INTRASPECIFIC GENETIC DIFFERENCES IN LINUM USITATISSIMUM AND INTERSPECIFIC GENETIC DIFFERENCES IN THE GENUS LINUM

By

CHRISTINA KATHERINE CAJIGAS

Submitted in partial fulfillment of the requirements

For the degree of Master of Science

Advisor: Dr. Christopher Cullis

Department of Biology

CASE WESTERN RESERVE UNIVERSITY

August, 2011

CASE WESTERN RESERVE UNIVERSITY

SCHOOL OF GRADUATE STUDIES

We hereby approve the thesis/dissertation of

Christina K. Cajigas

Candidate for the Masters of Science degree*

Dr. Christopher Cullis

Dr. Barbara Kuemerle

Dr. Emmitt Jolly

6/2/2011

*We also certify that written approval has been obtained for any proprietary material contained therein

Table of Contents

Table of Contents 3

List of Tables 7

List of Figures

Abstract 11

Chapter 1. Introduction 13

1. 1. Genomic variation within a species (intraspecific) or closely related species (interspecific) 13

1. 2. *Linum ussitassium*, unique genomic characteristics and heritable changes 14

1.2.1. Flax High A&T region 16

1.2.2. Heritable Genomic Changes 16

1. 3. Linum Insertion Sequence 1 21

1. 4. Linum Phylogeny and Linum species, wild relative species of domesticated flax 24

1.5. Primers used for PCR reactions 27

1.6. Rationale 27

1.7 Objectives 31

Chapter 2. Materials and Methods 32

2.1. Plant Material and DNA extraction from seeds and leaves 32

2.2. Determining flax actin gene homology 33

2.2.1. PCR using Actin Primers 33

2.2.2. Sequencing of amplified productions 34

2.2.3. Making alignment using sequence data and Clustalx 34

2.3. Making a phylogeny High A and T Primers 8 and 17 34

2.3.1. Shotgun sequencing of Bison flax variety 35

2.3.2. PCR using High A and T Primers 35

2.3.3. Sequencing of amplified products 36

2.3.4. Phylogeny using Clustalx 36

2.4 Bison contigs blasted against Bethune genome 36
(Intraspecific differences)

2.4.1. Dot Plots of repetitive sequences 37

2.5. LIS-1 Screening using Target L and R and Target L and 3' R primers 37
on different varieties of Linum bienne (intraspecific differences)

2.5.1. PCR using Target L and R primers on Linum bienne species 37

2.5.2. PCR using Target L and 3'R primers on Linum bienne species 37

2.6. LIS-1 screening on Linum species using primers Target L and R and 38
Target L and 3'

2.6.1. Target L and R PCR screening of all Linum species 38

2.6.2. Sequencing of PCR products 39

2.7. LIS-1 miRNA primers 1-9 screening using PCR

2.7.1. LIS-1 miRNA primers 1-9 run with Linum species 1-23 40

2.7.2. LIS-1 miRNA primers 1-9 run with Linum species 29-47 40

2.8. LIS-1 Primer 7 (miRNA primer) analysis through PCR and sequence 41

analysis

2.9. TOPO TA cloning of primer 7 for cloning 41

2.9.1. Agar broth plats for cloning 42

2.9.2 Generation of PCR products for cloning reaction 42

2.9.3. Setting up TOPO cloning reaction 42

2.9.4. Transformation of Competent Cells 43

2.9.5 Plasmid prep verification and PCR using M13 primers 43

2.9.6. Colony PCR 44

Chapter 3. Results: Determining homology of the actin gene 45

3.1. PCR and sequencing of Actin primers on all Linum species 45

3.2. Alignment using actin sequence data 48

Chapter 4: Results: Making a phylogeny using high A and T primer 8 50

4.1. Phylogeny using Primer 8 50

4.2. Analysis of PCR using high A and T primer 17 55

Chapter 5: Results: Bison and Bethune, comparing high A and T sequences between two varieties of the same species 57

5.1. Blast data: distribution of hits 57

5.2. Highly repetitive sequences: characterization using dot plot analysis 60

Chapter 6. Results: LIS-1 characterization of Linum bienne 60

6.1. Target Left and Right primers 60

6.2. Target Left and 3'Right primers

6.3. Sequence data of target left and right bands from two L. bienne varieties 62

Chapter 7. Results: LIS-1 characterization of different Linum species 65

7.1. Target left and right primers for PCR with Linum DNAs 65

7.2. Target left and 3' right primers for PCR with Linum DNAs 71

7.3. Sequencing of amplified bands 71

Chapter 8. Results: LIS-1 miRNA primers 1-9 screening 73

8.1. LIS-1 primers 1-9 run with species 1-23 73

8.2. LIS-1 primers 1-9 run with species 29-47 81

Chapter 9. Results: miRNA primer 7 analysis 89

9.1. PCR amplification with select species 89

9.2. TOPO TA cloning 92

Chapter 10. Discussion 95

Chapter 11. Future Directions 100

Appendix A: Actin sequences 101

Appendix B: High A and T Primer 8 sequences 103

Appendix C: Primer 8 Left and Right Alignments 107

Appendix D: Target, and Bienne sequences 109

Appendix E: Target 3' sequences from species 27 or Flax transformed 37 109

References 112

List of Tables

Table 1.1. List of *Linum* species used for PCR amplifications with select primers 25

Table 1.2. List of all the primers used for amplification of *Linum* DNA 26

Table 5.1. Table summarizing the BLAST between flax variety Bison 53
high A and T contigs with the whole of the Bethune genome

Table 10.1. *Linum* species that amplified with the flax actin gene 92

Table 10.2. *Linum* species grouping based on actin data and A and T primer 8 93

Table 10.3. Table showing which species amplified with A&T primer 17 93

Table 10.4. Table grouping those species with the same pattern of 94
expression when amplified with primer 7

List of Figures

Figure 1.1. Illustration of the various responses to nutrient stress occurring in the Pl flax line and the resultant stable genotrophs 17

Figure 1.2. Illustration showing LIS-1 and the target sequence 19

Figure 1.3. *Linum* phylogeny 24

Figure 1.4. Illustration of the target sequence showing site of LIS-1 insertion 27

Figure 3.1 A. Gel electrophoresis of PCR with flax actin primers used on *Linum* species 1-37 42

Figure 3.1 B. Gel electrophoresis of PCR with flax actin primers used on *Linum* species 38-S6 42

Figure 3.2. Agarose gel of PCR using Actin genes of select *Linum* species that amplified in Figure 3.1 A and B. 43

Figure 3.3. Aagrose gel of a re-PCR of the products seen in Figure 3.2. 43

Figure 3.4 A. Clustalx alignments for actin sequences left. 44

Figure 3.4B. Clustalx alignments for actin sequences right. 45

Figure 4.1. Agarose gel of primer 8 run with *Linum species* 1-49 48

Figure 4.2 A. Phylogenic tree created from the primer 8 left sequences Alignment on clustalx 49

Figure 4.2 B. Phylogenic tree created from the primer 8 right sequences Alignment on clustalx 50

Figure 4.3. PCR of high A and T primer 17 52

Figure 5.1. Dot plot analysis on High A and T contig 388 54

Figure 5.2. Dot plot for contig 48 55

Figure 5.3. Dot plot for contig 438 56

Figure 6.1. Gel run with *L. bienne* DNAs from Turkey 58

Figure 6.2. ClustalW2 alignment of *L. bienne* sequences from 608 and 620 with Target L and R sequence obtained from flax 60

Figure 7.1 A. PCR run with *Linum* species amplifying primer Target L and R using SpeedSTAR polymerase 63

Figure 7.1 B. PCR run with *Linum* species 1-41 amplifying primer Target L and R using SpeedSTAR polymerase 64

Figure 7.2. Gel of select species 9, 30, and 38 amplified with Target L and R 64

Figure 7.3. PCR of all *Linum* species amplified with Target L and R using goTaq green 65

Figure 7.4. Gel of select species 9, 14, 17, 30, 37, 38, 48, and S6 67

Figure 7.5. PCR products of select *Linum* species (9-S6) amplified with Target L and R 68

Figure 7.6 A. BLAST of sequence (left) obtained from amplifications of Target L and 3'R band from species 27, a flax variety 69

Figure 7.6 B. BLAST of sequence (right) obtained from amplification of Target L and 3'R band from species 27, a flax variety 70

Figure 8.1 A. Species 1-23 amplified with miRNA primer 2 71

Figure 8.1 B. *Linum* DNA species 1-23 amplified with miRNA primer 3 72

Figure 8.1 C. Species 1-23 amplified with miRNA primer 4 73

Figure 8.1 D. *Linum* DNA species 1-23 amplified with miRNA primer 5 74

Figure 8.1 E. *Linum* DNA species 1-23 amplified with miRNA primer 6 75

Figure 8.1 F. Species 1-23 amplified with miRNA primer 7 76

Figure 8.1 G. *Linum* species 1-23 amplified with miRNA primer 8 77

Figure 8.1 H. *Linum* species 1-23 amplified with miRNA primer 9 78

Figure 8.2 A and B. *Linum* DNA amplified with miRNA primers 1 and 2 79

Figure 8.2 C. *Linum* DNAs 29-47 amplified with miRNA primer 3 80

Figure 8.2 D. *Linum* DNAs 29-47 amplified with miRNA primer 4 81

Figure 8.2 E. *Linum* species 29-47 amplified with miRNA primer 5 82

Figure 8.2 F. *Linum* species 29-47 amplified with miRNA primer 6 83

Figure 8.2 G. *Linum* species 29-47 amplified with miRNA primer 7 84

Figure 8.2 H. *Linum* species 29-47 amplified with miRNA primer 8 85

Figure 8.2 I. *Linum* species 29-47 amplified with miRNA primer 9 86

Figure 9.1 A-C. *Linum* species 2, 5, 7, 9, 17, 29, 23, 27, 30, 38, and S6 87

Figure 9.5 A. Colony PCR of a 10^3 dilution 90

Figure 9.5 B. Colony PCR of a 10^4 dilution 91

INTRASPECIFIC GENETIC DIFFERENCES IN LINUM USITATISSIMUM AND INTERSPECIFIC GENETIC DIFFERENCES IN THE GENUS LINUM

Abstract

by

CHRISTINA KATHERINE CAJIGAS

Flax (*Linum usitatissimum*) has the ability to respond to environmental stress by altering its genome and passing these changes to subsequent generations. There are many genomic and phenotypic changes associated with these heritable modification, the most well characterized being the induction of Linum Insertion Sequence 1 (LIS-1) into the genome. In an effort to further understand the induction of LIS-1, this paper seeks to study the evolutionary relation between flax and its wild progenitors. Molecular studies will aid in determining, which if any species in the genus *Linum* have similar genomic characteristics of flax. The actin gene, which is expected to be highly conserved in these species, was used to make primers that did not amplify all *Linum* species studied. This displays the vast molecular differences seen in the genus. Using primers for the highly repetitive light satellite region (rich in Adenine and Thymine base comprising about 15% of the total flax genome) it was determined that this region may be conserved to some exist to some in the genus. LIS-1 is inserted in a target region between two known genes

(in flax). PCR results showed little homology of this region in the genus, showing that evolution of this region (or the ability for induction of LIS-1) could have been an independent event occurring only in flax. Potential miRNA sites were also studied in these *linum* species and found to not be homologous. Data from one miRNA primer (7) did show amplification hinting that is region is homologous but the different patterns of expression allowed for grouping of similar species. From these studies no species was found to contain LIS-1 fully inserted (6kb band) but there was homology found in some species with bands of interest, speaking to the existence of some of these regions in a few species.

Chapter 1. Introduction

1.1 Genomic variation within a species (intraspecific) or closely related species (interspecific)

Plant nuclear genomes exhibit great diversity in size, ranging from 130Mb in *Arabidopsis thaliana*, 430 Mb in *Oryza sativa* (rice), 750Mb in *Sorghum bicolor* (type of grass), 2400 Mb in *Zea mays* (maize), and 4900Mb in *Hordeum vulgare* (barley) (Bennetzen *et al*., 2005). Chromosome number, degree of gene clustering, and chromosome size can all differ by as much as an order of magnitude, even between closely related species. Along with size there is also variation in chromosome number, number and arrangement of genes and number of genome copies per nucleus. This variation is the result of highly active processes such as gene duplication and deletion, chromosomal duplication followed by gene loss, amplification of retrotransposons separating genes, and genome rearrangement following hybridization and/ or polyploidy (Kellogg and Bennetzen, 2004).

Nuclear genomes have been found to be largely similar among closely related species but rearrangements are observed with increasing phylogenetic distance. Variation in genomes can occur rapidly as in some varieties of maize where two different allelic versions of a chromosomal segment can be dissimilar in gene content and arrangement. Changes such as polyploidy and most gene duplications deletions are so frequent they can occur independently in multiple lineages. For this reason, it is important to remember that the correlations between rearrangements and time of divergence are not perfect (Kellogg and Bennetzen, 2004. Mechanisms like polyploidy and inversions and translocations, which result in genome rearrangements with new combinations of genomes or hybrid plants and

that can have changed patterns of gene expression, may be a major driving force in plant evolution (Kellogg and Bennetzen, 2004). Understanding the natures, rates and mechanisms of the various types of chromosomal rearrangements can provide insights into how these changes contribute to current evolved states, and how they can be used in phylogenetic analysis.

Recent sequencing of the maize genome (Schnable *et al*., 2009), has allowed for a comparison of the genomes of two inbred lines where several thousand sequences present in one line were absence in the other (Springer *et al*., 2009). Data like this shows that there are also large amounts of genomic variation occurring even within species, specifically between varieties that have been separated for only short periods of evolutionary time (Springer *et al*., 2009).

1.2. Linum usitatissimum, unique genomic characteristics and rapid heritable variations

Flax (*Linum usitatissimum*) is a vital part of the agriculture industry, in both the production of linen fiber and linseed oil. It has a diploid chromosome number of 30, and is comprised of 15 pairs of small similarly sized chromosomes (Oh and Cullis 2003). Work done by Cullis determined that the organization of the flax genome is very different from other plants of similar size and composition (1980). Molecular analysis shows that only 20% of the flax genome is actually basic coding function, or genes responsible for the production of proteins. The rest of the genome is repetitive sequences that can vary widely, but in flax are mainly arranged in tandem arrays along the genome (Cullis 1990). Tandemly repeated sequences usually fall into at three classes: centromeric satellite repeats which are located between each chromosome arm and span the centromere,

telomeric regions, and RNA (ribosomal) genes (Cullis 2004). In flax, additional families of highly repetitive or 100,000 or more base pairs tandemly arrayed sequences, comprise about 35% of the total nuclear DNA. The middle repetitive fraction of about 10,000-100,000 base pairs make up only about 15% of the total nuclear DNA, this is a less than in other higher plants, while the low copy number fraction which are about 1-10,000 base pairs comprises what is left of the genome. This means that flax has a larger fraction of its total genome made up of low copy number sequences (50%), than other plants of similar genome size (Cullis 2005)

In the maize genome, there are both tandem arrays and many dispersed families, which contrasts with peas where there are few tandemly arrayed families and the majority of the genome is composed of complex dispersed repetitive families (Cullis 1990). The origin of the diversity in organization between plant species is unknown, but it is hypothesized by Cullis that it could be due to differential rates of the processes that shape the genome itself. Translating this to flax, tandem amplification and deletion events occur more frequently than transpositions. In peas the opposite is true, and in maize the genome may come from an equal frequency of these events (1990).

Transpositions or transposable elements were first discovered in maize and have also been found in prokaryotes and eukaryotes, making their existence most likely ubiquitous in all genomes (Cullis 1990). The intermediately repetitive sequences are now thought to be derived from these transposable elements. There are two classes of transposable elements: class I are retrotransposons that replicate through an RNA intermediate and class II are transposons that move directly through a DNA form and change position without increasing in number as retrotransposons do (Cullis 2004). The deletion

associated with the movement of a class II transposon is frequently detected by the occurrence of unstable mutants, mutations that are observed in the presence of the transposon that disappear when the transposon is excised. They are again reactivated when another incompatible fusion occurs. The presence of transposable elements can also affect the rate at which organisms respond to changes in external environment (Cullis 1990). The flax genome also differs in that it has a longer period of interspersion pattern of lower and higher copy sequences, a characteristic not shared by other higher plants. This pattern could be the result of a much lower activity of mobile elements with fewer sites for insertion of such elements in the genome (Cullis 2005).

1.2.1. Flax High A&T region

The flax genome contains a light satellite region identified using cesium chloride gradients that makes up about 15% of the flax genome. This region is characterized by being rich in adenine and thymine nucleotides (Cullis, 1980).

1.2.2. Heritable Genomic Changes

Another unique characteristic of flax involves its ability to undergo heritable genomic changes in response to nutrient stress. These changes include variation in the total amount of nuclear DNA, in ribosomal DNA (rDNA) copy number, and in repetitive sequence copy number. The most notable change, also considered here, is the appearance of Linum Insertion Sequence 1 (LIS-1), a complex insertion event that occurs at a specific single copy site in a number of genotrophs. These changes occur before gamete production, during the period of vegetative growth, meaning that these characteristics can be passed

on to the next generation. Offspring resulting from these induced plants display differences in phenotypic characteristics such as plant height, weight, branching pattern and septa capsule hair number (Durrant, 1962, Cullis, 2005).

These genomic changes occurring in response to environment in which the plants have been grown have been observed in the inbred flax variety Stormont Cirrus' (Pl). Growing Pl in different environments can result in phenotypic and genotypic differences in first generation progeny. Resulting in lines which stable changes are observed, called genotrophs. It is important to note that stable inheritance of the subsequent generations resulting from self-fertilization does not occur in all growth conditions (Cullis, 2005). These different inducing environments consisted of different combinations of sulphate of ammonia (N), Triple superphosphate (P), muriate of potash or potassium chloride in fertilizer form (K), and calcium hydrate (Ca) (Durrant, 1962). In his experiments, Durrant sought to determine whether the environment in which the plant was grown was causing changes at the genomic level. Inbred plant lines were grown under each of these nutrient regimes and were allowed to self fertilize. F1 offspring resulting from this self-fertilization were then grown under 8 nutrient regimes that consisted of combinations of N, P, and K. By comparing plant weights at the end of this F1 growth period, the effects of both parental growth conditions and F1 growth conditions were characterized. It was determined that not only was F1 growth dependant on the growth conditions, but parental environment could also have a significant impact on plant weight. Therefore, F1 growth conditions were affecting growth by altering gene expression, but also some parental growth environments were creating heritable changes in the genome of these plants (Durrant, 1962).

Through his studies, Durrant determined that the npk treatments produced the largest plant weights whereas the nk treatments produced the smallest plant weights. Subsequent growth of many generations showed these large and small plant lines remained stable regardless of changes in the nutrient regimes (1962). These stable lines were termed genotrophs, with those being grown in the npk treatment being large or L, and those grown in nk being small or S. The original parental or inducible line Stormontnt Cirrus) was termed plastic, Pl, because of its ability to respond to stress with these heritable changes (Durrant, 1962; Cullis, 2005).

When first observing these changes in the flax variety Stormont Cirrus, Durrant described four aspects to the induction process that provide evidence that a novel phenomena is responsible for these observed genotypic and phenotypic changes and not conventional selection from a heterogeneous population of plants (1962; Cullis, 2005). First Stormont Cirrus has a biology resulting in it being a mostly a self-fertilizing plant, second, most of the seeds grown under inducing conditions survive and contribute to the next generation. Third, all the progeny grown under a certain inducing condition were identical to each other but different from all other progeny grown in other inducing conditions. The fourth and perhaps most compelling is that these induction changes have been repeated. Other flax varieties behave as Stormont Cirrus to these inducible environments, although not all to the same extent. They include Rembrandt, Hollandia, and Liral Monarch (Cullis, 2005). All the lines that have been identified to be responsive to stress, however, are varieties that are cultivated for fiber. No flax varieties cultivated for oil production have been shown to be responsive (Durrant, 1962). This correlation is consistent with the idea

that the genes controlling the stress response are closely linked to good fiber quality, resulting in those genes being inherited together.

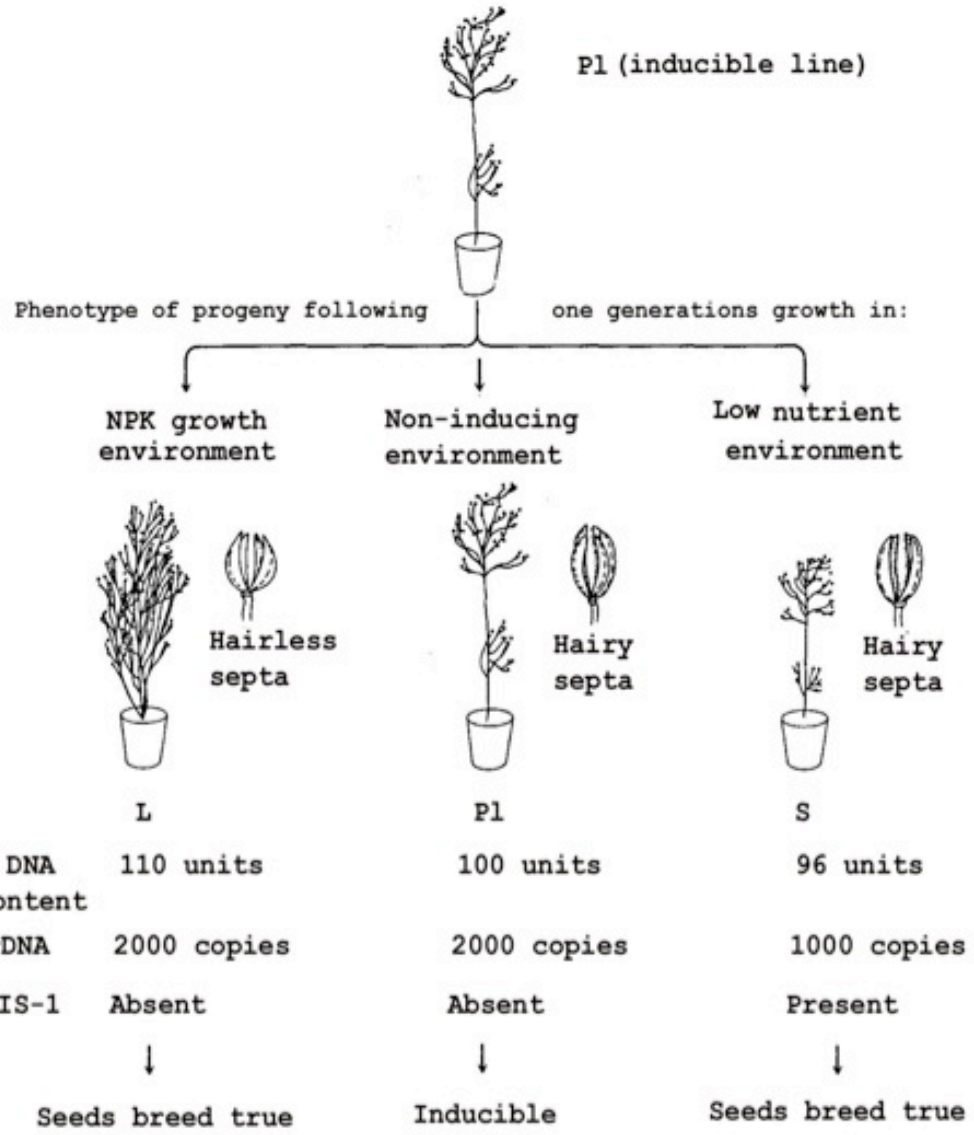
The stability of these genotrophs provide evidence that the changes seen are occurring not only at the genetic level, but that they can be permanent rather than the result of temporary (epigenetic) modifications. To further understand these changes, L and S genotrophs were studied at the molecular level. Work done by Evans *et al*. showed that stable lines experienced molecular changes correlated with changes in plant weight. This was seen when grown in the different nutrient regimes (1968). The DNA content of Pl, L and S was measured using Feulgen staining, a method for estimating relative amounts of DNA in individual nuclei. Results showed that when compared to PL, both L and S showed changes in DNA content. L displayed an increase in DNA content of 9% while S had a decrease of 6%. DNA content also changed in the Pl lines depending on the nutrient regime given. Evans noted that at 3 weeks DNA content began to increase or decrease depending on the conditions grown under, with DNA content reaching the levels seen in L and S genotrophs by week 5. This observation allowed for the conclusion that DNA changes occur within the first 5 weeks of growth (Evans 1968).

Another genomic variation was seen between the genotrophs and the Pl line in the large rDNA genes, which showed varying responses of rDNA copy number when compared with Pl. Studies done by Cullis using a hybridization technique with rDNA probes showed a relative decrease in rDNA copy number in S with L having a copy number not significantly different from Pl (Cullis 1977). Studies conducted in 2009 by Chen *et al*. found similar differences in rDNA copy number using qPCR data with the 18S portion of the large rDNA subunit. Here Pl plants showed copy number variation depending on

growth in L or S inducing conditions. L conditions displayed decreases in copy number and S displaying a variety of changes with decreases, no changes and a very small number showing an increase in rDNA. The small 5S rDNA gens also showed copy differences between Pl, L, and S in this study. S had a larger reduction in copy number when compared to the reduction seen in L. In contrast, when Pl was grown under low nutrient conditions it displayed both increases and decrease in 5S copy number. This is in contrast to Pl growing under nutrient -rich conditions where there was mostly a decrease in copy number seen (Chen *et al*., 2009). Biochemical studies conducted by Cullis and Kolodynska determined differences in peroxidase isozymes between Pl, L, ad S, that were attributed to differential expression of different isozyme genes rather than differences in isozyme modification (1975). A later study, however, showed that modified glycosolation of the enzyme was the cause of mobility differences (Tyson and Fields, 1982).

