± 2.4
189	GB190	B	50.2 ± 0.8
190	GB191	B	62.6 ± 0.9
191	GB192	B	62.6 ± 3.3
192	GB193	B	65 ± 0.7
193	GB194	Y	103.3 ± 4.3
194	GB195	Y	76.3 ± 2.9
195	GB196	B	59.3 ± 3
196	GB197	Y	98.3 ± 2.4
197	GB198	Y	90.5 ± 4.1

198	GB199	B	60.6 ± 4.4
199	GB200	B	54.1 ± 4.7
200	GB201	Y	93.7 ± 4.2
201	GB202	B	49.6 ± 0.7
202	GB203	B	56.2 ± 3.1
203	GB204	B	52 ± 3.2
204	GB205	B	64 ± 2.3
205	GB206	Y	87.9 ± 2.6
206	GB207	B	56.8 ± 1.3
207	GB208	Y	83.9 ± 2.1
208	GB209	B	59.1 ± 1.5
209	GB210	B	64.9 ± 3.4
210	GB211	Y	86.5 ± 5
211	GB212	B	57.2 ± 4.8
212	GB213	Y	91 ± 2.5
213	GB214	B	55.9 ± 1.5
214	GB215	B	57 ± 1.8
215	GB216	Y	105.1 ± 2.3
216	GB217	Y	99.5 ± 1.2
217	GB218	B	56.8 ± 0.5
218	GB219	B	58.2 ± 2.9
219	GB220	B	59.7 ± 1.6
220	GB221	Y	81.5 ± 1.9
221	GB222	B	57.9 ± 1.7
222	GB223	B	48 ± 5
223	GB224	B	50.2 ± 1.6
224	GB225	Y	88.5 ± 5.5
225	GB226	B	62.2 ± 3.5
226	GB227	Y	93 ± 2.3
227	GB228	B	53.5 ± 0.5
228	GB229	B	57.3 ± 5.6
229	GB230	Y	104.7 ± 2.7
230	GB231	Y	82.2 ± 6.6
231	GB232	B	62.7 ± 1.9
232	GB233	B	61.4 ± 1.9
233	GB234	B	60.7 ± 1.4
234	GB235	B	56.2 ± 3.2
235	GB236	B	53.9 ± 2.1
236	GB237	B	60.6 ± 1.3
237	GB238	B	56.4 ± 5.5
238	GB239	Y	92.9 ± 2.4
239	GB240	B	62.9 ± 2.5
240	GB241	B	64.1 ± 3
241	GB242	B	61.9 ± 2.6
242	GB243	B	55.8 ± 1.6
243	GB244	B	54.1 ± 2.4
244	GB245	B	60.1 ± 2.8
245	GB246	Y	84.8 ± 2.2
246	GB247	Y	96.4 ± 7.4
247	GB248	Y	89.4 ± 2
248	GB249	B	54.1 ± 1.8
249	GB250	B	62.2 ± 1.6
250	GB251	Y	82.1 ± 2.2

198	GB199	B	59.6 ± 1.5
199	GB200	B	52.9 ± 0.6
200	GB201	Y	97.2 ± 2.9
201	GB202	B	57.1 ± 2
202	GB203	B	57.8 ± 1.4
203	GB204	B	62.2 ± 2.5
204	GB205	B	59.1 ± 0.4
205	GB206	Y	87 ± 6.6
206	GB207	B	60.5 ± 1.5
207	GB208	Y	96.9 ± 4.1
208	GB209	B	56.2 ± 1.9
209	GB210	B	52.1 ± 0.2
210	GB211	Y	84 ± 5.1
211	GB212	B	59 ± 0.8
212	GB213	Y	91.2 ± 1.4
213	GB214	B	60.4 ± 2.8
214	GB215	B	65.7 ± 0.9
215	GB216	Y	92.4 ± 2.2
216	GB217	Y	84.7 ± 1.9
217	GB218	B	56.4 ± 2.5
218	GB219	B	61.5 ± 2.5
219	GB220	B	66 ± 2.2
220	GB221	Y	91.8 ± 2.6
221	GB222	B	60.6 ± 1.6
222	GB223	Blue	
223	GB224	B	57.4 ± 1.6
224	GB225	No flowers	
225	GB226	B	62.2 ± 2
226	GB227	Y	103.1 ± 3.5
227	GB228	B	54.3 ± 2.1
228	GB229	B	60.3 ± 2
229	GB230	Y	96.5 ± 2.7
230	GB231	Y	91.8 ± 2.4
231	GB232	B	64.3 ± 2
232	GB233	B	64.5 ± 2.1
233	GB234	B	58.9 ± 1.3
234	GB235	B	61.2 ± 3.5
235	GB236	B	57.2 ± 3.4
236	GB237	B	56.6 ± 1.2
237	GB238	B	65.2 ± 3.1
238	GB239	Y	88.7 ± 2.3
239	GB240	B	54.9 ± 1.8
240	GB241	B	60.9 ± 3.4
241	GB242	B	57.7 ± 2.1
242	GB243	B	46.8 ± 1.2
243	GB244	B	60.1 ± 2
244	GB245	B	54.8 ± 2.8
245	GB246	Y	88.6 ± 2.4
246	GB247	Y	98.5 ± 4.3
247	GB248	Y	89.5 ± 2.7
248	GB249	B	62 ± 3
249	GB250	B	59.4 ± 2.2
250	GB251	Y	70 ± 2.7

