QTL MAPPING AND NIRS ESTIMATION OF CYANOGENIC GLUCOSIDES IN FLAXSEED

BY

ADAM SCOTT CHIN-FATT

A Thesis

Submitted to the Faculty of Graduate Studies of the

University of Manitoba

in Partial Fulfilment of the Requirements

for the Degree of

MASTER OF SCIENCE

Department of Plant Science

University of Manitoba

Winnipeg, Manitoba, Canada

THE UNIVERSITY OF MANITOBA

FACULTY OF GRADUATE STUDIES

COPYRIGHT PERMISSION

QTL MAPPING AND NIRS ESTIMATION OF CYANOGENIC GLUCOSIDES IN FLAXSEED

BY

ADAM SCOTT CHIN-FATT

A Thesis submitted to the Faculty of Graduate Studies

of The University of Manitoba

in partial fulfilment of the requirements of the degree of

MASTER OF SCIENCE

ACKNOWLEDGEMENTS

To my family, for keeping me grounded and somewhat sane throughout this experience and for instilling in me a love of science and a curiosity for the unknown.

To an all-star team of lab technicians and post-docs. In particular, thanks to Debbie and Elsa, my surrogate work mothers, who were akin to walking manuals on 'how to be a good scientist.' To Andy, for his steady guidance and refreshingly dark humor. To Santosh and Raja, for being accommodating in their impressive wealth of experience and knowledge.

To the Chemistry experts of the Canadian Grain Commission. In particular, thanks to Ray, and his amazing attention to detail and his vast insight into the complexities of analysis and application. To Bert, whose experience has proved invaluable in navigating the infernal NIRS apparatus.

To my supervisor Sylvie, a bastion of efficiency, for keeping me on track with this project and inspiring me to be a better researcher. Finding a more capable mentor is unimaginable. Without her, this project submission would no doubt have been procrastinated for another couple of months.

To the aforementioned and all the other people I've met on this journey that have made this project possible. I wish them all the best and it has been an honor to collaborate with people of such great esteem and proficiency.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ALA	α-linolenic acid
ANOVA	Analysis of variance
BAC	Bacterial artificial chromosome
BAM	Binary alignment/map
CGs	Cyanogenic glucosides
CIM	Composite interval mapping
CDC	Crop Development Centre
CYP	Cytochrome P-450
DHA	Docosahexaenoic acid
DL	Double Low
DSC	Detrend scatter correction
E	Environment
EPA	Eicosapentaenoic acid
ER	Endoplasmic reticulum
EST	Expressed sequence tag
EU	European Union
G	Genotype
GC	Gas chromatograph(y)
GRAS	Generally Recognized As Safe
HCN	Hydrocyanic acid
HPLC	High performance liquid chromatography

HPTLC	High performance thin-layer chromatography
LG	Linkage group
LIN	Linustatin
LOD	Logarithm of odds
Max	Maximum
MD	AC McDuff
Min	Minimum
MPLS	Modified partial least squares
NEO	Neolinustatin
NIRS	Near infrared spectroscopy
PCR	Polymerase chain reaction
PLS	Partial least squares
QTL	Quantitative trait locus/loci
R^2	Coefficient of determination
RIL	Recombinant inbred line
RPD	Relative percent difference
SDG	Secoisolariciresinol diglucoside
SECV	Standard error of cross validation
SNP	Single nucleotide polymorphism
SNV	Standard normal variant
SSR	Simple sequence repeat
TUFGEN	Total utilization of flax genomics
UGT	Uridine diphosphate glycosyltransferase

ω-3	Omega 3
ω-6	Omega 6

LIST OF APPENDICES

ABSTRACT

Cyanogenic glucosides (CGs) are bioactive plant secondary metabolites that can release toxic hydrocyanic acid when hydrolyzed. The accumulation of CGs in flax seed (*Linum usitatissimum* L.) is a safety issue as a feed component and may contravene international trade restrictions. Here, we report the identification of major effect quantitative trait loci (QTL) for linustatin and neolinustatin, the two most abundant CGs in the mature flax seed. A genetic linkage map was developed with a recombinant inbred line population that segregated for CG content in mature seeds. The 15 linkage groups (LGs) of the genetic map included 155 simple sequence repeat markers, 39 single nucleotide polymorphism markers and one gene totalling 938 cM with an average marker interval of 5 cM. The population was grown in the field at three site-years from which linustatin and neolinustatin contents were measured by gas chromatography. Initial QTL mapping identified a 16.5 cM region on LG1 associated with linustatin content and a 22.5 cM region associated with neolinustatin content on LG6. Fine mapping using 82 recombinant lines shortened the linustatin associated region to 11.7 cM and the neolinustatin QTL to two regions of 3.5 cM and 15.2 cM. On the physical map, the linustatin QTL was anchored to two scaffolds and the two neolinustatin QTL were anchored to two and four scaffolds, respectively. *In silico* analysis of these scaffolds identified seven clusters of candidate sequences for future functional unit identification.

Current methods of CG analysis for screening and quality control are prohibitively expensive. Here, we also report on the development of a low cost, high throughput method of analysis using near infrared spectroscopy (NIRS) to estimate individual CGs and total

hydrocyanic acid equivalent based on a regression of reference data obtained by gas chromatography. A modified partial least squares model on a validation set produced standard error of prediction values ranging from 0.07 to 0.26 and ratio of performance to deviation values from 1.09 to 1.88. The genetic and physical mapping of the QTL and the NIRS calibration hold direct applications in the development of a flax breeding strategy for developing germplasm and cultivars with reduced CG content.

FOREWORD

This thesis is structured to conform to the guidelines for a paper style format set forth by the Faculty of Graduate Studies of the University of Manitoba. Its five chapters include a general introduction, a literature review, two manuscripts and a general discussion. The manuscripts were formatted as would be required for submission to the scientific journal Theoretical and Applied Genetics.

1.0 GENERAL INTRODUCTION

1.1 Background

The detection of QTL that associate with a significant difference in cyanogenic glucoside (CG) content is an effective strategy for developing a consistently low CG cultivar. Locating a QTL relies on the construction of a good genetic map combined to accurate phenotyping in multiple environments. Due to the phenomenon of linkage disequilibrium, that is, the non-random inheritance of a combination of genomic sequences, markers can pinpoint a particular locus in terms of relative genetic distance. The identification of tightly linked molecular markers provides breeders with a 'roadmap' of the relative positions of loci influencing certain traits. Knowledge of the complexity of the genetics and environmental sensitivity of a trait combined with knowledge of the location of a CG QTL in relation to other undesirable (or desirable) loci enable breeders to design the best breeding strategies to achieve improvement targets. Such strategies include the design of crosses, the selection strategies, the probability of a desirable recombination among loci of interest after crossing (linkage drag) and selection, and the required size of the selection population.

Even if a consistently low CG cultivar is developed, CG content in a single cultivar can vary greatly across environments and even within a single environment, depending on nutritional status, maturity date and other variables. A low cost, high throughput routine analytical method for screening bulk samples is needed to maintain quality control. We

currently do not have any analytical methods in flax that meet all of the requirements for large scale screening in a breeding program. Most, if not all, do not offer the accuracy and reproducibility required for effective selection. Near infrared spectroscopy (NIRS) is a technique that can estimate the concentration of an analyte based on a mathematical relationship between absorption spectra and reference values. Its development to estimate CG concentration would be beneficial to plant breeders for early generation screening.

NIRS is an appealing technique because it is straightforward, rapid and non-destructive. The obtained spectra confer molecular data, based on the sizes of overtone peaks from the harmonic oscillation of H bonds, for as many constituents as there are available calibrations. Minimal sample preparation, usually consisting of cleaning the seeds and possible grinding, is required and no hazardous reagents are involved. It is ideally suited to flax seed sampling because the penetration depth allows bulk material analysis and analysis through some translucent packaging is feasible. However, the selectivity of NIRS is comparatively low and is dependent on the quality of the calibration developed using chemometrics, i.e., where the most appropriate statistical model is selected based on the relationship between known concentrations of an analyte and the relevant wavelengths. The development of a reliable and stable calibration can be challenging because there are no standardized approaches to model development and troubleshooting calibration issues often require empirical approaches eased by experienced operators. The complexity of NIRS spectra combined to often overlapping absorption bands may obscure analysis unless deconvoluted by appropriate mathematical techniques. Water and physical attributes such

as sample size, shape and hardness can cause interference due to light scattering effects. The transfer of a calibration across instruments, even of the same type, is also often inconsistent and requires instrument standardization, a process consisting of various chemometric adjustments to either the existing calibration so that it works on a new instrument or to the spectral data to minimize wavelength and absorbance shifts. Nonetheless, as better chemometric techniques emerge, NIRS is becoming increasingly viable as an analytical method in the agricultural industry.

1.2 Research Objectives

1) To identify QTL of the flax genome that have a major effect on CG accumulation
2) To quantify the CG variation attributed to the QTL and their parental origins
3) To develop and validate a calibration using NIRS that can be used to estimate the concentration of CGs in samples of flax seeds

1.3 Basic Strategy

To identify an association between a particular DNA segment and CG content, genotypic and phenotypic datasets are needed. Using molecular markers that indicate the parental source of a segment of DNA, a binary score sheet can be constructed giving the relative regions of recombination along an individual's genome. These data are obtained for all members of a segregating population generated from parents that are phenotypically contrasting and genotypically divergent. The statistical power needed to pinpoint a QTL is

limited first by sample size and then by genetic marker coverage. The size of the population needed for sufficient statistical power depends on several factors, including complexity of the trait, environmental effects, precision of the analytical method as well as resources and objectives of the experiment. The more divergent the parents' genotypes are, the better the likelihood of identifying polymorphic markers that are well-distributed across the genome of the population. By partitioning the population into marker-based genotypic classes, a genetic linkage map indicating the relative positions of the markers can be constructed. The phenotypic dataset consists of the measured CG concentrations for each individual in the population. Because CG concentration is strongly environmentally influenced, measurements from multiple environments need to be made in order to partition the genetic component. For best results, a minimum sample size of 200 individuals and at least five environments are recommended though design of the experiment often requires taking the genetic architecture, heritability of the trait and available resources into consideration as well (Schön et al. 2004). The association between the two datasets is achieved through statistical modeling, the power of which critically depends on the accuracy and size of the datasets. The result is a correlation between an interval of molecular markers and a significantly different level of CG content. In summary, to map CG QTL, the salient requirements are: a suitable mapping population of a large enough size derived from parents that contrast for CG levels, a linkage map saturated with molecular markers, reliable measurements of the CG concentrations of the mapping population in several

environments and an appropriate statistical model to analyze the correlation between genotype and phenotype.

The concept behind an NIRS calibration is that most molecules demonstrate a unique optical profile upon exposure to sufficiently penetrant wavelengths of light. Estimating an unknown value relies on an association, created by an appropriate statistical model, between that unique profile and a reference dataset of known concentrations. To assess how successful the calibration is, the predicted values are compared to the known values in terms of specificity, linearity, range, accuracy, precision, robustness and limits of quantitation and detection.

The overall strategy for developing an NIRS calibration can be generalized in a few steps. First, a set of samples with known concentrations is selected. Second, holding as many other factors as possible constant, the response of the set to a dependent variable, such as the optical response to certain wavelengths of light, is measured. Third, an appropriate calibration model is developed to maximize the correlation and minimize the error between the known values and the dependent values. Last, this model is tested on a set of samples not included in the calibration to validate how accurate the estimations are. Its versatility as an analytical technique has already garnered widespread acceptance in the agricultural industry for raw material identification, physical trait classification and molecular quantification. Considering the rapidity of developments in chemometrics and NIRS hardware over the last two decades, it is anticipated that in the near future, NIRS will

progressively become the routine operating procedure for product evaluation at all stages of the agricultural production process.

2.0 LITERATURE REVIEW

2.1 The Utility of Flax

Flaxseed (*Linum usitatissimum* L.) is becoming a popular functional food as consumers become increasingly health conscious. Its potential as a functional food is mainly attributed to its high levels of α-linolenic acid (ALA), an ω-3 fatty acid, but also to its protein, dietary fiber and lignan contents. It is unique among oilseed crops in demonstrating a fatty acid composition favoring ω-3 over ω-6 content. Flaxseed oil typically consists of 59% ± 5% ALA and 15% ± 2% of the ω-6 fatty acid, linoleic acid (Siemens and Daun 2005). In comparison, canola oil usually consists of 9% ± 3% ALA and 20% ± 8% linoleic acid (Velasco and Becker 1998). Canada has established a nutrient intake recommendation of 0.5% energy or 1.1g ω-3 fatty acids per day based on a 2,000 kcal diet (Nutrition Recommendations 1990).The current Western diet is deficient in ω-3 fatty acids with an ω-6:ω-3 ratio of 15-20:1 far exceeding the 4-10:1 ratio recommended by Health Canada, likely due to lower prioritisation of ω-3 fatty acid consumption and widespread use of linoleic acid rich vegetable oils (Scientific Review Committee 1990; Eaton and Konner 1985). The ω-3 and ω-6 fatty acids compete for the same desaturase enzymes for metabolic conversion to a range of bioactive longer chain metabolites called eicosanoids with those of ω-6 origin being more pro-inflammatory than those of ω-3 (Lands et al. 1973). A diet overly rich in ω-6 fatty acids results in the elevation of thromboxane, an ω-6 based eicosanoid, which increases the risk of platelet aggregation and thrombosis (Hamburg and

Samuelsson 1975). The intake of ω-3 fatty acids has been shown to reduce risk of cardiovascular disease (Temple 1996), as well as inhibit growth of prostate and breast cancer (Pandalai et al. 1996; Rose 1997), slow the loss of immunological function in autoimmunity and aging (Fernandes 1995) and is required for fetalneural and optical development (Neuringer et al. 1998; reviewed by Ruxton et al. 2004). Aside from ALA, the only other sources of ω-3 fatty acids are the longer- chain eicosapentaenoic acid (EPA) and docosahexaenoic acid (DHA), both of which are primarily marine derived. Given that ALA requires conversion to EPA and DHA for metabolic activity and that this conversion is limited in humans, direct consumption of EPA and DHA via marine sources is more efficient for individual improvement of ω-3 levels (James et al. 2003). However, the production capacity of global fisheries cannot meet these demands on a population scale and the provision of alternative ω-3 dietary options is necessary (Pauly et al. 2001). An appealing, more sustainable alternative is the generation of functional foods, such as eggs, by inclusion of flaxseed in animal feed since chickens are more efficient than humans at converting ALA into DHA and EPA in significant quantities (Gregory et al. 2012).

Its high nutritional value and relatively low cost make linseed a prime choice as a feed component in animal nutrition. When fed ground flaxseed as a feed component, broiler chickens have shown increased ω-3 levels and significantly reduced ω-6:ω-3 ratios (Ajuyah et al. 1993). Flaxseed rations of 10% or 20% resulted in egg yolks with 5% and 9% higher ALA content, respectively (Caston et al., 1994). Scheideler and Froning (1996) have also reported ω-3 enriched eggs from flaxseed fed hens to contain 250mg ALA,

100mg DHA and an ω-6:ω-3 ratio of 2.6 compared to a normal egg, produced without flaxseed supplementation, containing 40mg ALA, 20mg DHA and an ω-6:ω-3 ratio of 13.0. Volunteers who consumed eggs from flaxseed-fed hens showed a 33% increase in total ω-3 fatty acids and a significant decrease in ω-6:ω-3 ratio in platelet phospholipids (Ferrier et al. 1995). Volunteers who consumed ω-3 enriched eggs also displayed decreased plasma triglyceride concentrations and lowered systolic and diastolic blood pressures (Ferrier et al. 1992; Oh et al. 1991)

Flaxseed also contains lignans, specifically secoisolariciresinol diglucoside (SDG), a phenolic compound that aids in bone development, delays the progression of lupus nephritis and other kidney diseases and, exhibits hydrogen donating antioxidant activity (Thompson 2003). Flaxseed is unique among grains in having up to a thousand-fold more lignan precursors than any other plant sources (Aldercreutz and Mazur 1997).

2.2 Anti-Nutrients in Flaxseed

The use of flaxseed as feed in amounts exceeding 5-10% of the daily intake is associated with depressed growth, a phenomenon attributed to the presence of antinutritional factors (Madhusudhan et al. 1986). CGs and linatine, an antipyroxide factor, are the major contributors and, to a lesser extent, phytic acid and trypsin inhibitors. Flaxseed rations of 10% or 20% caused hens to weigh 8% and 11% less respectively at 51 weeks (Caston et al., 1994).

Upon ingestion, CGs can undergo hydrolysis to produce hydrocyanic acid (HCN), which is toxic. Ruminants are more susceptible to cyanide poisoning than humans and other animals because their rumen has a mildly alkaline pH, high water content and, contains microfloral enzymes that favor hydrolysis (Majak et al. 1990). In North America, the Food and Drug Administration has designated flaxseed as being Generally Recognized As Safe (GRAS) for food and feed products up to a maximum of 12% intake (Food and Drug Act, 2009). Although the presence of anti-nutritionals is unlikely to pose a safety risk in feed products at levels beneath 12% intake, it diminishes the competitiveness of linseed compared to more commonly used cereal crops in feed products. When linseed substituted maize, soybean and sunflower oil in broiler diets, there was a fall in digestibility of nutrients and on dietary metabolisable energy content, presumably due to the presence of anti-nutrients (Rodríguez et al. 2001). In spite of high nutritional value, flaxseed has yet to be fully exploited as a feed component in part because of the presence of CGs. The prospect of reducing CG levels would increase its appeal as a viable feed product alternative for farmers looking to incorporate flaxseed's unique nutritional profile in livestock, particularly for niche marketing of an ω-3 rich value added product, without sacrificing growth.

The presence of CGs is not as high a risk to humans because the acute oral lethal dose of HCN is 0.5-3.5 mg/kg of body weight (Montgomery, 1969). The recommended daily intake of 1-2 tablespoons of ground flaxseed, which releases a total of 5-10 mg HCN, is well below the estimated acute toxic dosage for an adult bodyweight (Roseling, 1994).

An average adult would have to consume eight cups (1 kg) of ground flaxseed to achieve acute cyanide toxicity. Additionally, the highly acidic environment of the human stomach inhibits CG hydrolysis and HCN production (Majak et al. 1990). Acute poisoning in humans by HCN or its breakdown products due to flax ingestion has not been reported to date. The general acute toxicity profile of HCN is characterized by a steep rate-dependent dose-response curve for which lethality is generally observed and commonly precluded by nausea, cyanosis, ataxic movements, seizures and cardiovascular collapse (Way, 1984). The effects are achieved primarily by inhibition of the electron transport chain in cellular respiration and subsequent depression of the central nervous system (Way, 1984). Although cyanide has a preferential affinity for hemoglobin in red blood cells, it has a fairly rapid elimination half-life of 14 minutes and will not accumulate in tissues with chronic oral ingestion (Leuschner et al. 1991). No chronic dose-response studies of cyanide ingestion in humans have yet been reported. Epidemiological studies have indicated low dosage chronic exposure to be associated with goiter (Abuye et al. 1998), ataxic tropic neuropathy (Oluwole et al. 2003) and spastic parapesis (Tylleskär et al. 1992). However, the effects observed in these epidemiological studies were often confounded by dietary deficiencies, particularly protein, iodine, vitamin B_{12}, and overall malnutrition making it difficult to assess the hazard posed by chronic cyanide ingestion.

CGs in flax are subject to trade restrictions in international markets. In the European Union (EU) directive for undesirable substances, the maximum allowed content of HCN in animal feed products is 250 mg/kg (ppm) for linseed or 350 mg/kg (ppm) for linseed

cake relative to a feedstuff with a 12% moisture content (European Parliament and Council 2002). Canada is the world's largest producer and exporter of oilseed flax or linseed (FAO 2009). Over the last four years, Canadian flax production averaged approximately 480,000 tonnes, with 76% supplied by Saskatchewan and the remainder equally supplied by Manitoba and Alberta (Flax Council of Canada, 2010-2014). In 2009, trace level contaminations with some genetically modified flaxseed referred to as Crop Development Centre (CDC) Triffid (FP967) were detected in two Canadian shipments in Europe where 70 percent of the Canadian flaxseed was exported. This transgenic event approved in North America in the 1990s was not approved in the EU (FCC 2009) where there is a zero tolerance for non-approved events. This non-tariff trade barrier called low level presence, can restrict or shut down trade if trace amounts of a GM crop that has not received regulatory approval in an importing country is discovered. Subsequent to the discovery of the CDC Triffid contamination in EU shipments, flax prices fell as low as $6 per bushel and flax acreage planted in Canada declined from 1.71 million acres in 2009 to 0.915 million acres hectares in 2010 and to 0.74 milllion acres in 2011 (Statistics Canada, 2009-2011). This decline was primarily due to a sharp fall in export demand likely caused by both Europe's restriction of the product and the growth in flax production of countries in the Black Sea region, thereby resulting in diminished farmer bids. Since then, this export market has not fully recovered despite the implementation of stringent broad scale testing in Canada prior to planting, post-harvest, at receptor sites and at grain terminals prior to export. Canada's flaxseed shipments are under great scrutiny to conform to EU regulatory

standards. As such, exports to the EU have dramatically decreased but have been somewhat compensated for by export to the USA, China and by an expanding domestic demand for food and feed grade flaxseed. The non-tariff trade barrier surrounding CGs falls under the same category of sanitary and phytosanitary measures as the low level presence of GMOs. Although detection of CG levels exceeding the trade restriction limit is unlikely to generate the same level of alarmism as did detection of CDC Triffid because of public perception, the uncontrolled accumulation of CGs remains a liability that could limit trade in a similar manner.

2.3 Processing Methods to Remove Cyanogenic Glucosides

Detoxification of the CG content in flaxseed meal has been tried with variable success with many post-harvest methods including boiling, roasting, autoclaving, microwaving and extrusion. The more successful methods developed involve a fermentation step and/or thermal treatment. Thermal treatment using full-fat flaxseed extrusion has been shown to be an effective technique to reduce HCN content by up to 89.1% (Imran et al. 2013). Flaxseed normally contains decomposing enzymes but the process of heat pressing, a common process for oil extrusion, inactivates the enzymes causing significant retention of the CGs. A fermentation method has been proposed for commercial use involving the enzymatic hydrolysis of total CG content followed by steam evaporation of the incubated slurry (Yamashita et al. 2007). However, this method is relatively energy and cost intensive due to the operating utility costs of steam evaporation at 120°C for 2h. Alternatively, a

triple extraction using a methanol-ammonia-water/hexane solvent has been used to remove more than 90% of the HCN content albeit at the expense of several nutritive components (Wanasundara et al. 1993). A promising method using an enzymatic preparation of commercial β-glucosidase, which catalyzes hydrolysis of the CGs to HCN, and cyanide hydratase, which catalzyses hydrolysis of the cyanide ion to formamide, can reduce CGs by up to 99.3% without the use of steam heating to evaporate the HCN while retaining the beneficial nutrients, lignans and fatty acids at the same levels (Wu et al. 2012). The development of processing methods to be more cost effective and to minimize leaching of nutrients is promising. However, a cultivar bred to have low CGs is likely to be of greater appeal to processors seeking to minimize operational costs and to maintain quality control.

2.4 Accumulation of Cyanogenic Glucosides

There are four types of CGs in flax, consisting of the monoglucosides linamarin and lotaustralin and the diglucosides linustatin (LIN) and neolinustatin (NEO). In flax seeds, linamarin and lotaustralin levels fall to trace amounts during maturation while LIN and NEO levels simultaneously rise suggesting a conversion of monoglucosides to diglucosides (Frehner et al. 1990). However, following germination, the diglucosides are rapidly depleted and only cyanogenic monoglucosides are found in the seedling, leaf, flower, stem and root tissues (Krech and Fieldes 2003; Niedźwiedź-Siegień 1998). In the mature flax seed, CGs constitute approximately 0.1% of the dry seed weight but can reach as much as

5% of dry weight in seedlings. The total amount of CGs in flaxseed from Canadian cultivars typically ranges between 365-550 mg/100g seed (Oomah et al. 1992).

Cyanogenic content is known to vary considerably with seasonal, nutritional and genetic factors, with genotype being the most influential (Oomah et al. 1992; Mazza and Oomah 1995). Additionally, much variability exists in reported HCN content due to differences in extraction and detection methods (Bacala and Barthet 2007). In a study of ten Canadian flax cultivars grown at a single location in Manitoba, Oomah et al. (1992) used high performance liquid chromatography (HPLC) and reported that the main CG compound LIN ranged between 213-352 mg/100g of seed representing 54-76% of the total CG content. The second most abundant CG was NEO at 91-203 mg/100g seed. Linamarin was at trace levels of less than 32 mg/100g seed. Lotaustralin levels in seeds were beneath the limit of detection.