To better understand these genetic and phenotypic changes in the L and S genotrophs when compared with the original plastic (Pl) or parental line, Figure 1.1. displays a summary of these changes. Growth conditions with low or imbalanced nutrient content produce small or S genotrophs. Mature S genotrophs are shorter than the Pl line, have little branching, hairy capsule septa like Pl. They also have altered isozyme mobility with a reduction in the amount of nuclear DNA and fewer copies of rDNA genes compared to Pl. S also contains LIS-1, an insertion that will be discussed later.

Figure 1.1. Illustration of the various responses to nutrient stress occurring in the Pl flax line and the resultant stable genotrophs. Adapted by Cullis from Molecular Biology of the Cell, first edition.



High nutrient environments give rise to large or L genotrophs that are taller than S but somewhat shorter than Pl with more branching than Pl or S. L has a hairless capsule septa, and has altered isozyme mobility relative to both Pl and S, with greater amounts of nuclear DNA than Pl but having similar rDNA content as P (Durrant 1962; Cullis 1977). L does not have LIS-1, despite the fact that it is sometimes seen in the Pl parent when the plant is responding to high nutrient conditions, but it is not passed to L offspring under nutrient stress Chen *et al*., 2005, 2009).

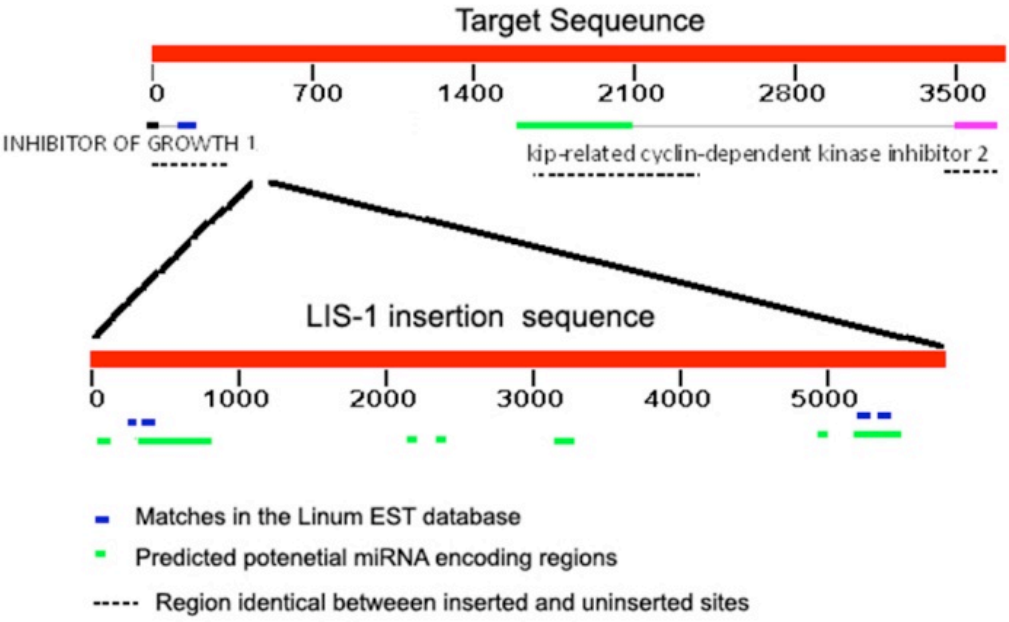
1.3. LIS-1 Linum Insertion Sequence 1

As previously discussed, the characteristics altered in the genotrophs compared to the parent Pl line include height, weight, branching pattern, presence of hairs on the seed capsule septa, mobility of isozymes that are hormonally regulated, total nuclear DNA content, and number of genes coding for the large rRNAs and 5S rRNAs. But perhaps most characterized genomic change occurring is the appearance of LIS-1. Linum Insertion Sequence 1 (LIS-1) is a single-copy 5.7 kilobase (kb) DNA fragment whose insertion is associated with environmentally induced heritable changes resulting in the stable genotroph lines. Because LIS-1 is absent in the parental Pl line, it is clear its insertion is the result of a targeted, highly specific, complex insertion event that occurs during the formation of some of the genotrophs. Its presence is also a natural occurrence in many flax and linseed varieties (Chen *et al*., 2005).

As is shown in Figure 1.2, LIS-1 is inserted into a specific single copy target sequence in the Pl genome when growing in low nutrient conditions, making the presence of LIS-1 stable in subsequent generations of the S genotrophs. Both LIS-1 and the target site have

been sequenced, and from this target sequence, primers for analysis in this paper were used to screen various *Linum* species for the presence of LIS-1. From this sequence, Chen and colleagues have determined certain characteristics of LIS-1 insertion (2005).

Figure 1.2. Illustration showing LIS-1 and the target sequence. There are two genes in the target sequence labeled, matches to the Linum EST database, and putative miRNAs. Note the sites within the target sequences that are unchanged during an LIS-1 insertion event. (Cullis, unpublished).



This event has only one possible feature in common with other genomic insertion events resulting from transposable elements insertions, which is a 3 bp duplication (TCC) at the target site. The duplication that occurs following the insertion of LIS-1 is a general signature of transposase activity. However, LIS-1 lacks direct and inverted repeat sequences at its termini that are present in most plant trasposons. Using database searches in GenBank, it was found that there is no homology to other transposable elements or any mobile DNA sequences known as well as no large open reading frames. These studies have also found some short matches to the Linum EST database, and many informatically identified putative miRNAs (Moss, personal communication). It is from these miRNA regions that primers have been made for this study to further understand these possible miRNA regions and their possible homology in other *Linum* species.

The origin of LIS-1 is still unknown. LIS-1 is present in S genotrophs but is absent from the original Pl line mwhich the genotrophs were derived. When PCR is done looking for parts of this insertion region in the Pl genome there were no amplifications from pairs of primers used for the original sequencing of LIS-1 which were about 400bp apart. This data has been confirmed by complete genome sequencing of Pl, L and S lines where LIS-1 was observed in S (as expected) but not in either Pl or L, or where any significantly long fragments similar to LIS-1 observed in either P or L (Cullis, personal communication). This data indicates that LIS-1 is assembled or perhaps rearranged from short sequences too small to span any set of primer pairs found spread throughout the genome (Chen *et al*., 2005). Recent data has shown primers made around the miRNA sites (primers used in this paper), gave products when used on Pl DNA that does not contain LIS-1 (Liu, unpublished data), indicating they are present even when LIS-1 is not

inserted in the genome. The target sequence, in which LIS-1 is inserted contains two genes on either side of the insertion site, inhibitor of growth 1, and kip-related cyclin-dependent kinase inhibitor 2 (Chen *et al*., 2005).

1.4. Linum Phylogeny and Linum species, wild relative species of domesticated flax

The Linaceae flax family consists of herbs or shrubs of about 14 genera and distributed widely geographically. The largest genus under Linaceae is *Linum*, with about 300 species scattered worldwide. Some of the species are shrubs usually found in tropical areas, with perennial and annual species are found mostly in temperate areas. There are three main centers of diversity for the genus *Linum*, each with large interspecific diversity. They include the Mediterranean area, southern North America and Mexico, and South America. There are, however, some populations found in temperate areas of Europe, Asia, and the Americas (Diederichsen and Richards, 2003). As previously mentioned, *L. usitatissimum* has been cultivated as a source of fibers for the production of linen, as well as seed oils that serve a medicinal value (rich in Omega-3-Fatty Acids), or as important components of paints, varnishes, and other industrial products. The seeds and leaves of many *Linum* species have been used in medicinal natural drugs as laxatives, anti-inflammatories, and treatments for burns (Diederichsen and Richards, 2003).
This wide diversity in the genus continues to challenge systemic taxonomic techniques, challenging botanists, who even propose the genus be divided into sections of like characteristics. Some *Linum* species still have yet to be fitted into a phylogeny. Proposals for grouping the species have been based on morphological characteristics, fatty oil compositions in seeds, even pollen morphology, all have been inconclusive with fatty oil

composition yielding the most sensible yet inconsistent results (Diederichsen and Richards, 2003).

Early work sought to classify the evolutionary origins of flax. Using cytogenetic relationships, Yermanos classified which *Linum* species are closely related to flax in his study. This paper found that among these species with n=15: *Linum angustifolium*, *L. usitatissimum* (flax), *Linum corymibiferum*, *Linum decumbens*, *Linum africanum*, *Linum pallescens*, and *Linum nervosum*; *Linum angustifolium* is the oldest species among this group. This study found *Linum usitatissimum* to be differentiated from *Linum angustifolium* by 1 chromosomal translocation, while *Linum africanum*, *Linum corymibiferum*, and *Linum decumbens* are different from *Linum angustifolium* by 2 translocations. This paper also found *Linum usitatissimum* and *Linum nervosum* to be chromosomally homologous (Yermanos, 1957).

This paper also cataloged the n=9 group concluding that their cytogenetic relationships are unclear. *Linum perenne* and *Linum alpinum* are closely related geographically and therefore might represent a pair of older species from which *Linum alpinum* might have been derived by 1 translocation and *Linum narbonense* by 2 translocations. It was also found that 3 species *Linum narbonense*, *Linum altaicum*, and *Linum julicum* differ in 1 translocation. Also *Linum austriacum*, *Linum narbonense*, and *Linum altaicum* also differ by 1 translocation (Yermanos, 1957). Figure 1.3 is a proposed phylogenic tree created through molecular data of the genus *Linum* (Project Description). Those species used in these experiments are listed marked with an arrow. For a complete list please of species and the numbers used for identification in each agarose gel, refer to Table 1.1.

1.5. Primers used for PCR reactions

Actin primers were used to amplify the actin bands expected to be highly conserved among *Linum* species. High A and T primers were used to determine conservation of these sequences between species. Primers were also used to screen *Linum* species for LIS-1 using Target L and R primers. Primers for the insertion region (Target L and 3'R) were used to determine the presence of this region in other species. Figure 1.3 illustrates where these primers are located in relation to LIS-1. Potential miRNA primers were used to determine how these sequences are represented in these species. It is expected for miRNA sites to be highly conserved, seeing that this is not the case will help determine similarities for grouping of these species. Table 1.2 is a list of the primers

1.6. Rationale

Flax has some unusual genomic characteristic, the most unusual being the ability to respond to environmental stress by altering its genome. Especially noteworthy is that these genomic changes can be heritable. In order to further understand this phenomenon, and specifically the induction of LIS-1, this paper seeks to study the evolutionary relation between flax and its wild progenitors to understand the phylogeny and to determine which species have similar genomic characteristics of flax. Using techniques such as PCR and Bioinformatics to study the interspecific differences between flax (*Linum usitatissimum*) and species in the same genus of *Linum*, as well as the intraspecific differences between different varieties of *Linum usitatissimum* were studied. Studying these molecular differences between flax varieties and different *Linum* species will give an indication of how these species and varieties differ at the molecular level. Using

programs like ClustalX will allow us to create phylogenic trees that may challenge previous phylogeny relationships.

Figure 1.3. Linum phylogeny adapted from http://www.sbs.utexas.edu/simpsonlab/josh%20files/mcdilldig.pdf)

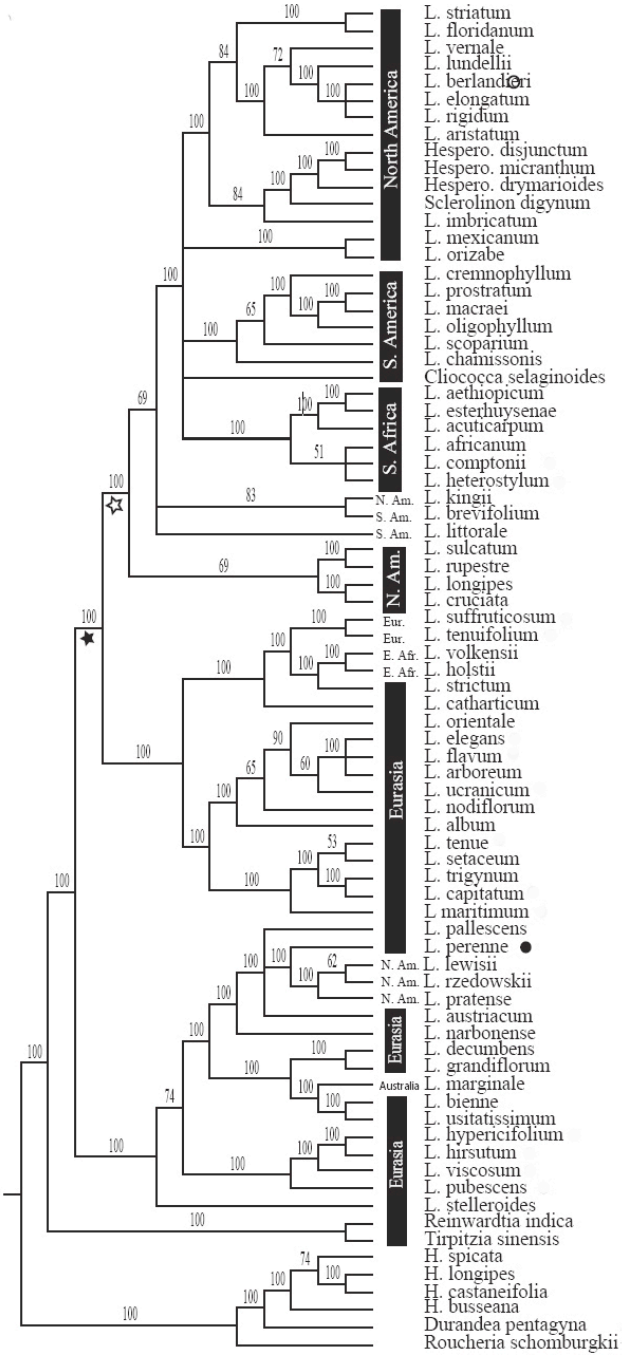


Table 1.1. List of the *Linum* species used for PCR amplifications with select primers. Those numbers that are crossed out signify those DNAs that throughout the study that were spent.

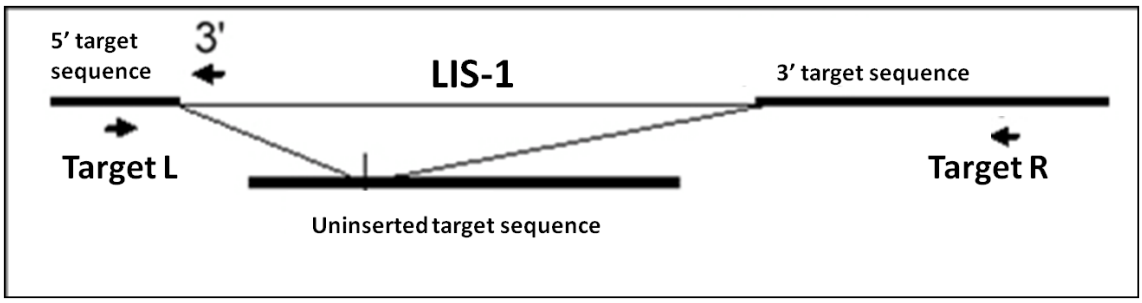
List of *Linum* species

1. 100 *Linum alpinum*
2. 103 *Linum austriacum*
3. ~~105 *Linum bienne*~~
4. 110 *Linum grandiflorum caeruleum*
5. 116 *Linum narbonense*
6. 117 *Linum nervosum*
7. 118 1 of 2m 2 of 2 *Linum perenne*
8. 119 *Linum perenne*
9. 120 *Linum perenne*
10. 121 *Linum perenne*
11. ~~126 *Linum trigynum*~~
12. 128 *Linum austriacum*
13. 132 *Linum bienne*
14. 133 *Linum grandiflorum reubrum*
15. ~~136 *Linum flavum*~~
16. 138 *Linum maritimum*
17. 141 *Linum perenne*
18. ~~142 *Linum punctatum*~~
19. ~~144 *Linum tenuifolium*~~
20. 145 *Linum flavum*
21. ~~147 *Linumcatharticum*~~
22. 154 *Linum altaicum*
23. 155 *Linum hologynum*
24. ~~165~~
25. ~~169~~
26. ~~BH164~~
27. 27 FT37 *Linum usitatissimum*
28. ~~PI650298 *Linum austriacum*~~
29. PI650310 *Linum campanulatum*
30. PI522292 *Linum corymibiferum*
31. PI650311 *Linum elegans*
32. PI650313 *Linum flavum*
33. PI650314 *Linum flavum*
34. Ames22631 *Linum grandiflorum*
35. PI650317 *Linum hirsutum*
36. Ames27614 *Linum lewisii*
37. Pi522308 *Linum marginale*
38. PI522309 *Linum maritimum*
39. PI650321 *Linum mucronatum*
40. PI650329 *Linum perenne*
41. PI502411 *Linum sp*
42. PI650330 *Linum stelleroides*
43. Ames 26261 *Linum strictum*
44. Ames23144 *Linum suffruticosum*
45. PI650334 *Linum tenuifolium*
46. PI650335 *Linum thracicum*
47. ~~PI650336 *Linum trigynum*~~
48. 114 *Linum grandiflorum desf*
49. 143 *Linum strictum subsp. Strichum*
50. 102 *Linum anglicum*
51. 106 *Linum bienne*
52. 104 *Linum austriacum*
53. 107 *Liunm campanolatum*
54. 151 *Linum trigynum*
55. 143 *Linum strictum subsp strictum*
56. 140 *Linum perenne*

Table 1.2. List of all the primers used for amplification of *Linum* DNA.

Primer Name	Primer Sequence
Actin L Actin R	CAGCAGATGTGGATCTCGAA TAGCATCCCAATTCCAGGAG
Target L Target R	CCCCCTTCTTCAGTTCTGCT GGCTAGGGTTAGGGTTTCCA
3'R	GAGGATGGAAGATGAAGAAGG
8 (388) L 8 (388) R	GACAAATTATTCACCGGCAGA AAGCACTAATCGGGCATTTC
17 (722) L 17 (722) R	TCGCTAATCAACCAATCGTG TGGAAGAAAGCAAAATGAATTG
Primer 1 (85) L 85 R	GAAATTTTGGGGGCATGTAA CATCCTCGAACCTCCCTGT
Primer 2 (594) L 594R	GGTTATATAGCGGATCAGAGATGG AACCGTTAACCGCTATGAATG
Primer 3 (801) L 801 R	GATTGCCAAGTCTAACCGTTG TCCGATTATAGCCTGACACG
Primer 4 (2696) L 2696 R	CTGCAACTGATCCCGATTTT GGATTTATCCAGGGCAACTG
Primer 5 (3240) L 3240 R	GCCTTGCAGTCTCTGTTTCC GTCCAGGAATGCCAAAATGT
Primer 6 (4894) L 4894 R	GCTCATTTGATAGGTGCGGTA TCCACACATTCATATTCATAAAACA
Primer 7 (5186) L 5186 R	CGACCGTACCCTCATCAAGT ACAGGTTGGGGATGGTTATG
Primer 8 (3190) L 3190 R	AATTTTGCTTGCAAGGGTTG CTTTCATTGAACTACATGGGCTGA
Primer 9 (683) L 683 R	CCTCAAAATATCCAGTCAACACG ATTACTTAACCGTTTGACAATCCT

Figure 1.4. Illustration of the target sequence showing site of LIS-1 insertion as well as location of primers in the LIS-1 target sequence. Target L and 3' R amplify the junction or insertion site of LIS-1. Target L and R amplifies the uninserted target sequence. The name 3' is merely a primer name and does not represent the direction or orientation of LIS-1



1.7. Objectives

The objectives of this thesis are outlined below.

1. To use flax actin primers to study the variation in the actin gene among flax and other *Linum* species

2. To use High A and T primers created from a 454 sequence reads of the flax variety Bison for use in PCR to compare how these sequences are conserved in other *Linum* species.

3. To use these High A and T sequences and BLAST them to the Flax variety Bethune genome (that is now sequenced) and study the homology of these highly repetitive sequences in two different flax varieties.

4. To use the LIS-1 primers Target L and R and Target L and 3'R to study whether LIS-1 exists in other *Linum* species or even shares any homology to this unique genomic insertion.

5. To look at the homology of miRNA sites (flax primers 1-9) in other *Linum* species to determine homology.

6. To determine how widespread sequences related to LIS-1 are in different *Linum* species. Other sequences of comparison include a gene highly conserved (actin gene), high A and T sequence that is not functional, a target primer for insertion of LIS-1 between two known genes, and potential miRNA sites that are believed to be conserved. Using a variety of sequences, we can get a better picture of the homology existing in these species, or varieties.

Chapter 2. Materials and Methods

2.1. Plant Material and DNA extraction from seeds and leaves

Linum species used in these experiments were split into two groups. The old and new *Linum* species as seen in Table 1.0 . The old sets of Linum DNAs (number 1-28, and 48, 49) were obtained from seeds soaked in a petri dish using a QUIAGEN mini prep kit for DNA extraction. The new *Linum* DNAs (29-47) were obtained from the USDA (United States Department of Agriculture) seed bank. They were grown and the leaves used to obtain DNA, done so through the use of a QUIAGEN mini prep DNA extraction kit, or the Denville Scientific mini prep DNA extraction kit, depending on availability. A

number system (1-49) was used to classify each species simplistically versus the three-digit system used originally or numbered/letter system used by the USDA.
Linum Bienne DNAs were obtained from Dr. Yong-Bi Fu, Plant Gene Resources of Canada, and originated from various sites in Turkey.

2.2. Determining flax actin gene homology

2.2.1. PCR using actin primers

Actin primers were designed by Cory Johnson, a PhD candidate in our lab, using the flax actin gene sequence found in the Expressed Sequence Tag (EST) database from the National Center for Biotechnology Information (NCBI) website. DNAs from 48 Linum species and two flax varieties, (Bison #27 and genotroph #S6) were amplified using the Actin primers and Promega GoTaq Green Master Mix (Promega Corp, Fitchburg, WI). The PCR reaction was as follows: an incubation time of 95°C for 2 minutes, followed by 35 cycles of 94°C for 20 seconds, 55°C for 30 seconds, and 72°C for 1 minute and 30 seconds, and finally a final step of 72°C for 5 minutes. PCR products were viewed on a 1.5% (w/v) agarose gel in 0.5 X TBE (1 X TBE = 100 mM Tris, 90 mM boric acid, 1 mM EDTA, pH 8.4) and run at 150 Volts for 35 minutes Those species that amplified (4, 5, 6, 9, 10, 16, 17, 20, 27, 30, 31, 38, 39, 49, positive control S6) were used in subsequent PCR amplifications. The PCR reaction was as follows: an incubation time of 95°C for 2 minutes, followed by 30 cycles of 94°C for 20 seconds, 60°C for 30 seconds, and 72°C for 1 minute and 30 seconds, and finally a final step of 72°C for 5 minutes. 10μL loaded onto a 1.7% gel and run at 75 Volts for 1 hour and 30 minutes. 1 μL of that PCR product was then used to re-amplify the fragment using the same PCR parameters but run for 20

cycles instead of 35 and an annealing temperature of 60°C instead of 55°C. 10μL of this reaction was again separated on a 1.7% agorose gel and run for 1 hour and 30 mins at 90 Volts.

2.2.2. Sequencing of amplified products

The re-amplified product was cleaned up using as Denville Scientific IsoPURE PCR purification kit (Denville Scientific, South Plainfield, New Jersey). The column was eluded using 50μL of distilled water, and the DNA dried in a vacuum desiccator. Once dried 15μL of water was added and the sample was sent to Eurofins MWG operon (Huntsville, Alabama) for sequencing with a 1:40 concentration of actin primers left and right.

2.2.3. Making alignment using sequence data and Clustalx

The sequences returned from MWG operon were not assigned a base pair (bp) due to the fact that the sequence run did not register at the beginning of the read. There was readable data and the sequence was manually read. The data obtained was crosschecked using the NCBI BLAST (Basic Local Alignment Search Tool). All the sequences matched to the flax actin gene.

Using clustalx (www.ebi.ac.uk/clustalw/) downloaded onto my computer two sequence alignments were performed one with the left actin sequences and another with the right actin sequences.

2.3 Making a phylogeny using High A and T Primers 8 and 17

2.3.1. Shotgun sequencing of Bison flax variety

Shotgun sequencing was done at Inqaba Biotechnical (Pretoria, South Africa) on a sample of L. *usitatissimum* DNA of the Bison variety. 1600 contigs were analyzed using a simple bioinformatic program created by an undergraduate in the lab, Sachin Gadani, called NUKE. This program counted the number of each base pair in each contig. From this, those sequences that were more than 69% Adenine (A) and Thymine (T) were used to design primers from these 32 contigs using the free online program Primer3 (http://frodo.wi.mit.edu/primer3/)

32 Primers were made from these contigs found to have an A and T content of greater than 69%.

2.3.2. PCR using High A and T Primers

These 32 high A and T primers were run on Flax Bison DNA (FT37) using Hot Start Promega GoTaq Green Master Mix (Promega Corp, Fitchburg, WI). The PCR reaction was as follows: an incubation time of 95°C for 1 minute 20 seconds, followed by 35 cycles of 94°C for 20 seconds, 60°C for 30 seconds, and 72°C for 90 seconds and finally a final step of 72°C for 5 minutes. PCR products were viewed on a 1.2% (w/v) agarose gel in 0.5 X TBE (1 X TBE = 100 mM Tris, 90 mM boric acid, 1 mM EDTA, pH 8.4) and run at 95 Volts for 1 hour and 30 minutes.