251	GB252	B	58.5 ± 1.4
252	GB253	Y	93.8 ± 2.6
253	GB254	Y	87.8 ± 1.1
254	GB255	B	59.2 ± 2.2
255	GB256	B	62.1 ± 1.5
256	GB257	Y	95.8 ± 1.7
257	GB258	Y	102.5 ± 2.7
258	GB259	Y	83.4 ± 2.6
259	GB260	B	62 ± 3
260	GB261	B	60.6 ± 2.1
261	GB262	B	65 ± 2.8
262	GB263	Y	95.5 ± 5.7
263	GB264	B	56.3 ± 2.6
264	GB265	Y	86 ± 2.5
265	GB266	B	65.6 ± 1.7
266	GB267	B	63.8 ± 3.6
267	GB268	B	55.9 ± 2.3
268	GB269	Y	95.1 ± 4.8
269	GB270	B	57.4 ± 2.2
270	GB271	B	51.4 ± 1.8
271	GB272	Y	81.8 ± 2.3
272	GB273	B	65.3 ± 1.7
273	GB274	B	47.8 ± 1.4
274	GB275	B	57.9 ± 1.8
275	GB276	Y	86.5 ± 5.6
276	GB277	B	59.2 ± 7.5
277	GB278	B	59.9 ± 1.2
278	GB279	B	56.2 ± 2
279	GB280	B	61.4 ± 2.7
280	GB281	B	63.1 ± 0.5
281	GB282	Y	92.5 ± 1.8
282	GB283	Y	90.1 ± 6.1
283	GB284	Y	85.6 ± 0.7
284	GB285	B	62.8 ± 2.8
285	GB286	B	58.4 ± 5
286	GB287	B	47.7 ± 2.3
287	GB288	Y	89 ± 3.2
288	GB289	B	60.3 ± 2.5
289	GB290	B	58.4 ± 1.9
290	GB291	B	61.5 ± 1.6
291	GB292	Y	90.3 ± 0.5
292	GB293	B	55.4 ± 2.9
293	GB294	Y	100 ± 3.7
294	GB295	B	57.8 ± 4.6
295	GB296	B	59 ± 3.5
296	GB297	Y	101.8 ± 3
297	GB298	Y	86.8 ± 4.9
298	GB299	B	55.9 ± 0.9
299	GB300	Y	87.9 ± 3.7
300	GB301	B	65.4 ± 0.9
301	GB302	Y	103.2 ± 5.1
302	GB303	B	52.8 ± 1.8
303	GB304	B	55.7 ± 0.7

251	GB252	B	64.7 ± 1.3
252	GB253	Y	94.2 ± 2.1
253	GB254	Y	74.1 ± 0.1
254	GB255	B	61.8 ± 1.8
255	GB256	B	57.4 ± 1.6
256	GB257	Y	89.4 ± 4.8
257	GB258	Y	99.6 ± 0.4
258	GB259	Y	79.9 ± 5.1
259	GB260	B	62.8 ± 2.4
260	GB261	B	60.2 ± 4.8
261	GB262	B	60.6 ± 1.1
262	GB263	Y	97.4 ± 0.3
263	GB264	B	55.2 ± 1.2
264	GB265	Y	83.6 ± 5.5
265	GB266	B	55.3 ± 1.3
266	GB267	B	55.9 ± 1.1
267	GB268	B	57.9 ± 0.6
268	GB269	Y	96.3 ± 6
269	GB270	B	62.9 ± 0.5
270	GB271	B	59.1 ± 1.2
271	GB272	Y	73.5 ± 2.2
272	GB273	B	65.1 ± 1.7
273	GB274	B	60.4 ± 3.3
274	GB275	B	62 ± 1.3
275	GB276	Y	80.6 ± 7.8
276	GB277	B	65.2 ± 4.5
277	GB278	B	55.5 ± 1.3
278	GB279	B	63.3 ± 2.5
279	GB280	B	52.1 ± 0.4
280	GB281	B	64.4 ± 4.4
281	GB282	Y	100.3 ± 5
282	GB283	Y	94.9 ± 3.6
283	GB284	Y	95.1 ± 3.7
284	GB285	B	53.6 ± 5.5
285	GB286	B	62.4 ± 2.9
286	GB287	B	60.8 ± 0.8
287	GB288	Y	104.3 ± 5.9
288	GB289	B	59.7 ± 3
289	GB290	B	62.3 ± 1
290	GB291	B	52.7 ± 2.6
291	GB292	Y	93.2 ± 1.8
292	GB293	B	53.9 ± 4.1
293	GB294	Y	99 ± 0.9
294	GB295	B	62.4 ± 2.2
295	GB296	B	66.4 ± 1.1
296	GB297	Y	102.6 ± 3.6
297	GB298	Y	98.5 ± 1.9
298	GB299	B	56.6 ± 1.7
299	GB300	Y	87.3 ± 3.1
300	GB301	B	51.8 ± 1.6
301	GB302	Y	92 ± 4.2
302	GB303	B	63 ± 1.9
303	GB304	B	53.5 ± 1.6

304	GB305	Y	94.4 ± 2
305	GB306	B	57.7 ± 2.9
306	GB307	Y	100.5 ± 1.6
307	GB308	Y	93.8 ± 2.5
308	GB309	Y	98.1 ± 3.2
309	GB310	Y	90 ± 0.8
310	GB311	B	58.9 ± 1.9
311	GB312	B	59.6 ± 1.6
312	GB313	B	55.7 ± 1.7
313	GB314	B	56.5 ± 1.8
314	GB315	B	62.9 ± 2.2
315	GB316	B	59.5 ± 0.8
316	GB317	B	57.5 ± 4.1
317	GB318	B	54.4 ± 2.5
318	GB319	B	63.1 ± 2.9
319	GB320	B	62.2 ± 1.3
320	GB321	B	50.9 ± 3.2
321	GB322	B	56.5 ± 1.4
322	GB323	Y	83.8 ± 4.6
323	GB324	B	53.2 ± 2.4
324	GB325	Y	79.7 ± 2.7
325	GB326	B	61.5 ± 6.8
326	GB327	B	53.3 ± 0.5
327	GB328	Y	89 ± 2.7
328	GB329	Y	90.2 ± 2
329	GB330	B	59.3 ± 2.4
330	GB331	B	59.4 ± 1.1
331	GB332	Y	83.3 ± 2.7
332	GB333	B	55.2 ± 1
333	GB334	Y	90.1 ± 6.5
334	GB335	Y	87.6 ± 0.6
335	GB336	Y	80 ± 4.2
336	GB337	B	52.3 ± 0.7
337	GB338	Y	91.6 ± 3.2
338	GB339	B	61.4 ± 0.5
339	GB340	B	57.8 ± 0.9
340	GB341	B	48.4 ± 1
341	GB342	B	63.8 ± 0.7
342	GB343	B	60.7 ± 0.8
343	GB344	B	48.9 ± 2.1
344	GB345	Y	84.8 ± 5
345	GB346	Y	76.7 ± 1.3
346	GB347	B	55.3 ± 2.1
347	GB348	B	56.6 ± 1.7
348	GB349	Y	81.2 ± 1.9
349	GB350	B	57.6 ± 4.1
350	GB351	Y	91.8 ± 4.7
351	GB352	B	61.4 ± 3.9
352	GB353	B	56.1 ± 1.1
353	GB354	B	54.7 ± 1.5
354	GB355	B	62.6 ± 0.9
355	GB356	B	54.8 ± 1.7
356	GB357	Y	88.6 ± 2.4