2.5 The Effect of Abiotic Factors on Cyanogenic Glucoside Accumulation

Aside from nitrogen availability, the effect of abiotic factors on CG accumulation appears to be variable across species. In general, plants with high nitrogen supply have an increased CG content. When *Sorghum almum* (Columbus grass) plants were supplied with increased nitrogen, the leaves contained significantly more cyanide per dry leaf weight and leaf total nitrogen (Kriedeman et al. 1964). In *Eucalyptus cladocalyx* (Sugar gum), high nitrogen supply consistently caused both an increase in prunasin concentration in absolute values and relative to leaf nitrogen and dry weight (Gleadow et al. 1998). Nitrate application to

Sorghum bicolor (Sorghum) shoots also increased CG content by as much as seven-fold (Busk and Møller 2002). High concentrations of atmospheric CO_2 are known to decrease the nitrogen pool in the leaf via higher photosynthetic activity. Despite a decreased nitrogen pool however, nitrogen allocation to cyanide is significantly increased when *E. cladocalyx* seedlings are grown at high CO_2 concentration albeit at the expense of protein concentration (Gleadow et al. 1998). CG content generally increases under stressful growth conditions though with variable response curves across species. Two possible reasons for this increase are that there is active transcriptional regulation in direct response to a stress or that there is indirect exaggeration of CG amounts caused by stress-induced delaying of maturity (Busk and Møller 2002; Miller et al. 2014). An *in vitro* study demonstrated directly enhanced CG biosynthesis in cell cultures of *Eschscoltzia californica* (California poppy) when placed under osmotic stress (Hösel et al. 1987). The HCN equivalent in water-stressed *Manihot esculenta* also significantly covaries with the accumulation of abscisic acid, a signaling hormone commonly associated with various plant stress responses, such as temperature, light intensity and drought, suggesting that the increase in CG is not simply an artefact of sampling at an earlier developmental stage (Alves and Setter 2004). CG content in flax increases with temperature, peaking at 30°C (Niedźwiedź-Siegień and Gierasimiuk 2001). In *Trifolium repens* (White clover) on the other hand, CGs accumulate approximately 50% more when grown at 19°C as opposed to 27°C (Hughes 1981). For flax seedlings, white light can double the CG content with increased intensity and time of exposure (Niedźwiedź-Siegień and Gierasimiuk 2001). However, HCN

decreases in *Hevea brasiliensis* (Rubber tree) leaves upon exposure to sunlight (Kongsawadworakul et al. 2009). In flax, water stress reduced CG levels to less than half (Niedźwiedź-Siegień and Gierasimiuk 2001). Mean cyanogenic content was found to be 30% higher in *E. cladocalyx* trees grown in a water stressed location where they also suffered less damage from herbivores (Woodrow et al. 2002).

2.6 Functions of Cyanogenic Glucosides

The most likely primary function of CGs in plants is to deter general herbivory (Nahrstedt, 1985). They are categorised as phytoanticipins because they are constitutively produced and stored in vesicles of plant tissues in preparation for herbivorous attack (VanEtten et al. 1994). Its capacity as a toxin is based on its ability to release toxic HCN and keto compounds upon maceration of host plant tissue (Conn, 1980). HCN is one of the most effective plant toxins. The cyanide anion (CN-) is toxic because it chelates di- and trivalent metal ions of a broad group of metalloenzymes. Cyanide inhibits aerobic respiration by competitively binding to the heme iron of the metalloenzyme cytochrome C oxidase thereby blocking the electron transport chain and resulting in chemical asphyxiation of the affected cells (Donato et al. 2007). Interestingly, the deterrent effect depends primarily on the concentration of the keto compound produced equimolarly with HCN following α-hydroxynitrile degradation rather than the concentration of CGs or HCN *per se* (Jones 1988). The capacity of CGs for herbivory resistance has been demonstrated: dhurrin biosynthesis genes from *S. bicolor* were transferred into non-cyanogenic *Arabidopsis*

thaliana to confer resistance to the flea beetle *Phyllotreta nemorum* (Tattersall et al. 2001). There is much variability in herbivores' reaction to the presence of CGs depending on their concentration in the host plant and the various strategies implemented to reduce the effective toxic exposure. Staying true to what is referred to as the 'chemical arms race', some insects, like *Heliconius sara*, can metabolize cyanogens thereby preventing HCN release and/or sequester the compounds (Engler et al. 2000). These adaptations open up feeding niches with fewer competitors. Some herbivores have developed a tolerance by reducing the effectiveness of CGs as a defense compound. Interestingly, for some CG specialized herbivores, CGs are a phagostimulant, that is, a means of identifying a feed source denied to generalist herbivores. CGs have also been shown to be synthesized and stored by various insects to aid in nitrogen metabolism or for protection against their own predators (reviewed by Gleadow and Woodrow 2002). *Zygaena filipendulae* (six-spot burnet) is able to store linamarin and lotaustralin produced by their feed plant *Lotus corniculatus* (bird's foot trefoil) and to biosynthesize the compounds (Zagrobelny et al. 2007a, b). Interestingly, the biosynthetic pathway in *Z. filipendulae* has identical reaction intermediates to its feed plant and highly similar insect versions of the two cytochrome P450 (CYP) genes, CYP405A2 and CYP332A3, and a uridine diphosphate glycosyltransferase (UGT) gene, UGT33A1 (Jensen et al. 2011).

Cyanide could play a role as a regulatory signal in the biotic stress response to bacterial, fungal and viral infection (Garcia et al. 2013). Mutant *A. thaliana* plants defective in endogenous cyanide detoxification displayed an altered transcriptome associated with

biotic stresses, accumulated reactive oxygen species and, had increased tolerance to biotrophic *Pseudomonas syringae* pv *tomato* DC3000 bacterium and *Beet curly top* virus and had increased susceptibility to necrotrophic *Botrytis cinerea* fungus (Garcia et al. 2013). The results suggest a pathway in which transient accumulation of cyanide during avirulent infection would mimic or induce salicylic acid defense signaling via reactive oxygen species. The expression and relative activity of the enzyme, formamide hydrolase, which can catalyze the detoxification of HCN into formamide, more strongly associates with fungi that are pathogenic to the cyanogenic plants *S, bicolor, L. corniculatus* or *M. esculenta,* compared to non-plant specific fungi or fungi that are pathogenic to non-cyanogenic plants (Fry and Evans 1977). This suggests that the ability to tolerate the HCN produced in cyanogenic plants is strongly selected for as a requirement for successful fungal pathogenesis in these plants.

CGs may function as transporters of reduced nitrogen and glucose at specific developmental stages for rosaceous and euphorbiaceous plants (Andersen et al. 2000; Jørgensen et al. 2005; Forslund and Jonsson 1997; Selmar et al. 1988). Following dehiscence in *Hevea brasiliensis,* it has been proposed that the stored linamarin in the endosperm is glucosylated to form LIN which is then transported upon germination to the seedling to be metabolized into asparagine and aspartic acid as an amino acid supply (Selmar et al. 1988). A similar role has been identified in *Prunus amygdalus* (almond) in which the monoglucoside prunasin is *de novo* synthesized in the tegument of the developing fruit and then transported to the developing cotyledons to be converted into the

diglucoside amygdalin (Sánchez-Pérez et al. 2008). CGs have also been suggested to act as a buffering nitrogen source for latex production in *H. brasiliensis* based on a positive correlation between leaf CG content and latex yield *(*Kongsawadworakul et al. 2009). In flax, the only tissue that has been found to contain diglucosides is the maturing seed. Therefore, they are not likely to be nitrogen transporters as was suggested for *H. brasiliensis*.

The free cyanide released from CGs may also act as a cell signal to regulate embryonal dormancy and the promotion of seed germination (Bogatek et al. 1991; Taylorson and Hendricks 1973). Although HCN is normally produced as a co-product of ethylene biosynthesis, in *Xanthium pennsylvaniicum* (Cocklebur) seeds, there is strong evidence that a small portion of the CGs amygdalin, linamarin and prunasin are sources of endogenously evolved HCN used to regulate germination (Esashi et al. 1991b; Yip and Yang 1988). The evolution of HCN during early periods of water imbibition to stimulate germination is a common mechanism in both cyanogenic and acyanogenic seeds (Esashi et al. 1991a). The induction of dormancy alleviation and embryo germination by HCN requires oxidative signaling via reactive oxygen species (Oracz et al. 2009). Since active ethylene receptors are required for induction of germination by HCN, there is likely considerable cross-talk between the two signaling pathways to coordinate germination (Gniazdowska et al. 2010; Oracz et al. 2008). Following germination in flaxseed, the diglucosides are hydrolyzed without an accompanying rise in monoglucosides until *de novo* synthesis of the monoglucosides after 36 hours (Krech and Fieldes 2003). A sharper

decrease of the diglucosides is also associated with better emergence and faster hypocotyl growth (Krech and Fieldes 2003). These observations hint at a role for CGs in the regulation of flax germination rate, possibly through transient HCN and ethylene signalling.

2.7 Evolution of Cyanogenic Glucoside Biosynthesis Pathway

There are more than 75 CGs known to date occurring in at least 2500 species of higher plants, representing 30 families of ferns, gymnosperms and both monocot and dicot angiosperms, implying that the cyanogenesis ability in plants is at least 300 million years old, when the clades are believed to have diverged (Poulton, 1990; Conn, 1994). They can be either aromatic, aliphatic or cyclopentenoid depending on the structure of the amino acid precursor. The expression of tyrosine or phenylalanine derived aromatic CGs in ferns and gymnosperms was likely the ancestral form that served as precursor for the aliphatic and aromatic forms derived from other amino acids found in angiosperms (Jaroszewskiet al. 2002). This hypothesis is supported by the occurrence of tyrosine derived CGs in Magnoliales, an ancestral angiosperm. The catalytic steps and the organization of the enzymes in the CG biosynthetic pathway are consistent across species but are not necessarily orthologous considering that there is low amino acid sequence identity of the respective enzymes (Takos et al. 2011). The CG biosynthetic pathway likely evolved independently in higher plant lineages by repeated duplication, recruitment and neofunctionalization of related primary metabolism genes (Takos et al. 2011).

2.8 Biosynthesis of Cyanogenic Glucosides

All CGs share a similar α hydroxynitrile derived O-β-glycoside chemical structure (Vetter 2000). Most are monosaccharides consisting of a single sugar residue that stabilizes a cyanohydrin moiety by glycosidic linkage. Alternatively, the disaccharide structure consists of two stabilizing sugar residues. They are all derived from one of six amino acids: L-valine, L-isoleucine, L-leucine, L-phenylalanine, L-tyrosine or a non-protein amino acid cyclopentenyl-glycine. L-valine is the precursor for linamarin and LIN while L-isoleucine is the precursor for lotaustralin and NEO.

The biosynthetic pathway consists of three main steps (Figure 2.1). Initially, a multifunctional cytochrome P450 (CYP) enzyme catalyzes a conversion from the precursor amino acid to a Z-aldoxime through two successive amino (N)-hydroxylations, followed by dehydration, decarboxylation and isomerization steps (Sibbesen et al. 1994). Next, another multifunctional cytochrome P450 catalyzes the conversion of the Z-aldoxime to a labile α-hydroxynitrile (cyanohydrin) through dehydration to a nitrile and subsequent (C)-hydroxylation of the α carbon (Kahn et al. 1997). Finally, a UGT catalyzes the glycosylation of the cyanohydrin moiety to generate the stabilized CG (Jones et al. 1999). Following biosynthesis, CGs are thought to be actively transported from the endoplasmic reticulum (ER) to the vacuole by a transporter not yet identified in plants.

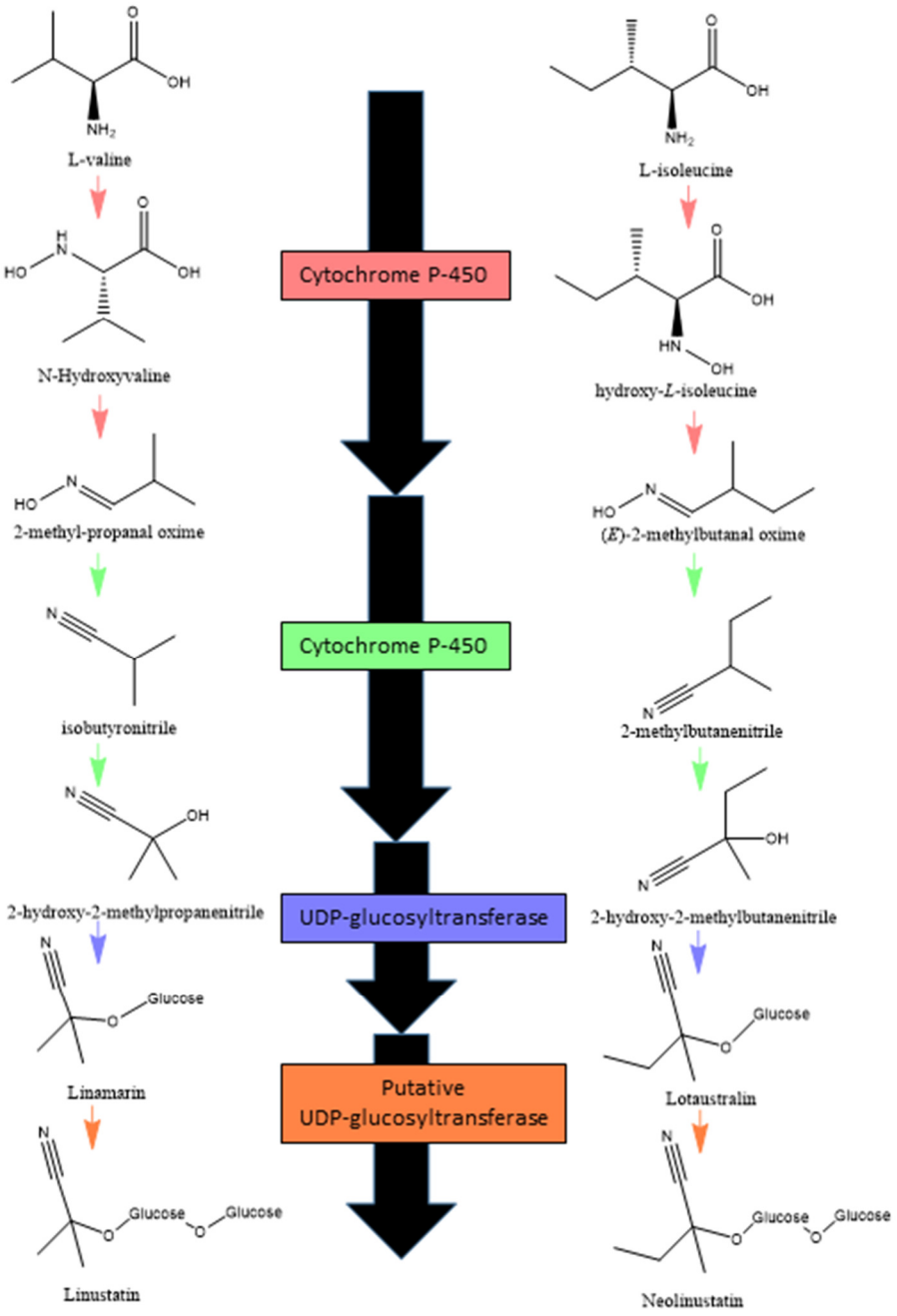

Figure 2.1 Schematic of the *in vivo* biosynthetic pathway of flax cyanogenic glucosides. Figure modified from Cutler and Conn (1981)

These three essential enzymes are organized in *S. bicolor*, and likely in other plants, as a multi-enzyme complex called a metabolon for efficient channeling (Nielsen et al. 2008, Møller and Conn, 1980). In *S. bicolor*, the CG dhurrin is catalyzed by two membrane bound CYP enzymes, CYP79A1 and CYP71E1, and a soluble UGT named UGT85B1 (Sibbesen et al. 1994, Bak et al. 1998, Jones et al. 1999). Transformation of these three biosynthetic genes from *S. bicolor* enabled production of dhurrin in normally acyanogenic *A.thaliana*. Coexpression of the two CYP enzymes as fluorescent fusion proteins in the transgenic *A. thaliana* plants prevented dhurrin formation although each enzyme retained its activity when expressed individually as fusion proteins, suggesting that a tight interaction between the enzymes is necessary for dhurrin formation but was prevented here by the stoichiometric hindrance of the fusion tags (Kristensen et al. 2005). Additionally, UGT85B1 localized in the cytoplasm in the absence of CYP79A1 or CYP71E1 but shifted toward the ER membrane surface of dhurrin producing cells when CYP79A1 and CYP71E1 were present. Active microsomal preparations for flax indicate that the biosynthetic scheme for linamarin and lotaustralin is likely similar to that for *S. bicolor* (Cutler et al. 1985).

Although CGs *per se* are non-toxic, as defense compounds that have the potential to hydrolyze and produce HCN, it is necessary for the plant to protect itself during their biosynthesis and storage. During biosynthesis, this protection is accomplished by the use of a metabolon that functions to increase catalytic efficiency by keeping the relevant active sites close to each other and to prevent the release of toxic intermediates (Zagrobelny et al.

2008). During storage, the common strategy of vacuole sub-localization prevents activation by β-glucosidases located in either the apoplast or cytoplasmic plasmids depending on the species (Saunders and Conn 1978; Nikus et al. 2001). Herbivore feeding causes a disruption of this compartmentalisation by membrane rupture. The aglycone moiety is cleaved and the released cyanohydrin can then be converted into HCN, sugar and a keto compound spontaneously or with greater efficiency by the activity of a hydroxynitrile lyase (Vetter 2000).

Linamarin and lotaustralin are likely biosynthesized by a single group of enzymes. Two observations give this theory credence. First, the ratio of glucosides formed from acetone cyanohydrin and butanone cyanohydrin following glucosylation remains constant. This indicates that the glucosylation of both acetone cyanohydrin and butanone cyanohydrin is catalyzed by the same glucosyl transferase (Hahlbrock and Conn 1971). Second, competition experiments with valine and isoleucine show HCN production rates that exclude two independent systems (Cutler et al. 1985).

The ability to synthesize CGs likely developed independently in plants and arthropods by convergent evolution, rather than horizontal gene transfer or divergent evolution (Jensen et al. 2011). The striking similarity between the CG pathways for plants and arthropods is highly unusual and unique. Both consist of two CYPs and a UGT that perform conversion from a precursor amino acid to the same intermediates, as demonstrated by radiolabeling experiments. The similarities between plant and arthropod CG pathways are likely due to an optimal organization of enzymes to efficiently produce

CGs while avoiding toxic exposure to intermediates. The evolution of a metabolon in both systems indicates the necessity of efficient channelling and rapid glucosylation to stabilize toxic intermediates and prevent HCN production in the cell. However, the organization of the metabolon is thought to be different in arthropods with the catalytic domain of the UGT facing the inside of the ER lumen rather than the cytosol as is the case in plants (Jensen et al. 2011).

2.9 Degradation of Cyanogenic Glucosides

In plants and herbivores, CGs undergo hydrolysis when in contact with β-glucosidases and α-hydroxynitrilases to produce HCN (Vetter 2000, Zagrobelny et al. 2004). Cyanogenic disaccharides may degrade via two distinct pathways (Fan and Conn, 1985). The 'simultaneous' mechanism involves hydrolysis at the aglycone-disaccharide bond to release an α-hydroxynitrile and a disaccharide as demonstrated in *Davallia trichomanoides*. Alternatively, the 'sequential' mechanism involves a stepwise hydrolysis of the two sugar residues with a cyanogenic monosaccharide intermediate by the same or different β-glycosidases. In *L. usitatissimum,* linustatinase and linamarase perform this function. Linustatinase hydrolyses β-(-bis-1,6) and β-(-bis-1,3) glucosides but is inactive toward the cyanogenic monosaccharides and cyanogenic disaccharides containing terminal xylose or arabinose moieties. This enzyme catalyzes the degradation of both LIN and NEO to linamarin and lotaustralin, respectively. Linamarase then deglucosylates linamarin and lotaustralin to acetone cyanohydrin and (2R)-2-hydroxy-2-methylbutanenitrile,

respectively. At neutral or mildly alkaline conditions, the α-hydroxynitrile compounds can spontaneously dissociate to form HCN, a sugar and a keto compound (Poulton 1990). In more acidic environments, further degradation by an α-hydroxynitrilase cleaves off acetone or butan-2-1, respectively, to give the same products.

2.10 Detoxification of Hydrocyanic Acid

For most animals, an HCN concentration of about 2 mg kg^{-1} is a lethal dose (Kingsbury 1964). For humans, the lethal dose is between 0.35 and 0.5 mg kg^{-1}(Kingsbury, 1964). However, at dosages under the toxic limit, cyanide can be effectively detoxified and metabolized via two distinct pathways. In plants and some lepidopteran species, this is mainly accomplished by β-cyanoalanine synthase and to a lesser extent thiosulfate-cyanide sulfur transferase, otherwise known as rhodanese and mercaptopyruvate sulfurtransferase. HCN can react with cysteine and be converted into β-cyanoalanine which is then converted into asparagine (Miller and Conn 1980). In animals and a few plants and insects, rhodanese catalyzes the majority of the detoxification by reacting cyanide and thiosulfate to produce sulfite and thiocyanate which can be excreted in the urine (Bordo and Bork 2002; Lang, 1933).

2.11 Regulation of Cyanogenic Glucoside Activity

In *S. bicolor,* dhurrin levels are maintained by the transcriptional regulation of enzymes involved in its synthesis and breakdown mechanisms (Bough and Gander 1971; Adewusi

1990). A strong correlation between activity of the biosynthetic enzymes CYP79A1 and CYP71E1 and their mRNA levels indicates that the transcriptional regulation of biosynthetic enzymes is likely the strongest determinant of dhurrin accumulation (Busk and Møller 2002). CYP79A1 coordinates the first dedicated step in CG biosynthesis and is the rate limiting step for dhurrin production (Sibbesen et al. 1995). Expression profiling in leaves of *Manihot esculenta* (Cassava) showed a pattern of increased expression of CYP79D1/D2 (biosynthesis), linamarase (catabolism), α-hydroxynitrile lyase (catabolism) and β-cyanoalanine synthase (reassimilation) with higher CG content (Echeverry-Solarte et al. 2013). Methyl jasmonate, a signaling molecule associated with wounding, has been found to induce *de novo* synthesis of β-glycosidases in *Medicago truncatula* upon tissue damage suggesting that the cyanogenic defense response is not only consistently active but can also be induced (Naoumkina et al. 2007). In *Hevea brasiliensis*, cyanogenic capacity can also be regulated at the post-translational level for degradation enzymes (Kadow et al. 2012). In directly affected leaf areas, the activity of linamarase and hydroxynitrile lyase can increase by up to ten-fold permitting a more immediate and local response to herbivorous attacks. However, linamarase activation was not observed in flax to a comparable degree (Kadow et al. 2012).

2.12 Cyanogenic Glucoside Biosynthetic Gene Organization

The biosynthetic genes for CGs tend to co-localize in clusters for *Lotus japonicus* (Birdsfoot trefoil), *M. esculenta* and *S. bicolor*, and likely for other cyanogenic plant

species (Figure 2.2) (Takos et al. 2011). Aside from the presence of the CG biosynthetic genes, the genomic regions for the three genomes show no other similarity and vary considerably in size and structure. The biosynthetic genes, consisting of the two CYPs and one UGT, are located within 160 kb in *L. japonicus*, 83 kb in *M. esculenta* and 104 kb in *S. bicolor*. The structural diversity across the regions suggests independent origins. Although the respective biosynthetic enzymes are related, they are not necessarily orthologous. Sequence alignment analysis indicated identity to varying degrees. The *L. japonicus* UGT85K3 shows 41% amino acid identity with the *S. bicolor* UGT85B1 and 56% identity with the *M. esculenta* UGT85K4 (Takos et al. 2011). The *S. bicolor* CYP71E1 shares 49% protein identity with its functional homologue CYP71E7 in *M.esculenta* but only 35% with CYP736A2 in *L. japonicas* (Jørgensen et al. 2011; Takos et al. 2011). The *M.esculenta* CYP79D1 shares approximately 55% identity with both the *S. bicolor* CYP79A1 and *L. japonicas* CYP79D3 (Andersen et al.2000; Forslund et al. 2004). The relatively low sequence identity along with the occurrence of different sub-families for each enzyme across species suggest that the CG biosynthetic pathway may have evolved independently in higher plant lineages by repeated recruitment of members from similar gene families.

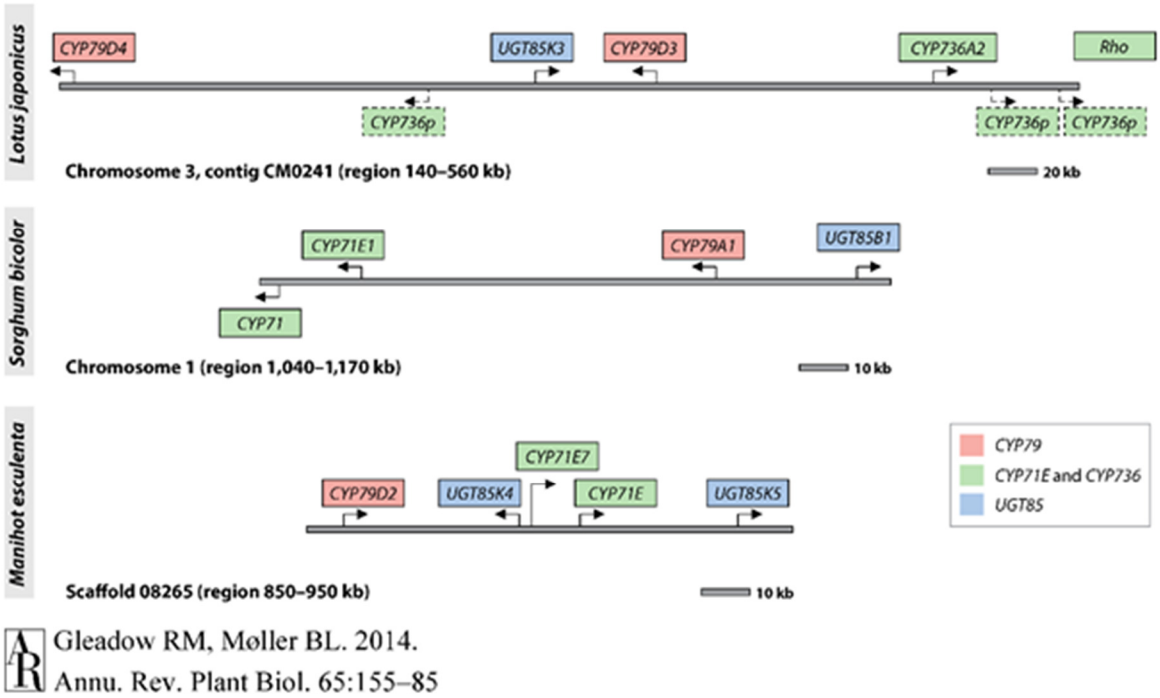

Figure 2.2 Genes for cyanogenic glucoside biosynthetic enzymes cluster in *L. japonicus, S. bicolor* and *M. esculenta.* Figure reused from Gleadow and Møller (2014).