Two primers were chosen for subsequent PCRs that amplified at strong intensities, primer 8 and 17 were run on 38 *Linum* species including two positive controls (S6 and species 27). PCR was also done using Promega GoTaq Green Master Mix (Promega Corp, Fitchburg, WI). The PCR reaction was as follows: an incubation time of 95°C for

2 minutes, followed by 35 cycles of 94°C for 20 seconds, 60°C for 30 seconds, and 72°C for 1 minute and 30 seconds, and finally a final step of 72°C for 5 minutes. PCR products were viewed on a 1.7% (w/v) agarose gel in 0.5 X TBE (1 X TBE = 100 mM Tris, 90 mM boric acid, 1 mM EDTA, pH 8.4) and run at 75 Volts for 2 hours. Primer 7 was run a second time with the same conditions but with an annealing temperature of 60°C.

2.3.3. Sequencing of amplified products

The PCR products of select species were cleaned up using as Denville Scientific IsoPURE PCR purification kit (Denville Scientific, South Plainfield, New Jersey) and sent for sequencing as previously described. For primer 8 the following species were sent: 4, 6, 13, 14, 16, 32, 33, 34, 37, 38, and 39. For primer 17 the following species were sent: 4, 5, 6, 10, 14, 34, 37, and 38.

2.3.4. Phylogeny using clustalx

For primer 8, using clustalx two sequence alignments were performed one with the left sequences and another with the right sequences. The alignment performed by the program created a ".dnd" file that was input into an online program called Phylodendron Phylogenetic tree printer or treeapp (http://iubio.bio.indiana.edu/treeapp/treeprint-form.html). This site automatically downloaded a pdf file of the phylogeny created. Sequence data for primer 17 was inconsistent. Therefore any phylogenic work done using this data was done absent of sequence data and based solely on the PCR product pattern.

2.4. Bison contigs blasted against Bethune genome (intraspecific differences)

The high A and T repetitive sequences from the flax variety Bison were blasted against the whole of the Bethune genome (another flax variety) (personal communication, M. Deyholos; www.Linum.ca). Results were divided into three categories: no hits, few hits, and many hits.

2.4.1. Dot Plots of repetitive sequences

Those contigs that had many hits were individually input into an online dot plot program to study their repetitive patterns (Nucleic Acid Dot Plots, 1998).

2.5. LIS-1 screening on Linum species using primers Target L and R and Target L and 3' R primers on different varieties of Linum bienne (intraspecific differences)

2.5.1 PCR using Target L and R primers on Linum bienne species

PCR was performed on *Linum bienne* DNA obtained from Turkey. PCR was done using Target L and R primers with TaKaRa SpeedSTAR HS DNA Polymerase, this polymerase was chosen because of its efficiency ad robust amplification of large DNA templates (like the LIS-1 Target site that is 6KB) (TaKaRa BIO INC. Otsu, Shiga, Japan). PCR reaction was as follows: an incubation time of 98°C for 1 minute, followed by 30 cycles of 95°C for 5 seconds, 60°C for 15 seconds, and 72°C for 1 minute and 30 seconds, and finally a final step of 72°C for 5 minutes.

2.5.2 PCR using Target L 3'R primers on Linum bienne species

Also performed with these *Linum bienne* DNAs was a PCR using Target L and 3'R primers using Promega GoTaq Green Master Mix (Promega Corp, Fitchburg, WI). The PCR reaction was as follows: an incubation time of 95°C for 2 minutes, followed by 35 cycles of 94°C for 20 seconds, 60°C for 30 seconds, and 72°C for 1 minute and 30 seconds, and finally a final step of 72°C for 5 minutes. Both PCR reactions were analyzed on a 1.2% gel run at 80 Volts for 2 hours. S6 was used as a positive control for the presence of LIS-1.

2.6. LIS-1 screening on Linum species using primers Target L and R and Target L and 3'

2.6.1. Target L and R PCR screening of all Linum species

PCR of Target L and R primers on all *Linum* species was performed using both TaKaRa SpeedSTAR HS DNA Polymerase (TaKaRa BIO INC. Otsu, Shiga, Japan), and Promega GoTaq Green Master Mix (Promega Corp, Fitchburg, WI) was performed. PCR using SpeedSTAR was done with an incubation time of 95°C for 2 minutes, followed by 35 cycles of 94°C for 20 seconds, 55°C for 30 seconds, and 72°C for 1 minute and 30 seconds, and finally a final step of 72°C for 5 minutes. PCR products were then run on a 1% agorose gel run at 120 Volts for 1 hour and 15 minutes. Based on this gel species 9, 14, 17, 27, 30, 37, 38, 48, 50, 51, 52, 53, 54, 55, 56 and S6 (run as a positive control) were chosen for a second PCR reaction run at the same parameters on a 1.2% gel for 1 hour at 100 V.

The PCR reaction using GoTaq Green and Target L and R primers was performed with a an incubation time of 95°C for 2 minutes, followed by 35 cycles of 94°C for 20 seconds,

55°C for 30 seconds, and 72°C for 2 minutes (in an effort to amplify larger sequences), and finally a final step of 72°C for 5 minutes. The products were analyzed using a 1.5% gel run at 120 Volts for 1 hour. From this gel select species (9, 14, 17, 27, 30, 37, 38, 48, 50, 51, 52, 53, 54, and positive control S6) were chosen for a second PCR reaction. For both the SpeedSTAR and goTaq species 50-53 were added in the second PCR runs because they were new species added to the entire list and not available when the first PCR reactions were run.

Another PCR on all the *Linum* species with Target L and 3'R was performed on speices (9, 14, 17, 27, 30, 37, 38, 48, 50, 51, 52, 53, 54, and positive control S6) using Promega GoTaq Green Master Mix (Promega Corp, Fitchburg, WI). The PCR reaction is as follows: an incubation time of 95°C for 2 minutes, followed by 35 cycles of 94°C for 20 seconds, 60°C for 30 seconds, and 72°C for 1 minute and 30 seconds, and finally a final step of 72°C for 5 minutes. The products were run on a 1.2% gel at 100 Volts for 1 hour.

2.6.2. Sequencing of PCR products

For Target L and R using SpeedSTAR bands obtained from species 9, 30, and 38 were sent for sequencing. Bands 1-L, 4-L, 3-L, and 5-L were sent, obtained from cut out bands excised from a 1.2% gel run at 20 Volts overnight. Bands were excised and purified using Denville Scientific IsoPURE gel extraction purification kit (Denville Scientific, South Plainfield, New Jersey).

For Target L and R using goTaq Green, another gel of select species (9, 14, 17, 30, 37, 38, 48, and S6) was run for band isolation that was purified as previously described. Bands from the following samples: species 14 (band 1 and 2), species 17 (band 3 and 4),

species 30 (band 6 and7), species 37 (band 8, 9, and 10), and finally species 38 (band 11, 12, and 13) were sequenced by Eurofins MWG Operon.

The only band to amplify (besides the positive control genotroph S6) for Target L and 3' was from species 27, a flax callus DNA (FT37).

2.7. LIS-1 miRNA primers 1-9 screening using PCR

2.7.1. LIS-1 miRNA primers 2-9 run with Linum species 1-23

Linum species 1-23 were run with miRNA primers 2-9 as well as a positive controls genotroph C3 and a flax callus DNA FT37 (speices 27). The PCR using Promega GoTaq Green Master Mix (Promega Corp, Fitchburg, WI), was as follows: an incubation time of 95°C for 2 minutes, followed by 35 cycles of 94°C for 20 seconds, 60°C for 30 seconds, and 72°C for 1 minute and 30 seconds, and finally a final step of 72°C for 5 minutes. PCR products for primer 2 were run on a 1.2% gel run at 100V for 1 hour. Products for primer 3 and 4 were run on a 1.2% gel at 100 Volts for 1 hour and 30 minutes. PCR products for primer 5, 6 and 7 were run on a 1.2% gel at 120 Volts for 1 hour and 30 minutes. Finally PCR products for primers 8 and 9 were run at 125 Volts for 1 hour and 15 minutes.

2.7.2. LIS-1 miRNA primers 1-9 run with Linum species 29-47

Linum species 29-47 were also amplified 1-9 using as above. Each agarose gel was run with a positive control (S6) that will amplify each primer. PCR products for primers 1-6 were run on a 1.2% gel at 100 Volts for 1 hour and 15 minutes. PCR products for primer

7 were also run on a 1.2% gel but at 120 Volts for 1 hour and 15 minutes. Primer 8 and 9 PCR products were run on a 1.5% gel run at 45 Volts for 3 hours and 30 minutes.

2.8. LIS-1 Primer 7 analysis through PCR and sequence analysis

Primer 7 was chosen for further analysis because of the varying patterns of amplification displayed in the various *Linum* species. Based on the original amplifications performed, the following DNAs (2, 5, 7, 9, 17, 20, 23, 27, 30, 38, and S6) were subsequently analyzed. The PCR was done using Promega GoTaq Green Master Mix (Promega Corp, Fitchburg, WI). The PCR reaction was as follows: an incubation time of 95°C for 2 minutes, followed by 35 cycles of 94°C for 20 seconds, 55° C for 30 seconds, and 72°C for 1 minute and 30 seconds, and finally a final step of 72°C for 5 minutes. These PCR products were run on at 1.2% gel for 2 hours at 75 Volts. Bands of interest were extracted and purified as before. The excised bands were labeled as such 7-1 (200bp), 7-2 (200bp), 7-3 (600bp), 7-4 (900bp), 7-5 (200bp), 7-6 (600bp), 7-7 (850), 7-8 (200bp), 7-9 (650), 7-10 (750bp), 7-11 (650), 7-12 (850), 7-13 (550), 7-14 (800bp), 7-15 (550), and 7-16 (700bp). The purified DNA was reamplified for 20 cycles instead of 35 under the previous conditions. Based on the results of this second reamplification PCR of the band cut outs, the bands were not sent for sequencing.

2.9. TOPO TA cloning of primer 7 bands.

TOPO TA Cloning (Invitrogen by Life Technologies Carlsbad, CA), was performed based on the protocol downloaded online at www.invitrogen.com. And all reagents came from the TOPO TA cloning kit purchased from Invitrogen.

2.9.1. Agar broth plates for cloning

Agar plates were made by weighing out 3.75 grams of agar and 6.25 grams of LB broth. 250 ml of distilled water were added. The solution was mixed by being heated in a microwave for 2 minutes. The solution was then put into an autoclave on a short liquid cycle (setting 4). A water bath was set to 65°C and once the autoclaved broth completed the cycle it was placed in the water bath too both cool and maintain a warm temperature. After 30 minutes 437.5 µl of Xgal, 100µl of IPTG and 250µl of Ampicillin were added to the broth mixture and 13 plates poured.

2.9.2. Generation of PCR product for cloning reaction

The TOPO protocol suggested using fresh PCR product for the cloning reaction. So using primer 7 and select *Linum species* (9, 17, 27, 30, 38, and S6), a PCR was done using Promega GoTaq Green Master Mix (Promega Corp, Fitchburg, WI). The PCR reaction was as follows: an incubation time of 95°C for 2 minutes, followed by 35 cycles of 94°C for 20 seconds, 55°C for 30 seconds, and 72°C for 1 minute and 30 seconds, and finally a final step of 72°C for 10 minutes. The protocol also suggested a final extension of at least 7 minutes at 72°C to ensure the PCR products is not only full length but 3' adenylated. Instead of the usual 5 minute final extension at 72°C, we opted for a 10 minute extension as suggested by the protocol (TOPO TA User Manual, 2006).

2.9.3. Setting up TOPO Cloning Reaction

Setting up the cloning reaction required 4ul of the PCR product (1.5µl of each PCR reaction tube and then extracting 4µl from that). 1µl of salt solution was added and 1µl of the TOPO® vector, making a final volume of 6µl. The reaction was mixed gently and incubated at room temperature for 10 minutes. The protocol suggests an incubation time between 30 seconds and 30 minutes, noting that 5 minutes should sufficient of colonies for analysis. The reaction was placed on ice to await transformation.

2.9.4. Transformation of Competent Cells

Following page 10 of the protocol (TOPO TA User Manual, 2006), a vial of One Shot® Chemically Competent *E. coli* cells were allowed to thaw on ice. Once thawed, 2µl of the cloning reaction was added and mixed gently (not by pipetting up and down). This mixture was incubated on ice for 20 minutes. The cells underwent heat shock by being placed in a 42°C waterbath for 30 seconds then immediately after placed on ice. 250µl of room temperature S.O.C. medium was added and the tube was placed in a shaker at 37°C for 1 hour.

Plates were pre-warmed for 20 minutes at 37°C and two different volumes were plated to ensure at least one plate would have well spaced colonies. One plate had 40µl of cells and 10µl of S.O.C medium and other had 10µl of cells and 40µl S.O.C medium added. These plates were incubated overnight (16 hours) at 37°C.

In addition to the plates, 5ml of LB broth containing 2.5µl of ampicillin was inoculated with 10µl of cells and incubated overnight in a 37°C shaker.

2.9.5. Plasmid prep verification and PCR using M13 primers

To ensure the presence of recombinant plasmids, the overnight culture was used to isolate plasmid using the Qiagen Plasmid Mini Kit (Qiagen Hilden, Germany). An aliquot of this isolated plasmid run on a 1% agorose gel run at 50 V for 1 hour and 30 minutes.
Once the presence of plasmid was confirmed, a PCR was performed on the purified plasmid (1µl) using the M13 primers included in the TOPO TA Cloning kit, and also primer 7. This reaction was done using Promega GoTaq Green Master Mix (Promega Corp, Fitchburg, WI). The PCR reaction was as follows: an incubation time of 95°C for 2 minutes, followed by 35 cycles of 94°C for 20 seconds, 55°C for 30 seconds, and 72°C for 1 minute and 30 seconds, and finally a final step of 72°C for 5 minutes. The products were fun on at 1.2% gel at 100 Volts for 1 hour. This gel confirmed the recombinant molecules were present in the transformed colonies.

2.9.6. Colony PCR

The overnight culture was re-plated using 10^3 and 10^4 dilution. These plates were grown overnight (along with the broth delusions in the shaker) at 37°C.
14 hours later, 40 colonies were picked from each plate and placed in a new plate to be grown. Each picked colony, once plated was also put into a PCR tube with of 50µl of water. These 40 tubes were placed in a water bath at 98°C for 10 minutes to lyse the cells for the colony PCR. Once this was done, 1µl of each 40 tubes (80 total with 2 dilutions) underwent a PCR reaction using the M13 primers. The PCR was done using Promega GoTaq Green Master Mix (Promega Corp, Fitchburg, WI). The PCR reaction was as follows: an incubation time of 95°C for 2 minutes, followed by 35 cycles of 94°C for 20 seconds, 55°C for 30 seconds, and 72°C for 1 minute and 30 seconds, and finally a final

step of 72°C for 5 minutes. Both were run on a 1.2% gel at 100 Volts for 1 hour and 10 minutes.

Chapter 3. Results: Determining homology of the actin gene

3.1. PCR and sequencing of Actin Primers on all Linum species

PCR using the flax actin gene primers was used to amplify *Linum* species DNA. Species 1—49 were used as well as S6 as a positive control, and the standard blank control (Figure 3.1 A and B). The species that amplified on this gel include: 4. *Linum grandiflorum caeruleum*, 5. *Linum narbonense*, 6. *Linum nervosum*, 9. *Linum perenne*, 10. *Linum perenne*, 16. *Linum maritimum*, 20. *Linum flavum*, 30. *Linum corymibiferum*, and 31. *Linum elegans*. Most did not amplify with the flax actin primers, suggesting that this gene is not as highly conserved in this genus as previously expected, and this information can be used to make inferences as to a grouping species with like molecular characteristics.

Based on these results, another PCR was done using the species (4, 5, 6, 9, 10, 16, 17, 20, 27, 30, 31, 38, 49, and S6) that amplified from this original gel (Figure 3.2), species 27, *Linum usitatissimum*, was added to this PCR as a different variety of flax to examine besides S6. A rePCR was done to get more intense bands for sequencing (Figure 3.3).

Figure 3.1 A. Gel electrophoresis of PCR with flax actin primers used on *Linum* species 1-37. The species to amplify (150bp) include 4, 5, 6, 9, 10, 16, 20, 30, and 31. Hyperladder 4 was used and 12µl was loaded. This is a 1.5% gel run at 150 Volts for 35 minutes. 5µl of marker IV was added.

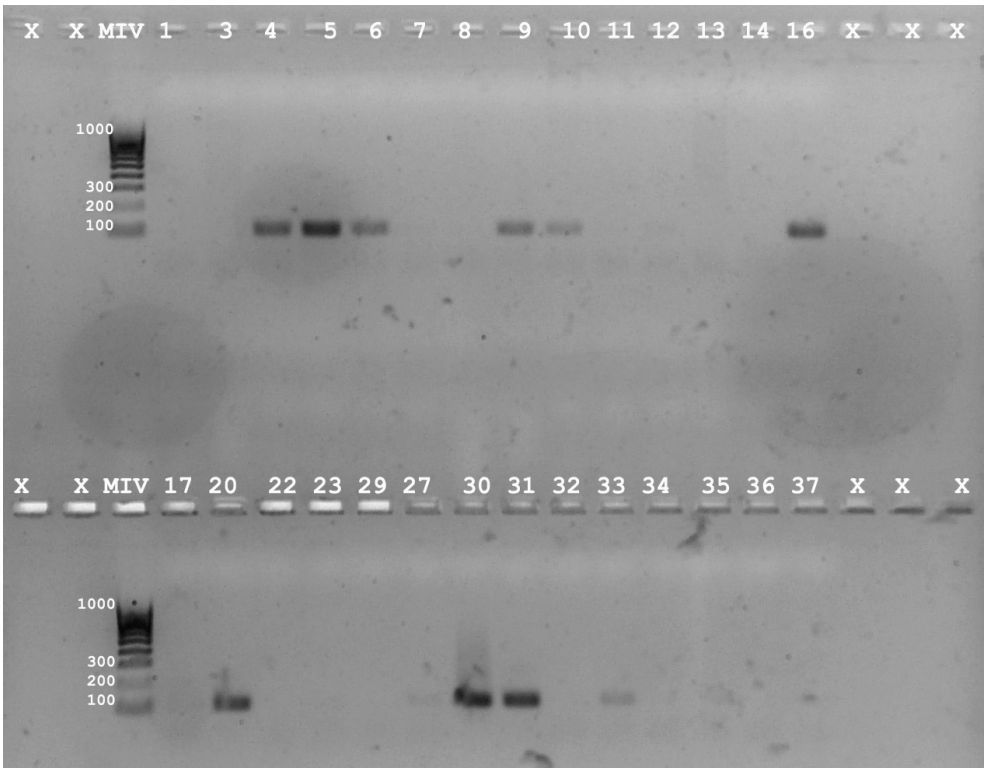


Figure 3.1 B. Gel electrophoresis of PCR with flax actin primers used on *Linum* species 38-S6. The species to amplify (150bp) include 38, 39, 49, S6 Hyperladder 4 was used and 12µl was loaded. This is a 1.5% gel run at 150 Volts for 35 minutes. 5µl of marker IV was added.

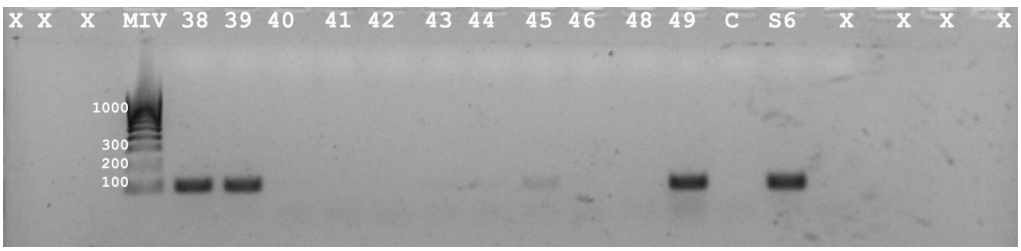


Figure 3.2. Agarose gel of PCR using Actin genes of select *Linum* species that amplified in Figure 3.1 A. and B. This PCR was done for 30 cycles instead of 35, which accounts for the faint bands. Hyperladder II was used as a marker and 10µl was loaded into this gel. 1.7% gel run at 75 Volts for 1 hour and 30 minutes. 5µl of hyperladder II was added

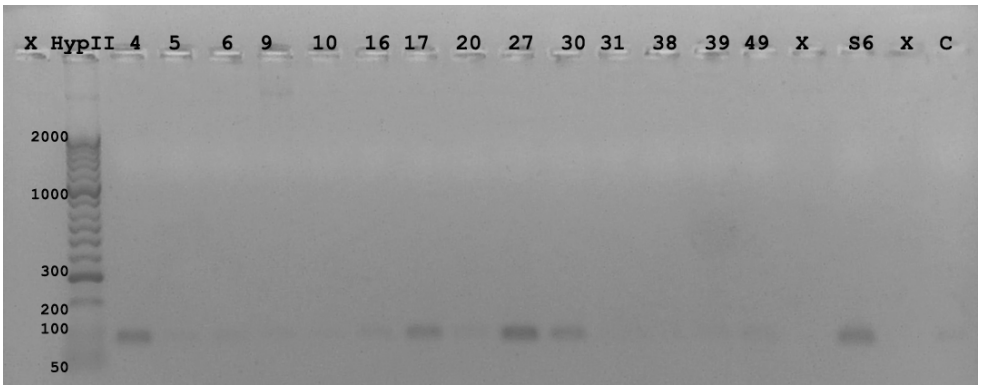
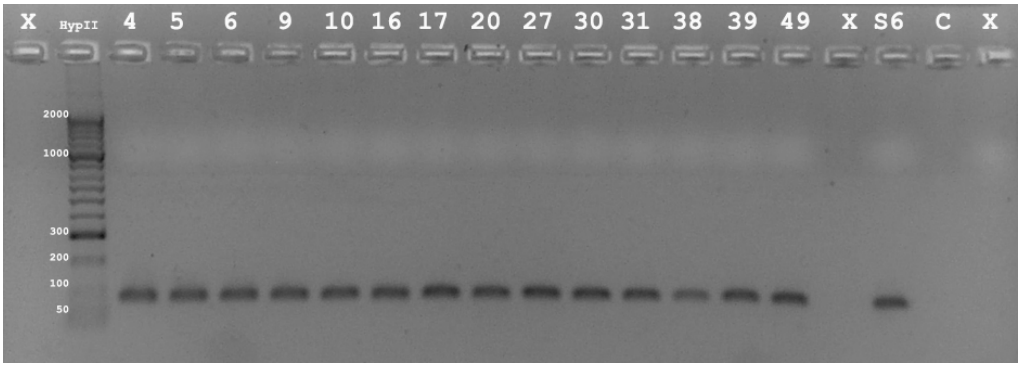


Figure 3.3. Agorose gel of a rePCR of the products seen in Figure 3.2. run for 20 cycles and 10µl were loaded into the wells. These 150bp bands were set for sequencing. This is a 1.7% gel run at 90 Volts for 1 hour and 30 minutes. 5µl of hyperladder II was used.



Using the PCR products seen in Figure 3.3, we did a clean up and sent them for sequencing (*Appendix A*).

3.2. Alignment using actin sequence data

From this sequencing we used the program clustalx to align the sequences. From this clustalx alignment, (Figure 3.4 A and B) for actin L and R sequences, it clear that these two sets of sequences are identical. This suggests the actin left and right sequences are highly conserved in the species that the primers amplify.