304	GB305	Y	94.9 ± 2.5
305	GB306	B	67 ± 0.8
306	GB307	Y	95.3 ± 1.8
307	GB308	Y	103.4 ± 0.7
308	GB309	Y	95.8 ± 1.4
309	GB310	Y	86.6 ± 0.6
310	GB311	B	63 ± 0.6
311	GB312	B	63.3 ± 1.5
312	GB313	B	53.8 ± 0.9
313	GB314	B	62.4 ± 1.9
314	GB315	B	58.1 ± 2.5
315	GB316	B	62.5 ± 2.7
316	GB317	B	63.6 ± 0.8
317	GB318	B	66.6 ± 0.9
318	GB319	B	51.3 ± 2.6
319	GB320	B	52.5 ± 1.4
320	GB321	B	58.2 ± 2.3
321	GB322	B	56.4 ± 1.3
322	GB323	Y	79.7 ± 3.8
323	GB324	B	65.6 ± 3.2
324	GB325	Y	98.1 ± 3
325	GB326	B	60.2 ± 4.6
326	GB327	B	59.2 ± 2
327	GB328	Y	88.1 ± 2
328	GB329	Y	96.6 ± 2.2
329	GB330	B	52.5 ± 1.1
330	GB331	B	58 ± 0.6
331	GB332	Y	87.4 ± 1.4
332	GB333	B	55.3 ± 1.4
333	GB334	Y	98.4 ± 2.6
334	GB335	Y	93 ± 2.7
335	GB336	Y	98.5 ± 2.4
336	GB337	B	55.6 ± 1
337	GB338	Y	87.5 ± 1.7
338	GB339	B	56.1 ± 2.2
339	GB340	B	61.1 ± 1.2
340	GB341	Blue	
341	GB342	B	64.7 ± 1.2
342	GB343	B	61.2 ± 1.7
343	GB344	B	55.6 ± 2
344	GB345	Y	91.6 ± 4.4
345	GB346	Y	89.5 ± 3.6
346	GB347	B	51.1 ± 2.9
347	GB348	B	63.1 ± 4.2
348	GB349	Y	77.5 ± 2.2
349	GB350	B	61.8 ± 4.6
350	GB351	Y	95.9 ± 9.8
351	GB352	B	58.2 ± 3.6
352	GB353	B	59.2 ± 3
353	GB354	B	60.8 ± 3.3
354	GB355	B	60.1 ± 0.9
355	GB356	B	62.1 ± 2.9
356	GB357	Y	97.6 ± 7.7

357	GB358	B	55.6 ± 1.5
358	GB359	Y	94.2 ± 3.1
359	GB360	B	54.1 ± 1
360	GB361	B	56 ± 2
361	GB362	B	65.8 ± 1.6
362	GB363	Y	85.8 ± 3.8
363	GB364	B	59.4 ± 2.4
364	GB365	B	58.4 ± 1.7
365	GB366	B	59.4 ± 2.8
366	GB367	Y	74.6 ± 1.5
367	GB368	B	53.6 ± 2.1
368	GB369	B	53.8 ± 1.9
369	GB370	Y	87.3 ± 1.3
370	GB371	Y	91 ± 6.2
371	GB372	B	54.7 ± 3.5
372	GB373	Y	104 ± 1.5
373	GB374	B	57.9 ± 1.2
374	GB375	Y	82.8 ± 3.1
375	GB376	B	61.7 ± 1.1
376	GB377	B	63.5 ± 0.6
377	GB378	B	56.9 ± 2
378	GB379	B	61 ± 0.7
379	GB380	B	61.8 ± 3.9
380	GB381	B	60.8 ± 1.4
381	GB382	B	63.1 ± 2.8
382	GB383	B	55.8 ± 0.5
383	GB384	Y	79 ± 4.9
384	GB385	B	48.8 ± 0.3
385	GB386	B	55.7 ± 1.4
386	GB387	Y	83.4 ± 1
387	GB388	Y	81.4 ± 2.5
388	GB389	Y	88.7 ± 5.6
389	GB390	B	60.6 ± 2.1
390	GB391	Y	88.3 ± 2.1
391	GB392	Y	84.2 ± 1.9
392	GB393	Y	90.6 ± 2.7
393	GB394	B	65.1 ± 1.8
394	GB395	Y	92.1 ± 1.1
395	GB396	Y	93 ± 3.2
396	GB397	Y	90.6 ± 2.5
397	GB398	B	64.3 ± 3.4
398	GB399	B	57.5 ± 2.8
399	GB400	B	59.7 ± 2.6
400	GB401	B	58.7 ± 1.2
401	GB402	B	60.4 ± 1.7
402	GB403	B	60.6 ± 1.1
403	GB404	Y	92.8 ± 1.3
404	GB405	Y	96.8 ± 1.8
405	GB406	Y	81 ± 1.3
406	GB407	Y	76.5 ± 2.5
407	GB408	Y	86.2 ± 0.8
408	GB409	Y	96.2 ± 1.2
409	GB410	B	63.1 ± 0.9