The occurrence of genetic and physical clustering of non-homologous genes involved in the same pathway is rare. However, it may be an intrinsic characteristic of defense compounds because independent recruitment also occurs for the initial steps of thalianol and avenacin biosynthesis pathways (Field and Osbourn 2008). Selective advantages for clustering are likely based on co-expression that would allow coordinated transcriptional regulation and co-inheritance, which in turn would serve to maintain a functional pathway (Chu et al. 2011b). The occurrence of highly toxic intermediates and the necessity of maintaining a highly conserved metabolon structure provide a plausible explanation.

Disrupted control of the component genes by recombination or aberrant transcription would be selected against by the release of harmful toxins, thus the strong selection pressure to maintain a complete functional unit as demonstrated for the dhurrin biosynthetic pathway in *A. thaliana* (Kristensen et al. 2005). Transgenic *A. thaliana* plants with *S. bicolor* genes CYP79A1 and CYP71E1 were stunted, had transcriptome alterations, accumulated numerous intermediate detoxification by-products and were unable to produce dhurrin and several brassicaceae-specific UV protectants. These pleiotropic effects were not observed when all three *S. bicolor* biosynthetic genes were transformed into *A. thaliana*.

2.13 Genetic Engineering of Cyanogenic Glucosides

Transgenic plants have been engineered with reduced or enhanced CG levels utilizing cloned genes of the CG biosynthetic pathway. The main objectives of these projects were the alteration of CGs in crop plants to reduce their toxic potential for consumption or to enhance or create CGs for the purpose of herbivore deterrence. RNA interference blocking the expression of CYP79D1/D2 has been used to develop transgenic *M. esculenta* with over 99% reduction in CG content stored in leaves and a 92% reduction in tubers (Jørgensen et al. 2005). However, lines transformed with the RNAi construct that had less than 25% of the average wild type CG levels had a frailer morphology with fewer leaves and grew more slowly indicating possible physiological functions of CGs.

Transgenic *M. esculenta* with 60-94% reduction in CG levels in the leaves were developed using a leaf-specific antisense expression of CYP79D1/D2 gene fragment (Siritunga and Sayre 2004). As mentioned previously, the CYP79D1/D2 enzymes are responsible for the first step of the biosynthesis of linamarin in *M. esculenta*. Interestingly, CGs were reduced by more than 99% in the roots of these transgenic plants. Anti-sense expression of CYP79D1/D2 under a root specific promoter produced normal root linamarin levels. Taken together, these two results indicate that linamarin synthesized in leaves is transported to the roots in *M. esculenta*.

Transgenic *A. thaliana* and *L. japonicus* plants have been developed with genes involved in the dhurrin biosynthetic pathway from *S. bicolor*. These heterologous systems could successfully produce and store dhurrin (Tattersall et al. 2001; Morant et al. 2003, 2007; Kristensen et al. 2005). The presence of accumulated dhurrin in transgenic *A. thaliana* plants conferred resistance to the flea beetle *Phyllotreta nemorum*, a natural pest for other cruciferous plants, without any observable pleiotropic effects (Tattersall et al. 2001). In bioassays using flea beetle larvae, significantly fewer leaf mines were initiated on transgenic plants and, for the larvae that did initiate mines, a higher mortality rate was observed on leaves containing dhurrin (Tattersall et al. 2001). There was also 80% less leaf tissue damage by the larvae in the dhurrin producing *A. thaliana* plants than for the control plants.

2.14 Analytical Methods of Cyanogenic Glucosides

The purpose of this section is to describe the analytical methods used for the detection of CGs in food samples, highlighting the strategy behind them and their practical applications. Analytical methods typically fall into two general categories: methods that directly measure the intact CGs and those that indirectly measure cyanogen content, usually through liberated HCN. Indirect methods generally require extraction of the total CGs and cyanohydrins, followed by hydrolysis, by either acid or enzymes, to produce total HCN which is then quantified using a chromatography method or a colorimetric assay. Direct methods do not have a hydrolysis step but may require derivitization following extraction followed by a chromatogaphy method to distinguish the individual CGs. The following is not an exhaustive list of analytical methods but represents the more well-established methods found in the literature that are often used as standard procedures.

2.15 Extraction

A wide variety of extraction methodologies exists across the literature and the choice thereof has been shown to affect extraction efficiency by up to 18% (Bacala and Barthet 2007). The extraction method must be accurate and repeatable because this directly limits the usefulness of the analytical method. Therefore, it is necessary to optimize any CG extraction method. Optimized reference and routine methods have been developed for individual CG extraction in flaxseed (Barthet and Bacala 2010). The reference method involves grinding using a high speed impact plus sieving mill at 18,000 rpm with a 1.0 mm sieve with triple pooled extraction in a sonicating water bath (40˚C, 30 min) using 75%

methanol. The routine method, approximately 88% as efficient as the reference method, differs only by the use of a coffee grinder to pulverize the samples and a single extraction with 75% methanol. The most common extractant for methods measuring total HCN is 0.02-0.2 M phosphoric acid, though other acids such as hydrochloric or sulfuric can also be used (European Commitee for Standardization, 2012). Regardless of the acid used, it must be at a low concentration and used at low extraction temperatures (<26°C) to prevent hydrolysis and HCN loss. The acidic conditions inactivate β-glucosidase activity that would hydrolyze the CGs (Bradbury et al. 1994).

2.16 Acid Hydrolysis

As its name suggests, the acid hydrolysis method refers to the hydrolysis of CGs using a mineral acid to directly produce HCN and a keto compound. Haque and Bradbury (2002) measured cyanide content for various CG substrates in 2M sulfuric acid at 100˚C in a glass stoppered test tube. Maximum HCN recovery was observed at 75 min for flaxseed meal. This method can be generally applied to the determination of cyanide content in any plant since the performance of the assay does not depend on the amount or the type of the CG or β-glucosidase. It is also comparatively inexpensive. However, this method suffers from difficulty of use, low accuracy and repeatability and, more importantly, has a tendency to underestimate HCN content. Clear evidence for the latter has not been established but HCN leakage, hydrolysis of cyanide to ammonia and formic acid and, oxidation to isocyanate have been hypothesized. Incubation times of up to six hours have been used as part of a

method to correct for cyanide losses (Haque and Bradbury 2002). Corrections for HCN leakage require that measurements be taken at different heating times and extrapolation to zero time be performed.

2.17 Enzymatic Hydrolysis

Enzymatic hydrolysis involves the use of enzymes to hydrolyze the CGs to obtain an indirect measurement from HCN liberation. In flax, β-glucosidases, linustatinase, linamarase and hydroxynitrile lyase are needed for the decomposition of all CGs to cyanide and keto compounds. When 28 flaxseed samples were hydrolyzed by crude enzyme preparations from flaxseed and subsequently measured for total cyanide by two colorimetric methods and HPLC, the values obtained were not statistically different (Kobaisy et al. 1996). The minimum amount of enzyme required for complete recovery of linamarin, LIN and NEO as cyanide was 1.5, 3 and 4 mg mL^{-1}, respectively. The standards indicated recoveries of 100% cyanide from linamarin and 97% from LIN and NEO. The conditions for hydrolysis can vary greatly and can be influenced by the genotype, location, stage of development, seed soundness and enzyme titer. Most methods use endogenous flax enzymes, also called autolysis, whether native or prepared separately, because of their high substrate specificity. However, this is not generally recommended due to considerable variation in β-glucosidase activity across cultivars. Yamashita et al. (2007) demonstrated that 5% or 10% raw ground flaxseed can be added as an enzyme source to ensure exhaustive hydrolysis provided that the cyanide content of the added flax is corrected. The

enzymatic hydrolysis method is generally accepted as a more accurate option to acid hydrolysis.

2.18 Measurement of Cyanide

The most common method of measuring total HCN is by spectrophotometry following reaction of the cyanide ion to generate a stable chromophore. The generation of the chromophore is usually accomplished by a König reaction, a Guignard reaction, a Feigl-Anger method or by use of nitrobenzaldehyde. Alternatively, total HCN is also commonly measured by titrimetric methods or chromatography.

2.19 König method

The basis of the König reaction is the conversion of the cyanide content into a light absorbing Schiff base that can be spectrophotometrically measured. Cyanide (CN^-) is first oxidized to a cyanogen halide using an oxidizing agent such as N-chlorosuccinimide. The halide then reacts with pyridine to produce glutaconic acid. A coupling reagent, such as pyridine-barbituric acid, is then used to convert glutaconic acid to produce a light absorbing Schiff base. Absorbance is measured with a spectrophotometer at 600 nm and the concentration is obtained by comparison with a calibration curve created using a standard KCN solution. Variation among choice of oxidizing, coupling reagents and assay conditions can affect the outcome. A simplified method replacing pyridine-barbituric acid with isonicotinic/barbituric acid as a color reagent was reported to be more effective

(Bradbury et al. 1994). Yamashita et al. (2007) proposed using a chloramine-T oxidizing agent and pyridine-pyrazolone coupling reagent König based method as part of a simple and convenient method that can be applied industrially.

A study comparing cyanogen content measurements using the pyridine-barbiturate colorimetric method and measurements using HPLC revealed no significant differences (Kobaisy et al. 1996). König-based spectrophotometric assays are well documented and accurate. Mean recovery is estimated to be approximately 98%±1% standard deviation (Harris et al. 1980).

2.20 Guignard method

The Guignard method, also called the picrate method, involves reaction of the cyanide ion with alkaline picrate. It was first developed as a qualitative indicator of gaseous HCN using picric acid soaked chromatography papers suspended above hydrolyzed CGs in a stoppered test tube. Exposure to cyanide reduces the picric acid to isopurpuric acid with a characteristic red-brown color on the paper. However, the Guignard reaction can cause the characteristic color change with non-cyanide carbonyl compounds as well. The interfering carbonyls are usually trapped in acidic 2,4-dinitrophenylhydrazine and HCN is volatilized in a gas stream (Alonso-Amelot and Oliveros 2000). A quantitative modification of the design involves eluting the picrate paper in water after cyanide exposure followed by the spectrophotometric measurement of the absorbance at 510 nm against a blank (Haque and Bradbury 2002). Recovery was reported to be 101.9% ± 0.64%. Bradbury (2009) increased

the sensitivity ten times with a modified design involving an elution in water. The method, simple to use, reasonably accurate and reproducible, is currently available in kit form for *M. esculenta*, *S. bicolor*and flax for simple field tests. Compared to König-based methods, picrate methods are considered less sensitive and accurate. Another drawback is the hazards associated with the handling of picric acid as it can be explosive.

2.21 Feigl Anger method

This method involves the use of a filter paper that has been soaked in a solution of copper (II)-ethylacetoacetate and di-(4-dimethylaminophenyl)-methane then dried (Feigl and Anger 1966). In the presence of cyanide, the paper turns blue with an intensity that is proportional to the cyanide concentration. This test is semi-quantitative at best. A modified high throughput design using microtiter plates has been used to quickly identify *L. japonicus* ethyl methyl sulfonate mutants that were defective in HCN release (Takos et al. 2010). Apical leaves were placed in wells of a 96-well microtiter plate and frozen overnight at -80˚C to disrupt the tissue and permit HCN release. The cyanide-sensitive Feigl Anger paper was placed over the wells and tightly sealed. The blue intensities were then correlated with the amount of HCN produced.

2.22 Reaction with Nitrobenzaldehyde

By reaction of the cyanide ion with *p*-nitrobenzaldehyde, an intermediate cyanohydrin is generated. This cyanohydrin reacts with *o*-dinitrobenzene to produce a purple dianion of *o*

-nitrophenyl hydroxylamine (Guilbault and Kramer 1966). Under optimum conditions for analysis, as little as 5×10^{-6} M cyanide can be detected. A simple test kit has been developed capable of detecting cyanide as low as 0.01 mg L^{-1} within 10 min (Untang et al. 2010).

2.23 Titration

The titration strategy used to determine cyanide concentration is to create either an acidic or alkaline distillate of HCN and then titrate it against a suitable acid or base using an indicator for the end point. The concentration of cyanide is determined based on the product of the titre value, titrant concentration and the molar ratio based on a balanced chemical equation corrected according to the weight of the samples used. Steam distillation is used on the hydrolysis mixtures and the HCN released is collected into an alcoholic sodium hydroxide solution (alkaline) or dilute acidified silver nitrate (acidic). The distillate can then be titrated against silver nitrate (known as the Volhard method) for the former or potassium thiocyanate for the latter. The alkaline method has been used to measure cyanide content for different plant parts of *M. esculenta* species (Etonihu et al. 2011). The acid method has been used to measure cyanide content of *Tacca leontopetaloides* tuber peels (Ubwa et al. 2011). Both of these methods have been criticized for lacking in specificity and sensitivity. Prior to 2012, the acid titration method was the European Union's official method for determination of acceptable levels of HCN in flaxseed derived feeding-stuffs (European Commission 1971).

2.24 Chromatography

A high performance liquid chromatography (HPLC) method for determination of total HCN has also been developed (Chadha et al. 1995). The method involves injecting the filtrate from autohydrolysed flaxseed into an HPLC system with an anion exchange column and an amperometric and/or oxidation detector. The homogenate is periodically analyzed until a maximum value is reached, typically after about three hours. The European Committee for Standardization has recently published its standard for HCN determination in animal feed by reverse phase HPLC of taurine/2,3-naphthalene dicarboxy aldehyde derivatised HCN (European Committee for Standardization, 2012). Here, the cyanide ion reacted with bromine produces cyanogen bromide in an acidic medium that is subsequently extracted into benzene. This benzene extract is then injected into a GC, separated along a column and the amount of cyanide bromide is detected by use of an electron capture detector. The cyanogen bromide peaks are integrated and cyanide amount determined based on a linear calibration. The mean recovery of cyanide when added to animal food stuff at concentrations of 10 and 20 mg Kg^{-1} was 98.2% ± 0.98% standard deviation with a limit of detection of 1 mg Kg^{-1} cyanide (Harris et al. 1980).

2.25 Direct Measurement of Intact Cyanogenic Glucosides by Gas Chromatography

Following extraction, derivatization is necessary because CGs are not volatile. Bacala and Barthet (2007) have developed a method for GC analysis for all four CGs in flax using trimethylsilyl ether derivatives. Following extraction, samples are dessicated using an inert

gas such as nitrogen or helium and anhydrous ethanol. A derivatizing cocktail consisting of 1-methylimidazole, N,O-bis-(trimethylsilyl) acetamideand trimethylsilyl chloride is then added and incubated prior to loading of the samples on the GC. The GC method benefits from high repeatability (>99%) and a superior detection limit and accuracy compared to other methods. GC and HPLC methods are generally preferred for direct quantitation due to their high resolving power and automation. GC has higher sensitivity, better detection and quantification limits, higher throughput since freeze drying is not necessary, and it is more economical with regards to standard usage.

2.26 Direct Measurement of Intact Cyanogenic Glucosides by High Performance Liquid Chromatography

The objective of this technique is to separate the CGs based on adsorption to hydrophobic stationary phases such as silica (straight phase) or octadecyl modified silica (reversed phase). The basic HPLC method for separating CGs in flax involves a C18 reverse phase column with detection by refractive index, evaporative light scattering or mass spectrometry. Varied ratios of mobile phase components are reported across the literature. The most common modifiers are methanol, acetonitrile and 2-propanol. All four CGs can be distinguished using a mobile phase of 5-10% methanol: 90-95% water: 0.05% phosphoric acid. Using a mobile phase of MeOH-H_2O-HOAc (6:93:0.05), Krech and Fieldes (2003) were able to resolve lotaustralin, linamarin and LIN and, using a 7:92:0.05

phase, they resolved NEO. The main drawback of the HPLC method is its detection and quantification limits for all CGs.

2.27 Direct Measurement of Intact Cyanogenic Glucosides by High Performance Thin-Layer Chromatography (HPTLC)

In contrast to the GC and HPLC, HPTLC uses a flat stationary phase for separation. Niedźwiedź-Siegień (1998) was able to resolve all four CGs in flax seeds during germination, in whole plants during development and in different organs. Plates of cellulose MN300 were used as the stationary phase and chromatograms were developed in a water saturated *n*-butanol mobile phase. At optimal performance, HPTLC can produce results that are comparable to the accuracy and reliability of HPLC and GC. The minimum amount detectable was reported to be 10 nmol. HPTLC benefits from being simple to operate and cost effective since many samples can be resolved on a single plate with low solvent usage. However, this method cannot distinguish between epimers and may be tedious for large numbers of samples. Although direct quantification using chromatographic methods are more accurate and sensitive, these methods are time consuming and prohibitively expensive for developing countries, small scale operations and routine application.

3.0 IDENTIFICATION OF QTL ASSOCIATED WITH THE ACCUMULATION OF LINUSTATIN AND NEOLINUSTATIN IN THE MATURE FLAX SEED

3.1 Abstract

A genetic linkage map of flax (*Linum usitatissimum* L.) was constructed for a recombinant inbred line population that segregated for cyanogenic content in mature seeds. The population consisted of 174 individuals and was produced by single seed descent from a cross between the line Double Low and the cultivar AC McDuff. The population was grown in the field at three site-years and linustatin (LIN) and neolinustatin (NEO) contents were measured by gas chromatography and near infrared spectroscopy. Seeds from Double Low contained a hydrocyanic acid equivalent of 219 ± 65 ppm while seeds from AC McDuff contained 293 ± 35 ppm over the three environments. The 15 linkage groups (LGs) of the genetic map included 155 simple sequence repeat markers, 39 single nucleotide polymorphism markers and one gene spanning 938 cM with an average marker interval of 5 cM. Initial QTL mapping identified a 16.5 cM region on LG1 associated with linustatin content and a 22.5 cM region associated with neolinustatin content on LG6. Fine mapping using 82 recombinant lines shortened the linustatin associated region to 11.7 cM and the neolinustatin QTL to two regions of 3.5 cM and 15.2 cM. On the physical map, the LIN

QTL was anchored to three scaffolds and the two NEO QTL anchored to two and one scaffolds, respectively. Candidate genes based on function prediction of transcripts parsed from these scaffolds were hypothesized.

3.2 Introduction

Flax (*Linum usitatissimum* L.) is an economically valuable oilseed crop because of its unique nutritional profile, low cost of production and usefulness in crop rotation. Aside from its multiple industrial applications, its value as a functional food and feed product is due primarily to its high level of alpha linolenic acid (ALA), which has been shown to be key to infant brain development, to reducing blood lipid levels and in the prevention and treatment of cardiovascular diseases (reviewed by Ruxton et al. 2004). However, flax seeds also contain anti-nutritional compounds such as cyanogenic glucosides (CGs) and linatin. The accumulation of CGs in the mature flax seed may restrict its usage in food products and animal feed by deterring livestock growth (Rodríguez et al. 2001). CG content also constitutes a trade barrier for certain countries, such as the EU's maximum allowed limit of 250 ppm hydrocyanic acid (HCN) equivalent in imported linseed (European Parliament and Council 2002). Many of Canada's flax varieties approach or exceed this limit (Oomah et al. 1992; Mazza and Oomah 1995). Reducing CGs in flax seed is an important goal in linseed breeding programs. Although yield, disease resistance and fatty acid composition will likely remain the key strategic priorities for making linseed a viable economic option, CG accumulation in current cultivars is a liability should enforcement of EU trade restrictions become stricter or should other markets adopt a similar policy. Understanding the genetic complexity of the trait, the environmental influence and the genetic by environment interactions are key to designing breeding strategies for reducing CGs in linseed. Such strategies include the design of crosses, selection strategies, probability of a

desirable recombination among loci of interest after crossing (linkage drag) and selection, and population size determination. Mutant acyanogenic varieties that would be unable to produce functional biosynthetic enzymes would be ideal but no such mutants have yet been found in germplasm collections or segregating populations, likely because either the recessive form is lethal or the presence of redundant functional gene copies. Markers associated with CG accumulation would also be useful to breeders for introgression of the low CG trait in breeding populations.

CGs are α-hydroxynitrile derived β-glucosides with a cyanohydrin moiety linked to a stabilizing glycosidic residue. In plants, they function as deterrents of general herbivory, intermediates of nitrogen metabolism and precursors for germination signalling (Nahrstedt 1985; Selmar et al. 1988; Esashi et al. 1991b). Hydrolytic cleavage of the glycosidic linkage by a β-glucosidase and subsequent degradation of the resulting cyanohydrin produce HCN (Cutler and Conn 1981). Maceration of tissue membranes, which occurs during feeding, enables this hydrolysis and, acute cyanide poisoning may occur depending on the effective HCN concentration. There are four types of CGs in flax: the monoglucosides linamarin and lotaustralin and the diglucosides linustatin (LIN) and neolinustatin (NEO). During seed maturation in flax, linamarin and lotaustralin levels fall to trace levels while LIN and NEO levels rise (Frehner et al. 1990).

Cyanogenic plants follow a common CG biosynthetic pathway consisting of three main steps. First, a multifunctional cytochrome P450 catalyzes the conversion from the precursor amino acid to a Z-aldoxime (Sibbesen et al. 1994). Next, another multifunctional

cytochrome P450 catalyzes the conversion of the Z-aldoxime to a labile α-hydroxynitrile (Kahn et al. 1997). Finally, a UDP-glycosyl transferase catalyzes the glycosylation of the cyanohydrin moiety to generate the stabilized CG (Jones et al. 1999). In *Sorghum bicolor*, these three essential enzymes form a multi-enzyme complex called a metabolon for efficient channeling (Nielsen et al. 2008). CG accumulation in *S. bicolor* is also monitored by transcriptional regulation of both biosynthesis and degradation enzymes (Bough and Gander 1971; Adewusi 1990).

The breakdown of CGs in flax is the result of the stepwise hydrolysis by linustatinase of LIN and NEO to linamarin and lotaustralin, respectively. Linamarase then deglucosylates the monoglucosides to cyanohydrin that subsequently dissociates to HCN, a sugar and a keto compound (Poulton 1990).

Mapping of the quantitative trait loci (QTL) controlling CG content offers the opportunity to understand the genetics of this trait and to identify closely associated markers which are beneficial for designing an efficient breeding strategy. Here, we mapped a population segregating for CG content using simple sequence repeat (SSR) and single nucleotide polymorphism (SNP) markers and phenotyped it by gas chromatography (GC) to identify QTL for LIN and NEO content in linseed that were subsequently fine-mapped by screening 864 F_2 plants.

3.3 Materials and Methods

Plant Materials and DNA isolation

An F_7-derived recombinant inbred line (RIL) population of 174 individuals was developed by single seed descent of a cross between a line known as Double Low (DL) and the cultivar AC McDuff (MD) (Kenashuk and Rashid 1994). Plants were grown in a growth cabinet under 16-hour light at 18°C and 8-hour darkness at 16°C, watered daily and fertilized weekly. Shoot tips and/or leaves from the RIL population and the two parents were harvested and lyophilized. Genomic DNA was extracted using the DNeasy 96 Plant kit (Qiagen, Mississauga, ON, Canada) according to the manufacturer's instructions, quantified by fluorometry, and resuspended to 6 ng/μL with distilled water.

Linkage Map

The extracted RIL DNA was assayed with 219 polymorphic SSRs to generate a linkage map. These markers had previously been developed for flax using expressed sequence tags (ESTs) and bacterial artificial chromosome (BAC) end sequences (Cloutier et al. 2009, 2012a). All polymerase chain reactions (PCR) for SSR markers consisted of 30 ng genomic DNA, 1× PCR buffer, 0.4 mM dNTPs, 1.5 mM $MgCl_2$, 0.3 pmole M13-tailed forward primer, 2 pmole reverse primer, 1.8 pmole M13-labelled primer, 0.875U *Taq* DNA polymerase in a final volume of 10 μL and amplified as previously described (Huang et al. 2006). The amplification products were resolved on an ABI 3130xl Genetic Analyzer (Applied Biosystems, Foster City, CA, USA). Data collected by fluorescent capillary electrophoresis was processed using GeneScan (Applied Biosystems) software and converted to a gel-like image with Genographer v2.1.4 software (Benham et al. 1999).