Figure 3.4 A. Clustalx alignments for actin sequences left; from this alignment it is clear these sequences share perfect homology.

```
CLUSTAL 2.1 multiple sequence alignment


A-4L_L_grandiflorum_caeruleum       GAACGGCTTCATCGACCACAGGAAGTGCTTCTAAAGCTCCAGGAATTGGG
A-5L_L_narbonense                   GAACGGCTTCATCGACCACAGGAAGTGCTTCTAAAGCTCCAGGAATTGGG
A-6L_L_nervosum                     GAACGGCTTCATCGACCACAGGAAGTGCTTCTAAAGCTCCAGGAATTGGG
A-9L_L_perenne                      GAACGGCTTCATCGACCACAGGAAGTGCTTCTAAAGCTCCAGGAATTGGG
A-10L_L_perenne                     GAACGGCTTCATCGACCACAGGAAGTGCTTCTAAAGCTCCAGGAATTGGG
A-16L_L_maritimum                   GAACGGCTTCATCGACCACAGGAAGTGCTTCTAAAGCTCCAGGAATTGGG
A-17L_L_perenne                     GAACGGCTTCATCGACCACAGGAAGTGCTTCTAAAGCTCCAGGAATTGGG
A-20L_L_flavum                      GAACGGCTTCATCGACCACAGGAAGTGCTTCTAAAGCTCCAGGAATTGGG
A-27L_L_usitatissimum               GAACGGCTTCATCGACCACAGGAAGTGCTTCTAAAGCTCCAGGAATTGGG
A-30L_L_corymibiferum               GAACGGCTTCATCGACCACAGGAAGTGCTTCTAAAGCTCCAGGAATTGGG
A-31L_L_elegans                     GAACGGCTTCATCGACCACAGGAAGTGCTTCTAAAGCTCCAGGAATTGGG
A-39L_L_mucronatum                  GAACGGCTTCATCGACCACAGGAAGTGCTTCTAAAGCTCCAGGAATTGGG
A-S6L_L_usitatissimum               GAACGGCTTCATCGACCACAGGAAGTGCTTCTAAAGCTCCAGGAATTGGG
A-49L_L_strictum_subsp_Strichu      GAACGGCTTCATCGACCACAGGAAGTGCTTCTAAAGCTCCAGGAATTGGG
                                    **************************************************

A-4L_L_grandiflorum_caeruleum       ATGCT
A-5L_L_narbonense                   ATGCT
A-6L_L_nervosum                     ATGCT
A-9L_L_perenne                      ATGCT
A-10L_L_perenne                     ATGCT
A-16L_L_maritimum                   ATGCT
A-17L_L_perenne                     ATGCT
A-20L_L_flavum                      ATGCT
A-27L_L_usitatissimum               ATGCT
A-30L_L_corymibiferum               ATGCT
A-31L_L_elegans                     ATGCT
A-39L_L_mucronatum                  ATGCT
A-S6L_L_usitatissimum               ATGCT
A-49L_L_strictum_subsp_Strichu      ATGCT
                                    *****
```

Figure 3.4 B. Clustalx alignments for actin sequences right; from this alignment it is clear these sequences share perfect homology.

```
CLUSTAL 2.1 multiple sequence alignment


A-4R_L_grandiflorum_caeruleum       CTCGGGAGATTGAGGCCGGATTCGTCGTACTCAGACTTCGAGATCCACAT
A-5R_L_narbonense                   CTCGGGAGATTGAGGCCGGATTCGTCGTACTCAGACTTCGAGATCCACAT
A-6R_L_nervosum                     CTCGGGAGATTGAGGCCGGATTCGTCGTACTCAGACTTCGAGATCCACAT
A-9R_L_perenne                      CTCGGGAGATTGAGGCCGGATTCGTCGTACTCAGACTTCGAGATCCACAT
A-10R_L_perenne                     CTCGGGAGATTGAGGCCGGATTCGTCGTACTCAGACTTCGAGATCCACAT
A-16R_L_maritimum                   CTCGGGAGATTGAGGCCGGATTCGTCGTACTCAGACTTCGAGATCCACAT
A-17R_L_perenne                     CTCGGGAGATTGAGGCCGGATTCGTCGTACTCAGACTTCGAGATCCACAT
A-20R_L_flavum                      CTCGGGAGATTGAGGCCGGATTCGTCGTACTCAGACTTCGAGATCCACAT
A-27R_L_usitatissimum               CTCGGGAGATTGAGGCCGGATTCGTCGTACTCAGACTTCGAGATCCACAT
A-30R_L_corymibiferum               CTCGGGAGATTGAGGCCGGATTCGTCGTACTCAGACTTCGAGATCCACAT
A-31R_L_elegans                     CTCGGGAGATTGAGGCCGGATTCGTCGTACTCAGACTTCGAGATCCACAT
A-38R_L_maritimum                   CTCGGGAGATTGAGGCCGGATTCGTCGTACTCAGACTTCGAGATCCACAT
A-39R_L_mucronatum                  CTCGGGAGATTGAGGCCGGATTCGTCGTACTCAGACTTCGAGATCCACAT
A-S6R_L_usitatissimum               CTCGGGAGATTGAGGCCGGATTCGTCGTACTCAGACTTCGAGATCCACAT
A-49R_L_strictum_subsp_Strichu      CTCGGGAGATTGAGGCCGGATTCGTCGTACTCAGACTTCGAGATCCACAT
                                    **************************************************

A-4R_L_grandiflorum_caeruleum       CTGCTG
A-5R_L_narbonense                   CTGCTG
A-6R_L_nervosum                     CTGCTG
A-9R_L_perenne                      CTGCTG
A-10R_L_perenne                     CTGCTG
A-16R_L_maritimum                   CTGCTG
A-17R_L_perenne                     CTGCTG
A-20R_L_flavum                      CTGCTG
A-27R_L_usitatissimum               CTGCTG
A-30R_L_corymibiferum               CTGCTG
A-31R_L_elegans                     CTGCTG
A-38R_L_maritimum                   CTGCTG
A-39R_L_mucronatum                  CTGCTG
A-S6R_L_usitatissimum               CTGCTG
A-49R_L_strictum_subsp_Strichu      CTGCTG
                                    ******
```

Chapter 4. Results: Making a phylogeny using High A and T Primers

4.1 Phylogeny using Primer 8

PCR products amplified from high A and T primer 8 (Figure 4.1) were run out on a gel and sent for sequencing. The amplified bands not only are of different intensities, but also different sizes (number of bp). Comparing species 4 (*L. grandiflorum caeruleum*) 200bp, 5 (*L. narbonense*) 250bp, and 6 (*L. nervosum*) 225bp, you can see that all three bands are of different sizes with species 4 matching the band pattern seen mostly with this primer. Species 13 (*L. bienne*) 225bp, 16 (*L. maritimum*) 225bp, and 39 (*L. mucronatum*) 225bp are also different than the 200bp pattern. Using this data we can make inferences as to which speices are more alike than others, creating groupings that showed like patterns.

The following bands were sent for sequencing: 4 (*L. grandiflorum caeruleum*), 6 (*L. nervosum*), 13 (*L. bienne*), 14 (*L. grandiflorum reubrum*), 16 (*L. maritimum*), 32 (*L. flavum* UK), 33 (*L. flavum* GER), 34 (*L. grandiflorum*), 37 (*L. marginale*), 38 (*L. maritimum*), and 39 (*L. mucronatum*). From this sequencing (*Appendix B*), Clustalx was used to do an alignment (*Appendix C)* and two phylogenic trees (Figure 4.2 A and B).

The analysis for these trees will be compared to data currently known of *Linum* phylogenies, mainly the tree seen in Figure 1.3. The tree for 8_L as expected had both *L. flavum* DNAs on the same branch, also as expected *L. grandiflorum desf.* and *L. grandiflorum reubrum*, which are closely related, exhibit that in the graph, whereas *L. grandiflorum caeruleum* which is very different from the two, is far them in the tree. *L. maritimum* and *L. perenne* are closely related species, and this is reflect in their sequences for 8_L. *L. marginale* and *L. maritimum* are also closely related, which is also

reflected in this tree, but *L. marginale* is also closely related to *L. bienne*, but as seen in the tree, they do not share such a close homology when it comes to the sequences obtained from primer 8_L amplification. *L. mucronatum* is not included in Figure 3, but according to this graph it is shares similarity in the 8_l sequence to *L. bienne*.

The tree for 8_R is quite different than the pattern seen for 8_L. *L. grandiflorum reubrum* and *L. grandiflorum* are on the same branch as before, but in these sequences *L. maritimum* and *L. mucronatum* are closer together in this tree where as *L. flavum* GER and *L. perenne* are both similar whose sequences seem to be evolutionary offspring of *L. bienne*.

Figure 4.1. Agarose gel of primer 8 run with *Linum species 1-49*. Note the different intensities of the bands as well as some being a different number of base pairs (bp). 1.7% gel, loaded 10µl, run at 75 Volts for 2 hours. 5µl of Hyperladder II was used.

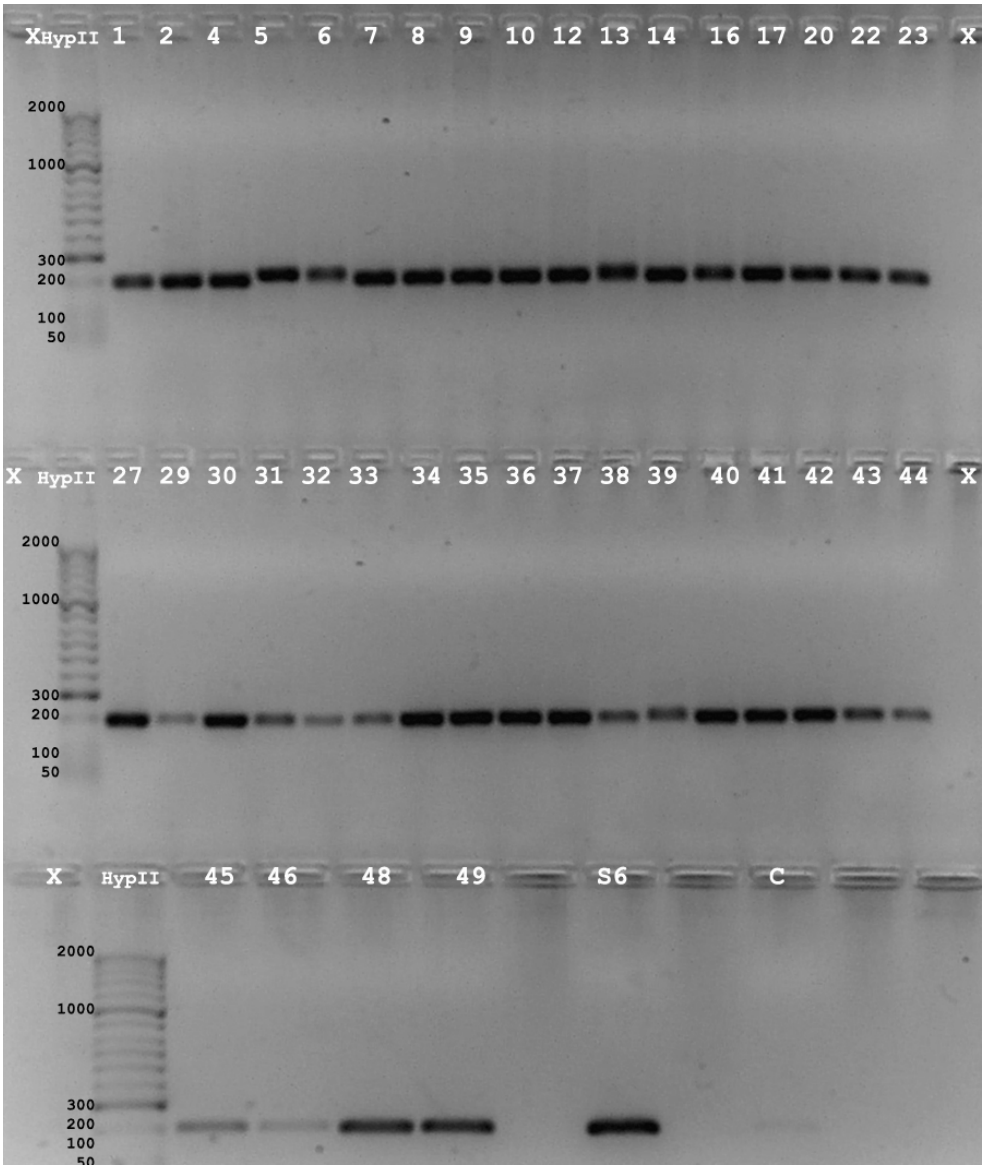
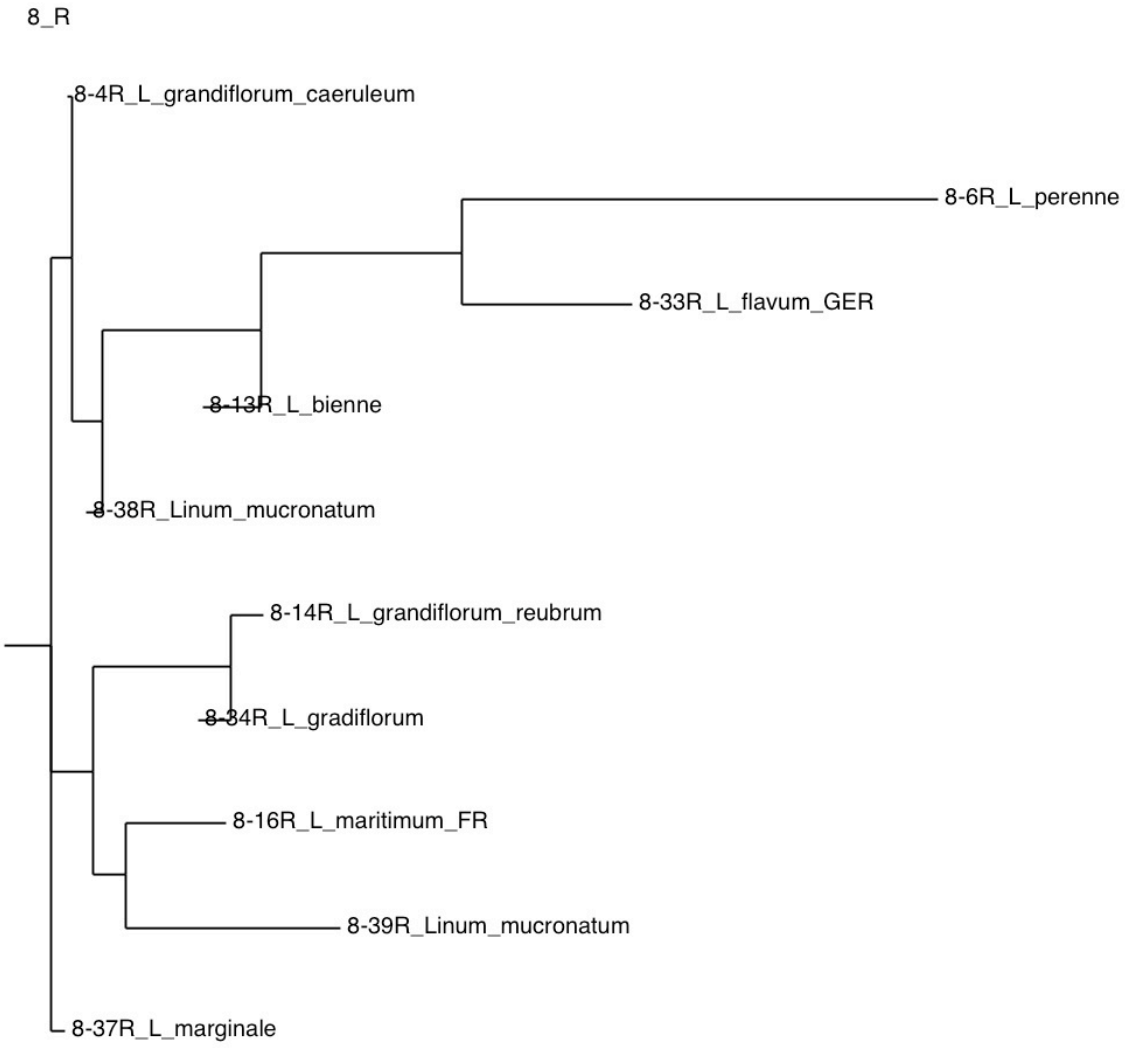


Figure 4.2 A. Phylogenic tree created from the primer 8 left sequences alignment on clustalx.

Figure 4.2 B. Phylogenic tree created from the primer 8 right sequences alignment on clustalx.

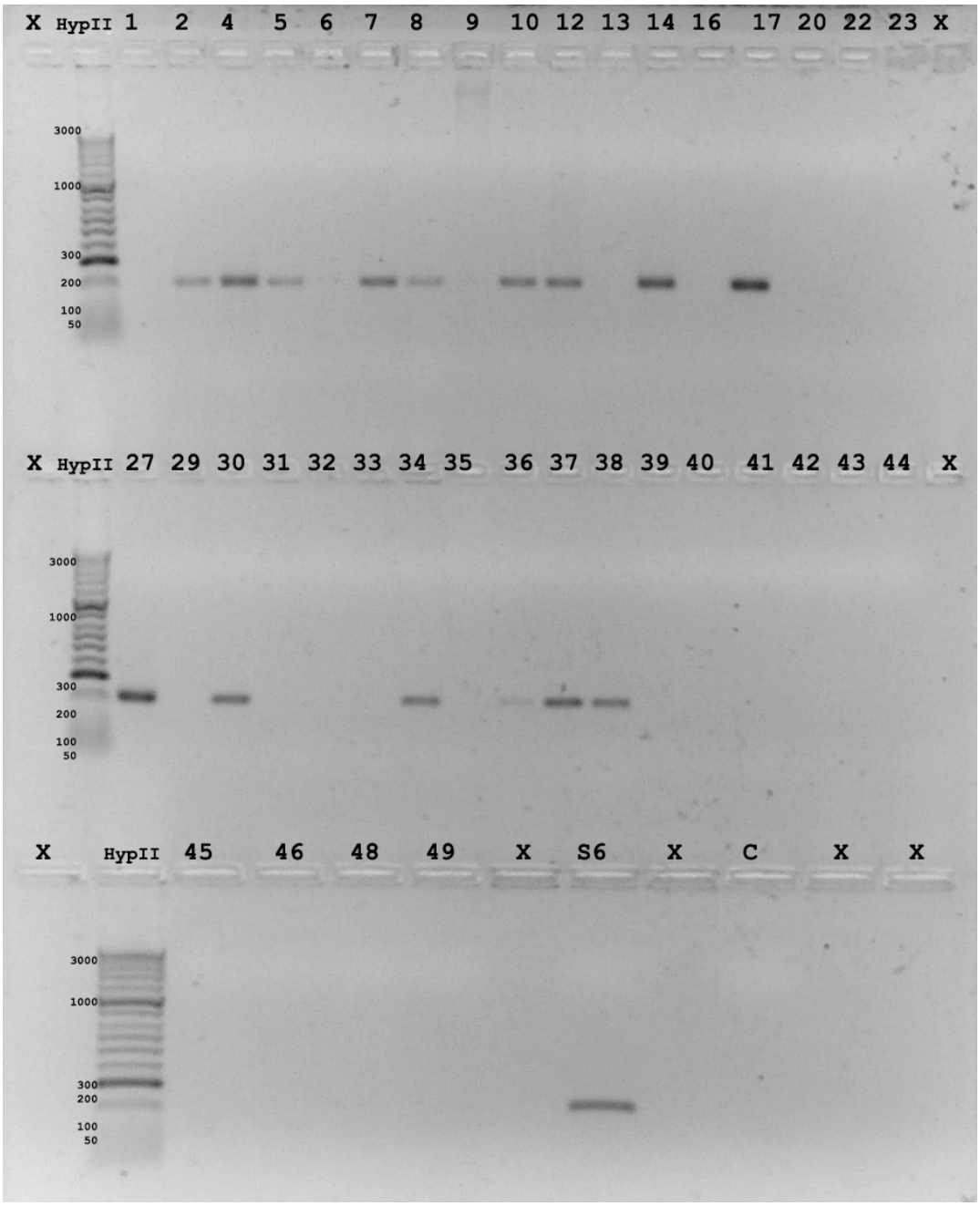


4.2. Anaylsis of PCR using High A and T Primer 17

PCR of primer 17 on all *Linum* species was done (Figure 4.3), and unlike primer 8, not all species amplified these primers. Those that did include: 2 (*L. austriacum*), 4 (*L. grandiflorum caeruleum*), 5 (*L. narbonense*), 7 (*L. perenne*), 8 (*L. perenne*), 10 (*L. perenne*), 12 (*L. austriacum*), 14 (*L. grandiflorum reubrum*), 17 (*L. perenne*), 27 (*L. usitatissimum),* 30 (*L. corymibiferum*), 34 (*L. grandiflorum*), 37 (*L. marginale*), 38 (*L. maritimum*), and S6 (*L. usitatissimum*). Species 4, 5, 6, 10, 14, 34, 37, and 38 were sent for sequencing.

The sequencing did not yield any data, but comparing the band intensities, one can make comparisons as to which species are most similar to flax when it comes to primer 17 relationship.

Figure 4.3. PCR of high A and T primer 17. Species that amplified include: 2, 4, 5, 7, 8, 10, 12, 14, 17, 27, 30, 34, 37, 38, and S6. Hyperladder II marker (5μl) was used and 10μl of PCR product was added to each well. 1.7% gel run at 75 Volts for 2 hours.



Chapter 5. Results: Bison and Bethune, comparing High A and T sequences between two varieties of the same species

5.1. Blast data: distribution of hits

The results of the Blast of High A and T sequence repetitive contigs from flax variety Bison against the whole of the Bethune genome are summarized in the Table 5.1 below. You can see there are 4 contigs that are not represented in the Bethune genome.

Table 5.1. Table summarizing the blast between flax variety Bison high A and T contigs and the whole of the Bethune genome (another flax variety). Most of the sequences had a few hits, but the focus of this analysis will be on the highly repetitive sequences (or Many hits).

No hits:	Few hits:		Many hits:
Contig 228	Contig 21	Contig 1034	Contig 17
Contig 368	Contig 42	Contig 1075	Contig 48
Contig 445	Contig 114	Contig 1158	Contig 145
Contig 1474	Contig 118	Contig 1175	Contig 167
	Contig 209	Contig 1229	Contig 202
	Contig 214	Contig 1248	Contig 309
	Contig 259	Contig 1325	Contig 388
	Contig 299	Contig 1343	Contig 438
	Contig 368	Contig 1353	Contig 488
	Contig 432	Contig 1370	Contig 515
	Contig 469	Contig 1375	Contig 985
	Contig 502	Contig 1398	Contig 1126
	Contig 568	Contig 1485	Contig 1207
	Contig 599	Contig 1497	Contig 1300
	Contig 601	Contig 1515	Contig 1500
	Contig 648	Contig 1537	
	Contig 664	Contig 1545	
	Contig 722		
	Contig 809		
	Contig 830		
	Contig 849		
	Contig 878		
	Contig 972		
	Contig 976		
	Contig 992		
	Contig 1002		

From this table note most of the contigs were represented with just a few hits, but for the purposes of this study focus is put on the highly repetitive sequences or those registering

"many hits". Contig 388 (76.3% As and Ts) was chosen first mainly because it is the sequences from which High A and T primer 8 was created. From the dot plot, note the main cluster of repetitive regions around base 97-119 (Figure 5.1). The other contigs chosen were 48 (72% As and Ts) and 438. They hit multiple times in the Bethune genome, over 50 hits each. This is reflected in the dot plot for contig 48 (Figure 5.2). Note the two main repetitive regions see diagonally across each end. Contig 438 (71% As and Ts) had four main regions of repetition (Figure 5.3).