357	GB358	B	58 ± 4.9
358	GB359	Y	90.9 ± 1.5
359	GB360	B	58.3 ± 2.8
360	GB361	B	59.8 ± 0.9
361	GB362	B	67.1 ± 2.7
362	GB363	Y	102.4 ± 13.4
363	GB364	B	59.9 ± 1.5
364	GB365	B	64.8 ± 3.1
365	GB366	B	59.4 ± 1
366	GB367	Y	76.1 ± 2.4
367	GB368	B	62.3 ± 1.2
368	GB369	B	67.1 ± 0.2
369	GB370	Y	93.5 ± 4
370	GB371	Y	82.6 ± 12.1
371	GB372	B	62.8 ± 3.9
372	GB373	Y	104.2 ± 6.6
373	GB374	B	59 ± 1.1
374	GB375	Y	77.2 ± 1.9
375	GB376	B	57.4 ± 4.9
376	GB377	B	56.4 ± 1.1
377	GB378	B	57 ± 1.4
378	GB379	B	54 ± 1.9
379	GB380	B	59.6 ± 0.9
380	GB381	B	54.7 ± 2.4
381	GB382	B	65.6 ± 1.6
382	GB383	B	63.2 ± 4.3
383	GB384	Y	74.4 ± 1.5
384	GB385	B	64.5 ± 0.7
385	GB386	B	62.7 ± 2.8
386	GB387	Y	93.4 ± 5.9
387	GB388	Y	84.7 ± 2
388	GB389	Y	92.6 ± 4
389	GB390	B	60 ± 2.3
390	GB391	Y	94.6 ± 3.2
391	GB392	Y	94.6 ± 1.8
392	GB393	Y	93.5 ± 1.5
393	GB394	B	62.9 ± 3.5
394	GB395	Y	75 ± 0.9
395	GB396	Y	100.7 ± 1.6
396	GB397	Y	83.2 ± 1.8
397	GB398	B	62.3 ± 6.6
398	GB399	B	51.2 ± 3.6
399	GB400	B	63.8 ± 0.7
400	GB401	B	63.4 ± 3.7
401	GB402	No flowers	
402	GB403	No flowers	
403	GB404	Y	94.3 ± 1.5
404	GB405	Y	102.8 ± 5.6
405	GB406	Y	80 ± 1.8
406	GB407	Y	76.9 ± 3.8
407	GB408	Y	87 ± 3.3
408	GB409	Y	103.2 ± 0.3
409	GB410	B	56.8 ± 1.2

410	GB411	Y	103.4 ± 2.9
411	GB412	B	57.8 ± 1.6
412	GB413	Y	78.3 ± 3.2
413	GB414	Y	83.7 ± 2.2
414	GB415	B	61.9 ± 1.4
415	GB416	B	47.5 ± 0.8
416	GB417	Y	87.9 ± 1.1
417	GB418	B	62.4 ± 0.4
418	GB419	Y	85.8 ± 1.8
419	GB420	B	50.9 ± 1
420	GB421	B	62 ± 1.5
421	GB422	B	61.7 ± 0.6
422	GB423	B	57.4 ± 2.3
423	GB424	Y	89.4 ± 1.4
424	GB425	Y	82.2 ± 1
425	GB426	B	52.9 ± 1.1
426	GB427	B	54.1 ± 1.9
427	GB428	Y	95.3 ± 0.4
428	GB429	Y	92.4 ± 1.1
429	GB430	Y	98.2 ± 1.7
430	GB431	B	51.3 ± 2
431	GB432	B	52.9 ± 1.6
432	GB433	Y	92.4 ± 2.3
433	GB434	Y	86.3 ± 2.3
434	GB435	Y	100.8 ± 3
435	GB436	B	60.4 ± 2.2
436	GB437	B	56.4 ± 3.4
437	GB438	B	51 ± 0.9
438	GB439	Y	85.4 ± 1.3
439	GB440	B	54.5 ± 1.7
440	GB441	B	53.9 ± 0.6
441	GB442	B	62 ± 0.8
442	GB443	Y	95.1 ± 0.6
443	GB444	B	60.7 ± 0.7
444	GB445	B	61.1 ± 2
445	GB446	B	64.1 ± 0.5
446	GB447	B	65.1 ± 0.9
447	GB448	Y	86.2 ± 0.5
448	GB449	Y	84.8 ± 3
449	GB450	B	48 ± 1.4
450	GB451	B	61.4 ± 1.3
451	GB452	Y	77.3 ± 4.1
452	GB453	B	48.7 ± 0.4
453	GB454	Y	92 ± 1.2
454	GB455	B	53.5 ± 2.2
455	GB456	Y	92.9 ± 0.7
456	GB457	Y	96.7 ± 0.7
457	GB458	Y	99.7 ± 0.9
458	GB459	Y	100.8 ± 0.8
459	GB460	B	53.5 ± 1.7
460	GB461	B	58.1 ± 0.9
461	GB462	B	57.7 ± 2.3
462	GB463	Y	95.3 ± 1.2

410	GB411	Y	94.7 ± 7
411	GB412	B	56.8 ± 3.3
412	GB413	Y	82.9 ± 9.5
413	GB414	Y	83.4 ± 1.4
414	GB415	B	68.5 ± 1.3
415	GB416	B	57.5 ± 1.2
416	GB417	Y	79.5 ± 2
417	GB418	B	67.6 ± 2.2
418	GB419	Y	100.7 ± 13.7
419	GB420	B	61.2 ± 2.8
420	GB421	B	61 ± 3.9
421	GB422	B	63.3 ± 4.8
422	GB423	B	55.8 ± 4.1
423	GB424	Y	99.4 ± 8.5
424	GB425	No flowers	
425	GB426	B	66.3 ± 3.4
426	GB427	B	63.5 ± 2.1
427	GB428	Y	101.6 ± 2.8
428	GB429	Y	87.9 ± 2.3
429	GB430	Y	87.6 ± 6.8
430	GB431	B	60.2 ± 2.5
431	GB432	B	58.4 ± 2
432	GB433	Y	89.9 ± 3.6
433	GB434	Y	87.8 ± 6.7
434	GB435	Y	92.2 ± 2.8
435	GB436	B	66.2 ± 1.8
436	GB437	B	61.2 ± 3.1
437	GB438	B	62.4 ± 2.6
438	GB439	Y	77.3 ± 0.9
439	GB440	B	61.4 ± 0.5
440	GB441	B	49.8 ± 1.1
441	GB442	B	60.5 ± 0.9
442	GB443	Y	93.7 ± 5.7
443	GB444	B	61.3 ± 4.9
444	GB445	B	60.9 ± 2.2
445	GB446	B	51.5 ± 0.6
446	GB447	B	60.5 ± 3.3
447	GB448	Y	78.6 ± 8.2
448	GB449	Y	96.2 ± 10.8
449	GB450	B	61.5 ± 1.4
450	GB451	B	59.5 ± 4.9
451	GB452	Y	77 ± 1.4
452	GB453	B	54 ± 1
453	GB454	Y	99 ± 3.9
454	GB455	Blue	
455	GB456	Y	89.3 ± 2.2
456	GB457	Y	100 ± 2.4
457	GB458	Y	96.1 ± 8.3
458	GB459	Y	95.3 ± 5.1
459	GB460	B	65.8 ± 2.4
460	GB461	B	58.7 ± 4.9
461	GB462	B	61.3 ± 2.8
462	GB463	Y	102.2 ± 3.9