Fragment sizes were estimated using the GeneScan ROX-500 internal size standard and scored for each individual. The construction of the genetic map was conducted with JoinMap4® (Van Ooijen 2006) using a logarithm of odds (LOD) score of 3.0 and a maximum recombination frequency of 40 cM. The expected Mendelian segregation ratios of 1:1 were tested using the Chi-square goodness-of-fit test. Recombination fractions were converted to centiMorgans (cM) with the Kosambi mapping function (Kosambi 1944). The map was verified against the previously described consensus map (Cloutier et al. 2012b). Graphical linkage maps were drawn with Mapchart 2.1 (Voorrips 2002).

Measurement of Cyanogenic Glucosides

The RIL population was grown in mini-plots in the field (as described in Soto-Cerda et al. 2014) at the following locations and years: Morden in 2011 and 2012 and, Regina in 2012, using a MAD type-2 experimental design in a 5 × 5 square (225 plots) (Lin and Poushinsky 1985; You et al. 2013). CDC Bethune was the main plot check and DL and MD were the sub-plot checks in all field trials. CGs were measured by GC for all RILs, plot checks and sub-plot checks from the Morden 2011 environment and for 60 and 40 RIL samples from the Morden 2012 and Regina 2012 locations, respectively, using previously described extraction and analysis protocols (Bacala and Barthet 2007; Barthet and Bacala 2010). These 80 and 62 RIL samples formed part of a calibration set for development of a near infrared spectroscopy (NIRS) method of analysis for phenotyping the remainder of the Regina and Morden 2012 samples. As an additional internal standard, 1% salicin was added

upon methanol extraction to control for volume errors. All samples from the three test sites were analyzed using the developed NIRS calibration. Due to insufficient harvested seed sample, no NIRS measurements were obtained for nine RILs from Morden 2011, and 15 each from Morden and Regina 2012. All plot and sub-plot check samples were measured for all environments. Phenotypic values were adjusted for row and column effects of the MAD type-2 experimental design using the pipeline described by You et al. (2013) and an analysis of variance (ANOVA) was performed to determine the level of significance associated with the genotype (G), the environment (E) and the G×E interaction.

QTL Analysis

QTL analyses for linamarin, lotaustralin, LIN, NEO, LIN:NEO ratio and total CG-derived HCN content were conducted using composite interval mapping (CIM) of Windows QTL Cartographer 2.5 software (Wang et al. 2007). A CIM standard model 6 was used with a window size of 10 cM and a 2 cM walk speed. Forward-backward regression was applied to determine the cofactors for background control. LOD thresholds were determined by 1,000 permutations at $P<0.05$. A QTL was declared when the overall LOD score and the LOD score of at least two of the locations-years exceeded the threshold level. The overall LOD was obtained using each individual's average over all three environments. QTL nomenclature is as follows: Q for QTL followed by a trait abbreviation (linustatin (lin); neolinustatin (neo)), a period, originating laboratory, a hyphen and the linkage group (LG) where the QTL is located. Estimates of phenotypic variance and coefficient of

determination (R^2) associated with the identified QTL were obtained from CIM analysis of all three datasets.

Anchoring to the physical map and fine-mapping

The consensus genetic map of flax (Cloutier et al. 2012b) was anchored to the physical map of flax cultivar CDC Bethune (Ragupathy et al. 2011) and its whole genome shotgun sequence (Wang et al. 2012). This information was used to pinpoint the scaffolds spanning the CG QTL identified. Genomic DNA from DL and MD was sequenced by the Michael Smith Genome Sciences Centre of the BC Cancer Agency (Vancouver, BC) using an Illumina HiSeq2000 (Illumina Inc, San Diego, USA). The 100-bp paired end tag sequence output was aligned against the whole genome shotgun sequence of CDC Bethune (Wang et al. 2012) as described in Kumar et al. (2012). The reads for the two parents were combined into a single binary alignment/map (BAM) file (Li et al. 2009) that was visualized in Tablet v1.13.07.31 (Milne et al. 2010).

A total of 40 SNPs polymorphic between DL and MD were selected from various scaffolds hypothesized to be physically mapped to the identified QTL. Twenty were from scaffolds 145 and 177 and another 20 were from scaffolds 297, 361, 112, 172, 1603, 618, 1491, 305, 176, 1578 and 1247. SNP assays designed and optimized by LGC Genomics (Massachusetts, USA) were tested on the DL/MD RIL population using a Fluorostar Omega Analyzer (BMG Labtech, North Carolina, USA) and the clusters were analyzed

and visualized with Klustercaller v2.21 (LGC Genomics) to score the SNPs. The SNPs were mapped to the SSR map using JoinMap with the criteria described above.

A total of 864 F_2 plants also derived from the DL/MD cross were grown. DNA was extracted and screened with the following flanking SSR markers: Lu2597, Lu2712, Lu 311, Lu2971, Lu2564 and Lu475. A total of 82 lines, recombinants at the LIN and/or NEO QTL, were identified and selfed to the F_4 generation. They were genotyped with the SNP markers to fine-map the recombination breakpoints. The F_4 families were grown in the field in Winnipeg in 2013 in single rows and CG profiles of the harvested F_5 seeds were obtained by GC as described above.

Candidate gene identification

A modified *in silico* approach has been used to identify candidate sequences for validation based on the available scaffolds on the physical map. These scaffolds were visualized using gbrowse on the phytozome website (http://www.phytozome.net) and all sequences for predicted proteins on these scaffolds were parsed into a text file. The CLC main workbench software v6.8.2 (http://www.clcbio.com) was then used to query these predicted protein sequences with the batch BLAST algorithm (Altschul et al. 1990). The BLAST output was a list of the hits for highest homology for each query. Gene clusters for secondary metabolic pathways typically consist of genes for 'signature enzymes' and genes for 'tailoring enzymes' on the same or a nearby scaffold (Osbourn, 2010). The signature enzymes for CG accumulation are UGT, CYP and glycoside hydrolase family enzymes.

Generic tailoring enzymes usually involved in secondary metabolic pathways are oxidoreductases, methyltransferases, acyltransferases and glycosyltransferases. The lists of hits were mined for clusters of these enzymes and recorded for future functional unit identification.

3.4 Results

Linkage Map

The DL/MD RIL linkage map had 15 LGs consisting of 195 polymorphic markers covering 938 cM and 31 markers remained unlinked (Figure 3.1). The number of markers per LG varied from three to 30 with an average distance of 5 cM. The order of markers was consistent with a consensus map developed using three other segregating populations (Cloutier et al. 2012b) but LG3 and LG5 had gaps greater than 40 cM causing them to split into two LGs each. LG9 could not be reconstituted because markers were either monomorphic or remained unlinked at LOD 3.0 and recombination fraction of 40 cM. Finally, when comparing the consensus map to the DL/MD map, a spurious linkage between LG6 and LG13 remained in the latter when using the same mapping criteria. Chi-square tests indicated that 41 of the 226 polymorphic markers (18.1%) significantly deviated from the expected 1:1 ratio ($P< 0.05$). Twenty-one of these markers were skewed toward DL and the remaining 20 were skewed toward MD (Table 3.1).

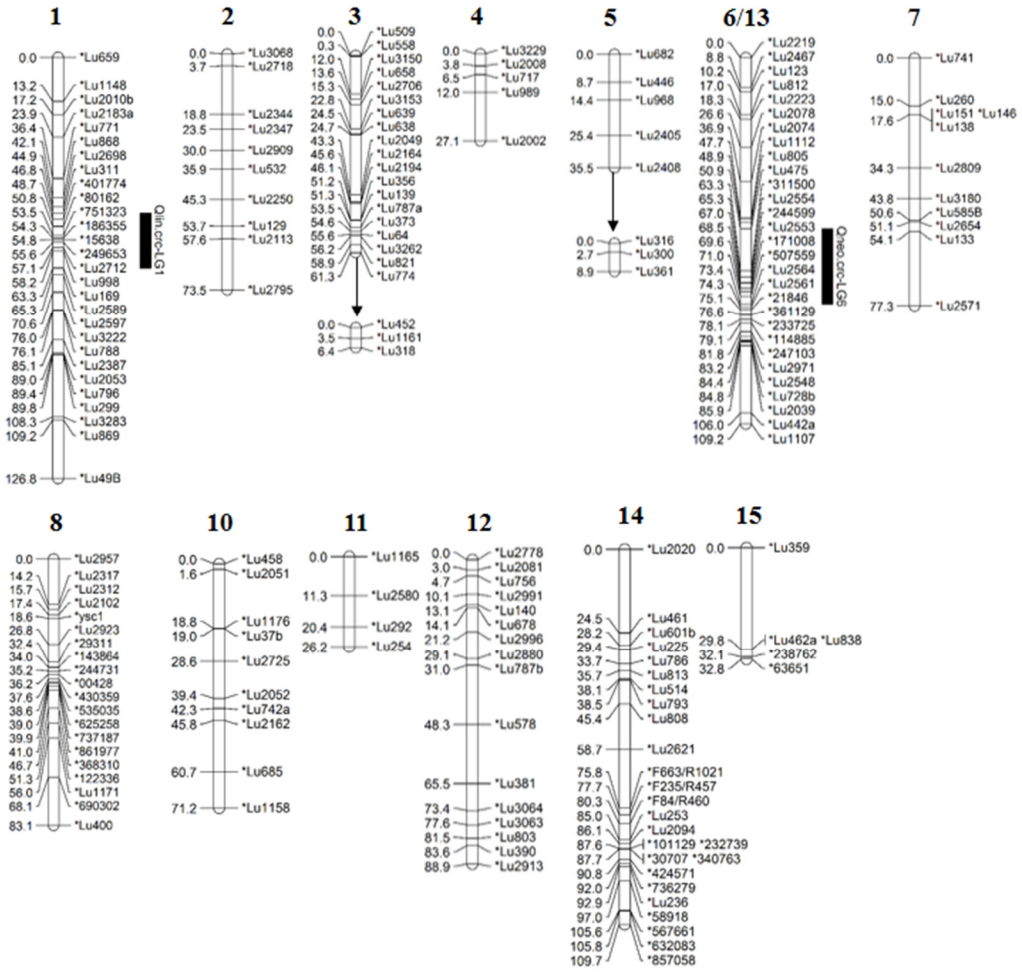

Figure 3.1 Linkage map of the Double Low/AC McDuff recombinant inbred line population. Gaps greater than 40 cM on linkage groups 3 and 5 are indicated by black arrows. Linkage group 6 and 13 (Cloutier et al. 2012b) collapsed into a single linkage group. QTL for linustatin on linkage group 1 and neolinustatin on linkage group 6 are indicated.

Table 3.1 Summary of linkage map for the Double Low/AC McDuff RIL population

Linkage Group	Distance (cM)	Total # of markers	SSRs	Genes	SNPs	Skewed markers	
						DL	MD
1	126.8	28	22	-	6	-	3
2	73.5	10	10	-	-	-	1
3a	61.3	19	19	-	-	-	1
3b	6.4	3	3	-	-	-	-
4	27.1	5	5	-	-	-	-
5a	35.5	5	5	-	-	-	-
5b	8.9	3	3	-	-	-	-
6/13	109.2	30	21	-	9	1	4
7	77.3	11	11	-	-	-	1
8	83.1	20	7	1 (*ysc1*)	12	4	1
10	71.2	10	10	-	-	-	2
11	26.2	4	4	-	-	-	-
12	88.9	16	16	-	-	3	-
14	109.7	26	16	-	10	-	4
15	32.8	5	3	-	2	4	
Sub-total	938	195	155	1	39	12	17
Unlinked		31	31	-	-	9	3
Total		226	186	1	39	21	20

Phenotypic Analysis

The concentrations of LIN and NEO were determined for the RILs, the parental lines (sub-plot controls) and multiple CDC Bethune samples (main plot control) grown in the field at three locations: Morden 2011, Morden 2012 and Regina 2012 (Figure 3.2; Table 3.2). DL had an average LIN concentration of 238.1 ± 91.7 mg/100g seeds and an average NEO concentration of 81.7 ± 16.4 mg/100g over the three environments while MD averaged 332.8 ± 86.4 mg/100g and 82.5 ± 14.0 mg/100g for the same two CGs, respectively (Table 3.2). Based on the Shapiro-Wilk test, LIN and NEO values were normally distributed for all environments except for LIN and NEO in Morden 2012 and NEO in Morden 2011 (Appendix I). Transgressive segregants were observed in all datasets and overall range was from 104.2 to 461.2 mg/100g for LIN and 32.8 to 147.3 mg/100g for NEO (Table 3.2). The ANOVA for row and column effects indicated no adjustments were necessary for measurements in any of the environments (Appendix II). The ANOVA for genotype, environment and G×E effects indicated significance for all partitions ($P<0.05$) for LIN and NEO concentration across locations with the environment (E) being the most significant (Table 3.3). Pearson correlation coefficients were positive (0.50, 0.32, 0.21) and significant ($P<0.01$) for LIN among all three environments and for NEO (0.28, 0.11, 0.22) among all except between Morden 2011 and Regina 2012 (Table 3.4). LIN and NEO were also positively and significantly correlated with each other in all three environments (0.24, 0.37, 0.19) (Table 3.4). This correlation lends support to the idea of related mechanisms of accumulation for both LIN and NEO.

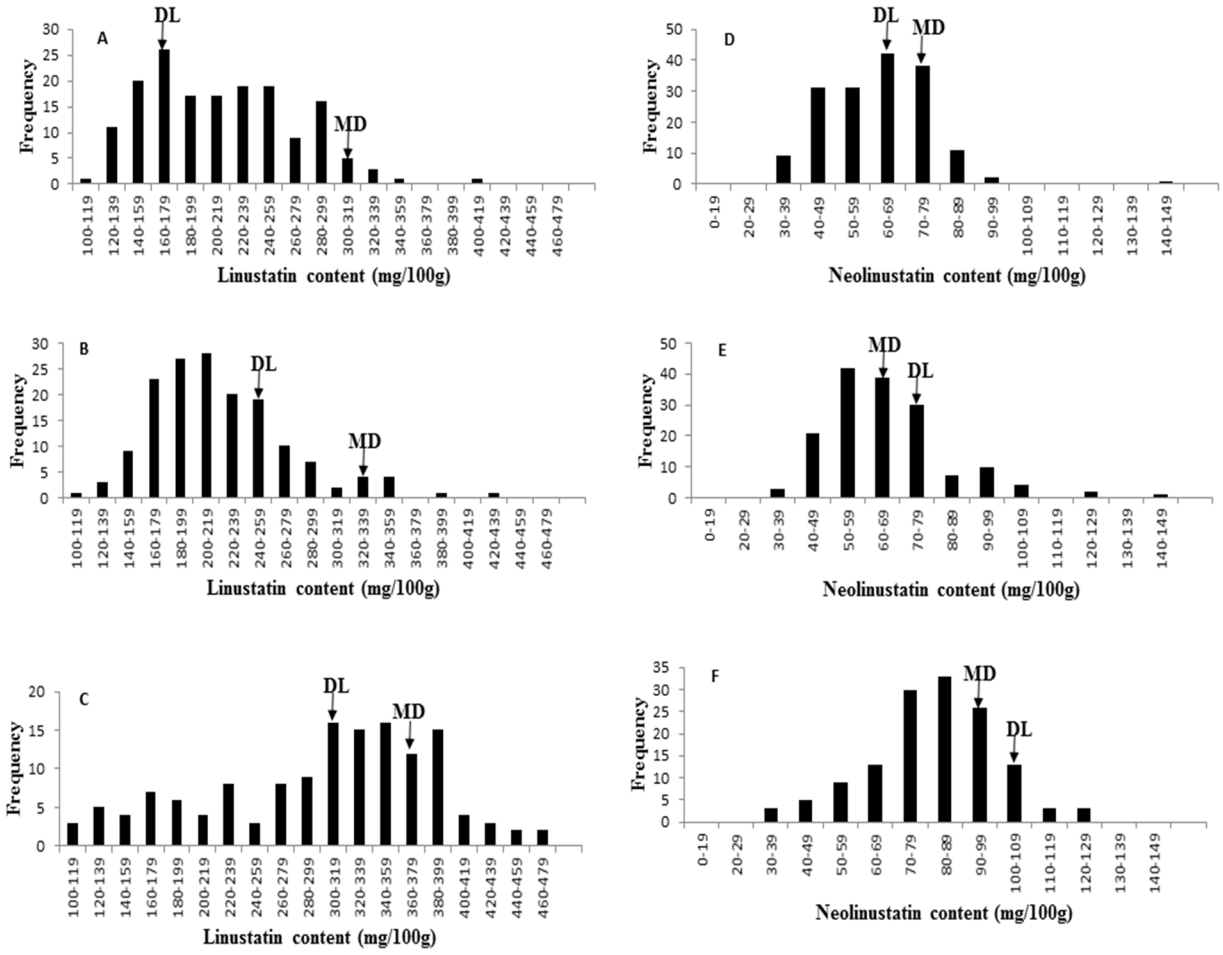

Figure 3.2 Distributions of linustatin and neolinustatin for the three environments **A**: Linustatin, Morden 2011 **B**: Linustatin, Morden 2012 **C**: Linustatin, Regina 2012 **D**: Neolinustatin, Morden 2011 **E**: Neolinustatin, Morden 2012 **F**: Neolinustatin, Regina 2012. **DL**: Double Low **MD**: AC McDuff

Table 3.2 Linustatin content (mg/100g seeds) distributions for the three environments

Linustatin	Morden 2011	Morden 2012	Regina 2012	Overall
RIL population				
N[1]	165	159	142	466
Mean ± SD[2]	213.4 ± 56.2	220.3 ± 53.3	300.6 ± 86.0	242.3 ± 76.3
Max	408.6	438.4	461.2	461.2
Min	112.9	116.1	104.2	104.2
Double Low				
N	12	12	12	36
Mean ± SD	160.3 ± 34.0	246.6 ± 27.6	307.5 ± 113.6	238.1 ± 91.7
Max	262.1	281.6	423.7	423.7
Min	135.2	205.9	164.9	135.2
AC McDuff				
N	12	13	14	39
Mean ± SD	300.4 ± 19.2	329.0 ± 87.4	364.2 ± 111.3	332.8 ± 86.4
Max	340.6	423.2	499.7	499.7
Min	263.9	138.6	151.7	138.6

[1]N: Number of samples; [2]SD: Standard deviation

Table 3.3 Neolinustatin content (mg/100g seeds) distributions for the three environments

Neolinustatin	Morden 2011	Morden 2012	Regina 2012	Overall
RIL population				
N[1]	165	159	142	466
Mean ± SD[2]	61.9 ± 15.2	66.4 ± 17.2	81.8 ± 18.5	69.4 ± 18.8
Max	147.3	143.8	129.0	147.3
Min	32.8	37.3	36.6	32.8
Double Low				
N	12	12	12	36
Mean ± SD	65.8 ± 10.9	78.2 ± 3.3	101.1 ± 5.2	81.7 ± 16.4
Max	88.1	83.5	108.0	108.0
Min	52.9	73.1	91.4	52.9
AC McDuff				
N	12	13	14	39
Mean ± SD	79.3 ± 4.7	68.0 ± 6.1	98.6 ± 4.6	82.5 ± 14.0
Max	87.8	81.8	105.4	105.4
Min	71.8	60.8	88.8	60.8

[1]N: Number of samples; [2]SD: Standard deviation

Table 3.4 ANOVA for linustatin and neolinustatin content in the RIL population across the three environments

Source	DF	Sum of Squares	F Value	Pr>F	R^2
Linustatin					
Environment (E)	2	2104799.55	2123.23	1.62E-106	0.56
Genotype (G)	176	1158504.59	13.28	4.31E-45	0.31
G×E	301	405066.38	2.72	6.58E-11	0.11
Error	142	70383.76	.	.	0.02
Neolinustatin					
Environment (E)	2	80079.15	561.52	3.66E-68	0.18
Genotype (G)	176	293335.22	23.37	8.05E-61	0.66
G×E	301	63144.26	2.94	2.66E-12	0.14
Error	142	10125.44	.	.	0.02

Table 3.5 Pearson correlation coefficients for linustatin and neolinustatin for the three environments

	Morden2011-LIN	Morden2012-LIN	Regina2012-LIN	Morden2011-NEO	Morden2012-NEO	Regina2012-NEO
Morden2012-LIN	*0.50					
Regina2012-LIN	*0.32	*0.21				
Morden2011-NEO	*0.24	*0.22	0.00			
Morden2012-NEO	0.13	*0.37	0.07	*0.28		
Regina2012-NEO	0.02	0.18	*0.19	0.11	*0.22	

* $P<0.01$

QTL Analysis

One region for LIN and one region for NEO met the defined criteria and were declared significant QTL across the 15 LGs (Table 3.5). Twenty-two other QTL were identified but not retained because they were not consistent across environments (Appendix III).

One QTL region associated with LIN content and spanning 16.5 cM was identified on LG1. The peak was significant for the Morden 2011 and Morden 2012 datasets. This QTL had an average LOD score of 11.9 and accounted for 20.9% of the phenotypic

variation based on the overall mean dataset. Its overall additive effect estimate of 28.7 indicates that MD contributes to increasing LIN content. The other QTL region of map distance 22.5cM was associated with NEO accumulation and identified on LG6. The peak was significant for all three site-years. This QTL had an average LOD score of 4.3 and accounted for 11.3% of the phenotypic variation based on the overall mean dataset. Its additive effect estimate of 6.1 was also indicative of MD's contribution to increasing NEO content. *Qlin.crc-LG1* also collocated with peaks for linamarin, LIN:NEO ratio and total HCN that were all significant in Morden 2011 while *Qneo.crc-LG6* collocated with significant peaks for LIN:NEO ratio in Morden 2011, lotaustralin in Morden 2012 and total HCN in Morden 2012 (Figure 3.3, Appendix III).

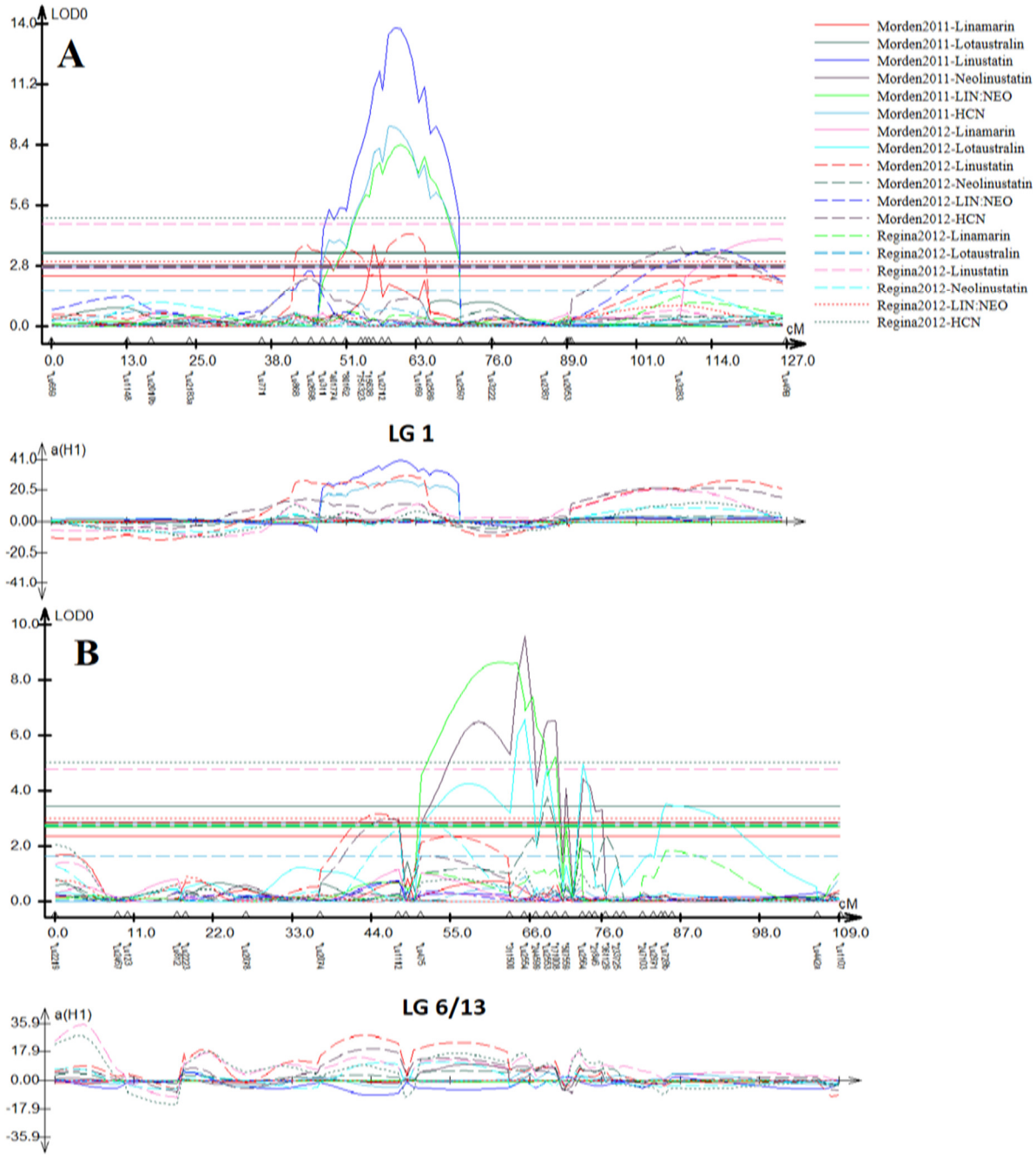

Figure 3.3 LOD thresholds and additive effects of QTL peaks for all traits from the three environments as seen in QTL cartographer v2.5 (Wang et al. 2007) for **A**: LG 1 and **B**: LG 6/13

Table 3.6 Details of QTL peaks that were present in at least two of the three environments

Trait	LG	Flanking markers	QTL	Location	LOD	Additive effect	R^2
Linustatin	1	Lu311-Lu2597	*Qlin.crc-LG1*	Morden 2011	13.8	39.9	29.0
				Morden 2012	3.8	27.7	7.0
				Regina 2012	2.0	2.4	0.3
				Overall mean	11.9	28.7	20.9
Neolinustatin	6	Lu475-Lu2564	*Qneo.crc-LG6*	Morden 2011	9.6	10.7	28.2
				Morden 2012	4.9	8.0	10.2
				Regina 2012	2.7	11.0	6.6
				Overall mean	4.3	6.1	11.3

Fine Mapping and Physical mapping of the CG QTL

The QTL regions were fine mapped using phenotypic data from the F_5 seeds and markers that span them. Of the 39 SNPs assayed, six mapped to LG1, nine mapped to LG6 and the remaining 25 mapped to LGs 8, 14 and 15. Mapping the SNPs on the 82 recombinant lines identified additional breakpoints and narrowed down the QTL for LIN to 11.7 cM, positioning it between SNP markers 80162 and 186355 (Figure 3.4). This QTL, anchored to scaffolds 112, 361 and 297 on the physical map, had a peak LOD score of 4.5 and

accounted for 20.8% of the phenotypic variation (Table 3.6). Similarly, fine mapping the region for NEO identified two QTL, one of 3.5 cM located between SNP markers 171008 and 244599 and one of 15.2 cM situated between SNP marker 311500 and SSR marker Lu475 (Figure 3.4). The former had a peak LOD score of 2.4, accounted for 10.4% of the phenotypic variation and was anchored to scaffolds 1491 and 305 on the physical map. The latter had a peak LOD score of 7.5, accounted for 38.4% of the phenotypic variation and anchored to scaffold 176 on the physical map (Table 3.6).