Figure 5.1. Dot plot analysis on High A and T contig 388. Parameters were as follows: Window=9, Mismatch=2

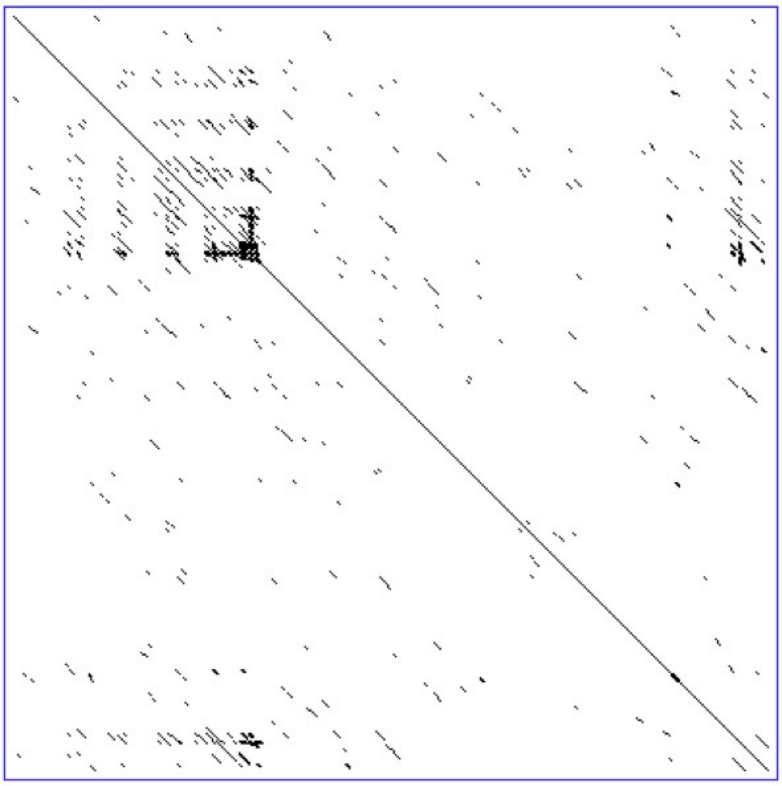


Figure 5.2. Dot plot for contig 48, note the repetitive regions seen diagonally on either end of the plot. This is expected because of the frequency of hits seen in the BLAST of this contig with the Bethune genome. Parameters are as follows: Window size=9, Mismatch Limit=1

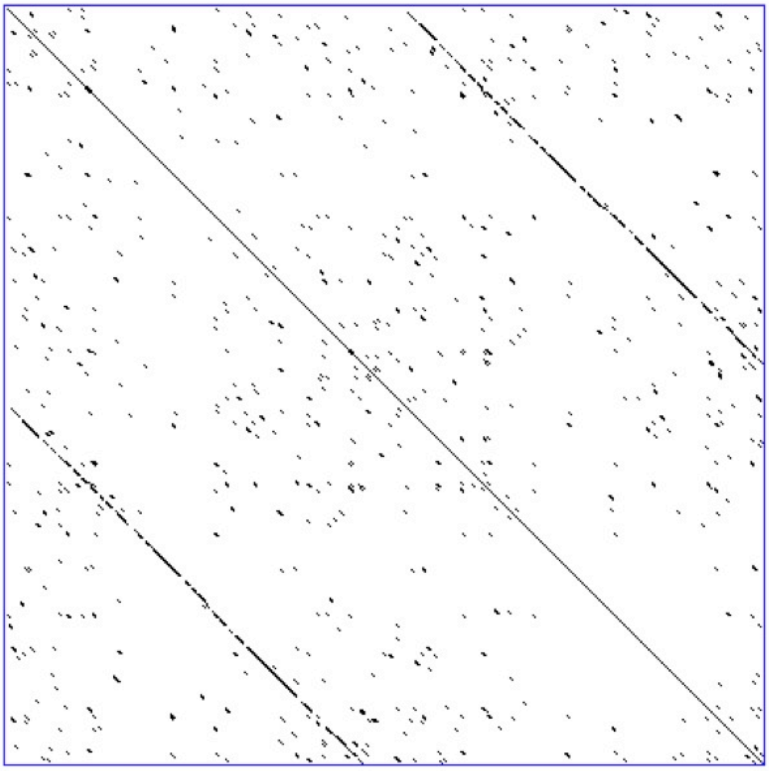
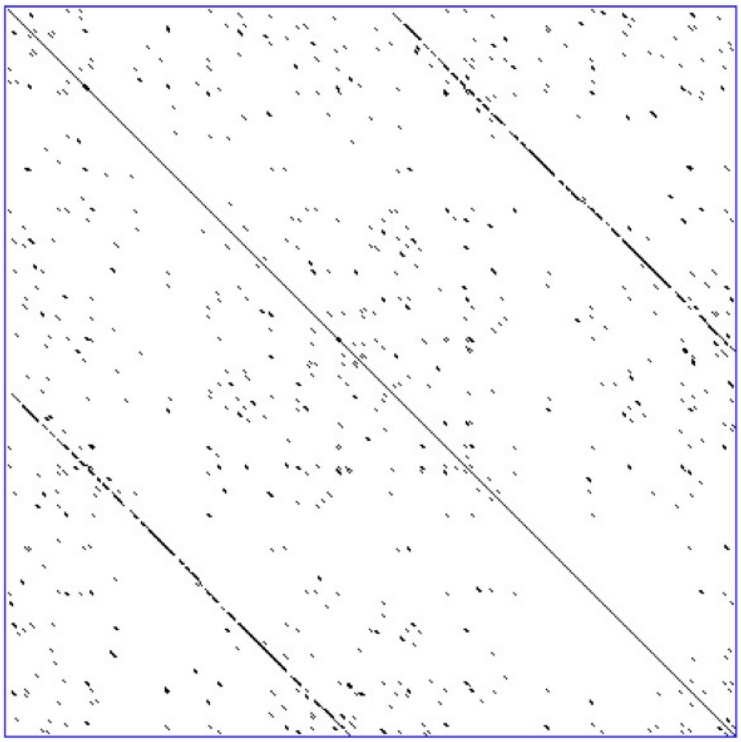


Figure 5.3. Dot plot of contig 438, note this plot has many more repetitive areas than the previous two. Parameters: Window size=9, Mismatch=1



5.2. Highly repetitive sequences: characterization using dot plot analysis

Dot plots are useful tool for sequence analysis, and for this project they will be useful in discovering repetitive regions within a sequence. "Window size" represents the number of bases in a sliding window that is moved along each sequence and compared to generate a single data point on the plot. "Mismatch Limit" will determine how similar two sequences are in a window to constitute a "match". Stringency is adjusted by

changing the window size and mismatch limit with a large window of comparison and a lower mismatch limit increasing the stringency. Repetitive regions will create diagonal lines from one corner to another, and the random scattering of dots represents small regions that match fairly closely by chance. Lowing the mismatch limit and lower the window size will reduce these scatterings (Nucleic Acid Dot Plots, 1998).

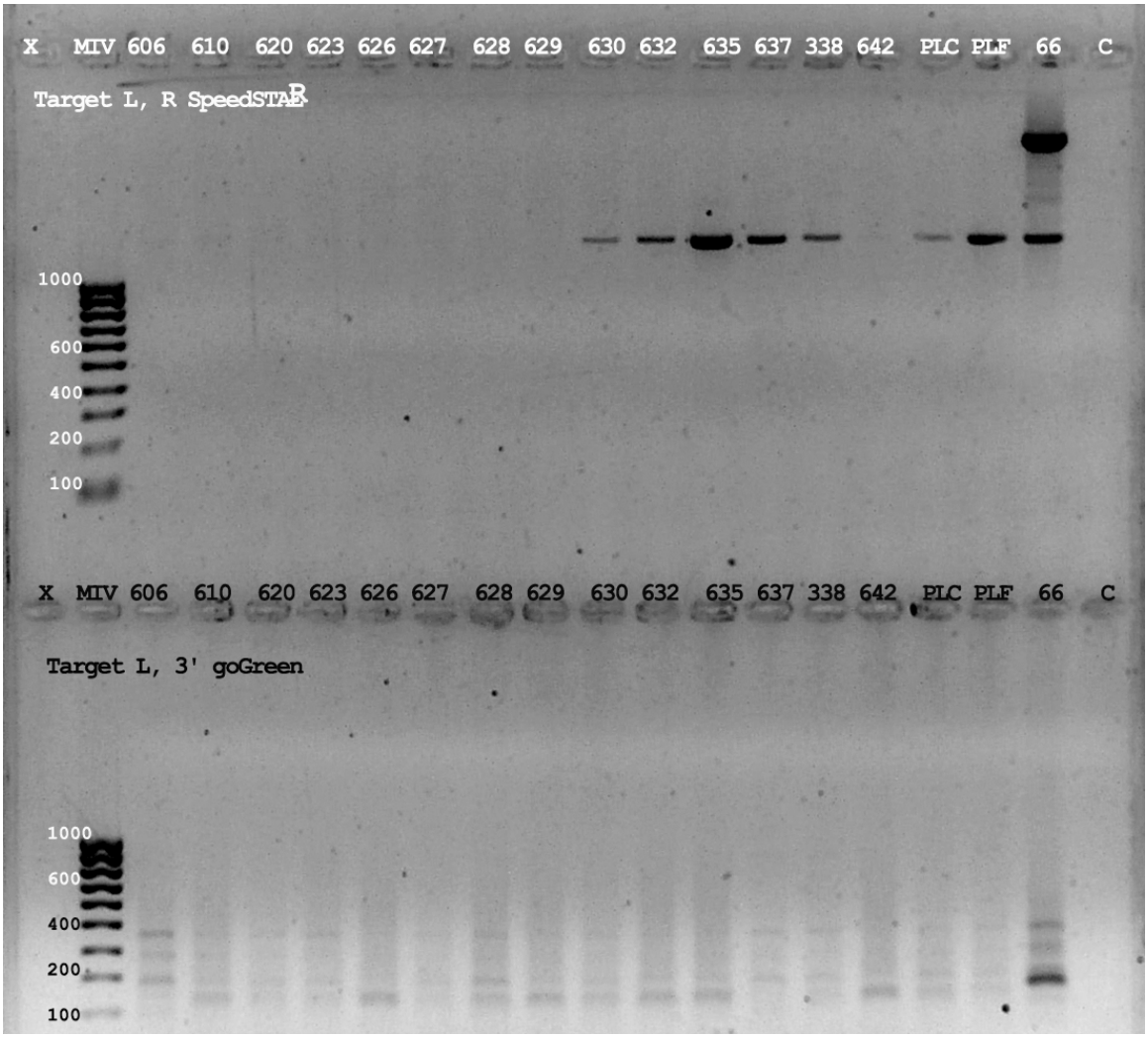
For the purposes of this study, four contigs were chose for dot plot analysis.

Chapter 6. Results: LIS-1 characterization of *Linum bienne*

6.1. Target left and right primers and Target left and 3'right primers

Bienne number 630, 632, 635, 637, and 638 (338) al amplified with Target L and R, they amplified a region associated with LIS-1 not being present. Looking at the positive control S6 (labelled 66) one can see the pattern associated with LIS-1 being present and inserted. The large fragment at about 6kb is LIS-1 inserted in the target region. The smaller band, the one corresponding with what is seen in the other bands that amplified is the insertion region. The pattern of expression for Target L and 3' R is most unique. Each bienne DNA is exhibiting a different pattern and with different size faint bands. The positive control S6 was the only one to amplify with intense bands.

Figure 6.1. Gel run with *L. bienne* DNAs from Turkey. Amplification with target L and R and Target L and 3'R. Species 630, 632, 635, 637, and 638 amplified with Target L and R as well as the positive control 66 (S6). (PLC and PLF were samples loaded for a different purpose and can be ignored in this analysis). For Target L and 3'R the only lane to amplify with strong bands was the positive control S6, but all other lanes did amplify with faint bands of different size and intensity.



6.2. Sequence data of target left and right bands from two L. bienne varieties

Sequence data obtained from Margaret Cullis (unpublished data) sheds some light on the characteristics of these bands appearing when select Bienne DNAs are amplified with Target L and R. These two bands were not included in the gel above, but they exhibit a similar band pattern when run with Target L and R. They come from Bienne number 608 and 621. This alignment shows that these amplified bands share significant homology to the Target L and R sequence from Flax (Figure 6.2).

Figure 6.2. ClustalW2 alignment of *L. bienne* sequences 608 and 620 with Target L and R sequence obtained from flax. Note the homology between these sequences.

```
CLUSTAL 2.1 multiple sequence alignment

621             GGGTTTCAGAACTGTAACGAATGAGTAAGGGGGGAAGAGGAGTTGAACAAATTAGTGAAG 60
608             GGGTTTCAGAACTGTAACGAATGAGTAAGGGGGGAAGAGGAGTTGAACAAATTAGTGAAG 60
Target_LR       GGGTTTCAGAACTGTAACGAATGAGTAAGGGGGGAAGAGGAGTTGAACAAATTAGTGAAG 60
                ************************************************************

621             CAAAGCAAGGAGCTGGGAGGAATCAGAGCTTACTGGTTTGAAACTCGTCAAGGAAAGACA 120
608             CAAAGCAAGGAGCTGGGAGGAATCAGAGCTTACTGGTTTGAAACTCGTCAAGGAAAGACA 120
Target_LR       CAAAGCAAGGAGCTGGGAGGAATCAGAGCTTACTGGTTTGAAACTCGTCAAGGAAAGACA 120
                ************************************************************

621             TTGGTTGAGATAATGGCGTCTTTGAATTTGATTTCTTCTGTTATCCGCCCCCTTCTTCAG 180
608             TTGGTTGAGATAATGGCGTCTTTGAATTTGATTTCTTCTGTTATCCGCCCCCTTCTTCAG 180
Target_LR       TTGGTTGAGATAATGGCGTCTTTGAATTTGATTTCTTCTGTTATCCGCCCCCTTCTTCAG 180
                ************************************************************

621             TTCTGCTCAACGTGATGACCCTTCCCGCCACCGATTTAGGCAACTCTGGCGTATCTGCAA 240
608             TTCTGCTCAACGTGATGACCCTTCCCGCCACCGATTTAGGCAACTCTGGCGTATCTGCAA 240
Target_LR       TTCTGCTCAACGTGATGACCCTTCCCGCCACCGATTTAGGCAACTCTGGCGTATCTGCAA 240
                ************************************************************

621             TTGGGGATTGAAAGGGAGGTATTGTAGAGACTCCTTTTTTTTTTCTTGTATATTGGTGATT 300
608             TTGGGGATTGAAAGGGAGGTATTGTAGAGACTCCTTTTTTTTTTCTTGTATATTGGTGATT 300
Target_LR       TTGGGGATTGAAAGGGAGGTATTGGAGAGACTCCTTTTTTTTTCCTTGTATATTGGTGGTT 300
                ************************ ***************** ************** **

621             AGTTAGGAGGAGCGATTATAATCGTAGCCGGGGGATATGGGGATGGAACGAAATCTATCA 360
608             AGTTAGGAGGAGCGATTATAATCGTAACCGGTGGATATAGGGATGGAACGAAATCTATCA 360
Target_LR       AGTTAAGAGGAGCGATTATAATCGTAATCGGTGGGTACAGGGATGGAACGAAATTTATCA 360
                ***** ********************  *** ** **  *************** *****

621             CCGAAATAGAAATGGTGATGATTATGGTTAATAACCGTCACAAAGGTGGATGAATAATGG 420
608             CCGAAATAGAAATGGTGATGATTATGGTTAATAACCGTCACAAAGGTGGATGAATAATGG 420
Target_LR       CCGAAATAGAAATAGTGATGATTATGGTTAATAACCGCCACAAAGGTGGATGAATAATGG 420
                ************* ********************** **********************

621             ATATGCCATACCCATCCCCGAACCTCCTCGTTGTCAAGCCTAAATCCGACCTGTTACAAG 480
608             ATATGCCATAACCATCCCCGAACCTCCTCGTTGTCAAGCCTAAATCCGACCTGTTACAAG 480
Target_LR       ATATGCCATAACCATCCCCGAACCTCCTCGTTGCCAAGTCTAAATCCAAGCTGTTACAAG 480
                ********** ********************** **** ******** * **********

621             ATAAACTTTATATTATTTTCACACCATATAATATATTACATACATTATTATAAAAAAAA- 539
608             ATAAACTTTGTATTATTTTCACACCATATAATATATTACATACATTATTATAAAAAAAAA 540
Target_LR       ATAAACTTTGTATTATTTTCACACCATATAATATACTACATACATTATTATAAAAAAA- 539
                ********* ************************ ***********************

621             GTTATTTATGAGGACAATATAATACTTTAACTTTTATTTAATCAAATATAACTCATTATT 599
608             GTTATTTATGAGGACAATATAATACTTTAACTTTTATTTAATCAAATATAACTCATTATT 600
Target_LR       GTTATTTATGAGGGCAATATAATACTTTAACTTTTATTTAATCAAATATAACTCTTTATT 599
                ************* ****************************************** *****

621             TTGAGTTAATATAGCAGAATAATTTCTATAAAATCCAAAATATTTTATAAGTATCATCTT 659
608             TTGAGTTAATATAGCAGAATAATTTCTATAAAATCCAAAATATTTTATAAGTATCATCTT 660
```

```
Target_LR       TTGAGTTAATATAGCAGAATAATTTTTATAAAATCCAAAATATTTTATAAGTATCACCTT 659
                ************************* ****************************** ***

621             TTCAAATAATATATTATTTTATTTTAAAATACTACATAT-ACAAAATAACTCATAAAAAA 718
608             TTCAAATAATATATTATTTTATTTTAAAATACTACATAT-ACAAAATAACTCATAAAAAA 719
Target_LR       TTCAAATAATATATTATTTTATTTTAAAATACTACATATTACAAAATAACTCATAAAACA 719
                ************************************ ****************** *

621             CACCAAAATTAATCTCATTGAACTAAAAATTAAAAAAAACAAATAAAATTTGCGAAGGAA 778
608             CACCAAAATTAATCTCATTGAACTAAAAATTAAAAAAAACAAATAAAATTTGCGAAGGAA 779
Target_LR       CATCAAAATTAATCTCATTGAACTAAAAATTAAAAAAAACAAATAAAATTTGCGAAGGAA 779
                ** ************************************************************

621             ACATGATTTAAAGTTTAAACTGTAGAGACATGATCACAAAATCTAAAGAAAAA-TAAAAT 837
608             ACATGATTTAAAGTTTAAACTGTAGAGACATGATCACAAAATCTAAAGAAAAA-TAAAAT 838
Target_LR       ACATGATTTAAAGTTTAAACTGTAGAGACATGATCACAAAATCTAAAGAAAAAATAAAAT 839
                ************************************************** ******

621             TAAGAAGAGAGAAAAATGAAAGAAAAATATCACAGAAAAAGAAAGAAGAGCGAGTAGGAA 897
608             TAAGAAGAGAGAAAAATGAAAGAAAAATATCACAGAAAAAGAAAGAAGAGCGAGTAGGAA 898
Target_LR       TAAGAAGAGAGAAAAATGAAAGAAAAATATCACAGAAAAAGAAAGAAGAGCGAGTAGGAA 899
                ************************************************************

621             ACACGTGCGGCTGTGGGCCTCCGAAGCCCGTCGCATCAATGGCTGAGATTTAGTTTCAGT 957
608             ACACGTGCGGCTGTGGGCCTCCGAAGCCCGTCGCATCAATGGCTGAGATTTAGTTTCAGT 958
Target_LR       ACACGTGCGGCTGTGGGC-TCCGTAGCC-GTCGAATCAATGGCTGAGATTTAGTTTCAGT 957
                ****************** **** **** **** ************************

621             GGTTGTTTGGGATTCGAAATCCGACGCAGCTTCCCGTTTCCAATTTCAGTTTTACAAATA 1017
608             GGTTGTTTGGGATTCGAAATCCGACGCAGCTTCCCGTTTCCAATTTCAGTTTTACAAATA 1018
Target_LR       GGTTGTTTGGGATTCGAAATC-GACGCAGCTTCC-GTTTCCAATTTCAGTTTTACAA-TA 1014
                ******************** ************ ********************** **

621             CGAAGACTTCGCCCCCCGTTTCCCAT--CTCTCCTTCTTCTTCTTCCCCGTTTTGCTCTC 1075
608             CGAAGACTTCGCCCCCCGTTTCCCAT--CTCTCCTTCTTCTTCTTCCCCGTTTTGCTCTC 1076
Target_LR       CGAAGACTTT-CGCCCCGTTTCCCATTTCTCTCCTTCTTCTTCTTCCCCGTTTTGCTCTC 1073
                *********  * *************  ********************************

621             TCTTTCTGTAACTCTCTCTCTCTCTCT--------CTCTCACTCCACCAACTGAAATAAC 1127
608             TCTTTCTGTAACTCTCTCTCTCTCTCT--------CTCTCACTCCACCAACTGAAATAAC 1128
Target_LR       TCTTTCTGTAACTCTCTCTCTCTCTCTTTCTCTCTCTCTCACTCCACCAACTGAAATAAC 1133
                **************************        ************************

621             CCACCTCCACGCGGCCATTAACCGAAGCAAATCCAAACCATTTCCCTTGTTTCTCCCCTT 1187
608             CCACCTCCACGCGGCCATTAACCGAAGCAAATCCAAACCATTTCCCTTGTTTCTCCCCTT 1188
Target_LR       CCACCTCCACGCGGCCATTAACCGAAGCAAATCCAAACCATTTCCCTTGTTTCTCCCCTT 1193
                ************************************************************

621             TGAATGGCTTTCTTCCGGAGCTAAAAAACAAAACCCGGCCGCCACCGTTGCGCACCGCCG 1247
608             TGAATGGCTTTCTTCCGGAGCTAAAAAACAAAACCCGGCCGCCACCGTTGCGCACCGCCG 1248
Target_LR       TGAATGGCTTTCTTCCGGAGCTAAAAAACAAAACCCGGCCGCCACCGTTGCGCACCGCCG 1253
                ************************************************************

621             CCGTGTTGAATGGGAAGGTACATGGCCAAGACAAAGAGAACCGGCCACGTCGCCGGCATG 1307
608             CCGTGTTGAATGGGAAGGTACATGGCCAAGACAAAGAGAACCGGCCACGTCGCCGGCATG 1308
Target_LR       CCGTGTTGAATGGGAAGGTACATGGCCAAGACAAAGAGAAACG-CCA-GCTGCCGGCATG 1311
                ****************************************** ** *** *  *********

621             GACCTCTCCCTCGGCGGCGTCCGTACCAGAGCCAAAA-CCCTAGCTCTCCTCCGCCTTGA 1366
608             GACCTCTCCCTCGGCGGCGTCCGTACCAGAGCCAAAA-CCCTAGCTCTCCTCCGCCTTGA 1367
Target_LR       GACCTCTCCCTCGA-----TCCGTACCAGAGCCAAAAACCCTAGCTCTCCTCCGCCTTGA 1366
                *************      ****************** ********************

621             ACAGCAGCAAACCACCCCGAAATTGCATGCCTCTCCTCTCTCTTGTCCTCCGCCAGATGC 1426
608             ACAGCAGCAAACCACCCCGAAATTGCATGCCTCTCCTCTCTCTTGTCCTCCGCCAGATGC 1427
Target_LR       ACAGCAGCAAACCACCCCGAAATTGCATGCCTCTCCTCTCTCTTATCCTCCGCCAGATGC 1426
                ******************************************* ****************



621             TGCTGCTCCTGCCTCCGCCGGATATCTGCAGCTTCGGAGCCGCCGTCTTGAGAAGCCTCC 1486
608             TGCTGCTCCTGCCTCCGCCGGATATCTGCAGCTTCGGAGCCGCCGTCTTGAGAAGCCTCC 1487
Target_LR       TGCTGCTCCTGCCTCCGCCGGATATCTGCAGCTTCGGAGCCGCCGTCTTGAGAAGCCTCC 1486
                ************************************************************

621             CGTTGTGTTGTACGGCTCTAACCGGCGACAGTCCAGGCAGGCTGATGATGTTGCTGACCA 1546
608             CGTTGTGTTGTACGGCTCTAACCGGCGACAGTCCAGGCAGGCTGATGATGTTGCTGACCA 1547
Target_LR       CGTTGTGTTGTACGGCTCTAACCGGCGACAGTCCAGGCAGGCTGATGATGTTGCTGACCA 1546
                ************************************************************

621             AGTGGCTGGTGGTGGAAACCCTAACCCTAGCC- 1578
608             AGTGGCTGGTGGTGGAAACCCTAACCCTAGCC- 1579
Target_LR       AGTGGCTGGTGGTGGAAACCCTAACCCTAGCCC 1579
                ********************************
```

Chapter 7. Results: LIS-1 characterization of different *Linum* species

7.1. Target left and right primers for PCR with Linum DNAs

Amplification of all *Linum* speices was done using Target L and R primers with SpeedSTAR polymerase. From this gel we note that species 2 (L. autriacum), 6 (L. nervosum), 9 (L. perenne) 14 (L. grandiflorum reubrum), 17 (L. perenne), 30 (L. corymibiferum), 37 (L. marginale), 38 (L. maritimum), 4 (L. grandiflorum desf), and positive control S6 amplified (with speices 2 and 6 smearing).

Another gel (Figure with species 9, 30 and 38 was run on a small gel for excision of bands, the numbers labelled represent those extracted bands sent for sequencing. Note the unique pattern in species 9, with a band at 6kb (band 7), one at about 2.1kp (band 8), and the most intense being band 1, which is at about 3.4kb. Species 30 also has a unique pattern with 2 bands being most noticeable, band 2 at 2.2kb and band 3 at 920bp. Species 38 had just two bands to be seen, the most intense being 4 at about 1.4kb and band 5 at 920kb. No two *Linum* species have the same pattern with only band 3 and 5 being of the same size between species 30 and 38. These different patterns can be used to make inferences as to how closely related these species are.