463	GB464	Y	80.5 ± 0.8	463	GB464	Y	87.8 ± 11
464	GB465	Y	96.2 ± 1.8	464	GB465	Y	92.1 ± 8.7
465	GB466	B	52.8 ± 0.5	465	GB466	B	54 ± 3.9
466	GB467	Y	85.4 ± 1.8	466	GB467	Y	75.3 ± 3.5
467	GB468	B	63.1 ± 2	467	GB468	B	64 ± 2.5
468	GB469	B	63.2 ± 0.9	468	GB469	B	57.2 ± 1.4
469	GB470	B	56.5 ± 2.3	469	GB470	B	54.3 ± 1.1
470	GB471	Y	86.9 ± 1.4	470	GB471	Y	86 ± 4.9
471	GB472	Y	79.6 ± 1.3	471	GB472	Y	80 ± 10.3
472	GB473	Y	83.6 ± 1.8	472	GB473	Y	93.7 ± 4.4
473	GB474	Y	87 ± 1.3	473	GB474	Y	93 ± 3.6
474	GB475	Y	83.9 ± 2.1	474	GB475	Y	87.6 ± 1.5
475	GB476	Y	74.8 ± 0.5	475	GB476	Y	87.5 ± 4.6
476	GB477	Y	82.7 ± 0.4	476	GB477	Y	84.7 ± 2.8
477	GB478	Y	103.2 ± 6.9	477	GB478	Y	95.9 ± 5.4
478	GB479	B	61.9 ± 0.4	478	GB479	B	59.4 ± 2.8
479	GB480	Y	75.3 ± 3.8	479	GB480	Y	83.7 ± 5.2
	Maximum		108.1		Maximum		106.1
	Minimum		46.8		Minimum		43.9
	CDC Bethune	B	56.7 ± 0.4		CDC Bethune	B	63.6 ± 1.8
	G1186/94	Y	99.3 ± 3		G1186/94	Y	99.7 ± 3.1

Appendix 4: The list of 94 RILs with their seed colour values (RGB) used for genotyping. Seed colour values of two growth conditions represented by GC 2008 (growth chamber, 2008) and GH 2011 (greenhouse, 2011) were used in CIM analyses.