Data mining of the scaffolds identified by fine mapping indicated a total of seven gene clusters found on scaffolds 112, 361, 1491 and 176 (Table 3.7, Appendix IV). A total of five clusters predicted to encode UGTs were on scaffolds 361, 1491 and 176 while scaffold 112 had one cluster predicted to encode CYPs and one predicted to encode transcription factors of the GRAS family.

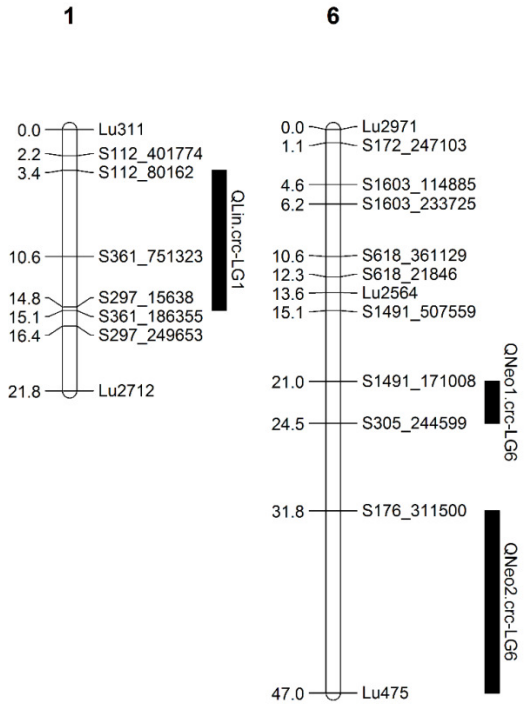

Figure 3.4 Location of fine mapped QTL for linustatin and neolinustatin accumulation

Table 3.7 Summary of the QTL identified after fine mapping

Trait	LG	Flanking markers	QTL	LOD	Additive effect	R^2
Linustatin	1	s112_80162-s361_186355	*Qlin.crc-LG1*	4.5	24.9	20.8
Neolinustatin	6	s1491_171008-s305_244599	*QNeo1.crc-LG6*	2.4	4.9	10.4
	6	s176_311500-Lu475	*QNeo2.crc-LG6*	7.5	9.1	38.4

Table 3.8 A list of all gene clusters found on the scaffolds that anchor the QTL identified by fine mapping

Scaffold	Description	Accession numbers
112	GRAS family transcription factor	XP_002316033, XP_002316033, XP_002316033, XP_002316034
112	Cytochrome P450	XP_002521476, XP_002329123, XP_004237151, XP_003525679, XP_002306255
361	UDP-glycosyltransferase 1	AFJ52912, AFJ52909, AFJ52914
1491	UDP-glycosyltransferase 1 (cluster 1)	AFJ52991, AFJ52989
1491	UDP-glycosyltransferase 1 (cluster 2)	AFJ52945, AFJ52948, AFJ52946, AFJ52929, AFJ52929, AFJ52944, AFJ52931
176	UDP-glycosyltransferase 1 (cluster 1)	AFJ52932, AFJ52935, AFJ52938, AFJ52937
176	UDP-glycosyltransferase 1 (cluster 2)	AFJ52941, AFJ52940

3.5 Discussion

Genetic Linkage Map

In this study, a RIL population from a cross between a linseed line called Double Low and variety AC McDuff was used to develop a linkage map using SSR markers. The RIL population segregated for LIN and NEO accumulation in seeds. Major effect QTL for this accumulation were identified based on correlation between the allelic state of linked markers and phenotype.

With the exception of a few local inversions, two gaps and one spurious linkage, the genetic map reported here was collinear with the consensus map previously obtained from three segregating populations (Cloutier et al. 2012b). With 938 cM, the DL/MD map covers 60% of the total consensus map. Six LGs have low marker density consisting of only three to five markers because they harbored mostly monomorphic markers. Ten markers belonging to LG9 were resolved but were not mapped to this LG at a LOD of 3.0. While the marker density and coverage of the DL/MD SSR map was partial, it proved adequate to identify QTL for LIN and NEO content in flax seeds but not sufficient to guarantee that all CG QTL potentially segregating in this cross were detected.

Marker Segregation Distortion

Segregation distortion is the deviation of observed genetic ratios from expected Mendelian ratios. It is a common phenomenon in mapping populations that may occur due to various factors including genetic elements, mapping population, marker type, parental relationship

and cytoplasmic components. A total of 18.1% of the polymorphic markers deviated from the expected 1:1 Mendelian segregation ratio with equal numbers of loci skewed toward each parent. This percentage is similar to those obtained in other mapping studies using SSRs on RILs (Sa et al. 2012). The distorted markers were found in all but three LGs.

Distorted markers are important to consider because their impact on recombination frequencies impair mapping precision and possibly QTL detection, although, unless the mapping population is small, the distortion effect can generally be ignored (Zhang et al. 2010). The degree of impact also depends on the distance between the putative QTL from the segregated marker. Because the QTL identified herein were not adjacent to any distorted markers, the estimated position or effect size of the QTL should not be significantly affected by them.

QTL Analysis

QTL detection can be limited by the population size. With 174 individuals, the DM population is neither small nor large but may be insufficient to detect all QTL and may overestimate the phenotypic variance due to censoring, a phenomenon known as the Beavis effect (Beavis 1994; Xu, 2003). Previous studies have indicated that QTL with an explained variance of 10% have an 80% chance of being detected in a population of 200 individuals with the probability of detection reducing more or less linearly with decreasing population size. Despite the incomplete map coverage and the biased markers, the 174 individual RIL population combined with the F_2 fine-mapping permitted the detection of one QTL for LIN

accumulation (*Qlin.crc-LG1)* on LG1 and two for NEO accumulation (*Qneo1.crc-LG6; Qneo2.crc-LG6)* on LG6.

Results of the ANOVA had *P*-values for the F statistic less than 0.001 for genotype, environment and G×E for both LIN and NEO providing strong evidence against their null hypotheses. Comparison of the LIN R^2 values indicates that the environment (56%) contributes to variation almost twice as much as the genotype (31%). In contrast, the R^2 values for NEO showed that genotype (66%) was more important than environment (18%). In comparison, Oomah et al. (1992) reported that for ten Candian flax cultivars, genotype contributed to about 80% of LIN content variability while NEO content variability was almost completely determined by an equal contribution of genotype and environment. The G×E effects for LIN (11%) and NEO (14%) were fairly similar and may provide a general estimate of the stability of the response curve a breeder might expect across environments. Based on the relative environmental contributions, cultivar development using marker assisted selection of the NEO associated QTL will be more robust across environments compared to the LIN associated QTL. However, considering that LIN is the major linseed CG, even small reductions can contribute significantly to reducing the overall CG content in the seed.

Phenotyping accuracy is perhaps the biggest limiting factor of most QTL analyses, hence the choice of method and the use of controls are paramount to minimize the sources of errors are important to consider (Cobb et al. 2013). Among the many published methods of analysing CGs, the GC method is considered superior in terms of high repeatability,

detection limit and accuracy. In addition, a series of control measures were taken with every run. Prior to analysis, system suitability controls consisting of blanks were used to check for contamination and five replicate injections were analyzed to ensure sensitivity and repeatability. Multiple point calibration curves and intercepts were also checked for linearity. With every set of experimental samples, two different internal standards were used to control for volume errors in sample preparation and injection. Quality control samples of known concentrations were included to ensure that the instrument performance did not change over time. Check samples of Crop Development Centre (CDC) Bethune were included with every run and charted to establish repeatability of the assay and to set limits for acceptable data points, i.e., within two σ of the mean. The final values were the averages of three technical replicates to estimate precision.

The identification of separate QTL peaks for LIN and NEO accumulation suggests a degree of independence in the biosynthesis pathways or differential biosynthetic regulation for the two glucosides. A shared QTL peak would have indicated a correlation in their accumulation and a likely shared regulatory pathway. Through a series of steps, the starting amino acid L-valine is converted to linamarin and then to LIN while L-isoleucine is converted to lotaustralin and then to NEO. Previous *in vitro* experiments have indicated that linamarin and lotaustralin biosynthesis is catalyzed by a shared set of enzymes (Hahlbrock and Conn, 1971; Cutler et al. 1985). Accumulation of the intermediates is cytotoxic and the highest specificity factor limiting production is the first CYP enzyme in the pathway. Because of the high degree of channeling, accumulated CGs should therefore

be a product of the activity of the biosynthesis enzyme system and degradation. In *L. japonicus*, two separate CYP79 enzymes, CYP79D4 and CYP79D3, are key for the differential accumulation of LIN and NEO (Forslund et al. 2004). Though these enzymes share 94% identity, they are substrate-specific for either L-valine or L-isoleucine respectively and have little similarity in the promoter regions suggesting that the different pathways may have emerged by duplication and neofunctionalization based on the specificity of key residues. In cassava, two cytochrome P450 cDNA clones have been isolated, CYPD1 and CYPD2, that can both metabolize L-valine and L-isoleucine except that CYPD1 has a higher maximum reaction rate with L-valine as substrate than with L-isoleucine (Andersen et al. 2000). The possibility of alternate CYP encoding enzymes with differential affinities in flax is also plausible considering that flax also experienced at least one ancient genome duplication event which may have given rise to functional paralogs (Wang et al. 2012, Mühlhausen and Kollmar 2013). Krech and Fieldes (2003) have reported differential responses of the L-valine and L-isoleucine derived CGs in flax lines that contrast for germination emergence and hypocotyl growth rate. No differences for the L-valine CGs were detected between the two lines but the more vigorous of the two lines significantly accumulated more NEO, hydrolysed it faster and initiated *de novo* synthesis of lotaustralin earlier (Krech and Fielded 2003). These differential responses suggest that L-valine and L-isoleucine derived CGs may be regulated separately and may have separate functional niches in development. Although linustatinase has been shown to hydrolyze both LIN and NEO, the possibility of a separate degradation regulatory mechanism for the

two glucosides cannot be ruled out. The activity of β-glucosidases has been demonstrated to be a significant factor in HCN release from CGs (Ballhorn et al. 2006). However, our data do not provide any evidence for this assumption because no β-glucosidases were located in the scaffolds overlapping the QTL. This is consistent with previous reports that the cyanogenic β-glucosidases do not cluster with the biosynthetic genes suggesting an independence to the two pathways (Takos et al. 2011).

Linamarin and lotaustralin have been suggested to be formed by the same glucosyltransferase in flax seedlings based on consistent linamarin:lotaustralin ratios and feeding experiments in which there was a significant drop in the incorporation of a radioactive starting amino acid for the respective glucoside when preincubated with the other starting amino acid (Hahlbrock and Conn 1971). However, our results indicated highly variable ratios for both linamarin:lotaustralin (1.1 to 12.9) and LIN:NEO (0.7 to 6.1) in the RIL population. Although LIN and NEO positively and significantly correlate with each other, distinct QTL for LIN and NEO were identified suggesting that although their accumulation mechanisms tend to be related, loci exist that can cause a separate and significant difference in accumulation for the two glucosides in possibly distinct or redundant biosynthesis pathways. Alternatively, the distinct LIN and NEO QTL may correspond to different putative glucosyltransferases in the transition between monoglucosides to diglucosides.

QTL or genes for CG accumulation or HCN equivalent have been identified in a number of other crop species with the most extensively studied being cassava (*Manihot*

esculenta) and sorghum (*S. bicolor).* A positional mapping study for HCN in cassava identified two QTL on two different LGs (Kizito et al. 2007). The genes for CYP79D1/D2 (Zhang et al. 2003; Andersen et al. 2000), CYP71E7 (Jørgensen, 2011), linamarase (Liddle et al. 1998) and α-hydroxynitrile lyase (Wang et al. 2004) have been identified and partially characterized. Cassava produces the CGs linamarin and lotaustralin but not LIN and NEO. Interestingly, the genes for dhurrin accumulation in sorghum were identified via isolation of the enzyme pathway in a microsomal system and subsequent cloning using a PCR approach on a cDNA library as opposed to the more traditional map-based cloning approach (Møller and Conn 1980; Bak et al. 1998). These genes include: CYP71E1 (Kahn et al. 1997), CYP79 (Koch et al. 1992), UDPGT (Jones et al. 1999), dhurrinase (Verdoucq et al. 2004) and α-hydroxynitrile lyase (Wajant et al. 1994). Both cassava and sorghum exhibit the same respective elements of the CG biosynthesis and degradation pathways. These previous reports of QTL and cloned CG genes, along with enzyme purification and microsomal experiments in flax (Hahlbrock and Conn 1971; Cutler et al. 1985), reinforce the notion of a small, conserved metabolic system across species and lend support to the likelihood of a small number of QTL with large effects in linseed.

Linamarin and lotaustralin levels fall to trace amounts during maturation while LIN and NEO levels rise suggesting a conversion of monoglucosides to diglucosides during seed maturation (Frehner et al. 1990). At anthesis, the inflorescences contain two to three μmol HCN, a significantly higher amount than the 0.4-1.3 μmol in the almost mature seed (35-40 days after flowering). To minimize the total HCN equivalent in harvested seed, it

is recommended that the seed be allowed to mature as fully as possible before harvesting. This pattern of CG accumulation affects the power of our QTL detection because the RILs ranged in maturity. Indeed MD is an earlier maturing variety than DL, hence part of the environmental effect observed may be attributable to the variable maturity of the population, affecting the measured concentrations for LIN and NEO. However, the monoglucoside levels can be used to gauge the relative maturity of lines, and, in virtually all samples, the total monoglucoside concentration never exceeded 15 mg/100g seeds. There were no consistently significant QTL peaks for either linamarin or lotaustralin across the three environments, likely because their phenotypic measurements were frequently beneath the limit of quantitation for GC analysis.

Physical Mapping and Candidate Sequences

The classical procedure for identifying a manageable list of candidate sequences has been to obtain a genetic resolution sufficiently high enough by virtue of extensive screening to identify additional recombinants and marker development at the QTL. Combined with the *in silico* approach, the strategy used effectively in other mapping studies to pinpoint the candidate QTL/gene can be effective (El-Assal et al. 2001; Werner et al. 2005). Examination of the scaffolds that correspond to the putative QTL regions show five predicted UGT clusters, one CYP cluster and a GRAS transcription factor cluster. The genes that encode CGs have been so far demonstrated to occur in clusters (Takos et al. 2011; Nützmann and Osbourn 2014). The physical clustering of relevant genes for

specialized metabolic pathways is a common phenomenon in plants and may be an intrinsic characteristic of defense compounds because co-expression could allow coordinated transcriptional regulation and co-inheritance in order to maintain a functional pathway (Chu et al. 2011b). A common clustering feature for secondary metabolites is the existence of 'signature' metabolic genes that encode the skeleton structures for the various metabolite pathways and a variable number of other 'tailoring' genes that modify the skeleton structure (Osbourn 2010). These signature genes appear to have evolved from genes for primary metabolism via gene duplication and neofunctionalisation and are presumably the origins of the gene clusters (Chu et al. 2011b). Clustering may be necessary in the CG pathway because leakage of pathway intermediates could prove toxic to the cell. Very little has been published about the transcriptional regulation of CG biosynthetic enzymes, although it has been suggested that it is likely the strongest determinant of CG accumulation based on a strong correlation between enzyme activity and corresponding mRNA levels (Busk and Møller 2002). No links have been established between CGs and GRAS transcription factors or giberellin signaling. However, investigation of the coexpression of this GRAS cluster on scaffold 112 may be warranted because of its colocalization.

Conclusion

The identification of an allele with a high R^2 in one genetic background does not necessarily guarantee the path to a successfully bred cultivar with altered CGs via marker

assisted selection or genetic engineering. The challenges of gene interactions, metabolic fluxes, developmental regulation and integrated responses to environmental cues are sources of complexities that may distort the observed phenotypic effect across breeding lines. However, if the relatively simple pathway of CG accumulation holds true in linseed as it does in sorghum and cassava, optimism for the utility of the loci in breeding programs is warranted.

This study is the first to map QTL for CG content in linseed. Validation will lead to marker-assisted selection and help in developing low CG linseed varieties that will ensure that Canadian flaxseed complies with trade limitations. Fine mapping by saturation of the regions with SNPs and the development of additional recombinants narrowed down the QTL for LIN and NEO, paving the way for the map-based cloning of the functional units, assisted by the candidate gene list generated based on function prediction. Ultimately, the map and loci will provide useful tools towards the improvement of flaxseed safety and marketability.

4.0 A NEAR INFRARED SPETROSCOPY CALIBRATION FOR CYANOGENIC GLUCOSIDES IN FLAX SEED

4.1 Abstract

Cyanogenic glucosides (CGs) are secondary metabolites found in flax seed that hydrolyze to produce toxic hydrocyanic acid (HCN) when ingested. In this study, we developed a near infrared spectroscopy calibration to estimate individual CGs and total HCN equivalent. Reference values were obtained using gas chromatography analysis for linamarin, lotaustralin, linustatin (LIN), neolinustatin (NEO) and total HCN equivalent. The calibration used a modified partial least squares regression approach with a 3,5,5,1 pretreatment on 252 wavelengths over the full spectral range (408-2492 nm). When applied to a validation set, the calibration produced standard error of prediction values of 0.14, 0.26, 0.08, 0.11 and 0.07, and ratio of performance to deviation values of 1.72, 1.09, 1.37, 1.88 and 1.75 for linamarin, lotaustralin, LIN, NEO and total HCN equivalent, respectively. This calibration is intended as a tool for large scale screening in flax breeding programs. At the moment, the utility of the calibration is limited to a cursory screening of breeding material providing a general ranking of high to low values and requires the further integration of samples with a greater range and of more variable composition to improve its robustness.

4.2 Introduction

Near infrared spectroscopy (NIRS) is a routine analytical method for many seed-related traits for post-harvest quality control (Williams, 1975). Spectra-structure correlations are made based on a measured spectral response of a seed sample and a calibration curve developed using a reference dataset for a particular seed component. It enables qualitative and quantitative assessment of seed composition based on spectral data and appropriate multivariate calibration models. In the grain and oilseed industry, these models are typically used to determine moisture, protein, oil, starch and fatty acid profiles (Pandford and Williams 1988; Siemens and Daun 2005). Quick and reliable monitoring of seed components can be done at all stages of production, storage and transport to maintain seed quality and market value.

NIRS is advantageous because it is cost effective, high throughput, non-destructive and able to estimate concentrations of multiple components simultaneously. NIRS calibrations accurate enough to screen breeding material have been developed to estimate HCN potential in sorghum (*Sorghum bicolor)* using a partial least squares (PLS) regression approach *(*Fox et al. 2012; Goff et al. 2011).

Rapid, cost effective and non-destructive methods for the determination of CG content in seeds have been developed for a limited number of crops. Such methods can be extremely useful to plant breeders for early generation screening. Unfortunately, the majority, if not all of these methods, do not offer the accuracy and reproducibility required

for effective selection. In flax, analytical methods that meet all of the requirements for large scale screening in a breeding program have yet to be developed. Gas chromatography (GC) quantification methods are precise and reproducible but they require a considerable amount of seed and, they are time-consuming, relatively expensive and destructive.

The HCN of flaxseed can vary by as much as 200 ppm within a single population depending on growing conditions, harvest time and environmental stresses (Schilcher and Wilkens-Sauter 1986). Various post-harvest processing methods to reduce CG content exist but they are not a cost effective long term option. The European Union is one of Canada's biggest linseed export markets. However, it has set a trade restriction limit for effective HCN concentration of 250 ppm in imported linseed for feed products (European Parliament Council 2002). The CG content of most Canadian linseed cultivars oscillate around this limit and some are consistently higher (Oomah et al. 1992; Mazza and Oomah 1995). The development of an NIRS calibration to monitor CG content to ensure compliance with accepted limits would prove useful.

4.3 Materials and Methods

Flax Samples

A total of 360 seed samples were pooled across three environments: Morden 2011, Morden 2012 and Regina 2012. Samples included the cultivars CDC Bethune, AC McDuff, Double Low, E1747, Viking and a recombinant inbred line (RIL) population of 174 individuals derived from a cross between line Double Low and cultivar AC McDuff (Kenaschuk and

Rashid 1994). Forty samples were randomly removed from the overall set to be used as a validation dataset. The remaining 320 samples constituted the total calibration dataset, of which 310 samples formed the effective dataset for linamarin, LIN, NEO and total HCN after outlier or missing data removal and 218 for lotaustralin. Prior to analysis, all seed samples were cleaned using a seed blower to remove any foreign material, chaff and damaged seeds, and then further cleaned by hand to ensure minimal interfering material in spectra measurement.

Reference dataset

Individual CGs were extracted from the calibration and validation seed samples and analyzed using a previously described validated method for GC analysis (Bacala and Barthet 2007; Barthet and Bacala 2010). As an additional internal standard, 1% salicin was added to the 75% methanol to control for volume errors. Total HCN equivalent was calculated based on the total amount of CGs (linamarin, lotaustralin, LIN and NEO) detected. Concentration values were corrected for moisture content determined using a previously described method (Pandford and Williams 1988) on a NIRSystems 6500 scanning monochromator (FOSS NIRSystems Inc., Silver Spring, Maryland). All values were $\log_{10}$-transformed and normality was assessed using a Shapiro-Wilk test.

NIRS estimation

Reflectance spectra using absorbance at a log 1/R function for all calibration models for flax were measured at 2-nm intervals for two segments: 408-1092 (visible spectrum) and 1108-2492 (near infrared spectrum) using a NIRSystems 6500 scanning monochromator (FOSS NIRSystems Inc., Silver Spring, Maryland) and analysed using the WinISI v1.05 software. Approximately 30 g of whole flax seeds were scanned for each sample using a coarse sample cell. Scans were performed in triplicate and values at each wavelength were averaged to control for fluctuations across reads. Spectral outliers were removed from the calibration set by the creation of a score file on the WinISI software using a PL2 method loading type with a global H of three and neighborhood H of 0.60 with factory defaults for wavelengths, math treatments and data transformations. Biological outliers were kept in the calibration dataset. H outliers are based on Mahalanobis distance measurements such that the lower the H value, the closer the fit. PL2 is a PLS method for spectral data with multiple reference values. The global calibration option was used to test modified partial least squares (MPLS), PLS and principal component regression models. The math pretreatments considered were the recommended ones in the WinISI manual following the format D,G,S1,S2 where D is the derivative number, G is the gap between points used to calculate the difference and, S1 and S2 are the number of data points used to smooth the data. The pretreatments considered included a 0,1,1,1 setting without standard normal variant (SNV) and detrend scatter correction (DSC), a 1,4,4,1 setting with SNV and DSC, a 2,4,4,1 setting with SNV and DSC, a 2,6,4,1 setting with SNV and DSC and a 3,5,5,1 setting with SNV and DSC. SNV is a preprocessing technique that corrects for baseline

shifts and variations in signal intensity by normalizing the spectra to minimize scattering effects. DSC applies a correction factor for the baseline shift variation using a second degree polynomial. The constituents tested were linamarin, lotaustralin, LIN, NEO and total hydrocyanic acid (HCN) equivalent concentrations.