Figure 7.1 A. PCR run with *Linum* species 1-41 amplifying primer Target L and R using SpeedSTAR polymerase

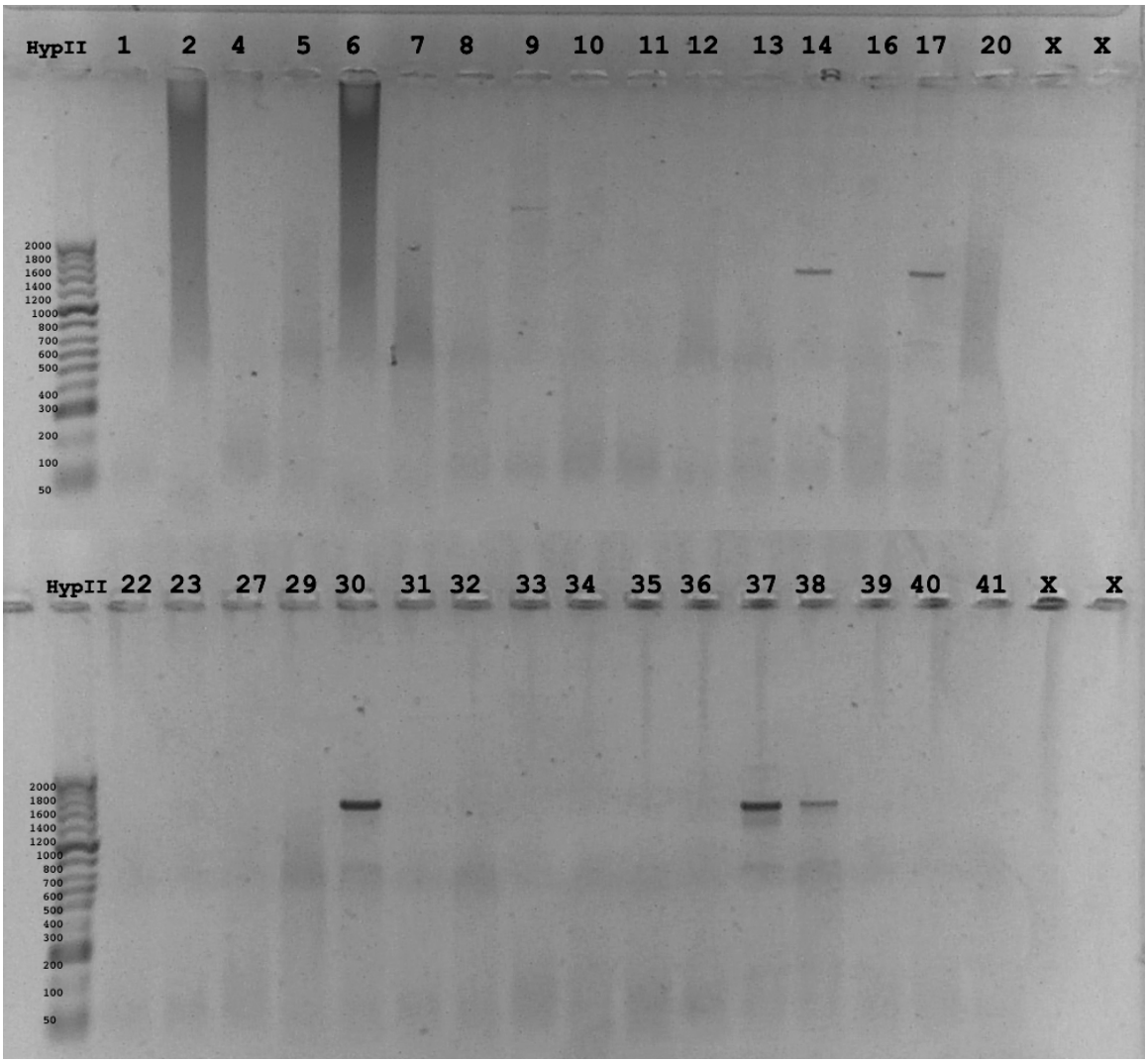


Figure 7.1 B. PCR run with *Linum* species 42-S6 amplifying primer Target L and R using SpeedSTAR polymerase

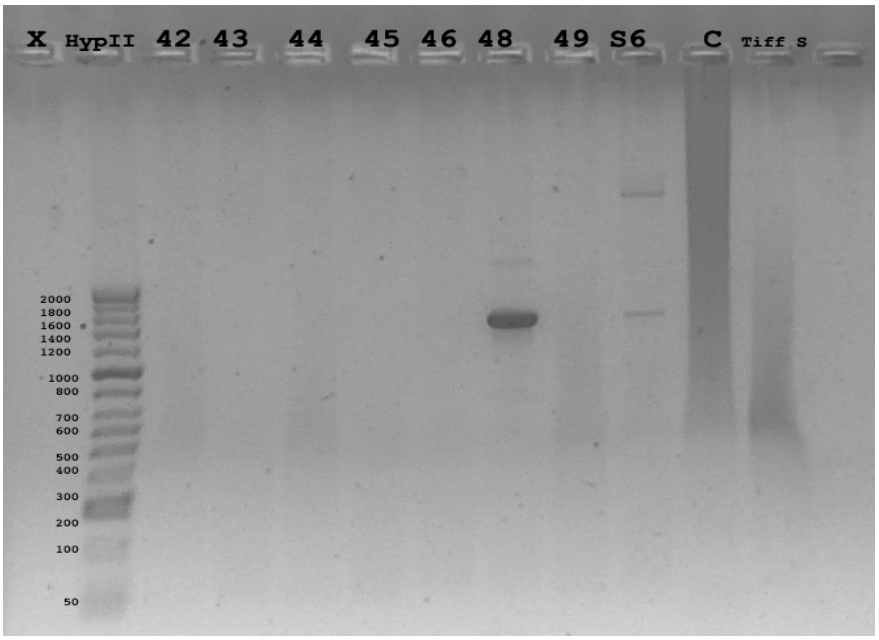
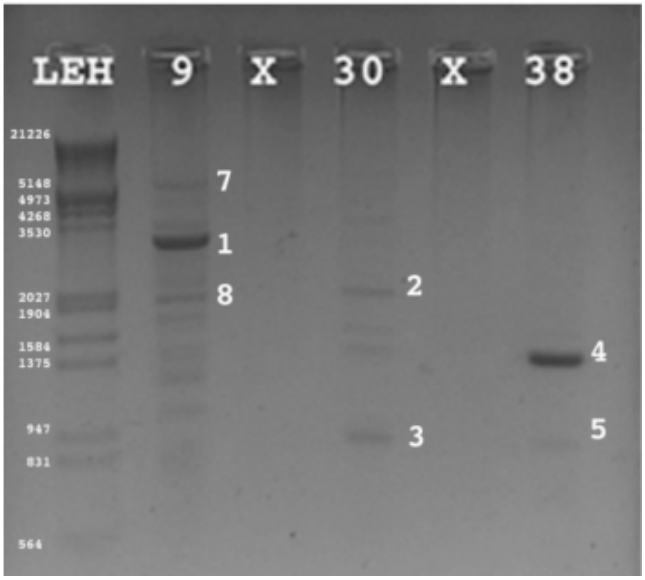


Figure 7.2. Gel of select species 9, 30, and 38 amplified with Target L and R. Note the different band patterns with each species. Those bands that are numbered were excised and sent for sequencing



PCR using all *Linum* species amplified with Target L and R using GoTaq green was done yielding a different set of amplified products (Figure 7.3). Species 9 (*L. perenne*), 14 (*L. grandiflorum reubrum*), 17 (*L. perenne*), 30 (*L. corymibiferum*), 37 (*L. marginale*), 38 (*L. maritimum*), and 48 (*L. grandiflorum desf*) all amplified. These species were run again on three smaller gels for excision of bands (Figure 7.4). Looking at those species that amplified with these primers and with each polymerase, this data can be used to make inferences about phylogenic relationships.

Figure 7.3. PCR of all *Linum* species amplified with Target L and R using GoTaq green. Note the different patters of expression seen compared to amplification with SpeedSTAR. Species 9, 14, 17, 30, 37, 38, 48 and S6 amplified.

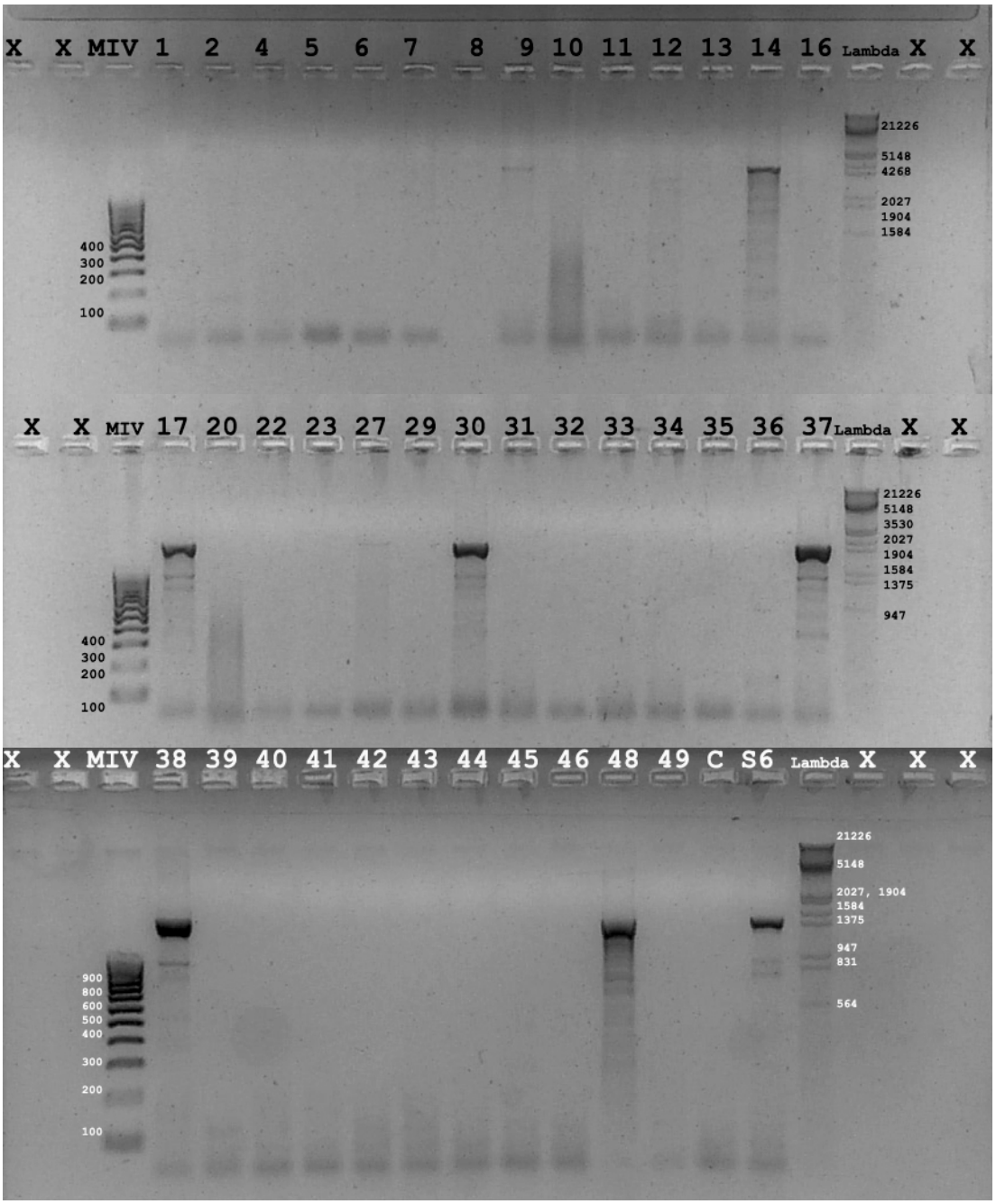
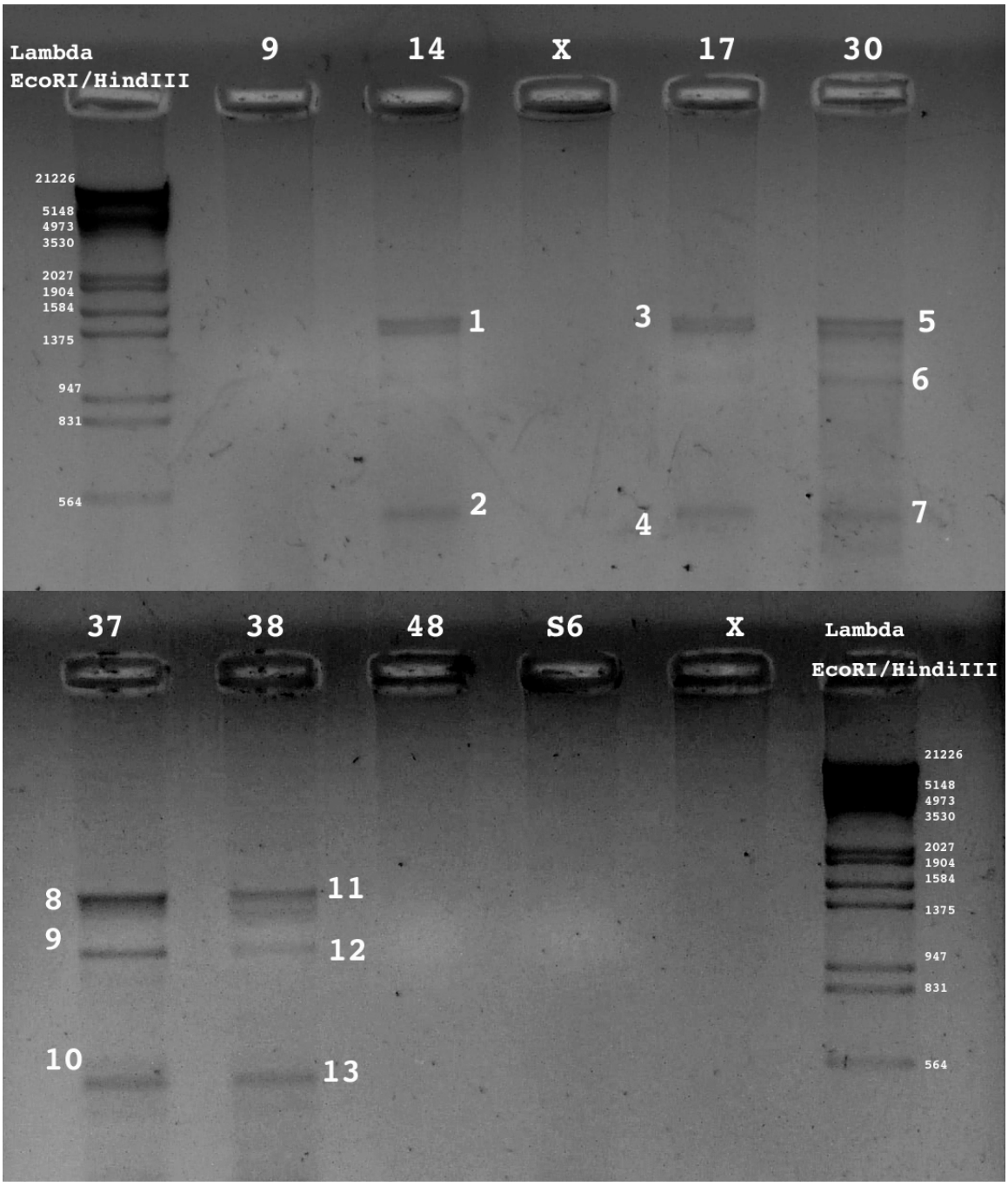


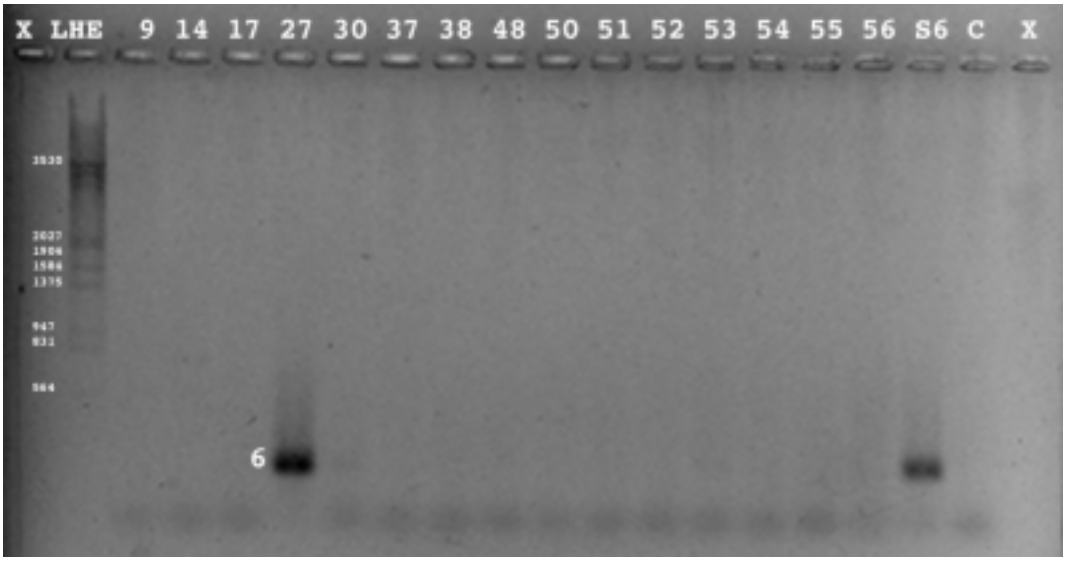
Figure 7.4. Gel of select species 9, 14, 17, 30, 37, 38, 48, and S6) amplified with Target L and R. using GoTaq green. Note the different band patterns with each species. Those bands that are numbered were excised and sent for sequencing



7.2. Target left and 3' right primers for PCR with Linum DNAs

Target L and 3' R were run with select speices 9, 14, 17, 30, 37, 38, 48, 59, 51, 52, 53, 54, 55, 56, and S6. Besides the positive control S6, FT37 or species 27 was the only other one to amplify (Figure 7.5).

Figure 7.5. PCR products of select *Linum* species (9, 14, 17, 30, 37, 38, 48, 59, 51, 52, 53, 54, 55, 56, and S6) amplified with Target L ad 3'R. Note species 27 is the only one to amplify.



7.3. Sequencing of amplified bands

From all the sequences sent for Target L and R with GoTaq green and SpeedSTAR and the one sent with Target L and 3'R, only one had a successful sequence and that was

Target L and 3'R. When that sequence (*Appendix C)* was Blasted using NCBI BLAST (http://blast.ncbi.nlm.nih.gov/Blast.cgi). The left sequence had a 93% max identity match to LIS-1 insertion sequence in genotrophs induced, and LIS-1 insertion sequence. (Figure 7.6 A). The right sequence showed 100% homology to the 5' LIS-1 flanking region as well as 98% homology to the LIS-1 insertion region. (Figure 7.6 B).

Figure 7.6 A. BLAST of sequence (left) obtained from amplification of Target L and 3'R band from species 27, a Flax variety.

```
>emb|AJ131994.1|  Linum usitatissimum insertion sequence in genotrophs induced
by the environment
Length=5792

 Score =  223 bits (246),  Expect = 4e-55
 Identities = 139/148 (94%), Gaps = 1/148 (1%)
 Strand=Plus/Plus

Query  54   GGATCTGGAGCCAGGAAATCAAATTCTGGATCTGATTCGCATTGGCACAACCCATCCGAA  113
            |||||||||||||||||||||||||||||||||||||||| | ||||||||||||||| |||
Sbjct  1    GGATCTGGAGCCAGGAAATCAAATTCTGGATCTGATTCGGAATGGCACAACCCATCTGAA  60

Query  114  TCCTCCCAAAAATGGTACCCTTGATCTAACTCCTGAAATTTTGGGGGGCATGTAAATAGTC  173
            || ||  |||||||||||||||||||||| |||||||||||||||||||||||||| |||||
Sbjct  61   TCGTCA-AAAAATGGTACCCTTGATCTAACGCCTGAAATTTTGGGGGGCATGTAAGTAGTC  119

Query  174  TCCTTCCTTCTTCATCTTCCATCTTCAA  201
            |||||||||||||||||||||| ||||
Sbjct  120  TCCTTCCTTCTTCATCTTCCATCCTCAA  147


>gb|AF104351.1|AF104351  Linum usitatissimum LIS-1 insertion sequence
Length=5792

 Score =  223 bits (246),  Expect = 4e-55
 Identities = 139/148 (94%), Gaps = 1/148 (1%)
 Strand=Plus/Plus

Query  54   GGATCTGGAGCCAGGAAATCAAATTCTGGATCTGATTCGCATTGGCACAACCCATCCGAA  113
            |||||||||||||||||||||||||||||||||||||||| | ||||||||||||||| |||
Sbjct  1    GGATCTGGAGCCAGGAAATCAAATTCTGGATCTGATTCGGAATGGCACAACCCATCTGAA  60

Query  114  TCCTCCCAAAAATGGTACCCTTGATCTAACTCCTGAAATTTTGGGGGGCATGTAAATAGTC  173
            || ||  |||||||||||||||||||||| |||||||||||||||||||||||||| |||||
Sbjct  61   TCGTCA-AAAAATGGTACCCTTGATCTAACGCCTGAAATTTTGGGGGGCATGTAAGTAGTC  119

Query  174  TCCTTCCTTCTTCATCTTCCATCTTCAA  201
            |||||||||||||||||||||| ||||
Sbjct  120  TCCTTCCTTCTTCATCTTCCATCCTCAA  147
```

Figure 7.6 B. BLAST of sequence (right) obtained from amplification of Target L and 3'R band from species 27, a Flax variety.

```
>emb|AJ131992.1|  Linum usitatissimum 5' flanking sequence of LIS-1 insertion in
genotroph LH
Length=613

 Score =  192 bits (212),  Expect = 7e-46
 Identities = 106/106 (100%), Gaps = 0/106 (0%)
 Strand=Plus/Minus

Query  94   GGACTCCTCCAAGACCTCCCTTTCAATCCCCAATTTCAGATTCGCCAGAGTTGCCTAAAT  153
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  613  GGACTCCTCCAAGACCTCCCTTTCAATCCCCAATTTCAGATTCGCCAGAGTTGCCTAAAT  554

Query  154  CGGTGGCGGGAAGGGTCATCACATTGAGCAGAACTGAAGAAGGGGG  199
            |||||||||||||||||||||||||||||||||||||||||||||
Sbjct  553  CGGTGGCGGGAAGGGTCATCACATTGAGCAGAACTGAAGAAGGGGG  508


>gb|AF104351.1|AF104351  Linum usitatissimum LIS-1 insertion sequence
Length=5792

 Score =  159 bits (176),  Expect = 4e-36
 Identities = 92/93 (99%), Gaps = 1/93 (1%)
 Strand=Plus/Minus

Query  1    GGCGTTAGATCAAGGGTACCATTTTTTGGACGATTCAGATGGGTTGTGCCATTCCGAATC  60
            |||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Sbjct  92   GGCGTTAGATCAAGGGTACCATTTTTTG-ACGATTCAGATGGGTTGTGCCATTCCGAATC  34

Query  61   AGATCCAGAATTTGATTTCCTGGCTCCAGATCC  93
            |||||||||||||||||||||||||||||||||
Sbjct  33   AGATCCAGAATTTGATTTCCTGGCTCCAGATCC  1
```

Chapter 8. Results: LIS-1 miRNA primers 1-9 screening

8.1. LIS-1 primers 2-9 run with species 1-23

Linum species 1-23 were run with miRNA primers 2-9. They are shown below in Figure 8.1 A-H. Primer 2 amplified species 4 (*L. grandiflorum caeruleum*), 13 (*L. bienne*), 14 (*L. bienne*), 17 (*L. perenne*), 21 (*L. catharticum*), 18 (*L. punctatum*), and 21 (*L. catharticum*). Primer 3. With primers 3, 6, 8 and 9 only the positive controls amplified. Primer 4 amplified species 16 (*L. maritimum*) and primer 5 amplified species 21 (*L. catharticum*). Amplifications occurring with Primer 7 displayed the most amplification between the different species. Multiple bands and different sizes and intensities with each

species were seen (Figure 8.1 G). From these amplifications with these primers we can make inferences as to the homology of these sequence sin these species. Using it as a method of creating phylogenic relationships.

Figure 8.1 A. Species 1-23 amplified with miRNA primer 2. Note species 4, 13, 14, 17, 18, and 21 along with positive controls C3 and 27 (FT37) all amplified.

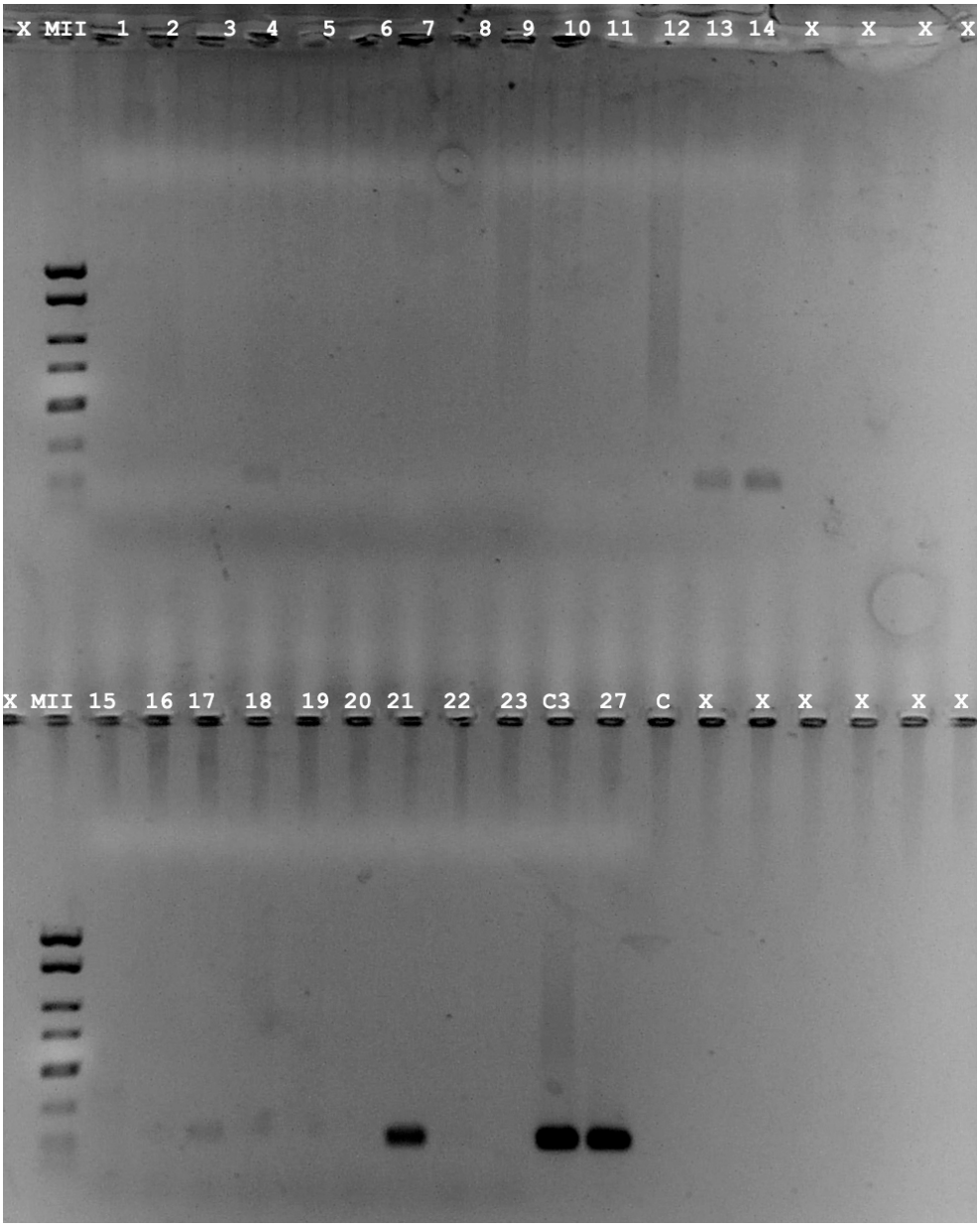


Figure 8.1 B. *Linum* DNA species 1-23 amplified with miRNA primer 3. Only the positive controls C3 (flax genotroph) and FT37 (species 27) amplified with this primer.