		Mean seed colour values (RGB)				Mean seed colour values (RGB)				Mean seed colour values (RGB)	
#	RILs	GC, 2008	GH, 2011	#	RILs	GC, 2008	GH, 2011	#	RILs	GC, 2008	GH, 2011
1	**GB2**	63.4 ± 4.7	63.2 ± 1.1	36	**GB321**	50.9 ± 3.2	58.2 ± 2.3	71	**GB230**	104.7 ± 2.7	96.5 ± 2.7
2	**GB9**	65.2 ± 3.7	70.2 ± 1.4	37	**GB333**	55.2 ± 1	55.3 ± 1.4	72	**GB246**	84.8 ± 2.2	88.6 ± 2.4
3	**GB18**	61 ± 3.6	65.3 ± 2.1	38	**GB343**	60.7 ± 0.8	61.2 ± 1.7	73	**GB247**	96.4 ± 7.4	98.5 ± 4.3
4	**GB26**	51.4 ± 2	59.6 ± 1.4	39	**GB350**	57.6 ± 4.1	61.8 ± 4.6	74	**GB253**	93.8 ± 2.6	94.2 ± 2.1
5	**GB39**	60.5 ± 3.1	58.6 ± 2.6	40	**GB356**	54.8 ± 1.7	62.1 ± 2.9	75	**GB257**	95.8 ± 1.7	89.4 ± 4.8
6	**GB48**	60.2 ± 3.5	53 ± 3	41	**GB362**	65.8 ± 1.6	67.1 ± 2.7	76	**GB272**	81.8 ± 2.3	73.5 ± 2.2
7	**GB72**	63.7 ± 2.5	60 ± 3.5	42	**GB364**	59.4 ± 2.4	59.9 ± 1.5	77	**GB282**	92.5 ± 1.8	100.3 ± 5
8	**GB83**	58.8 ± 3.2	44 ± 0.1	43	**GB366**	59.4 ± 2.8	59.4 ± 1	78	**GB292**	90.3 ± 0.5	93.2 ± 1.8
9	**GB92**	62.1 ± 4.8	61.8 ± 2.2	44	**GB374**	57.9 ± 1.2	59 ± 1.1	79	**GB298**	86.8 ± 4.9	98.5 ± 1.9
10	**GB97**	59.7 ± 4.6	55.5 ± 0.9	45	**GB380**	61.8 ± 3.9	59.6 ± 0.9	80	**GB309**	98.1 ± 3.2	95.8 ± 1.4
11	**GB103**	56.4 ± 1.2	61.6 ± 0.6	46	**GB385**	48.8 ± 0.3	64.5 ± 0.7	81	**GB323**	83.8 ± 4.6	79.7 ± 3.8
12	**GB110**	56.4 ± 5.5	56.4 ± 5.5	47	**GB386**	55.7 ± 1.4	62.7 ± 2.8	82	**GB338**	91.6 ± 3.2	87.5 ± 1.7
13	**GB126**	54.1 ± 0.9	54.1 ± 0.9	48	**GB6**	84.3 ± 9.2	90.1 ± 2.8	83	**GB345**	84.8 ± 5	91.6 ± 4.4
14	**GB143**	50.1 ± 0.5	51.3 ± 1.3	49	**GB13**	94.1 ± 4.7	98.4 ± 3	84	**GB357**	88.6 ± 2.4	97.6 ± 7.7
15	**GB148**	54.8 ± 4.6	55.1 ± 1.2	50	**GB35**	94.4 ± 3.6	99.9 ± 2	85	**GB371**	91 ± 6.2	82.6 ± 12.1
16	**GB155**	55.9 ± 1.4	58.7 ± 2	51	**GB37**	80.8 ± 2.7	82.7 ± 0.3	86	**GB387**	83.4 ± 1	93.4 ± 5.9
17	**GB160**	61.4 ± 2.6	63.5 ± 1.4	52	**GB44**	95.6 ± 2.1	95.9 ± 1.8	87	**GB388**	81.4 ± 2.5	84.7 ± 2
18	**GB168**	56.8 ± 3.1	60.8 ± 1.1	53	**GB54**	91 ± 2.9	92.1 ± 1.5	88	**GB393**	90.6 ± 2.7	93.5 ± 1.5
19	**GB183**	56.7 ± 2.6	59.1 ± 1.5	54	**GB70**	81 ± 3.3	81 ± 3.3	89	**GB404**	92.8 ± 1.3	94.3 ± 1.5
20	**GB193**	61.2 ± 2.2	65 ± 0.7	55	**GB79**	89.2 ± 2.9	106.2 ± 0.9	90	**GB408**	86.2 ± 0.8	87 ± 3.3
21	**GB207**	56.8 ± 1.3	60.5 ± 1.5	56	**GB88**	83.1 ± 1.9	79.1 ± 0.9	91	**GB413**	78.3 ± 3.2	82.9 ± 9.5
22	**GB218**	56.8 ± 0.5	56.4 ± 2.5	57	**GB98**	92.5 ± 2.9	83.8 ± 0.9	92	**GB428**	95.3 ± 0.4	101.6 ± 2.8
23	**GB229**	57.3 ± 5.6	60.3 ± 2	58	**GB104**	103.3 ± 1.3	98.6 ± 2.4	93	**GB429**	92.4 ± 1.1	87.9 ± 2.3
24	**GB244**	54.1 ± 2.4	60.1 ± 2	59	**GB115**	78.2 ± 3.2	77.5 ± 1.2	94	**GB439**	85.4 ± 1.3	77.3 ± 0.9
25	**GB262**	65 ± 2.8	60.6 ± 1.1	60	**GB117**	96.6 ± 1.3	102 ± 0.5				
26	**GB264**	56.3 ± 2.6	55.2 ± 1.2	61	**GB124**	89.3 ± 2.3	91.3 ± 1.1				
27	**GB267**	63.8 ± 3.6	55.9 ± 1.1	62	**GB131**	85.6 ± 1.2	94.4 ± 3.3				
28	**GB268**	55.9 ± 2.3	57.9 ± 0.6	63	**GB137**	89.1 ± 5.5	83.1 ± 3.1				
29	**GB271**	51.4 ± 1.8	59.1 ± 1.2	64	**GB145**	79.2 ± 6.1	80.5 ± 5.7				
30	**GB281**	63.1 ± 0.5	64.4 ± 4.4	65	**GB158**	91 ± 2.4	91.4 ± 3.5				
31	**GB291**	61.5 ± 1.6	52.7 ± 2.6	66	**GB164**	89.4 ± 1.8	88.7 ± 4.3				
32	**GB304**	55.7 ± 0.7	53.5 ± 1.6	67	**GB177**	95.7 ± 3.4	99.7 ± 0.9				
33	**GB313**	55.7 ± 1.7	53.8 ± 0.9	68	**GB180**	93.8 ± 2	88.5 ± 0.8				
34	**GB315**	62.9 ± 2.2	58.1 ± 2.5	69	**GB185**	90.1 ± 4.7	84.9 ± 4.8				
35	**GB316**	59.5 ± 0.8	62.5 ± 2.7	70	**GB195**	86.5 ± 6.7	76.3 ± 2.9				

Appendix 5: The list of SSR markers with their genetic distance, forming 19 linkage groups in the mapping population. The markers are shown with their corresponding LGs published in Cloutier et al. (2012a).

#	Locus	linkage group	Corresponding linkage group in Cloutier et al. (2012a)	Number of markers per linkage group
1	Lu125	LG 1	LG 2	4
2	d	LG 1		
3	Lu209	LG 1	LG 2	
4	Lu257	LG 1	LG 2	
5	Lu344	LG 1	LG 2	
6	Lu2746	LG 2	LG 10	3
7	Lu371	LG 2	LG 10	
8	Lu2268	LG 2	LG 8	
9	Lu2317	LG 3	LG 8	5
10	Lu857	LG 3	LG 8	
11	Lu2957	LG 3	LG 8	
12	Lu3059	LG 3	LG 8	
13	Lu2840	LG 3	LG 8	
14	Lu390	LG 4		4
15	Lu2913	LG 4	LG 12	
16	Lu3064	LG 4	LG 12	
17	Lu381	LG 4	LG 12	
18	Lu2779	LG 5	LG 12	4
19	Lu896	LG 5	LG 12	
20	Lu220	LG 5	LG 12	
21	Lu140	LG 5	LG 12	
22	Lu2054	LG 6	LG 4	8
23	Lu587	LG 6	LG 4	
24	Lu3113	LG 6	LG 4	
25	Lu3229	LG 6	LG 4	
26	Lu2087	LG 6	LG 4	
27	Lu2073	LG 6	LG 4	
28	Lu2006	LG 6	LG 4	
29	Lu2966	LG 6	LG 4	
30	Lu3053	LG 7	LG 1	2
31	Lu2010a	LG 7	LG 15	
32	Lu299	LG 8	LG 1	2