4.4 Results and Discussion

Sample selection

The optimum number of samples for NIRS calibration depends on the trait, the complexity of the spectra corrections and the algorithms. While the smallest number of samples required for a robust calibration is generally accepted to be approximately 100, robust NIRS calibration models are often constructed using hundreds to thousands of samples from multiple years and locations to capture as much variability as possible (Caoa 2013). However, too many duplicate samples can reduce the prediction accuracy due to noise. Sample discrimination using the score file removed nine spectral outliers to give a working total of 311 samples in the calibration dataset. One and 93 samples were missing from the reference datasets for linamarin and lotaustralin respectively due to their values being below the limit of quantitation for GC analysis. To develop a robust calibration method, the concentration of chemical components should ideally be uniformly distributed over the entire range. Table 4.1 summarizes the distribution of the reference values obtained for the samples used in the calibration and validation datasets. A $\log_{10}$ transformation was applied

to the raw reference values and the normality of the transformed values verified using a Shapiro-Wilk test (Appendices V, VI and VII).

Table 4.1 Summary of reference values for the untransformed calibration and validation datasets

	Linamarin (mg/100g seed)	**Lotaustralin (mg/100g seed)**	**Linustatin (mg/100g seed)**	**Neolinustatin (mg/100g seed)**	**Total HCN (ppm)**
Calibration					
Min	0.92	0.44	104.20	32.82	99.67
Max	47.75	22.30	471.53	294.88	494.00
Mean	6.67	2.56	231.16	89.91	219.13
SD[1]	5.27	2.32	69.20	54.73	65.07
N[2]	310	218	311	311	311
Validation					
Min	1.81	0.61	113.38	37.56	109.29
Max	17.43	15.17	391.81	224.08	340.35
Mean	5.29	2.84	222.62	89.62	211.51
SD[1]	3.09	2.97	57.70	51.30	58.77
N[2]	40	21	40	40	40

[1]SD: standard deviation; [2]N: number of samples

The completely mature flax seed contains virtually all diglucosides and only traces of monoglucosides although this ratio may fluctuate depending on the maturity of the seed samples (Frehner et al. 1990). At anthesis, the inflorescences contain two to three µmol HCN, a significantly higher amount than the 0.4-1.3 µmol in the almost mature seed (35-40 days after flowering). Following germination, the diglucosides are rapidly depleted and cyanogenic monoglucosides are *de novo* synthesized in the developing seedling, leaf, flower, stem and root tissues, peaking at 5 days after germination (Krech and Fieldes 2003; Niedźwiedź-Siegień 1998). Even though the mature seed diglucosides can be resolved at a high accuracy, reproducibility and repeatability by GC, the monoglucosides are frequently below the limit of quantitation in this tissue. It is impossible from a practical standpoint to harvest the seeds at an even physiological maturity which would produce more uniformly low levels of monoglucosides. The inaccuracy of the monoglucoside reference values creates a bias in the reliability of their calibrations. Consequently, the calibrations for linamarin and lotaustralin are likely overestimated. Ultimately, achieving low levels of total HCN is the primary breeding objective to adhere to trade restrictions and to minimize the anti-nutritional effect. However, the levels of monoglucosides need to be considered since their levels are much higher in the immature or germinating seed. If the breeding lines are harvested too prematurely or if harvested seed is accidentally germinated, effective comparison of total HCN levels may be distorted.

Calibration model development

The models and pretreatments tested were assessed based on the standard error of cross validation (SECV) values obtained when generating an equation from the calibration set. The SECV is the square root of the mean square of the residuals with the lowest SECV value indicating a better calibration. The best calibration for estimating the values of the validation dataset was obtained using an MPLS regression model for full spectrum (408-2492 nm) on 252 wavelengths with a 3,5,5,1 pretreatment plus SNV and DSC based on the lowest SECV values obtained prioritising LIN, NEO and total HCN equivalent. The statistics for this model and its performance in an independent validation set are summarized in Table 4.2.

Validation of the calibration

The SEP and RPD are the parameters generally found in literature to assess the goodness of fit of a calibration model. The SEP is comparable to the expected accuracy since it represents the averaged error for the validation set. Since SEP scales with measurement range, the RPD, which is the ratio of performance to deviation (SD/SEP), is used as a standardized measure to compare the quality of calibrations. Assessment is based on a scale provided by Chang et al. (2001) defining RPD values into three categories: >2 are excellent models, 1.4<RPD<2 are considered fair and <1.4 are non-reliable.

Table 4.2 Peformance statistics for the most successful calibration model applied to the calibration and validation datasets

	Linamarin	Lotaustralin	Linustatin	Neolinustatin	Total HCN
Calibration					
N	297	211	304	302	299
Mean	0.73	0.33	2.34	1.90	2.32
SD*	0.21	0.19	0.13	0.21	0.12
Min	0.10	0.00	1.97	1.28	1.95
Max	1.36	0.91	2.72	2.51	2.69
SEC*	0.14	0.14	0.07	0.09	0.06
R^{2*}	0.58	0.48	0.67	0.80	0.79
SECV*	0.14	0.18	0.10	0.12	0.08
1-VR*	0.55	0.18	0.33	0.69	0.56
Validation					
R^{2*}	0.67	0.17	0.51	0.73	0.68
SEP*	0.14	0.26	0.08	0.11	0.07
SD*	0.17	0.14	0.10	0.19	0.11
Bias	-0.03	0.04	-0.01	0.01	-0.01
Slope	1.10	0.89	0.78	0.92	0.91
RPD*	1.72	1.09	1.37	1.88	1.75

*SD: standard deviation; SEC: standard error of calibration; R^2: coefficient of determination; SECV: standard error of cross validation; VR: variance ratio; SEP: standard error of prediction; RPD: relative percent difference

The calibration was most successful for linamarin, NEO and total HCN where the validation set gave RPD values of 1.72, 1.88 and 1.75, respectively, suggesting a fair calibration. Scatter plots comparing actual and predicted values of the calibration on the validation dataset are shown in Figure 4.1. Fox et al. (2012) have reported a successful NIRS calibration for the HCN equivalent of dhurrin accumulation using a 2,4,4,1 MPLS model. They obtained a SEP of 0.057 and an RPD of 3.4. We obtained similar SEP values but with a much lower RPD, possibly attributable to a much larger available range (520-2450ppm) and better uniformity of their calibration sample set. The calibration by Fox et al. (2012) also used ground rather than whole seed samples, another difference that may explain the variance between our results since physical characteristics, such as surface reflectance, can cause higher absorption peaks to be more compressed than small peaks thereby reducing sensitivity. Bias (intercept) is an estimate of the systematic error and can occur due to test samples having variation not captured in the calibration, drifts in instrument performance, wet chemistry, environment of the instrument or changes in sample preparation. Since none of the bias values for any of the components were significant, the calibrations were not significantly affected by systematic error. Residual values assess how well a linear model describes the data. Standardized residual plots for LIN, NEO, diglucosides and total HCN are shown in Figure 4.2. Random patterning around zero and even spread of the residuals indicate homoscedasticity denoting that transformation of the predicted values was not necessary.

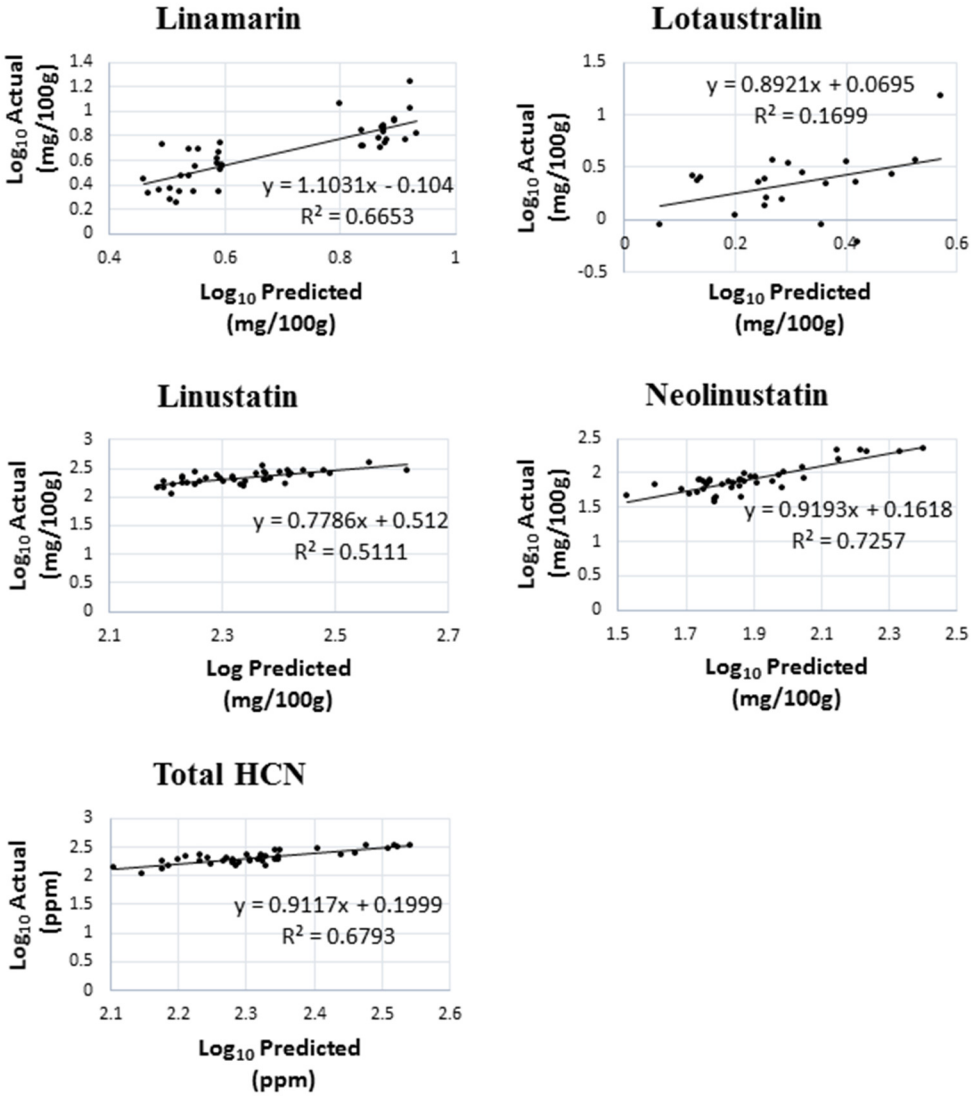

Figure 4.1 Scatter plots for actual versus predicted $\log_{10}$ values of the calibration upon application to the validation dataset

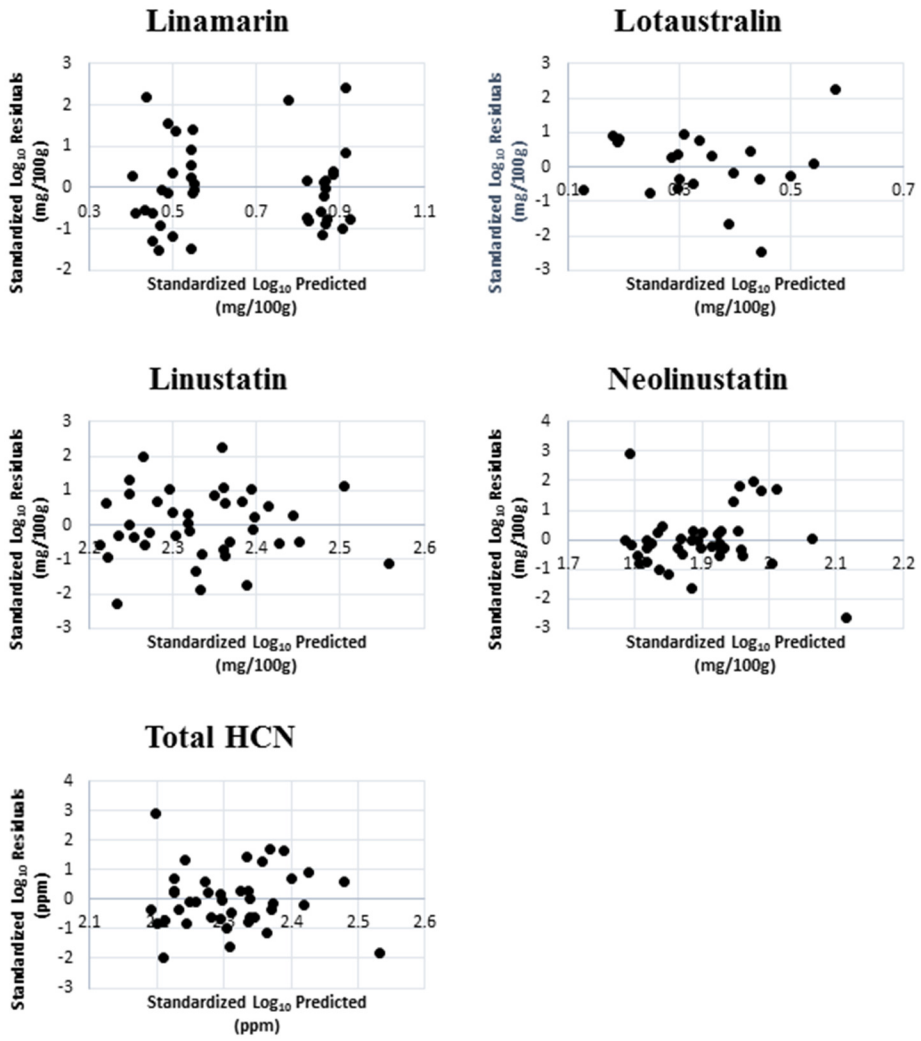

Figure 4.2 Standardized residual plots for predictions of the constituents

Conclusion

The primary objective of validating any analytical method is to demonstrate suitability for its intended purpose. At the moment, the utility of the calibration is limited to a cursory screening of breeding material providing a general ranking of high to low values, which

for breeding objectives, may be sufficient in selecting the lines for genetic advance. As such, this calibration remains "under development" until additional experimental data can complement the existing ones to enhance the calibration. Future research should be geared towards incorporating more samples and further validation should focus on demonstrating specificity, linearity, range, accuracy and precision of the calibration. A robust multivariate NIRS calibration is highly useful as a rapid analysis tool for determining CG content along with other relevant traits for quality control in the grain industry. The robustness of the calibration created critically depends on the selection of samples. Even when a robust calibration has been developed, recalibration on a year to year basis may be necessary due to the comparatively unstable variability of CGs in various environments. Previous research on triticale, a highly phenotypically temperamental crop, has suggested that a spectral adaptation strategy relying more on the most recent samples rather than a database expansion strategy offers the best precision and accuracy, indicating capture of variability is more important than the absolute number of samples (Igne et al. 2007). Our future sampling strategy will prioritize variability because the current key limitation to this calibration may be the narrow range of concentrations. Incorporation of more extreme values into the calibration is anticipated to improve its robustness. Germplasm containing more extreme values of CGs may be developed through selection of transgressive segregants in a CG segregating population such as Double Low x AC McDuff for multiple environments. Alternatively, if the genes for accumulation are known, then RNA interference or an overexpressing promoter may be used, as has been done successfully in

M. esculenta (Jørgensen et al. 2005). The mark of a truly robust calibration is the ability to predict data from different environments. Here, we used samples from three environments but the limited number of samples required that our calibration set be a random selection across all three.

This study demonstrates that CG content can be estimated using an NIRS calibration at an accuracy level suitable for screening purposes. The calibration model presented here is a good starting point. Integration of samples covering a greater range and of more variable composition will increase the model's robustness.

5.0 GENERAL DISCUSSION AND CONCLUSION

5.1 Overall assessment

The overarching objective of this project at its inception was to generate a better understanding of the genetic and environmental factors that affect cyanogenic glucoside (CG) content in linseed and to equip Canadian flax breeders with the knowledge and tools to develop breeding strategies to reduce CG content and consequently improve the marketability of linseed. This research was a component of the Total Utilization of Flax GENomics (TUFGEN) project, a Genome Canada scientific research project designed to improve the utility of flax. The accumulation of CGs contravenes trade regulations in international markets and poses a safety risk in feed products, hence the development of a consistently low CG linseed cultivar is an important breeding objective. Monitoring and exploitation of this trait in breeding strategies is difficult because of the strong environmental effect and the prohibitive cost of screening. Here, molecular markers strongly linked to major effect QTL were identified and a first NIRS calibration was developed - both have immediate applications in breeding.

The identified markers can be used to introgress the major QTL from Double Low into an elite variety. Since Double Low is an EMS mutagenic line, a large segregating population will likely be required to remove unwanted mutations. The low CG transgressive segregants in the DLxMD population may also be selected and utilized in the program. The availability of markers is of great use since selection can be performed at a

juvenile stage and run in parallel with markers for other traits of interest, saving a considerable amount of time, labour and field space. An NIRS calibration directly impacts a plant breeding strategy becase of its time and cost effectiveness. CG estimation can be performed simultaneously as with calibrations already developed for a variety of flaxseed components, including oil content, protein and moisture (Siemens and Daun 2005). It will also be able to gauge pleiotropic changes in gene expression for these traits in the candidate lines selected for advancement. Plant breeding to obtain superior genotypes is often considered to be a numbers game. The superiority of a newly developed cultivar is determined by the combination of genes it contains. To obtain the full range of potential genotypic combinations from a cross, a large segregating population is necessary. In addition, a large number of crosses are often needed to break linkage. Often, the breeder is forced to compromise the number of recombinants desired with the resources available. Both molecular markers and the NIRS calibration reduce the resource constraints for selection and testing allowing a greater chance of obtaining a superior genotype.

It is tempting to speculate on the identities of the QTL. The CG biosynthesis and degradation pathways are unusually simple and straightforward, lacking most of the complexities of interactive effects that other metabolic pathways demonstrate. Transformation of the three *Sorghum bicolor*-derived biosynthesis enzymes into *A. thaliana* was sufficient to enable dhurrin production without any pleiotropic effects (Kristensen et al. 2005). Likewise, removal of all these components seems to disable CG production without hampering other pathways (Siritunga and Sayre, 2004). Considering

that the regulation of CG levels appears to be a stand-alone system, the identities of the detected QTL are hypothesized to be biosynthesis-related genes or modifying agents of these genes, such as transcription factors, epigenetic factors or cis-regulatory elements. Since the QTL peak locations for linustatin (LIN) and neolinustatin (NEO) were unlinked, the underlying functional units are not likely to concern the degradation pathway which is indiscriminatory to LIN or NEO. These discrete peaks suggest a degree of independence in the biosynthesis pathways for LIN and NEO. The repeated occurrence of the CG biosynthetic genes in clusters along the genomes of cyanogenic plants allow for intelligent guesses on possible regions to target. A cursory visual inspection of the sequence alignment for these clusters using the Tablet software showed that all were mutational hot spots. The maintenance of both functional and non-functional CG alleles within a cluster is thought to be due to the effect of balancing selection from opposing selective forces (Ballhorn et al. 2010, Lankau 2007). Despite a conserved pathway for the three biosynthesis enzymes, the sequences of the enzymes are highly variable across species. On a more general level, the presence of hypermutable gene copies and the existence of different chemotypes may reflect adaptation to pathogens and herbivores that have specialised to circumvent cyanogenic defense (Bjarnholt et al. 2008, Engler et al. 2000, Nielsen et al. 2006, Zagrobelny et al. 2007b).

The results of the ANOVA indicated a stronger contribution to variation of environment compared to the genotype for LIN and vice-versa for NEO, differing from previous findings by Oomah et al. (1992) who indicated that the genotype was the most

influential in determining variability in LIN content and that genotype and environment contributed equally to NEO content for ten Canadian flax cultivars. Based on our ANOVA, selection for the AC McDuff alleles for these two traits in a plant breeding program will mean the resulting phenotypic change for NEO will be more sustained across environments as compared to LIN. The G×E effects for LIN (10.5%) and NEO (12.6%) were fairly similar and may provide a general estimate of the stability of the response curve a breeder might expect across environments. The use of an NIRS for approximately half of the Morden and Regina 2012 samples may be the cause of an overestimation of the environmental sum of squares values and an exaggeration of actual proportional significance between environment and genotype.

In testing out multiple approaches to an NIRS calibration, we found that LIN consistently gave lower R^2 and higher SEP values than NEO despite changes in model, pretreatment, calibration sample number and distribution. This was unexpected because LIN is more abundant than NEO in the seed and the two glucosides share almost an identical chemical structure aside from NEO having an ethyl group instead of a methyl group on C2. The differential optical properties may be due to the chirality of the centre bearing the nitrile group or possibly the existence of epimers like dhurrin and taxiphyllin (Zilg and Conn 1974). Fox et al. (2012) have reported a successful NIRS calibration for the HCN equivalent of dhurrin accumulation using a 2,4,4,1 MPLS model. They obtained a validation R^2 of 0.824, SEP of 0.057 and an RPD of 3.4. We obtained similar SEP values but lower RPD, possibly attributable to the larger range (520-2450ppm) and better

uniformity of their calibration sample set. Considering that dhurrin is identical to linamarin except for a phenol instead of a methyl group on C2, we are hopeful that similar success may be achieved with improvement of our calibration.

It is possible that sample morphological differences such as the shape, size and roughness of seed surface may have led to sample-to-sample variations in the overall pathlength in photon travel distance to the detector resulting in light scattering effects that would influence the spectra affecting baseline shifts and scaling variations. Although this is somewhat corrected with SNV and DSC, the altered spectra will still contain some degree of error. The calibration by Fox et al. (2012) for dhurrin in *S. bicolor* used ground rather than whole seed samples, another difference that may explain the variance between our results. However, grinding is destructive and the resultant homogeneity of sample is a luxury that cannot be afforded in early generations of a breeding program. Perhaps grinding could be an option outside breeding programs such as in the grain handling system.

5.2 Future work

The next obvious step is the narrowing down of the LIN and NEO QTL to identify the functional units and improvement of the accuracy of the NIRS calibration to enable quality control and maintenance. More specifically, future research objectives would entail:

- Further fine mapping with additional SNPs and recombinants and, validation of the QTL in additional environments with the aim of map-based cloning the functional units underlying low LIN and NEO content in linseed. Alternatively, the candidate

gene approach can be investigated directly using targeted SNPs and, again, additional recombinants. Either way, validation will be required through functional analysis in a transient system, a transgenic approach or through other alternative strategies.

- Application of the existing markers in breeding and validation through field trial assessment.
- Optimization of the NIRS calibration by inclusion of new variations, model update and assessing the 'noise' effect of external perturbations.
- Adjustment of the calibration model(s) for instrument standardization.

5.3 Concluding remarks

This research is the first to identify QTL for CGs in linseed and the first to develop an NIRS calibration that can estimate individual CGs. These findings provide a strong foundation towards developing better quality germplasm and a reliable, cost-effective analytical method for large scale screening. The culmination of these results will prove highly useful in flax breeding programs and will hopefully improve the utility of flax.