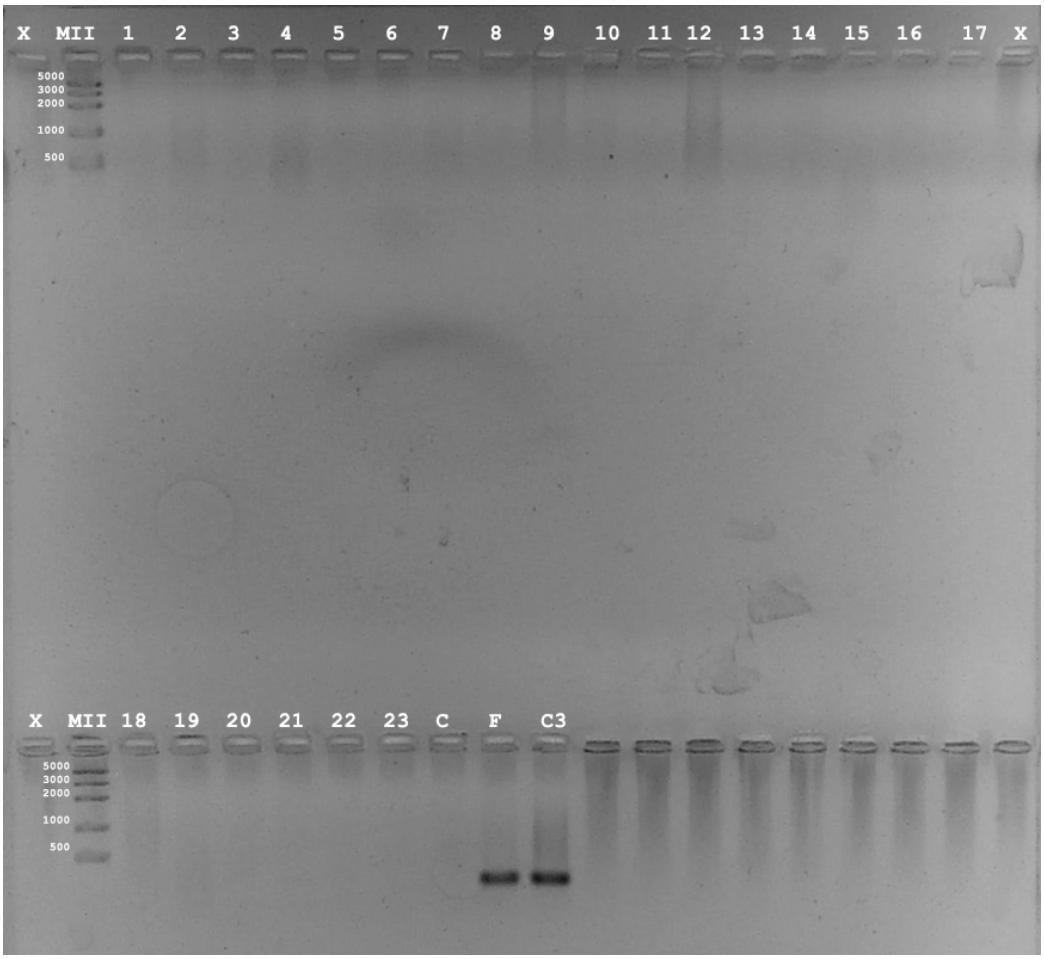


Figure 8.1 C. Species 1-23 amplified with miRNA primer 4. Note species 16 (L. maritimum) along with positive controls C3 and 27 (FT37) all amplified.

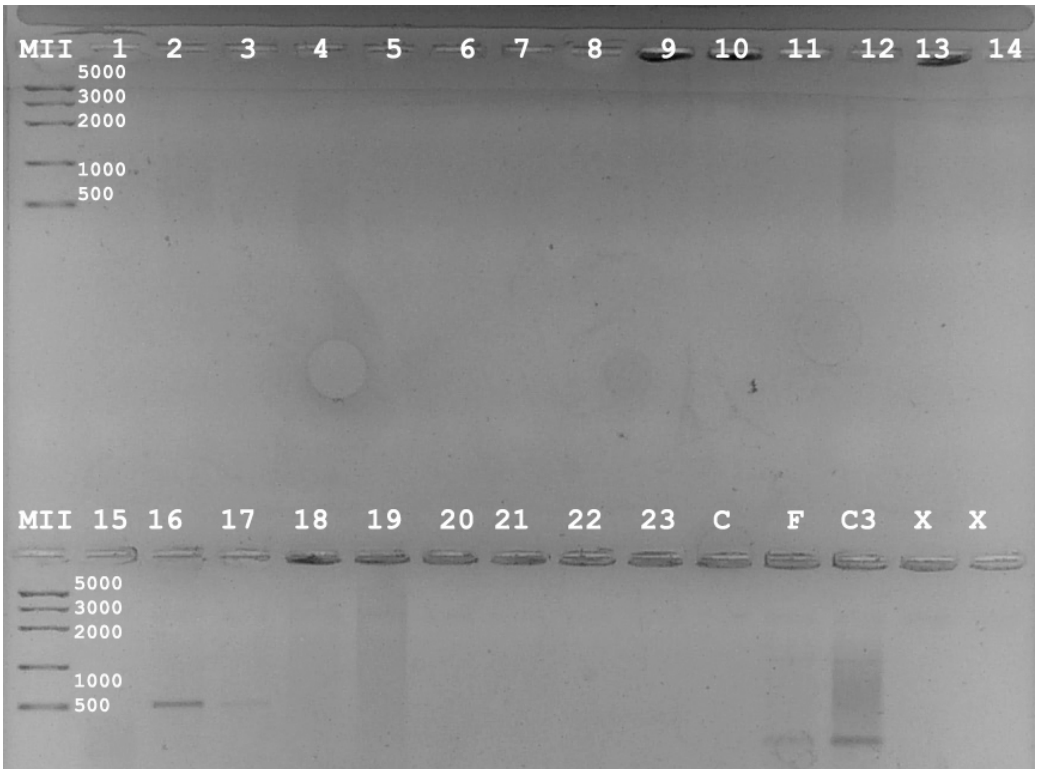


Figure 8.1 D. *Linum* DNA species 1-23 amplified with miRNA primer 5. Only species 21 (L. catharticum) positive FT37 (species 27) amplified with this primer.

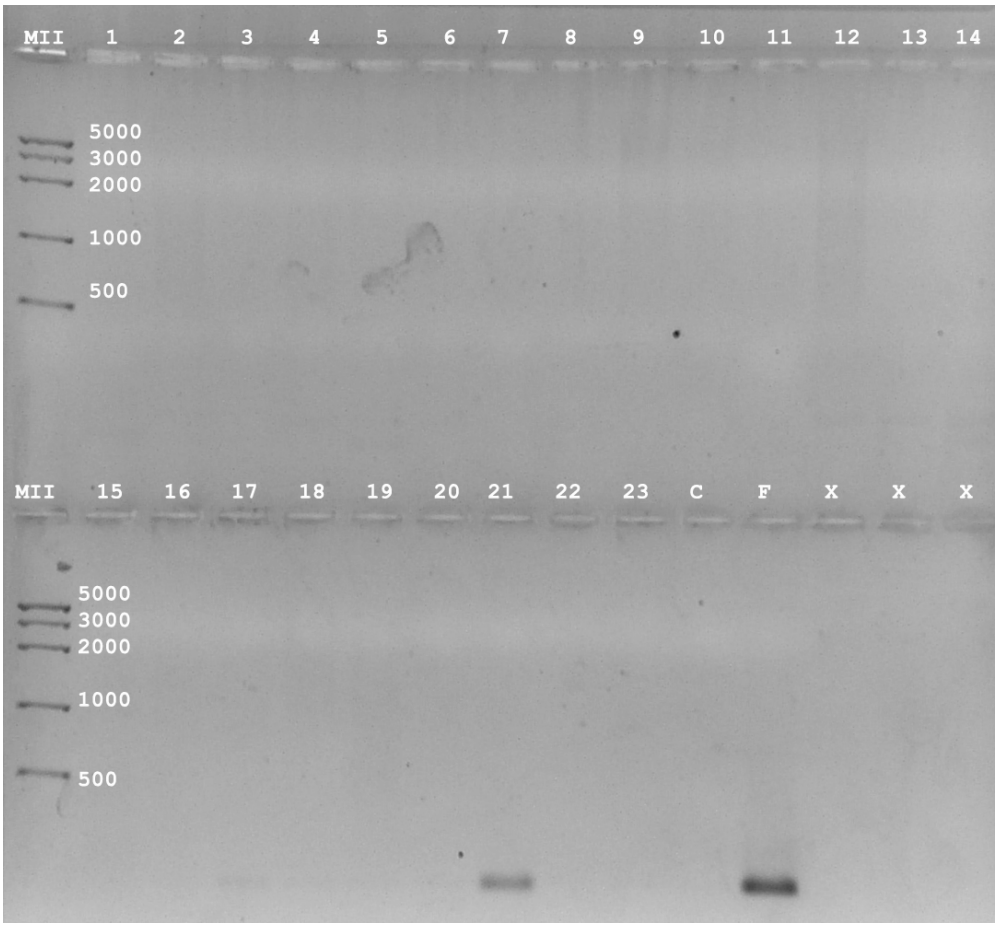


Figure 8.1 E. *Linum* DNA species 1-23 amplified with miRNA primer 6. Only positive control FT37 (species 27) amplified with this primer.

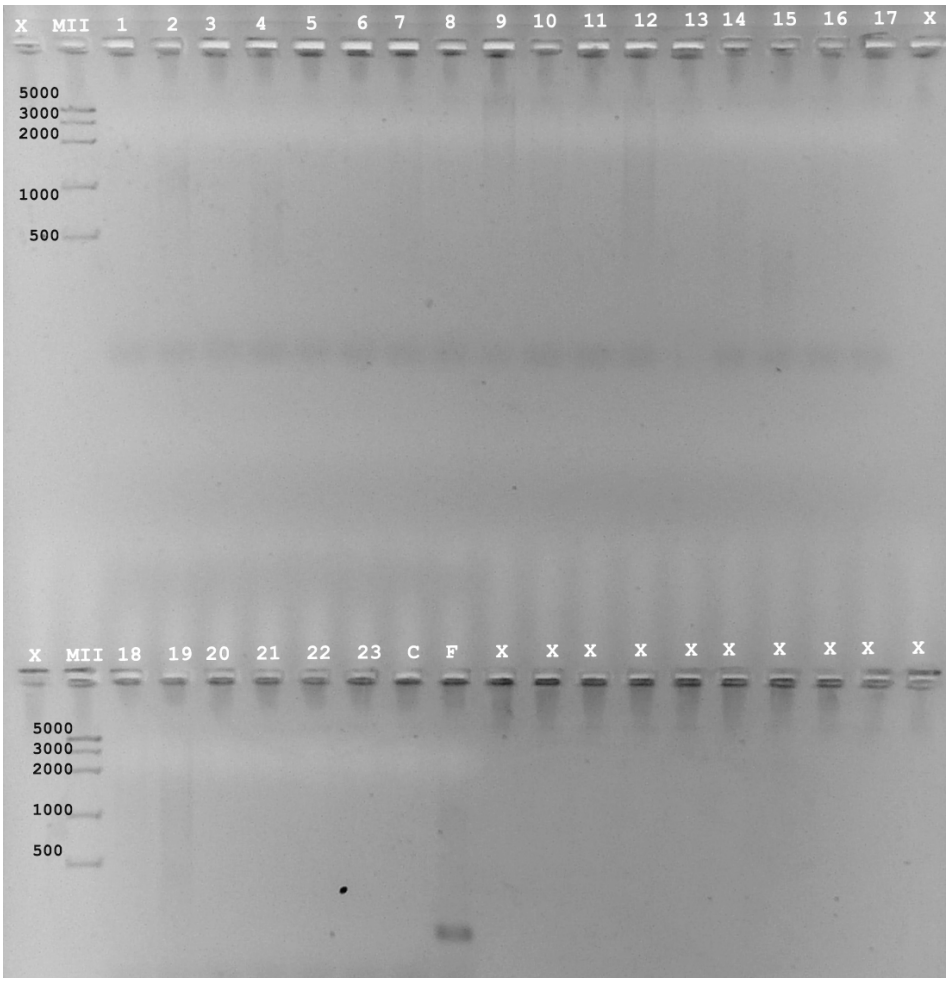


Figure 8.1 F. Species 1-23 amplified with miRNA primer 7. All the species along with positive control 27 (FT37) all amplified. Note the different pattern seen for each *Linum* species as well as different intensities.

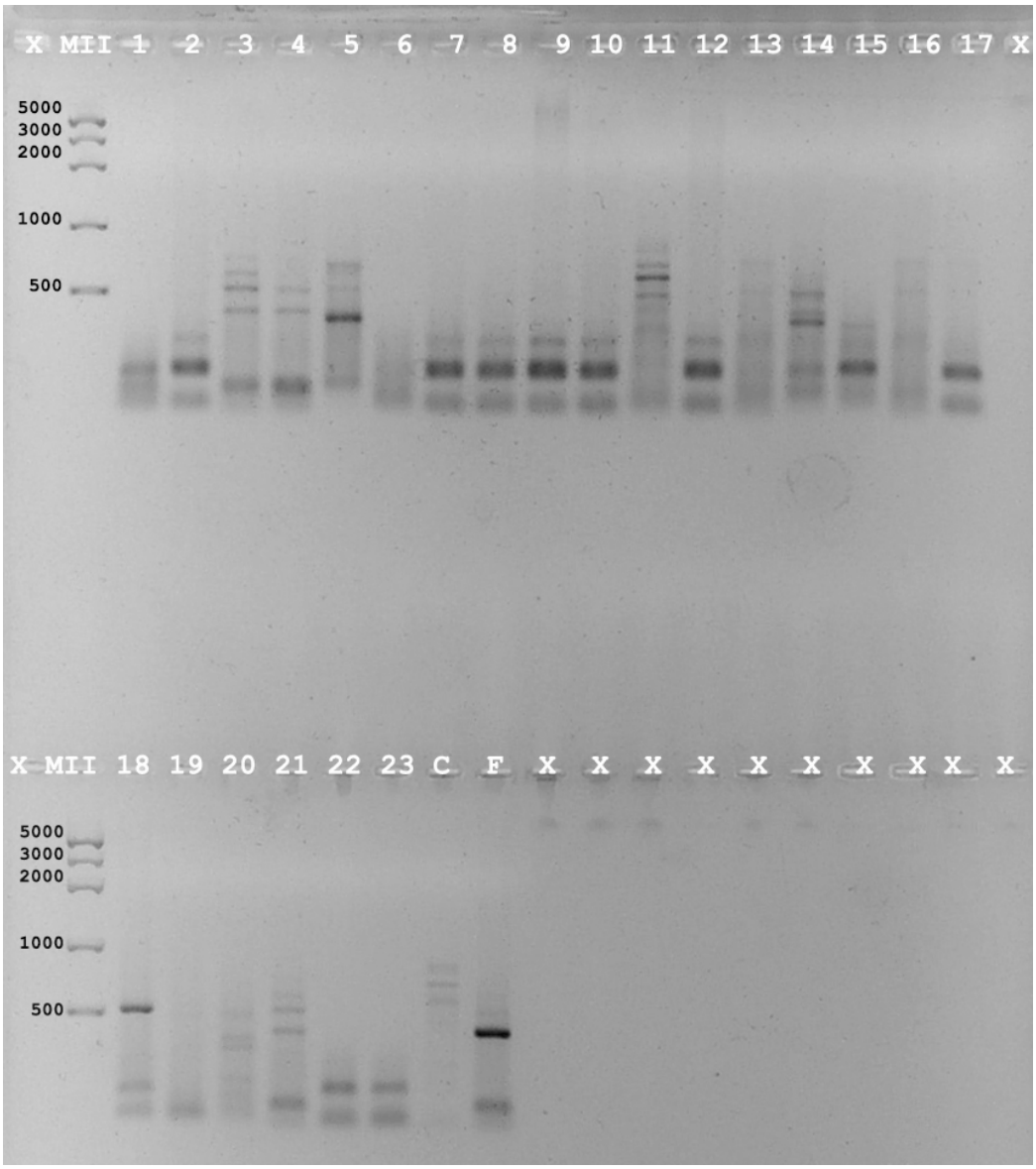


Figure 8.1 G. *Linum* species 1-23 amplified with miRNA primer 8. Only the positive control 27 (FT37) amplified.

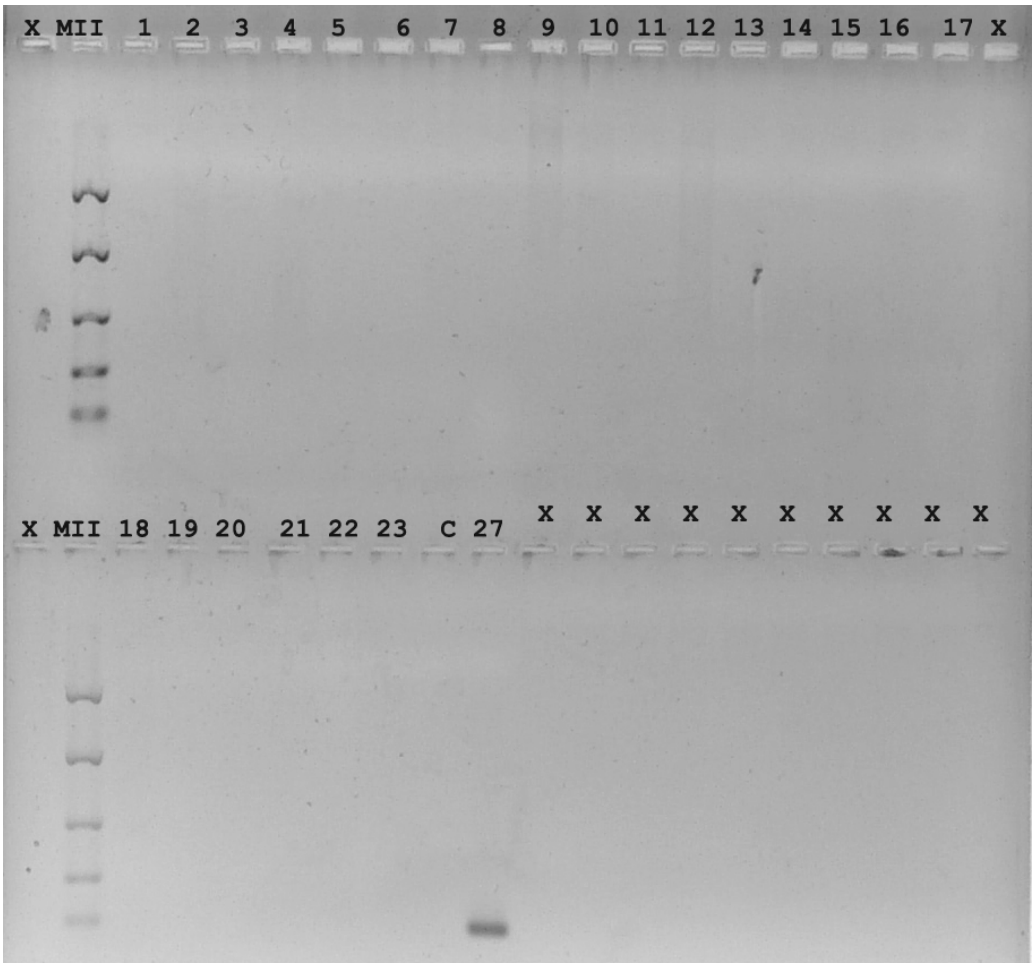
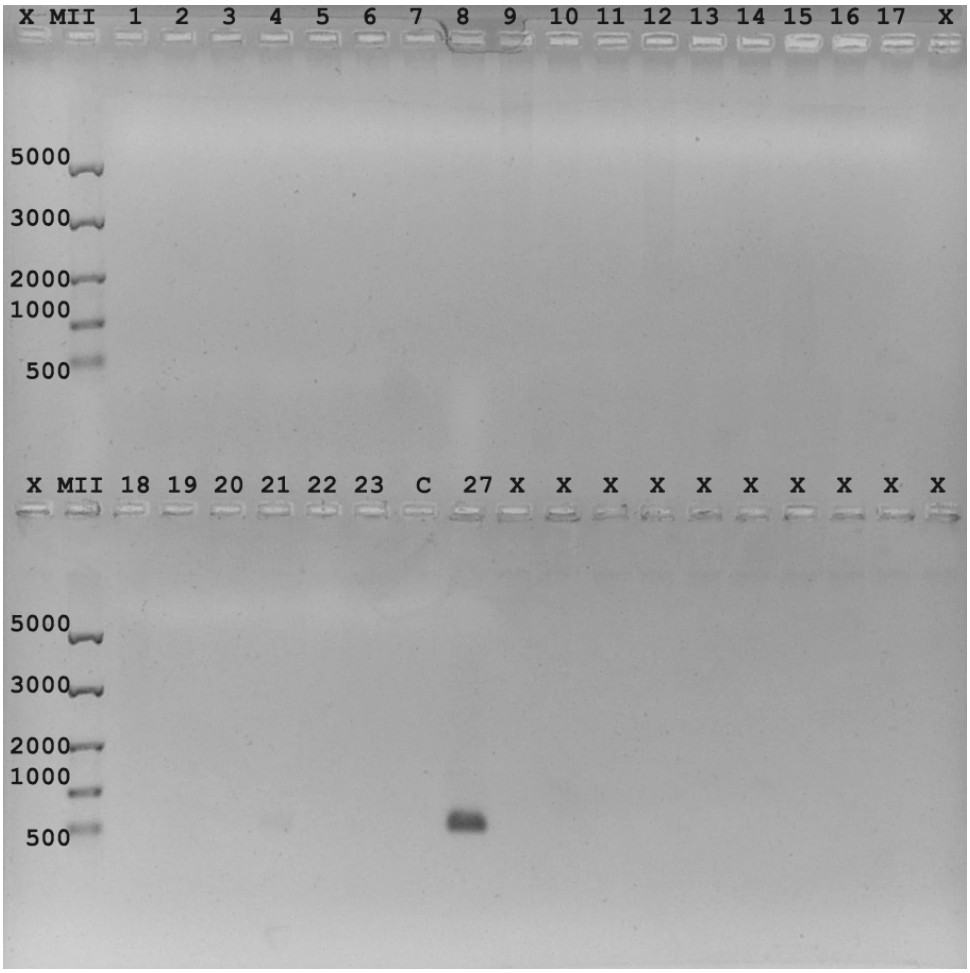


Figure 8.1 H. *Linum* species 1-23 amplified with miRNA primer 9. Only the positive control 27 (FT37) amplified.



8.2. LIS-1 primers 1-9 run with species 29-47

Linum species 29-47 were run with miRNA primers 1-9. They are shown below in Figure 8.2 A-I. Primer 1 did not amplify with any DNA but primer two did amplify DNA from species 38 (*L. maritimum*) and species 43 (*L. strictum*) amplified. The only other primer to amplify with these *Linum* species was primer 7. These results speak to the homology of these potential miRNA sites in other *Linum* species.

Figure 8.2 A and B. *Linum* DNA amplified with miRNA primers 1 and 2. Note nothing amplified with primer 1. But with primer 2 species 38 (L. maritimum) and species 43 (L. strictum) amplified. The positive controls for this gel were run on Figure 8.2 C.