33	Lu56	LG 8	LG 1	
34	Lu318	LG 9	LG 3	2
35	Lu2628	LG 9	LG 3	
36	Lu774	LG 10	LG 3	7
37	Lu64	LG 10	LG 3	
38	Lu2164	LG 10	LG 3	
39	Lu3111	LG 10	LG 3	
40	Lu933	LG 10	LG 3	
41	Lu658	LG 10	LG 3	
42	Lu422a	LG 10	LG 3	
43	Lu2557	LG 11	LG 6	5
44	Lu3057a	LG 11	LG 6	
45	Lu442a	LG 11	LG 6	
46	Lu2971	LG 11	LG 6	
47	Lu69	LG 11	LG 6	
48	Lu699	LG 12	LG 6	3
49	Lu502	LG 12	LG 6	
50	Lu2917	LG 12	LG 6	
51	Lu672	LG 13	LG 7	2
52	Lu585	LG 13	LG 7	
53	Lu176	LG 14	LG 5	2
54	Lu223	LG 14	LG 5	
55	Lu650	LG 15	LG 13	2
56	Lu805	LG 15	LG 13	
57	Lu458	LG 16	LG 10	3
58	Lu273	LG 16	LG 10	
59	Lu685	LG 16	LG 10	
60	Lu3033	LG 17	LG 14	3
61	Lu3038	LG 17	LG 14	
62	Lu476	LG 17	LG 14	
63	Lu785	LG 18	LG 11	2
64	Lu292	LG 18	LG 11	
65	Lu462	LG 19	LG 15	5
66	Lu2695	LG 19	LG 15	
67	Lu2697b	LG 19	LG 15	
68	Lu1172	LG 19	LG 15	
69	Lu2696	LG 19	LG 15	

Appendix 6: Segregation of SSR and CAPS marker alleles in $F_{8:9}$ RIL populations of CDC Bethune X G1186/94 and its reciprocal cross of *Linum usitatissimum.*

Marker	Total plants	Observed marker genotype		Segregation 1:1	
		No. Brown type	No. Yellow type	χ^2	p-value
Lu943	71	39	32	0.7	0.41
Lu3053	92	66	26	17.4	< 0.05
Lu2055	91	57	34	5.8	< 0.05
Lu299	90	47	43	0.2	0.67
Lu56	92	44	48	0.2	0.68
Lu49b	67	26	41	3.4	0.07
Lu125	93	39	54	2.4	0.12
Lu209	91	45	46	0.01	0.92
Lu257	89	48	41	0.6	0.46
Lu344	86	52	34	3.8	0.051
LuCAPS_110	90	44	46	0.04	0.83
Lu2351	94	44	50	0.4	0.54
Lu2346	92	38	54	2.8	0.1
LuM566	91	44	47	0.1	0.75
LuM568	91	44	47	0.1	0.75
LuM569	92	43	49	0.4	0.53
LuM588	92	47	45	0.04	0.83
LuM592	92	44	48	0.2	0.68
LuM593	92	44	48	0.2	0.68
LuM595	91	45	46	0.01	0.92
LuM597	91	45	46	0.01	0.92
LuM71	90	47	43	0.2	0.67
Lu2347	80	32	48	3.2	0.07
Lu2341	83	35	48	2.04	0.15
Lu318	89	42	47	0.3	0.6
Lu2628	91	45	46	0.01	0.92
Lu774	92	42	50	0.7	0.4
Lu64	93	40	53	1.8	0.18
Lu2164	88	45	43	0.05	0.83
Lu3111	91	41	50	0.9	0.35
Lu933	93	44	49	0.3	0.6
Lu658	93	48	45	0.1	0.76
Lu422	87	46	41	0.3	0.6
Lu587	88	44	44	0	1

Lu2966	88	44	44	0	1
Lu2006	87	51	36	2.6	0.11
Lu3229	82	47	35	1.8	0.19
Lu2087	92	49	43	0.4	0.53
Lu3113	92	50	42	0.7	0.4
Lu2054	85	50	35	2.6	0.1
Lu2024	90	58	32	7.5	< 0.05
Lu2073	92	50	42	0.7	0.4
Lu682	92	39	53	2.1	0.14
Lu652	86	40	46	0.4	0.52
Lu176	91	46	45	0.01	0.92
Lu223	91	44	47	0.1	0.75
Lu69	91	24	67	20.3	< 0.05
Lu442	92	23	69	23	< 0.05
Lu699	93	48	45	0.1	0.75
Lu502	89	49	40	0.9	0.34
Lu2917	91	49	42	0.5	0.46
Lu2971	90	24	66	19.6	< 0.05
Lu2557	86	41	45	0.2	0.67
Lu3057a	94	44	50	0.4	0.54
Lu146	93	47	46	0.01	0.91
Lu672	92	38	54	2.8	0.1
Lu585	92	37	55	3.5	0.06
Lu566	90	42	48	0.4	0.53
Lu2587	90	45	45	0	1
Lu2317	92	41	51	1.1	0.3
Lu857	94	49	45	0.2	0.68
Lu2957	89	40	49	0.9	0.34
Lu3059	91	46	45	0.01	0.92
Lu2840	92	35	57	5.3	< 0.05
Lu91	74	40	34	0.5	0.49
Lu757	88	54	34	4.5	< 0.05
Lu801	89	54	35	4.1	< 0.05
Lu2746	92	43	49	0.4	0.53
Lu371	87	47	40	0.6	0.45
Lu458	89	50	39	1.4	0.24
Lu273	90	43	47	0.2	0.7
Lu685	93	45	48	0.1	0.75
Lu291	94	47	47	0	1