6.0 LITERATURE CITED

Abuye C, Kelbessa U, Wolde-Gebriel S (1998) Health effects of cassava consumption in south Ethiopia. East Afr Med J 75:166–170

Adewusi SR (1990) Turnover of dhurrin in green sorghum seedlings. Plant Physiol 94:1219-1224

Adlercreutz H, Mazur W (1997) Phyto-oestrogens and Western diseases. Ann Med 29:95-120

Ajuyah AO, Ahn DU, Hardin RT, Sim JS (1993) Dietary antioxidants and storage affect chemical characteristics of ω-3 fatty acid enriched broiler chicken meats. J Food Sci 58:43-46. doi: 10.1111/j.1365-2621.1993.tb03206.x

Alonso-Amelot ME, Oliveros A (2000) A method for the practical quantification and kinetic evaluation of cyanogenesis in plant material. Application to *Pteridium aquilinum* and *Passiflora capsularis*. Phytochem Anal 11:309-316. doi: 10.1002/1099-1565(200009/10

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403-410. doi: 10.1016/S0022-2836(05)80360-2

Alves A, Setter (2004) Abscisic acid accumulation and osmotic adjustment in cassava under water deficit. Environ Exp Bot 51:259-271

Andersen MD, Busk PK, Svendsen I, Møller BL (2000) Cytochromes P-450 from cassava (*Manihot esculenta* Crantz) catalyzing the first steps in the biosynthesis of the cyanogenic glucosides linamarin and lotaustralin. Cloning, functional expression

in *Picha pastoris*, and substrate specificity of the isolated recombinant enzymes. J Biol Chem 275:1966-1975

Bacala R, Barthet V (2007) Development of extraction and gas chromatography analytical methodology for cyanogenic glycosides in flaxseed (*Linum usitatissimum*). J AOAC Int 90:153-161

Bak S, Kahn RA, Nielsen HL, Møller BL, Halkier BA (1998) Cloning of three A-type cytochromes P450, CYP71E1, CYP98, and CYP99 from *Sorghum bicolor* (L.) Moench by a PCR approach and identification by expression in *Escherichia coli* of CYP71E1 as a multifunctional cytochrome P450 in the biosynthesis of the cyanogenic glucoside dhurrin. Plant Mol Biol 36:393-405

Ballhorn DJ, Heil M, Lieberei R (2006) Phenotypic plasticity of cyanogenesis in lima bean Phaseolus lunatus—activity and activation of β-glucosidase. J Chem Ecol 32:261-275

Ballhorn DJ, Pietrowski A, Lieberei R (2010) Direct trade-off between cyanogenesis and resistance to a fungal pathogen in lima bean (Phaseolus lunatus L.). J Ecol 98:226-236

Barthet VJ, Bacala R (2010) Development of optimized extraction methodology for cyanogenic glycosides from flaxseed (*Linum usitatissimum*). J AOAC Int 93:478-484

Beavis WD (1994) The power and deceit of QTL experiments: Lessons from comparative QTL studies. In: Proc 49th Annu Corn and Sorghum Res Conf. American Seed Trade Association, Washington, pp 250-266

Benham J, Jeung J-U, Jasieniuk M, Kanazin V, Blake T (1999) Genographer: a graphical tool for automated fluorescent AFLP and microsatellite analysis. J Agric Genomics 4

Bjarnholt N, Rook F, Motawia MS, Cornett C, Jørgensen C, Olsen CE, Jaroszewski JW, Bak S, Møller BL (2008) Diversification of an ancient theme: hydroxynitrile glucosides. Phytochemistry 69:1507-1516

Bogatek R, Dziewanowska K, Lewak S (1991) Hydrogen cyanide and embryonal dormancy in apple seeds. Physiol Plant 83:417-421. doi: 10.1111/j.1399-3054.1991.tb00114.x

Bordo D, Bork P (2002) The rhodanese/Cdc25 phosphatase superfamily. EMBO Rep 3:741-746. doi: 10.1093/embo-reports/kvf150

Bough WA, Gander JE (1971) Exogenous L-tyrosine metabolism and dhurrin turnover in sorghum seedlings. Phytochemistry 10:67-77. doi: 10.1016/S0031-9422(00)90252-8

Bradbury JH (2009) Development of a sensitive picrate method to determine total cyanide and acetone cyanohydrin contents of gari from cassava. Food Chem 113:1329-1333. doi: 10.1016/j.foodchem.2008.08.081

Bradbury JH, Bradbury MG, Egan SV (1994) Comparison of methods of analysis of cyanogens in cassava. .Acta Hortic 375:87-96

Busk PK, Møller BL (2002) Dhurrin synthesis in sorghum is regulated at the transcriptional level and induced by nitrogen fertilization in older plants. Plant Physiol 129:1222-1231. doi: 10.1104/pp.000687

Butler GW, Conn EE (1964) Biosynthesis of the cyanogenic glucosides linamarin and lotaustralin I. Labeling studies in vivo with *Linum usitatissimum*. J Biol Chem 239:1674-1679

Caoa N (2013) Calibration sample selection from a large data pool of soybean in near-infrared spectroscopy. In: Calibration optimization and efficiency in near infrared spectroscopy, pp 84-102

Caston L, Squires E, Leeson S (1994) Hen performance, egg quality, and the sensory evaluation of eggs from SCWL hens fed dietary flax. Can J of Anim Sci 74:347-353

Chadha RK, Lawrence JF, Ratnayake WM (1995) Ion chromatographic determination of cyanide released from flaxseed under autohydrolysis conditions. Food Addit Contam 12:527-533

Chang C-W, Laird DA, Mausbach MJ, Hurburgh CR (2001) Near-infrared reflectance spectroscopy–principal components regression analyses of soil properties. Soil Sci Soc Am J 65:480-490

Chu C, Niu Z, Zhong S, Chao S, Friesen TL, Halley S, Elias EM, Dong Y, Faris JD, Xu SS (2011a) Identification and molecular mapping of two QTLs with major effects for resistance to Fusarium head blight in wheat. Theor Appl Genet 123:1107-1119

Chu HY, Wegel E, Osbourn A (2011b) From hormones to secondary metabolism: the emergence of metabolic gene clusters in plants. Plant J 66:66-79

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

Cloutier S, Miranda E, Ward K, Radovanovic N, Reimer E, Walichnowski A, Datla R, Rowland G, Duguid S, Ragupathy R (2012a) Simple sequence repeat marker development from bacterial artificial chromosome end sequences and expressed sequence tags of flax (*Linum usitatissimum* L.). Theor Appl Genet 125:685-694. doi: 10.1007/s00122-012-1860-4

Cloutier S, Ragupathy R, Miranda E, Radovanovic N, Reimer E, Walichnowski A, Ward K, Rowland G, Duguid S, Banik M (2012b) Integrated consensus genetic and physical maps of flax (*Linum usitatissimum* L.). Theor Appl Genet 125:1783-1795. doi: 10.1007/s00122-012-1953-0

Cobb JN, DeClerck G, Greenberg A, Clark R, McCouch S (2013) Next-generation phenotyping: requirements and strategies for enhancing our understanding of genotype–phenotype relationships and its relevance to crop improvement. Theor Appl Genet 126:867-887. doi: 10.1007/s00122-013-2066-0

Conn EE (1979) Biosynthesis of cyanogenic glycosides. Naturwissenschaften 66:28-34

Conn EE (1980) Cyanogenic compounds. Annu Rev Plant Physiol 31:433-451

Conn EE (1994) Cyanogenesis-a personal perspective. Acta Hortic 375:31-44

Cutler AJ, Conn EE (1981) The biosynthesis of cyanogenic glucosides in *Linum usitatissimum* (linen flax) in vitro. Arch Biochem Biophys 212:468-474

Cutler AJ, Sternberg M, Conn EE (1985) Properties of a microsomal enzyme system from *Linum usitatissimum* (linen flax) which oxidizes valine to acetone cyanohydrin and isoleucine to 2-methylbutanone cyanohydrin. Arch Biochem Biophys 238:272-279

Donato DB, Nichols O, Possingham H, Moore M, Ricci PF, Noller BN (2007) A critical review of the effects of gold cyanide-bearing tailings solutions on wildlife. Environ Int 33:974-984

Eaton SB, Konner M. (1985) Paleolithic nutrition. A consideration of its nature and current implications. New Engl J Med 312:283–289

Echeverry-Solarte M, Ocasio-Ramirez V, Figueroa A, González E, Siritunga D (2013) Expression profiling of genes associated with cyanogenesis in three cassava cultivars containing varying levels of toxic cyanogens. Am J Plant Sci 4:1533-1545. doi: 10.4236/ajps.2013.47185

El-Din El-Assal S, Alonso-Blanco C, Peeters AJ, Raz V, Koornneef M (2001) A QTL for flowering time in Arabidopsis reveals a novel allele of CRY2. Nat Genet 29:435-440

Engler HS, Spencer KC, Gilbert LE (2000) Insect metabolism: Preventing cyanide release from leaves. Nature 406:144-145

Esashi Y, Isuzugawa K, Matsuyama S, Ashino H, Hasegawa R (1991a) Endogenous evolution of HCN during pre-germination periods in many seed species. Physiol Plant 83:27-33

Esashi Y, Matsuyama S, Ashino H, Ogasawara M, Hasegawa R (1991b) β-Glucosidase activities and HCN liberation in unimbibed and imbibed seeds, and the induction of cocklebur seed germination by cyanogenic glycosides. Physiol Plant 83:34-40. doi: 10.1111/j.1399-3054.1991.tb01278.x

Etonihu AC, Olajubu O, Ekanem EO, Bako SS (2011) Titrimetric evaluation of cyanogens in parts of some Nigerian cassava species. Pak J Nutr 10:260-263

European Commission (1971) Issue: 31971L0250. First Commission Directive 71/250/EEC of 15 June 1971 establishing Community methods of analysis for the official control of feeding-stuffs.

European Parliament Council (2002) Directive 2002/32/EC of the European Parliament and of the Council of 7 May 2002 on undesirable substances in animal feed - Council statement. p 10

Fan TW, Conn EE (1985) Isolation and characterization of two cyanogenic β-glucosidases from flax seeds. Arch Biochem Biophys 243:361-373

FAO (2009) FAOSTAT database. Agricultural crops: linseed: area harvested/yield. http://faostat.fao.org. Accessed: September 2013.

Feigl F, Anger V (1966) Replacement of benzidine by copper ethylacetoacetate and tetra base as spot-test reagent for hydrogen cyanide and cyanogen. Analyst 91:282-284

Fernandes G (1995) Effects of calorie restriction and omega-3 fatty acids on autoimmunity and aging. Nutr. Rev. 53:S72-S79

Ferrier LK, Caston LJ, Leeson S, Squires EJ, Celi B, Thomas L Holub BJ (1992) Changes in serum lipids and platelet fatty acid composition following consumption of eggs enriched in alpha-linolenic acid. Food Res Int 25:263-268

Ferrier LK, Caston LJ, Leeson S, Squires EJ, Weaver BJ, Holub BJ (1995) Alpha-linolenic acid and docosahexaenoic acid-enriched eggs from hens fed flaxseed: influence on blood lipids and platelet phospholipid fatty acids in humans. Am J Clin Nutr 62:81-86

Field B, Osbourn AE (2008) Metabolic diversification—independent assembly of operon-like gene clusters in different plants. Science 320:543-547

Flax Council of Canada (2009) GMO flax update, 28 September 2009 (corrected). http://www.flaxcouncil.ca/english/index.jsp?p=section1&mp=main

Flax Council of Canada (date accessed: July, 2014). [Website]. Production Statistics. http://www.flaxcouncil.ca/english/index.jsp?p=statistics2&mp=statistics

Food and Drug Act (2009) GRAS Notice 000280: Whole and Milled Flaxseed. Washington, DC: US Food and Drug Administration

Forslund K, Jonsson L (1997) Cyanogenic glycosides and their metabolic enzymes in barley, in relation to nitrogen levels. Physiol Plant 101:367-372. doi: 10.1111/j.1399-3054.1997.tb01010.x

Forslund K, Morant M, Jørgensen B, Olsen CE, Asamizu E, Sato S, Tabata S, Bak S (2004) Biosynthesis of the nitrile glucosides rhodiocyanoside A and D and the cyanogenic glucosides lotaustralin and linamarin in *Lotus japonicus*. Plant Physiol 135:71-84

Fox GP, O'Donnell NH, Stewart PN, Gleadow RM (2012) Estimating hydrogen cyanide in forage sorghum (*Sorghum bicolor*) by near-infrared spectroscopy. J Agric Food Chem 60:6183-6187

Frehner M, Scalet M, Conn EE (1990) Pattern of the cyanide-potential in developing fruits implications for plants accumulating cyanogenic monoglucosides (*Phaseolus lunatus*) or cyanogenic diglucosides in their seeds (*Linum usitatissimum*, *Prunus amygdalus*). Plant Physiol 94:28-34

Fry WE, Evans PH (1977) Association of formamide hydro-lyase with fungal pathogenicity to cyanogenic plants. Phytopathology 67:1001-1006

García I, Rosas T, Bejarano ER, Gotor C, Romero LC (2013) Transient transcriptional regulation of the CYS-C1 gene and cyanide accumulation upon pathogen infection in the plant immune response. Plant Physiol 162:2015-2027. doi: 10.1104/pp.113.219436

Gleadow RM, Foley WJ, Woodrow IE (1998) Enhanced CO2 alters the relationship between photosynthesis and defence in cyanogenic *Eucalyptus cladocalyx* F. Muell. Plant Cell Environ 21:12-22. doi: 10.1046/j.1365-3040.1998.00258.x

Gleadow RM, Woodrow IE (2002) Mini-review: constraints on effectiveness of cyanogenic glycosides in herbivore defense. J Chem Ecol 28:1301-1313

Gleadow RM, Møller BL (2014) Cyanogenic Glycosides: Synthesis, Physiology, and Phenotypic Plasticity. Annu Rev Plant Bio 65: 155-185

Gniazdowska A, Krasuska U, Czajkowska K, Bogatek R (2010) Nitric oxide, hydrogen cyanide and ethylene are required in the control of germination and undisturbed development of young apple seedlings. Plant Growth Regul 61:75-84. doi: 10.1007/s10725-010-9452-2

Goff BM, Moore KJ, Fales SL, Pedersen JF (2011) Comparison of gas chromotography, spectrophotometry and near infrared spectroscopy to quantify prussic acid potential in forages. J Sci Food Agri 91:1523-1526. doi: 10.1002/jsfa.4366

Gleadow R, Moller BL (2014) Cyanogenic Glycosides: Synthesis, Physiology and Phenotypic Plasticity. Annu Rev Plant Bio 65:155-185

Gregory M, Geier M, Gibson R James M (2013) Functional characterization of the chicken fatty acid elongases. J Nutr 143:12-16

Guilbault GG, Kramer DN (1966) Ultra sensitive, specific method for cyanide using p-nitrobenzaldehyde and o-dinitrobenzene. Anal Chem 38:834-836

Hahlbrock K, Conn EE (1971) Evidence for the formation of linamarin and lotaustralin in flax seedlings by the same glucosyltransferase. Phytochemistry 10:1019-1023. doi: 10.1016/S0031-9422(00)89932-x

Hamberg M Samuelsson B (1975) Thromboxanes: a new group of biologically active compounds derived from prostaglandin endoperoxides. Proc Natl Acad Sci 72:2994-2999

Haque MR, Bradbury JH (2002) Total cyanide determination of plants and foods using the picrate and acid hydrolysis methods. Food Chem 77:107-114. doi: 10.1016/S0308-8146(01)00313-2

Harris JR, Merson GHJ, Hardy MJ, Curtis DJ (1980) Determination of cyanide in animal feeding stuffs. Analyst 105:974-980. doi: 10.1039/AN9800500974

Hisano H, Sato S, Isobe S, Sasamoto S, Wada T, Matsuno A, Fujishiro T, Yamada M, Nakayama S, Nakamura Y, Watanabe S, Harada K, Tabata S (2007) Characterization of the soybean genome using EST-derived microsatellite markers. DNA Res 14:271-281. doi: 10.1093/dnares/dsm025

Hösel W, Berlin J, Hanzlik TN, Conn EE (1987) In-vitro biosynthesis of 1-(4′-hydroxyphenol)-2-nitroethane and production of cyanogenic compounds in osmotically stressed cell suspension cultures of *Eschscholtzia californica.* Cham Planta 166:176-181

Huang XQ, Cloutier S, Lycar L, Radovanovic N, Humphreys DG, Noll JS, Somers DJ, Brown PD (2006) Molecular detection of QTLs for agronomic and quality traits in a doubled haploid population derived from two Canadian wheats (*Triticum aestivum* L.). Theor Appl Genet 113:753-766

Hughes M (1981) The genetic control of plant cyanogenesis. In:Vennesland B et al. (eds) Cyanide in biology. Academic Press, New York, pp 495-508

Igne B, Gibson LR, Rippke GR, Hurburgh Jr CR (2007) Influence of yearly variability of agricultural products on calibration process: a triticale example. Cereal Chem 84:576-581

Imran M, Anjum FM, Butt MS, Siddiq M, Sheikh MA (2013) Reduction of cyanogenic compounds in flaxseed (*Linum usitatissimum* L.) meal using thermal treatment. Int J Food Prop 16:1809-1818

James MJ, UrsinVM Cleland LG (2003) Metabolism of stearidonic acid in human subjects: comparison with the metabolism of other n-3 fatty acids. Am J Clin Nutr 77:1140-1145

Jaroszewski JW, Olafsdottir ES, Wellendorph P, Christensen J, Franzyk H, Somanadhan B, Budnik BA, Jørgensen LB, Clausen V (2002) Cyanohydrin glycosides of *Passiflora*: distribution pattern, a saturated cyclopentane derivative from *P. guatemalensis*, and formation of pseudocyanogenic α-hydroxyamides as isolation artefacts. Phytochemistry 59:501-511

Jensen NB, Zagrobelny M, Hjernø K, Olsen CE, Houghton-Larsen J, Borch J, Møller BL, Bak S (2011) Convergent evolution in biosynthesis of cyanogenic defence compounds in plants and insects. Nat Commun 2:273. doi: 10.1038/ncomms1271

Jones DA (1988) Cyanogenesis in animal-plant interactions. In: Evered D, Harnett S (eds) Ciba Foundation Symposium 140 - Cyanide compounds in biology. J. Wiley, Chichester, UK, pp 151-165

Jones PR, Møller BL, Høj PB (1999) The UDP-glucose: p-hydroxymandelonitrile-O-glucosyltransferase that catalyzes the last step in synthesis of the cyanogenic glucoside dhurrin in *Sorghum bicolor*: isolation, cloning, heterologous expression, and substrate specificity. J Biol Chem 274:35483-35491

Jørgensen K, Bak S, Busk PK, Sørensen C, Olsen CE, Puonti-Kaerlas J, Møller BL (2005) Cassava plants with a depleted cyanogenic glucoside content in leaves and tubers. Distribution of cyanogenic glucosides, their site of synthesis and transport, and blockage of the biosynthesis by RNA interference technology. Plant Physiol 139:363-374

Jørgensen K, Morant AV, Morant M, Jensen NB, Olsen CE, Kannangara R, Motawia MS, Møller BL, Bak S (2011) Biosynthesis of the cyanogenic glucosides linamarin and lotaustralin in cassava: isolation, biochemical characterization, and expression pattern of CYP71E7, the oxime-metabolizing cytochrome P450 enzyme. Plant Physiol 155:282-292

Kadow D, Voß K, Selmar D, Lieberei R (2012) The cyanogenic syndrome in rubber tree *Hevea brasiliensis*: tissue-damage-dependent activation of linamarase and hydroxynitrile lyase accelerates hydrogen cyanide release. Ann Bot 109:1253-1262. doi: 10.1093/aob/mcs057

Kahn RA, Bak S, Svendsen I, Halkier BA, Møller BL (1997) Isolation and reconstitution of cytochrome P450ox and in vitro reconstitution of the entire biosynthetic pathway of the cyanogenic glucoside dhurrin from sorghum. Plant Physiol 115:1661-1670

Kenaschuk EO, Rashid KY (1994) AC McDuff flax. Can J Plant Sci 74:815-816 doi: 10.4141/cjps94-146

Kingsbury JM (1964) Poisonous plants of the United States and Canada. Soil Sci 98:349

Kizito EB, Rönnberg-Wästljung AC, Egwang T, Gullberg U, Fregene M, Westerbergh A (2007) Quantitative trait loci controlling cyanogenic glucoside and dry matter content in cassava (*Manihot esculenta* Crantz) roots. Hereditas 144:129-136

Kobaisy M, Oomah B, Mazza G (1996) Determination of cyanogenic glycosides in flaxseed by barbituric acid-pyridine, pyridine-pyrazolone, and high-performance liquid chromatography methods. J Agric Food Chem 44:3178-3181

Koch B, Nielsen VS, Halkier BA, Olsen CE, Møller BL (1992) The biosynthesis of cyanogenic glucosides in seedlings of cassava (*Manihot esculenta* Crantz). Arch Biochem Biophys 292:141-150

Kongsawadworakul P, Viboonjun U, Romruensukharom P, Chantuma P, Ruderman S, Chrestin H (2009) The leaf, inner bark and latex cyanide potential of *Hevea brasiliensis*: Evidence for involvement of cyanogenic glucosides in rubber yield. Phytochemistry 70:730-739. doi: 10.1016/j.phytochem.2009.03.020

Kosambi DD (1944) The estimation of map distances from recombination values. Ann Eugenics 12:172-175. doi: 10.1111/j.1469-1809.1943.tb02321.x

Krech MJ, Fieldes MA (2003) Analysis of the developmental regulation of the cyanogenic compounds in seedlings of two lines of *Linum usitatissimum* L. Can J Bot 81:1029-1038. doi: 10.1139/b03-097

Kriedeman PE, Neales TF, Ashton DH (1964) Photosynthesis in relation to leaf orientation and light interception. Aust J Biol Sci USA 17:591-600. doi: 10.1071/B19640591

Kristensen C, Morant M, Olsen CE, Ekstrøm CT, Galbraith DW, Møller BL, Bak S (2005) Metabolic engineering of dhurrin in transgenic Arabidopsis plants with marginal inadvertent effects on the metabolome and transcriptome. Proc Natl Acad Sci 102:1779-1784

Kumar S, Banks TW, Cloutier S (2012) SNP discovery through next-generation sequencing and its applications. Int J Plant Genomics:831460. doi: 10.1155/2012/831460

Lands WE, LeTellier PR, Rome LH Vanderhoek JY (1973) Inhibition of prostaglandin biosynthesis. Adv Biosci 9:15-27

Lang K (1933) Thiocyanate formation in the animal body. Biochem Z 259:243-256

Lankau RA (2007) Specialist and generalist herbivores exert opposing selection on a chemical defense. New Phytol 175:176-184

Leuschner J, Winkler A, Leuschner F (1991) Toxicokinetic aspects of chronic cyanide exposure in the rat. Toxicol Lett 57:195–201.

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. doi: 10.1093/bioinformatics/btp352

Liddle S, Keresztessy Z, Hughes J, Hughes MA (1998) A genomic cyanogenic beta-glucosidase gene from cassava (accession no. X94986) (PGR 98-148). Plant Physiol 117:1526

Lin C-S, Poushinsky G (1985) A modified augmented design (type 2) for rectangular plots. Can J Plant Sci 65:743-749

Madhusudhan KT, Ramesh HP, Ogawa T, Sasaoka K, Singh N (1986) Detoxification of commercial linseed meal for use in broiler rations. Poult Sci 65:164-171

Majak W, McDiarmid RE, Hall JW, Cheng KI (1990) Factors that determine rates of cyanogenesis in bovine ruminal fluid in vitro. J Anim Sci 68:1648-1655

Mazza G, Oomah B (1995) Flaxseed, dietary fiber, and cyanogens. In: Cunnane S, Thompson LU (eds) Flaxseed in human nutrition, AOCS Press, Champaign, Illinois, pp56-81

Miller JM, Conn EE (1980) Metabolism of hydrogen cyanide by higher plants. Plant Physiol 65:1199-1202

Miller RE, Gleadow RM, Cavagnaro TR (2014) Age versus stage: Does ontogeny modify the effect of phosphorous and arbuscular mycorrhizas on above- and below- ground defence in forage sorghum? Plant Cell Environ 37:929-42

Milne I, Stephen G, Bayer M, Cock PJ, Pritchard L, Cardle L, Shaw PD, Marshall D (2013) Using Tablet for visual exploration of second-generation sequencing data. Briefings Bioinf 14:193-202

Møller BL, Conn EE (1980) The biosynthesis of cyanogenic glucosides in higher plants. Channeling of intermediates in dhurrin biosynthesis by a microsomal system from *Sorghum bicolor* (linn) Moench. J Biol Chem 255:3049-3056

Montgomery R (1969) Cyanogens. In: Liener I (ed) Toxic constituents of plant foodstuffs. Academic Press, pp 143-157

Morant AV, Jørgensen K, Jørgensen B, Dam W, Olsen CE, Møller BL, Bak S (2007) Lessons learned from metabolic engineering of cyanogenic glucosides. Metabolomics 3:383-398

Morant M, Bak S, Møller BL, Werck-Reichhart D (2003) Plant cytochromes P450: tools for pharmacology, plant protection and phytoremediation. Curr Opin Biotechnol 14:151-162

Mühlhausen S, Kollmar M (2013) Whole genome duplication events in plant evolution reconstructed and predicted using myosin motor proteins. BMC Evol Biol 13:202

Nahrstedt A (1985) Cyanogenic compounds as protecting agents for organisms. Plant Syst Evol 150:35-47

Naoumkina M, Farag MA, Sumner LW, Tang Y, Liu C-J, Dixon RA (2007) Different mechanisms for phytoalexin induction by pathogen and wound signals in *Medicago truncatula*. Proc Natl Acad Sci USA 104:17909-17915

Neuringer M, Andersen GJ Conner WE (1998) The essentiality of ω-3 fatty acids for the development and function of the retina and brain. Annu. Rev. Nutr. 8:517-541

Niedźwiedź-Siegień I (1998) Cyanogenic glucosides in *Linum usitatissimum*. Phytochemistry 49:59-63. doi: 10.1016/S0031-9422(97)00953-9

Niedźwiedź-Siegień I, Gierasimiuk A (2001) Environmental factors affecting the cyanogenic potential of flax seedlings. Acta Physiol Plant 23:383-390

Nielsen KA, Hrmova M, Nielsen JN, Forslund K, Ebert S, Olsen CE, Fincher GB, Møller BL (2006) Reconstitution of cyanogenesis in barley (Hordeum vulgare L.) and its implications for resistance against the barley powdery mildew fungus. Planta 223:1010-1023

Nielsen KA, Tattersall DB, Jones PR, Møller BL (2008) Metabolon formation in dhurrin biosynthesis. Phytochemistry 69:88-98

Nikus J, Daniel G, Jonsson LM (2001) Subcellular localization of β-glucosidase in rye, maize and wheat seedlings. Physiol Plant 111:466-472

Nutrition Recommendations (1990) Scientific Review Committee. Minister of National Health and Welfare, Ottawa, Canada.

Nützmann H-W, Osbourn A (2014) Gene clustering in plant specialized metabolism. Curr Opin Biotechnol 26:91-99. doi: 10.1016/j.copbio.2013.10.009

Oh SY, Ryue J, Hseih CH, Bell DE (1991) Eggs enriched in omega-3 fatty acids and alterations in lipid concentrations in plasma and lipoproteins and in blood pressure. Am J Clin Nutr 54:689-695

Oluwole O, Onabolu A, Cotgreave I, Rosling H, Persson A, Link H (2003) Incidence of endemic ataxic polyneuropathy and its relation to exposure to cyanide in a Nigerian community. J Neurol Neurosurg Psychiatry 74:1417-1422

Oomah BD, Mazza G, Kenaschuk EO (1992) Cyanogenic compounds in flaxseed. J Agric Food Chem 40:1346-1348

Oracz K, El-Maarouf-Bouteau H, Bogatek R, Corbineau F, Bailly C (2008) Release of sunflower seed dormancy by cyanide: cross-talk with ethylene signalling pathway. J Exp Bot 59:2241-2251. doi: 10.1093/jxb/ern089

Oracz K, El-Maarouf-Bouteau H, Kranner I, Bogatek R, Corbineau F, Bailly C (2009) The mechanisms involved in seed dormancy alleviation by hydrogen cyanide unravel the role of reactive oxygen species as key factors of cellular signaling during germination. Plant Physiol 150:494-505. doi: 10.1104/pp.109.138107

Osbourn A (2010) Gene clusters for secondary metabolic pathways: an emerging theme in plant biology. Plant Physiol 154:531-535

Pandalai PK, Pilat MJ, Yamazaki K, Naik H Pienta KJ (1996) The effects of omega-3 and omega-6 fatty acids on *in vitro* prostate cancer growth. Anticancer Res. 16:815-820.

Pandord J, Williams P (1988) Analysis of oilseeds for protein, oil, fiber and moisture by near-infrared reflectance spectroscopy. Journal of the American Oil Chemists' Society 65:1627-1634

Pauly D, Palomares ML, Froese R, Sa-a P, Vakily M, Preikshot D Wallace S (2001) Fishing down Canadian aquatic food webs. Can J Fish Aquat Sci 58:51-62

Poulton JE (1990) Cyanogenesis in plants. Plant Physiol 94:401-405

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

Rodríguez ML, Alzueta C, Rebolé A, Ortiz LT, Centeno C, Treviño J (2001) Effect of inclusion level of linseed on the nutrient utilisation of diets for growing broiler chickens. Brit Poul Sci 42:368-375

Rose DP (1997) Dietary fatty acids and prevention of hormone-responsive cancer. PSEBM. 216:224-233

Roseling H (1994) Measuring effects in humans of dietary cyanide exposure to sublethal cyanogens from Cassava in Africa. Acta Hortic 375:271-283

Ruxton CH, Reed SC, Simpson MJ, Millington KJ (2004) The health benefits of omega-3 polyunsaturated fatty acids: a review of the evidence. J Hum Nutr Diet 17:449-459

Sa KJ, Park JY, Park K-C, Lee JK (2012) Analysis of genetic mapping in a waxy/dent maize RIL population using SSR and SNP markers. Genes Genomics 34:157-164

Sánchez-Pérez R, Jørgensen K, Olsen CE, Dicenta F, Møller BL (2008) Bitterness in almonds. Plant Physiol 146:1040-1052

Saunders JA, Conn EE (1978) Presence of the cyanogenic glucoside dhurrin in isolated vacuoles from Sorghum. Plant Physiol 61:154-157

Schilcher H, Wilkens-Sauter M (1986) Quantitative Bestimmung cyanogener Glykoside in *Linum usitatissimum*. Mit Hilfe der HPLC. Fette, Seifen, Anstrichmittel 88:287-290

Schön CC, Utz H, Groh S, Truberg B, Openshaw S, Melchinger A (2004) Quantitative trait locus mapping based on resampling in a vast maize testcross experiment and its relevance to quantitative genetics for complex traits. Genetics 167: 485-498

Scientific Review Committee (1990) Nutrition recommendations. Minister of National Health and Welfare. Ottawa, ON, Canada

Selmar D, Lieberei R, Biehl B (1988) Mobilization and utilization of cyanogenic glycosides the linustatin pathway. Plant Physiol 86:711-716

Sibbesen O, Koch B, Halkier BA, Møller B (1994) Isolation of the heme-thiolate enzyme cytochrome P-450TYR, which catalyzes the committed step in the biosynthesis of the cyanogenic glucoside dhurrin in *Sorghum bicolor* (L.) Moench. Proc Natl Acad Sci USA 91:9740-9744

Sibbesen O, Koch B, Halkier BA, Møller BL (1995) Cytochrome P-450TYR is a multifunctional heme-thiolate enzyme catalyzing the conversion of L-tyrosine to p-hydroxyphenylacetaldehyde oxime in the biosynthesis of the cyanogenic glucoside dhurrin in *Sorghum bicolor* (L.) Moench. J Biol Chem 270:3506-3511

Siemens BJ, Daun JK (2005) Determination of the fatty acid composition of canola, flax, and solin by near-infrared spectroscopy. J Am Oil Chem Soc 82:153-157

Siritunga D, Sayre R (2004) Engineering cyanogen synthesis and turnover in cassava (*Manihot esculenta*). Plant Mol Biol 56:661-669

Soto-Cerda BJ, Duguid S, Booker H, Rowland G, Diederichsen A, Cloutier S (2014) Genomic regions underlying agronomic traits in linseed (*Linum usitatissimum* L.) as revealed by association mapping. J Integr Plant Biol 56:75-87 doi: 10.1111/jipb.12118

Statistics Canada. (date accessed: October, 2013). [Website]. http://statcan.gc.ca/

Takos A, Lai D, Mikkelsen L, Abou Hachem M, Shelton D, Motawia MS, Olsen CE, Wang TL, Martin C, Rook F (2010) Genetic screening identifies cyanogenesis-deficient mutants of *Lotus japonicus* and reveals enzymatic specificity in hydroxynitrile glucoside metabolism. Plant Cell 22:1605-1619. doi: 10.1105tpc.109.073502

Takos AM, Knudsen C, Lai D, Kannangara R, Mikkelsen L, Motawia MS, Olsen CE, Sato S, Tabata S, Jørgensen K, Møller BL, Rook F (2011) Genomic clustering of cyanogenic glucoside biosynthetic genes aids their identification in *Lotus japonicus* and suggests the repeated evolution of this chemical defence pathway. Plant J 68:273-286

Tattersall DB, Bak S, Jones PR, Olsen CE, Nielsen JK, Hansen ML, Høj PB, Møller BL (2001) Resistance to an herbivore through engineered cyanogenic glucoside synthesis. Science 293:1826-1828

Taylorson RB, Hendricks SB (1973) Promotion of seed germination by cyanide. Plant Physiol 52:23-27

Temple (1996) Dietary fats and coronary heart disease. Biomed. Pharmacother. 50: 261-268

Thompson LU (2003) Analysis and bioavailability of lignans. In: Cunnane S, Thompson LU (eds) Flaxseed in human nutrition. AOCS Press, Champaign, Illinois, pp 92-116

Tylleskär T, Rosling H, Banea M, Bikangi N, Cooke R, Poulter N (1992) Cassava cyanogens and konzo, an upper motoneuron disease found in Africa. The Lancet 339:208-211

Ubwa ST, Anhwange BA, Chia JT (2011) Chemical analysis of *Tacca leontopetaloides* peels. Am J Food Technol 6:932-938

Untang M, Shiowatana J, Siripinyanond A (2010) A simple cyanide test kit for water and fruit juices. Anal Methods 2:1698-1701Van Ooijen J (2006) JoinMap 4. Software for the calculation of genetic linkage maps in experimental populations. Kyazma BV, Wageningen, The Netherlands

VanEtten HD, Mansfield JW, Bailey JA, Farmer EE (1994) Two classes of plant antibiotics: Phytoalexins versus" Phytoanticipins". Plant Cell 6:1191. doi: 10.1105/tpc.6.9.1191

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

Verdoucq L, Morinière J, Bevan DR, Esen A, Vasella A, Henrissat B, Czjze M (2004) Structural determinants of substrate specificity in family 1 β-glucosidases. Novel insights from the crystal structure of sorghum dhurrinase-1, a plant β-glucosidase with strict specificity, in complex with its natural substrate. J Biol Chem 279:31796-31803

Vetter J (2000) Plant cyanogenic glycosides. Toxicon 38:11-36

Voorrips RE (2002) MapChart: software for the graphical presentation of linkage maps and QTLs. J Hered 93:77-78

Wajant H, Riedel D, Benz S, Mundry K-W (1994) Immunocytological localization of hydroxynitrile lyases from *Sorghum bicolor* L. and *Linum usitatissimum* L. Plant Sci 103:145-154

Wanasundara PKJPD, Amarowicz R, Kara MT, Shahidi F (1993) Removal of cyanogenic glycosides of flaxseed meal. Food Chem 48:263-266

Wang D, Li W, Wan J, Wang C (2004) Cloning and preliminary expression of alpha-hydroxynitrile lyase gene from cassava. Chinese J Appl Environ Biol 10:428-431

Wang S, Basten C, Zeng Z (2007) Windows QTL cartographer 2.5. Department of Statistics, North Carolina State University, Raleigh, NC

Wang Z, Hobson N, Galindo L, Shilin Z, Daihu S, 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 A, GKS Wong, Wang J,

Deyholos MK (2012) The genome of flax (*Linum usitatissimum*) assembled de novo from short shotgun sequence reads. Plant J 72:461-473

Way JL. (1984) Cyanide intoxication and its mechanism of antagonism. Annu Rev Pharmacol Toxicol 24:451–481

Werner JD, Borevitz JO, Warthmann N, Trainer GT, Ecker JR, Chory J, Weigel D (2005) Quantitative trait locus mapping and DNA array hybridization identify an FLM deletion as a cause for natural flowering-time variation. Proc Natl Acad Sci USA102:2460-2465

Williams PC (1975) Application of near infrared reflectance spectroscopy to analysis of cereal grains and oilseeds. Cereal Chemistry 52:561-576

Woodrow IE, Slocum DJ, Gleadow RM (2002) Influence of water stress on cyanogenic capacity in *Eucalyptus cladocalyx*. Funct Plant Biol 29:103-110

Wu C-F, Xu X-M, Huang S-H, Deng M-C, Feng A-J, Peng J, Yuan J-P, Wang J-H (2012) An efficient fermentation method for the degradation of cyanogenic glycosides in flaxseed. Food Addit Contam: Part A 29:1085-1091. doi 10.1080/19440049.2012.680202

Xu S (2003) Theoretical basis of the Beavis effect. Genetics 165:2259-2268

Yamashita T, Sano T, Hashimoto T, Kanazawa K (2007) Development of a method to remove cyanogen glycosides from flaxseed meal. Int J Food Sci Technol 42:70-75. doi: 10.1111/j.1365-2621.2006.01212.x

Yip W-K, Yang SF (1988) Cyanide metabolism in relation to ethylene production in plant tissues. Plant Physiol 88:473-476

You FM, Duguid SD, Thambugala D, Cloutier S (2013) Statistical analysis and field evaluation of the type 2 modified augmented design (MAD) in phenotyping of flax ('*Linum usitatissimum*') germplasms in multiple environments. Aust J Crop Sci 7:1789-1800

Zagrobelny M, Bak S, Rasmussen AV, Jørgensen B, Naumann CM, Møller BL (2004) Cyanogenic glucosides and plant–insect interactions. Phytochemistry 65:293-306

Zagrobelny M, Bak S, Olsen CE, Møller BL (2007a) Intimate roles for cyanogenic glucosides in the life cycle of *Zygaena filipendulae* (Lepidoptera, Zygaenidae). Insect Biochem Mol Biol 37:1189-1197. doi:10.1016/j.ibmb.2007.07.008

Zagrobelny M, Bak S, Ekstrøm CT, Olsen CE, Møller BL (2007b) The cyanogenic glucoside composition of *Zygaena filipendulae* (Lepidoptera: Zygaenidae) as effected by feeding on wild-type and transgenic lotus populations with variable cyanogenic glucoside profiles. Insect Biochem Mol Biol 37:10-18. doi:10.1016/j.ibmb.2006.09.008

Zagrobelny M, Bak S, Møller BL (2008) Cyanogenesis in plants and arthropods. Phytochemistry 69:1457-1468. doi:10.1016/j.phytochem.2008.02.019

Zhang L, Wang S, Li H, Deng Q, Zheng A, Li S, Li P, Li Z, Wang J (2010) Effects of missing marker and segregation distortion on QTL mapping in F_2 populations. Theor Appl Genet 121:1071-1082. doi: 10.1007/s00122-010-1372-z

Zhang P, Bohl-Zenger S, Puonti-Kaerlas J, Potrykus I, Gruissem W (2003) Two cassava promoters related to vascular expression and storage root formation. Planta 218:192-203. doi: 10.1007/s00425-003-1098-0

Zilg H, Conn E (1974) Stereochemical aspects of lotaustralin biosynthesis. J Bio Chem 249: 3112-3115

7.0 APPENDICES

Appendix I Summary of Shapiro-Wilk test for normality for linustatin and neolinustatin values for all three environments

Linustatin	**Morden 2011**	**Morden 2012**	**Regina 2012**	**Overall mean**
mean	213.36	220.26	300.64	242.05
SD[1]	56.23	53.29	85.96	50.45
variance	3162.06	2839.57	7389.67	2544.83
kurtosis	-0.30	1.62	-0.66	-0.09
calculted shapiro-wilk statistic w	0.97	0.94	0.96	0.99
calculated shapiro-wilk p-value	0.00	0.00	0.00	0.27
critical value of w (5%)	0.95	0.95	0.95	0.95
null hypothesis	accept	reject	accept	accept
Neolinustatin				
mean	61.95	66.45	81.78	70.80
SD	15.17	17.19	18.51	14.43
variance	230.03	295.45	342.50	208.13
kurtosis	4.50	2.61	0.06	4.30
calculted shapiro-wilk statistic w	0.93	0.93	0.99	0.93
calculated shapiro-wilk p-value	0.00	0.00	0.21	0.00
critical value of w (5%)	0.95	0.95	0.95	0.95
Null hypothesis	reject	reject	accept	reject

[1] SD: Standard Deviation

Appendix II ANOVA for row and column effects for the MAD II experimental design

Source	DF	SS	MS	F	Prob>	Significance
Morden 2011 linustatin						
Row	4	3980.10	995.02	1.79	0.18	ns
Column	4	165.77	41.44	0.07	0.99	ns
Row x Column	15	8330.40	555.36	1.13	0.39	ns
Whole plot	11	13662.85	1242.08	2.53	0.03	*
Control	2	125921.50	62960.74	128.27	0.00	**
Subplot error	22	10798.87	490.86			
Adjustment unecessary						
Morden 2011 neolinustatin						
Row	4	1521.31	380.33	1.61	0.22	ns
Column	4	364.01	91.00	0.38	0.82	ns
Row x Column	15	3548.09	236.54	1.92	0.08	ns
Whole plot	11	1866.37	169.67	1.38	0.25	ns
Control	2	49529.05	24764.53	200.89	0.00	**
Subplot error	22	2711.97	123.27			
Adjustment unecessary						
Morden 2012 linustatin						
Row	4	0.00	0.00	0.00	1.00	ns
Column	4	0.00	0.00	0.00	1.00	ns
Row x Column	16	4767.07	297.94	0.80	0.67	ns
Whole plot	11	4333.01	393.91	1.06	0.43	ns
Control	2	11604.68	5802.34	15.64	0.00	**
Subplot error	23	8531.35	370.93			
Adjustment unecessary						

Source	DF	SS	MS	F	Prob>	Significance
Morden 2012 neolinustatin						
Row	4	189.46	47.37	1.66	0.21	ns
Column	4	65.61	16.40	0.58	0.68	ns
Row x Column	16	455.95	28.50	1.45	0.20	ns
Whole plot	11	377.85	34.35	1.75	0.12	ns
Control	2	9219.58	4609.79	234.94	0.00	**
Subplot error	23	451.29	19.62			
Adjustment unecessary						
Regina 2012 linustatin						
Row	4	1252.41	313.10	1.18	0.35	ns
Column	4	1690.30	422.58	1.60	0.22	ns
Row x Column	18	4760.60	264.48	0.53	0.92	ns
Whole plot	13	7126.44	548.19	1.10	0.40	ns
Control	2	36515.44	18257.72	36.63	0.00	**
Subplot error	25	12460.75	498.43			
Adjustment unecessary						
Regina 2012 linustatin						
Row	4	195.77	48.94	1.05	0.41	ns
Column	4	179.71	44.93	0.96	0.45	ns
Row x Column	18	838.33	46.57	1.36	0.24	ns
Whole plot	13	460.76	35.44	1.03	0.45	ns
Control	2	11398.52	5699.26	165.85	0.00	**
Subplot error	25	859.08	34.36			
Adjustment unecessary						

Appendix III All QTL detected for the three environments

Trait	LG	Flanking Markers	Location	Position(cM)	LOD	Additive effect	R^2
Linamarin	1	751323-Lu2712	Morden2011	55.61	3.83	1.68	0.09
	1	Lu3283-Lu49B	Morden2012	124.21	4.09	0.50	0.10
Lotaustralin	6	Lu475-507559	Morden2012	65.31	6.60	0.40	0.15
	6	Lu2564-21846	Morden2012	73.41	5.02	0.40	0.11
	6	Lu2971-Lu442a	Morden2012	84.81	3.57	-0.29	0.08
	15	Lu462a-238762	Morden2012	30.81	32.44	-0.81	0.70
Linustatin	1	Lu311-Lu2597	Morden2011	59.21	13.83	39.53	0.28
	1	Lu868-Lu2589	Morden2012	44.11	3.80	27.68	0.07
	2	Lu2344-Lu532	Morden2012	27.51	5.10	34.05	0.10
	13	Lu2074-Lu1112	Morden2012	44.91	3.19	27.70	0.08
Neolinustatin	6	Lu475-507559	Morden2011	65.41	9.62	9.07	0.20
	6	Lu2564-361129	Morden2011	73.41	4.46	6.43	0.10
	2	Lu2344-Lu2909	Morden2012	24.51	2.99	8.46	0.06
	8	Lu2957-Lu2923	Morden2012	17.41	4.30	10.38	0.08
	6	Lu2553-507559	Morden2012	68.51	3.81	9.54	0.07
	6	Lu475-Lu2564	Regina2012	50.91	2.95	10.88	0.07

Trait	LG	Flanking Markers	Location	Position(cM)	LOD	Additive effect	R^2
Lin:Neo	1	Lu311-Lu2597	Morden2011	60.21	8.42	0.57	0.16
	8	ysc1-29311	Morden2011	25.61	3.26	-0.35	0.06
	6	Lu475-507559	Morden2011	61.91	8.66	-0.59	0.17
	1	Lu3283-Lu49B	Morden2012	114.21	3.61	0.40	0.11
	14	Lu461-Lu601b	Morden2012	24.51	3.19	0.31	0.07
	14	Lu225-Lu514	Morden2012	35.81	4.02	0.34	0.08
TotalHCN	1	Lu311-Lu2597	Morden2011	58.21	9.29	26.42	0.18
	1	Lu3283-Lu49B	Morden2012	108.41	3.76	21.29	0.07
	2	Lu2344-Lu532	Morden2012	28.51	4.51	22.48	0.09
	2	Lu532-Lu2250	Morden2012	39.01	2.93	19.36	0.06
	8	Lu2317-29311	Morden2012	17.41	4.12	21.12	0.07
	6	Lu2074-Lu1112	Morden2012	46.91	3.03	18.49	0.06
	14	Lu514-Lu808	Morden2012	42.51	2.63	17.45	0.05

Appendix IV List of all clusters found on the scaffolds identified after fine mapping the linustatin and neolinustatin QTL

Query	Position	Strand	Identity %	Accession	Description
Scaffold 112					
Lus10018098	8935-11021	+	50	XP_002316033	GRAS family transcription factor [*Populus trichocarpa*] >gi|222865073|gb|EEF02204.1|
Lus10018099	32617-34194	-	47	XP_002316033	GRAS family transcription factor [*Populus trichocarpa*] >gi|222865073|gb|EEF02204.1|
Lus10018100	65641-67521	-	49	XP_002316033	GRAS family transcription factor [*Populus trichocarpa*] >gi|222865073|gb|EEF02204.1|
Lus10018101	69518-70864	-	45	XP_002316034	GRAS family transcription factor [*Populus trichocarpa*] >gi|222865074|gb|EEF02205.1|

Query	Position	Strand	Identity %	Accession	Description
Lus10018157	494506-495165	-	42	XP_002521476	cytochrome P450, putative [*Ricinus communis*] >gi|223539375|gb|EEF40966.1|
Lus10018158	496124-496450	-	54	CAN80536	hypothetical protein VITISV_035975 [*Vitis vinifera*]
Lus10018159	496831-497355	-	54	XP_002329123	cytochrome P450 [*Populus trichocarpa*] >gi|222869792|gb|EEF06923.1|
Lus10018160	499786-500364	-	69	XP_004237151	PREDICTED: cytochrome P450 86B1-like [*Solanum lycopersicum*]
Lus10018161	504328-504846	-	59	XP_003525679	PREDICTED: cytochrome P450 86B1-like [*Glycine max*]
Lus10018162	505153-506178	-	74	XP_002306255	cytochrome P450 [*Populus trichocarpa*] >gi|222855704|gb|EEE93251.1|

Query	Position	Strand	Identity %	Accession	Description
Scaffold 361					
Lus10026793	715866-717486	-	99	AFJ52912	UDP-glycosyltransferase 1 [*Linum usitatissimum*]
Lus10026795	732913-734313	-	100	AFJ52909	UDP-glycosyltransferase 1 [*Linum usitatissimum*]
Lus10026796	740872-741063	-	86	AFJ52914	UDP-glycosyltransferase 1 [*Linum usitatissimum*]
Scaffold 1491					
Lus10019808	111547-112953	-	100	AFJ52991	UDP-glycosyltransferase 1 [*Linum usitatissimum*]
Lus10019809	115896-117269	-	100	AFJ52989	UDP-glycosyltransferase 1 [*Linum usitatissimum*]
Lus10019829	232014-233001	-	91	AFJ52945	UDP-glycosyltransferase 1 [*Linum usitatissimum*]

Query	Position	Strand	Identity %	Accession	Description
Lus10019830	236623-237705	+	61	XP_002313864	predicted protein [*Populus trichocarpa*] >gi|222850272|gb|EEE87819.1|
Lus10019831	239254-240633	-	100	AFJ52948	UDP-glycosyltransferase 1 [*Linum usitatissimum*]
Lus10019832	244311-245801	+	100	AFJ52946	UDP-glycosyltransferase 1 [*Linum usitatissimum*]
Lus10019833	248838-250732	+	100	AFJ52929	UDP-glycosyltransferase 1 [*Linum usitatissimum*]
Lus10019834	251196-251929	+	78	AFJ52929	UDP-glycosyltransferase 1 [*Linum usitatissimum*]
Lus10019835	252854-254350	-	100	AFJ52944	UDP-glycosyltransferase 1 [*Linum usitatissimum*]
Lus10019836	255149-255451	-	63	AFJ52931	UDP-glycosyltransferase 1 [*Linum usitatissimum*]
Scaffold 176					
Lus10014399	241081-244184	+	100	XP_002326389	cytochrome P450 [*Populus trichocarpa*] >gi|222833582|gb|EEE72059.1|

Query	Position	Strand	Identity %	Accession	Description
Lus10014400	244706-246899	+	86	CBI30187	unnamed protein product [*Vitis vinifera*]
Lus10014401	250247-253718	-	100	AFJ52932	UDP-glycosyltransferase 1 [*Linum usitatissimum*]
Lus10014402	258256-259749	-	100	AFJ52935	UDP-glycosyltransferase 1 [*Linum usitatissimum*]
Lus10014403	269709-271175	-	100	AFJ52938	UDP-glycosyltransferase 1 [*Linum usitatissimum*]
Lus10014404	273264-274808	-	100	AFJ52937	UDP-glycosyltransferase 1 [*Linum usitatissimum*]
Lus10014437	481889-483577	-	100	AFJ52941	UDP-glycosyltransferase 1 [*Linum usitatissimum*]
Lus10014438	485272-486045	-	94	AFJ52940	UDP-glycosyltransferase 1 [*Linum usitatissimum*]

Appendix V Summary of the reference values for the log_{10} transformed calibration and validation datasets

	Linamarin	Lotaustralin	Linustatin	Neolinustatin	Total HCN
Calibration					
Min	-0.04	-0.36	2.02	1.52	2.00
Max	1.68	1.35	2.67	2.47	2.69
Mean	0.75	0.34	2.34	1.90	2.32
SD	0.25	0.23	0.13	0.21	0.13
N	310	218	311	311	311
Validation					
Min	0.26	-0.21	2.05	1.57	2.04
Max	1.24	1.18	2.59	2.35	2.53
Mean	0.66	0.34	2.33	1.90	2.31
SD	0.23	0.29	0.11	0.20	0.12
N	40	21	40	40	40

Appendix VI Summary of Shapiro-Wilk test statistics for untransformed values of the calibration set

	Linamarin	**Lotaustralin**	**Linustatin**	**Neolinustatin**	**HCN**
Mean	6.67	2.56	231.16	89.91	219.13
SD	5.27	2.32	69.20	54.73	65.07
Variance	27.76	5.38	4789.04	2995.51	4233.60
Kurtosis	30.07	-0.79	0.18	1.90	0.07
Shapiro-Wilk statistic W	0.66	0.98	0.96	0.93	0.95
Critical value of W (5%)	0.95	0.95	0.95	0.95	0.95
Null Hypothesis	reject	accept	accept	reject	accept

Appendix VII Summary of Shapiro-Wilk statistics for the log10 transformed values of the calibration set

	Linamarin	**Lotaustralin**	**Linustatin**	**Neolinustatin**	**HCN**
Mean	0.75	0.34	2.34	1.90	2.32
SD	0.25	0.23	0.13	0.21	0.13
Variance	0.06	0.05	0.02	0.04	0.02
Kurtosis	0.72	-0.49	-0.63	0.01	-0.83
Calculated Shapiro-Wilk statistic W	0.98	0.96	0.99	0.99	0.98
Critical value of W (5%)	0.95	0.95	0.95	0.95	0.95
Null Hypothesis	accept	accept	accept	accept	accept