Lu575	78	57	21	16.6	< 0.05
Lu785	84	48	36	1.7	0.19
Lu292	89	45	44	0.01	0.92
Lu390	89	48	41	0.6	0.46
Lu2913	90	47	43	0.2	0.67
Lu3064	87	45	42	0.1	0.75
Lu381	94	43	51	0.7	0.41
Lu2779	93	50	43	0.5	0.47
Lu896	90	48	42	0.4	0.53
Lu220	86	43	43	0	1
Lu140	79	44	35	1	0.31
Lu628	57	33	24	1.4	0.23
Lu650	88	53	35	3.7	0.06
Lu805	92	50	42	0.7	0.4
Lu3033	89	35	54	4.1	< 0.05
Lu3038	87	30	57	8.4	< 0.05
Lu476	92	35	57	5.3	< 0.05
Lu959	93	41	52	1.3	0.25
Lu462	86	45	41	0.2	0.67
Lu2010a	93	55	38	3.1	0.08
Lu1172	89	43	46	0.1	0.75
Lu2696	91	41	50	0.9	0.35
Lu2695	91	41	50	0.9	0.35
Lu2697b	92	44	48	0.2	0.68
Lu2268	71	38	33	0.4	0.6
Lu2921	70	15	55	22.9	< 0.05

Appendix 7: The list of ESTs and their sequences retrieved from NCBI *Linum* EST database which matched with the putative candidate gene F3'H.

1) >gi|324710483|gb|JG242995.1| LUSTC1NG-RP-147_F12_11MAR2007_086 LUSTC1NG *Linum usitatissimum* cDNA, mRNA sequence

ACCACCACGTCCACGTACCCTAGCCTGAGGTATATTAGGGGTCCGTACCTCCTAGCCAGGGCGG
CTATAGTGTGGTGGGGCATTGGACCCAAGTGGAGAAGGTTCCCCACAATCGGCCACGGAGTGGG
GCCCGGTAGGAGAGGCTTTCTGCCACGTTGGCGGAGGAGGTGGATGAGGCTGAGAAGGAGGTAG
ATGGCGGTGGCGATGATAATGCCGTCGCACATGATGGCCGTCGACGTAGACATGTTTGTGTTTA
TGTATGTTATTTGTTAGGGGATGAAGGGTC

2) >gi|324642576|gb|JG057957.1| 03-LUSEN1NG-RP-226_G04_05JUN2007_020 LUSEN1NG *Linum usitatissimum* cDNA, mRNA sequence

ACAAATAATAATTCTATTATTTTATATAAGTGATGTATGAAATATTACAATGAAGCCACCCAGA
TTTGTTACAATCACATTACTACCATTCATACGCAGCCATCTGTCACCCAACGTTTTTATTCATA
CATAAAACAAACGAAAAACTCATCCCACGAGGATCCCTGGGGGCTTGGTTGGTTCCCTTCACCC
TCGATAAACATGCTCCGCCAATCGTGCTTTTGGCCGCACGACCAGCGGCACAGCCCGTTGTAAC
GAAATCCCGAAAACCTCATCCATATTCAGCTCCTCCGCCGAGACTCCGTCTTTAAGCTCCCAAT
CAAACCCATGGGCAAGAACAGCCGTCATCAGCTGAACCGTCCGGAGCCCCAAGCTCATACCCGA
GCAGATCCTCCTACCCGCCC

3) >gi|324837394|gb|JG042539.1| 03-LUSEN1NG-RP-049_E03_10MAR2007_023 LUSEN1NG *Linum usitatissimum* cDNA, mRNA sequence

CGGTTACAAATAATAATTCTATTATTTTATATAAGTGATGTATGAAATATTACAATGAAGCCAC
CCAGATTTGTTACAATCACATTACTACCATTCATACGCAGCCATCTGTCACCCAACGTTTTTAT
TCATACATAAAACAAACGAAAAATTCATCCCACGAGGATCCCTGGGGGCTTGGTTGGTTCCCTT
CACCCTCGATAAACATGCTCCGCCAATCGTGCTTTTGGCCGCACGACCAGCGGCACAGCCCGTT
GTAACGAAATCCCGAAAACCTCATCCATATTCAGCTCCTCCGCCGAGACTCCGTCTTTAAGCTC
CCAATCAAACCCATGGGCAAGAACAGCCGTCATCAGCTGAACCGTCCGGAGCCCC

4) >gi|324866540|gb|JG079443.1| LUSFL2AD-WB-027_I05_27NOV2008_023 LUSFL1AD *Linum usitatissimum* cDNA, mRNA sequence

ACTTAAAACAAACGAAAAATTCATCCCAAGAGGATCCCTGGTGGCTTGGTTGGTTCCCTTCACC
CTCGATAAACATGCTCCGCCAATCGTGCTTTTGGCCGCACGACCAGCGGCACAGCCCGTTGTAA
CGAAATCCCGAAAACCTCATCCATATTCAGCTCCTCCGCCGAGACTCCGTCTTTAAGCTCCCAA
TCCATAACCGAGAACTTGCAATTGGTCGGCTGGTCAGCAAAGAACGGCCAGCACAGCACCGGAA
C

5) >gi|324766924|gb|JG058418.1| 03-LUSEN1NG-RP-231_H10_05JUN2007_066 LUSEN1NG *Linum usitatissimum* cDNA, mRNA sequence

CAATAATAATTCTATTATTTTATATAAGTGATGTATGAAATATTACAATGAAGCCACCCAGATT
TGTTACAATCACATTACTACCATTCATACGCAGCCATCTGTCACCCAACGTTTTTATTCATACA
TAAAACAAACGAAAAATTCATCCCACGAGGATCCCTGGGGGCTTGGTTGGTTCCCTTCACCCTC
GATAAACATGCTCCGCCAATCGTGCTTTTGGCCGCACGACCAGCGGCACAGCCCGTTGTAACGA
AATCCCGAAAACCTCATCCATATTCAGCTCCTCCGCCGAGACTCC

6) >gi|324866485|gb|JG075410.1| LUSFL1AD-WB-018_B08_10NOV2008_032 LUSFL1AD *Linum usitatissimum* cDNA, mRNA sequence

ACATAAAACAAACGAAAAATTCATCCCACGAGGATCCCTGGTGGCTTGGTTGGTTCCCTTCACC
CTCAATAAACATGCTCCGCCAATCGTGCTTTTGGCCGCACGACCAGCGGCACA