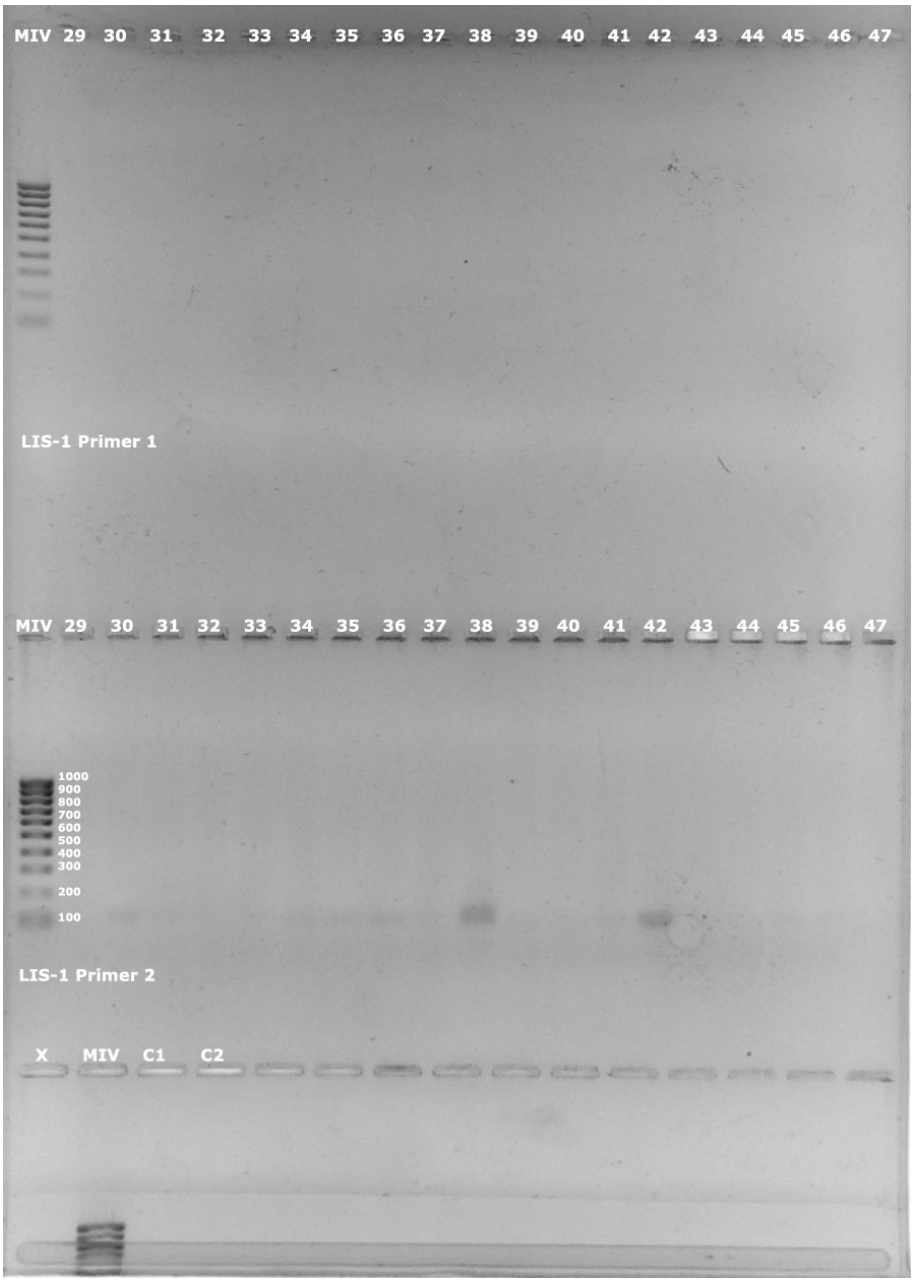


Figure 8.2 C. *Linum* DNAs 29-47 amplified with miRNA primer 3. The only DNAs to amplify were the three positive controls (S63, S61, and S62). With S6 1 and 2 being PCRs run with primer 1 and 2 but where not put with the other gel (Figure 8.1 A and B) because of error, so they were run here.

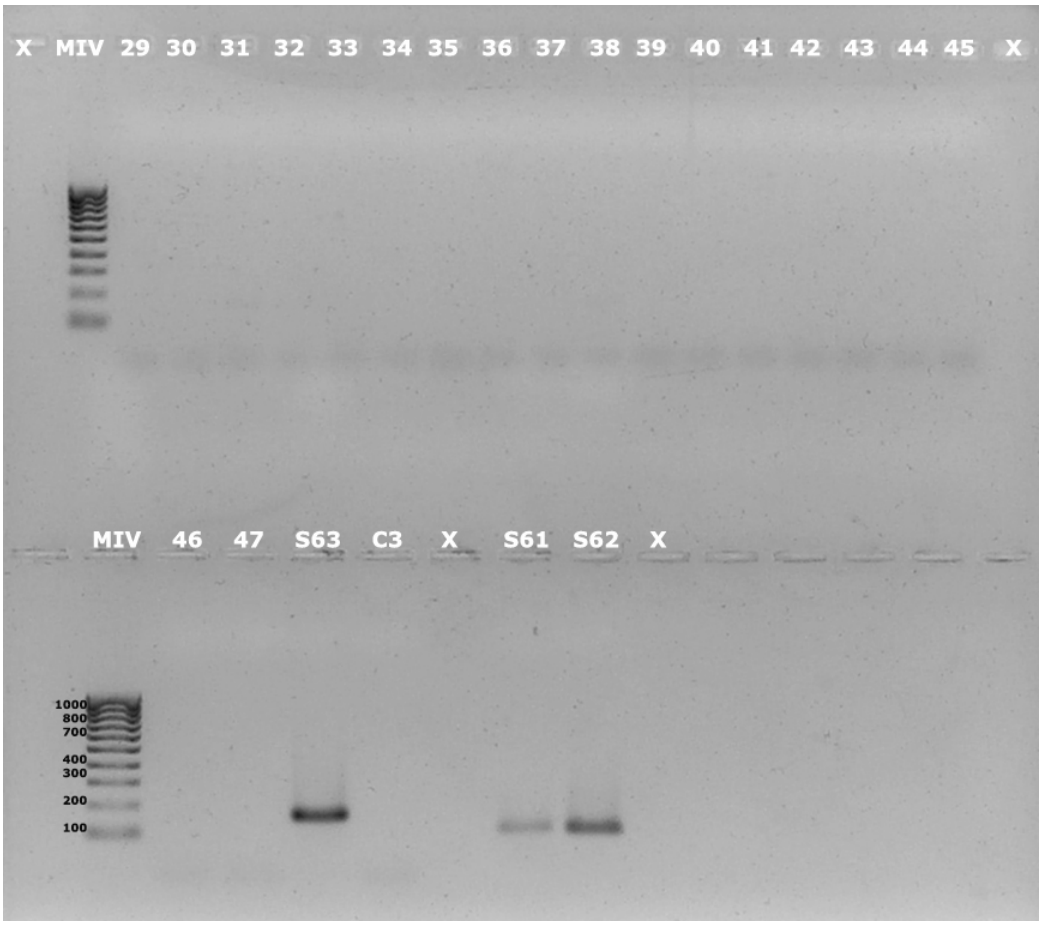


Figure 8.2 D. *Linum* DNAs 29-47 amplified with miRNA primer 4. The only DNA to amplify was the positive control S6.

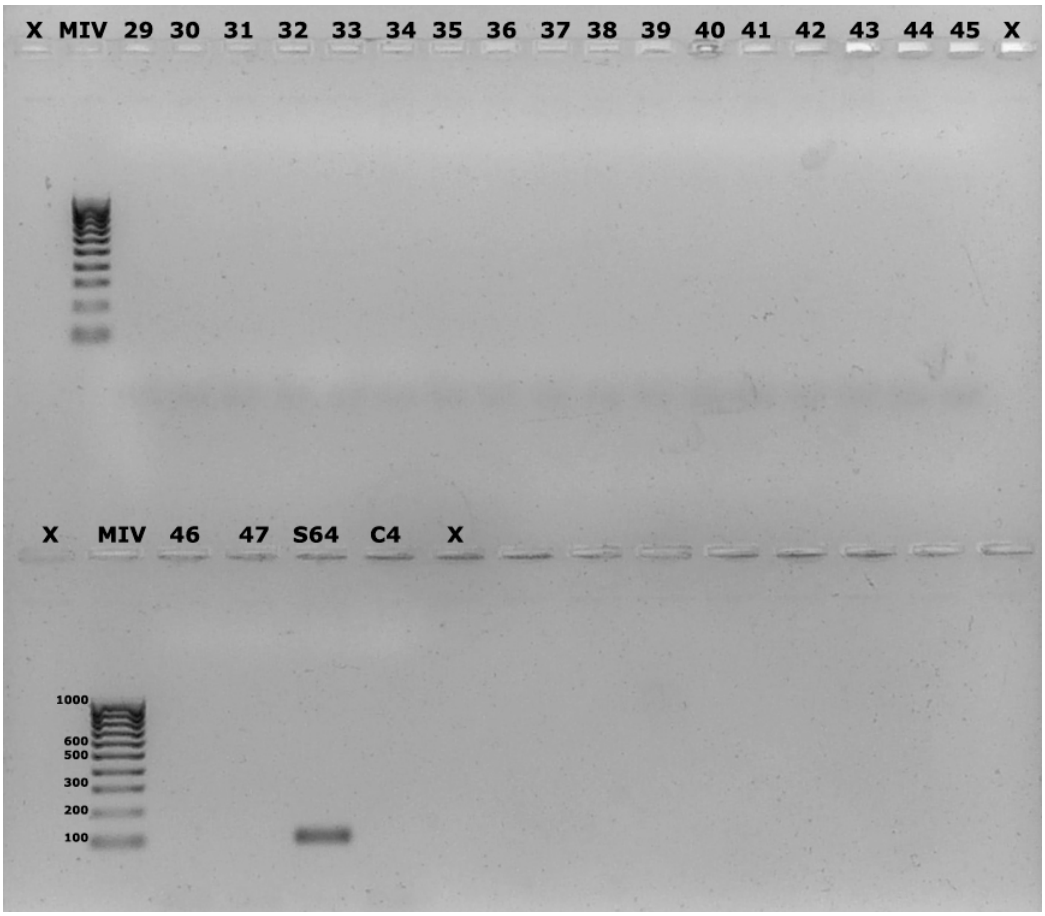


Figure 8.2 E. *Linum* species 29-47 amplified with miRNA primer 5. The positive control was the only one to amplify in this gel.

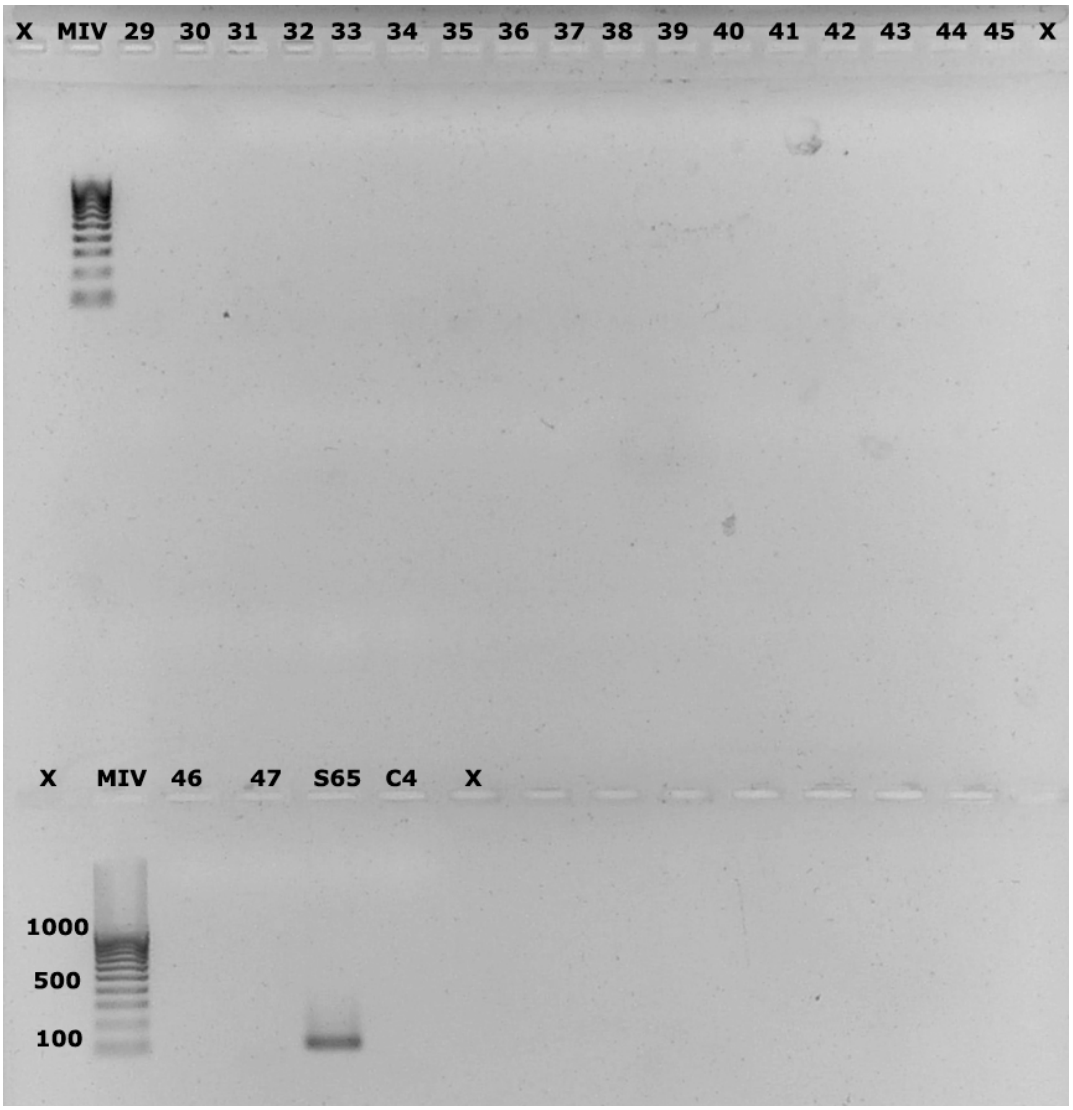


Figure 8.2 F. *Linum* species 29-47 amplified with miRNA primer 6. No DNAs amplified but the positive control.

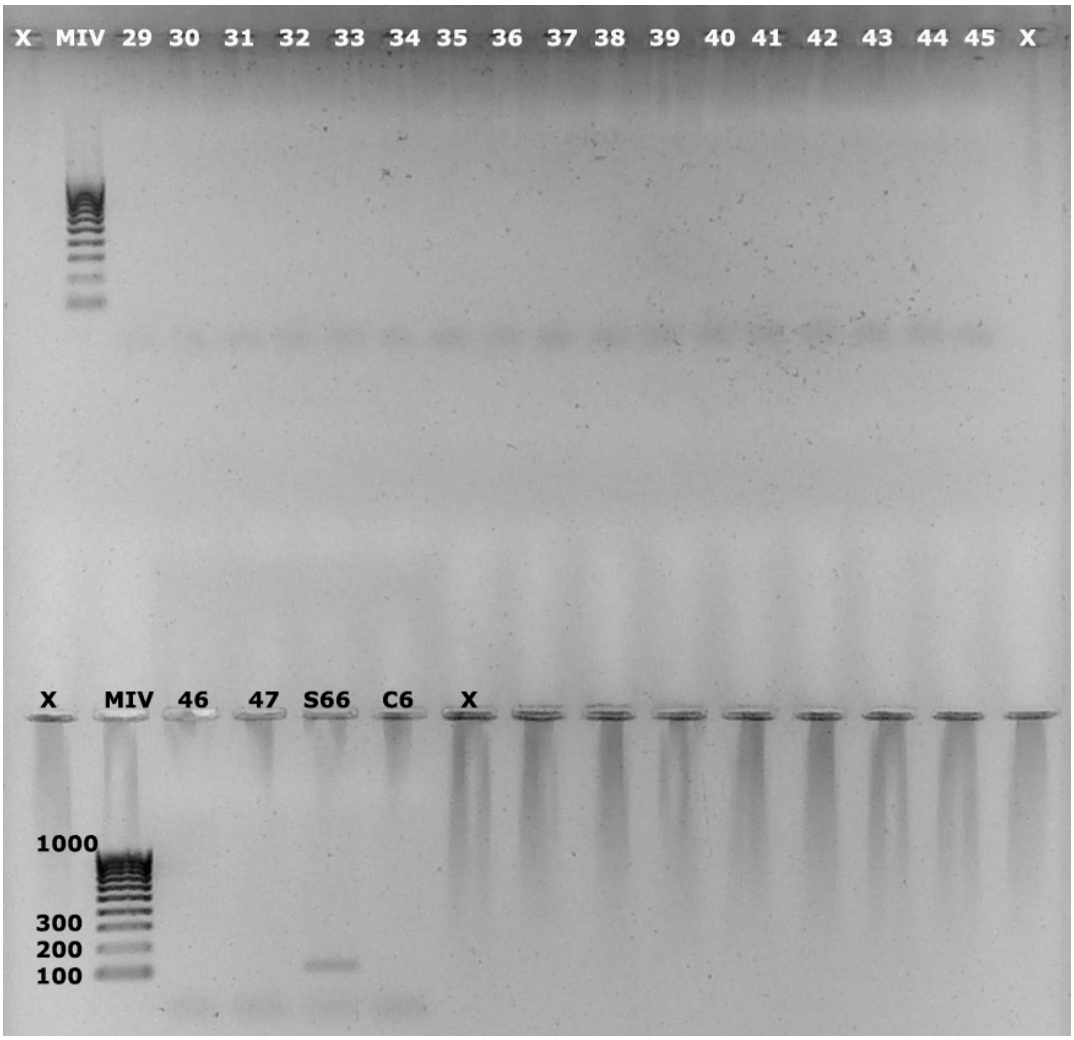


Figure 8.2 G. *Linum* species 29-47 run with miRNA primer 7. Here all the species amplified with this primer and with multiple bands of different size and intensity. A pattern very similar to what was seen in Figure 8.1F.

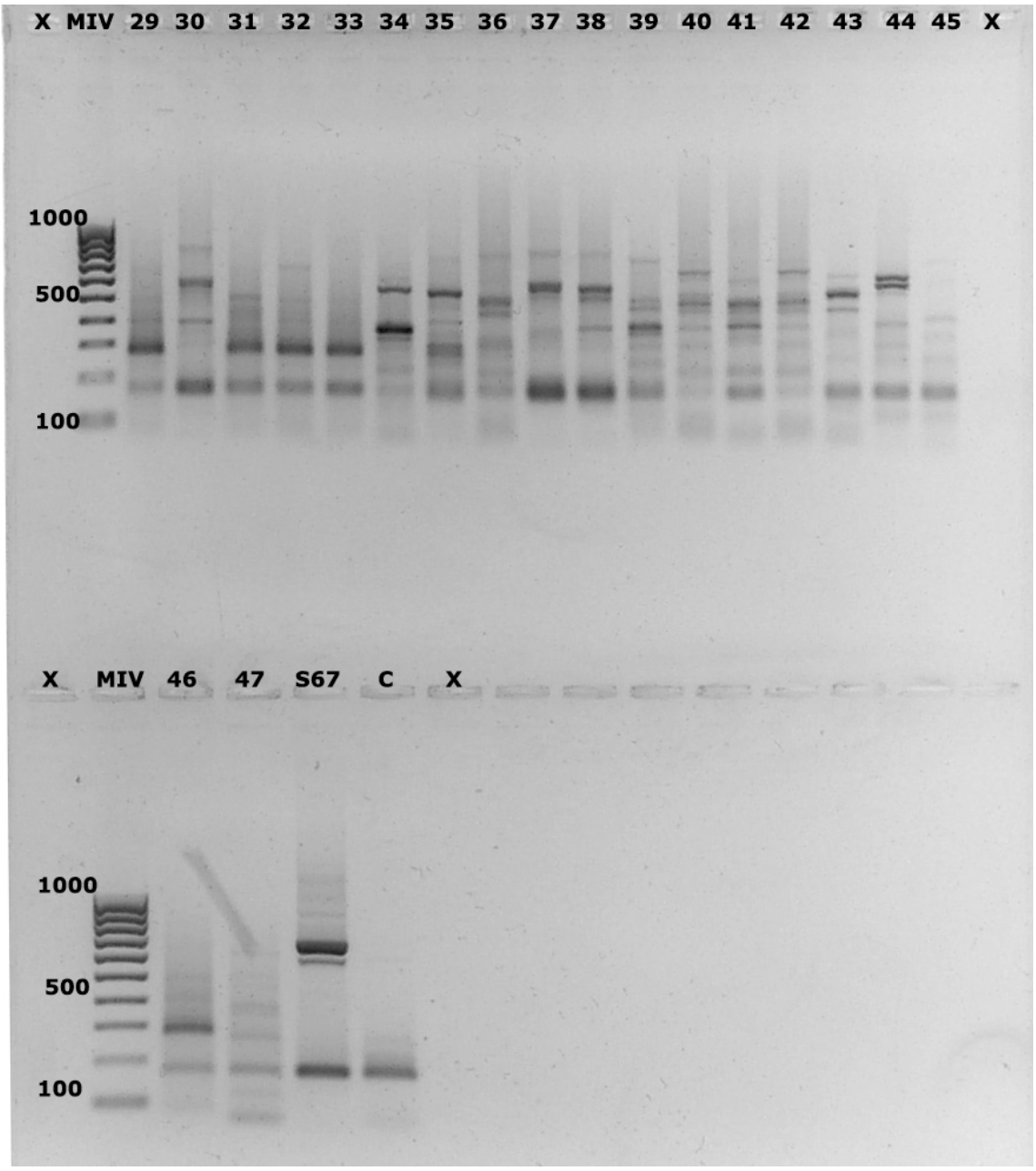


Figure 8.2 H. *Linum* DNAs 29-47 amplified with miRNA Primer 8. Note positive control S6 was the only DNA to amplify

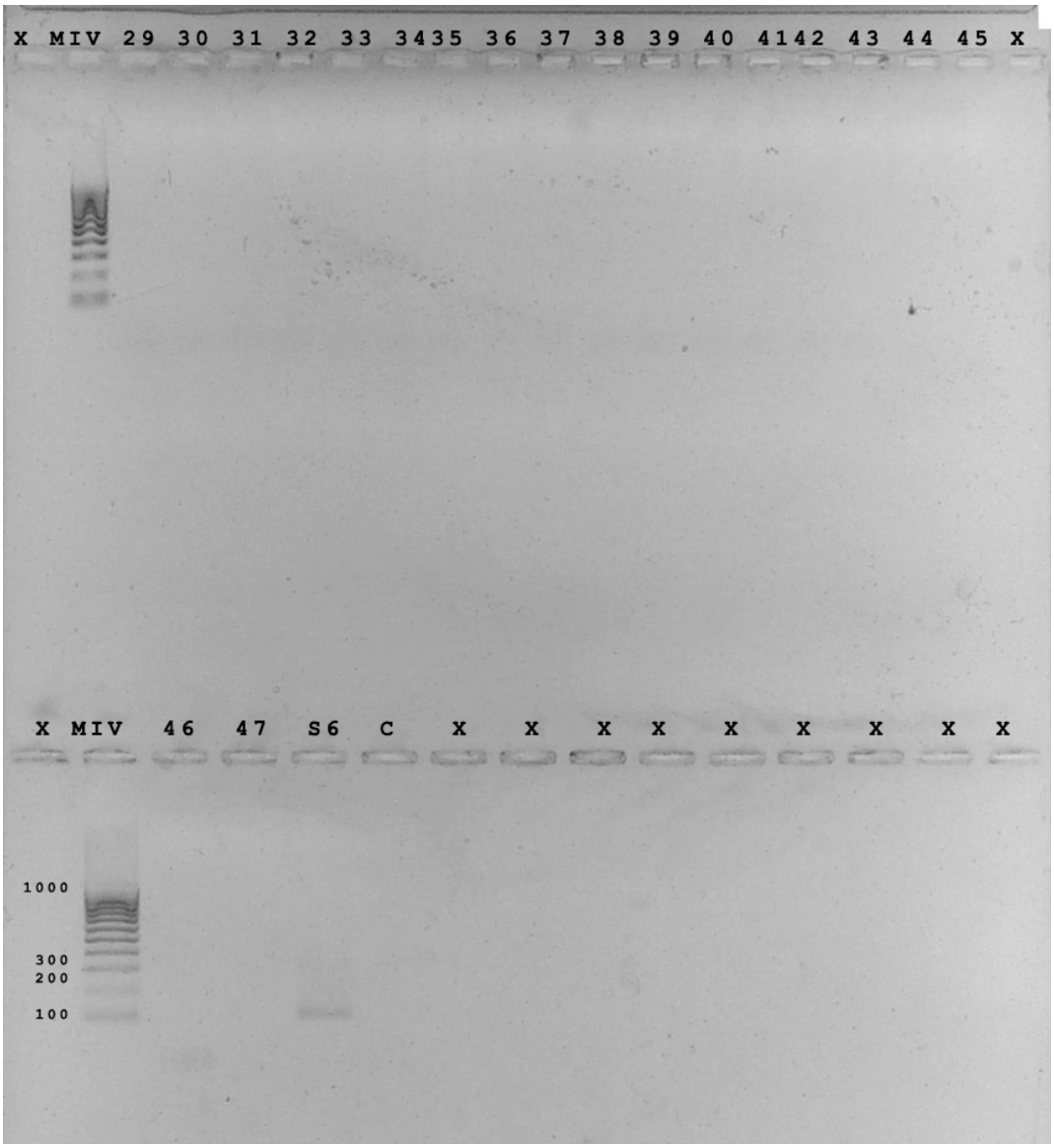
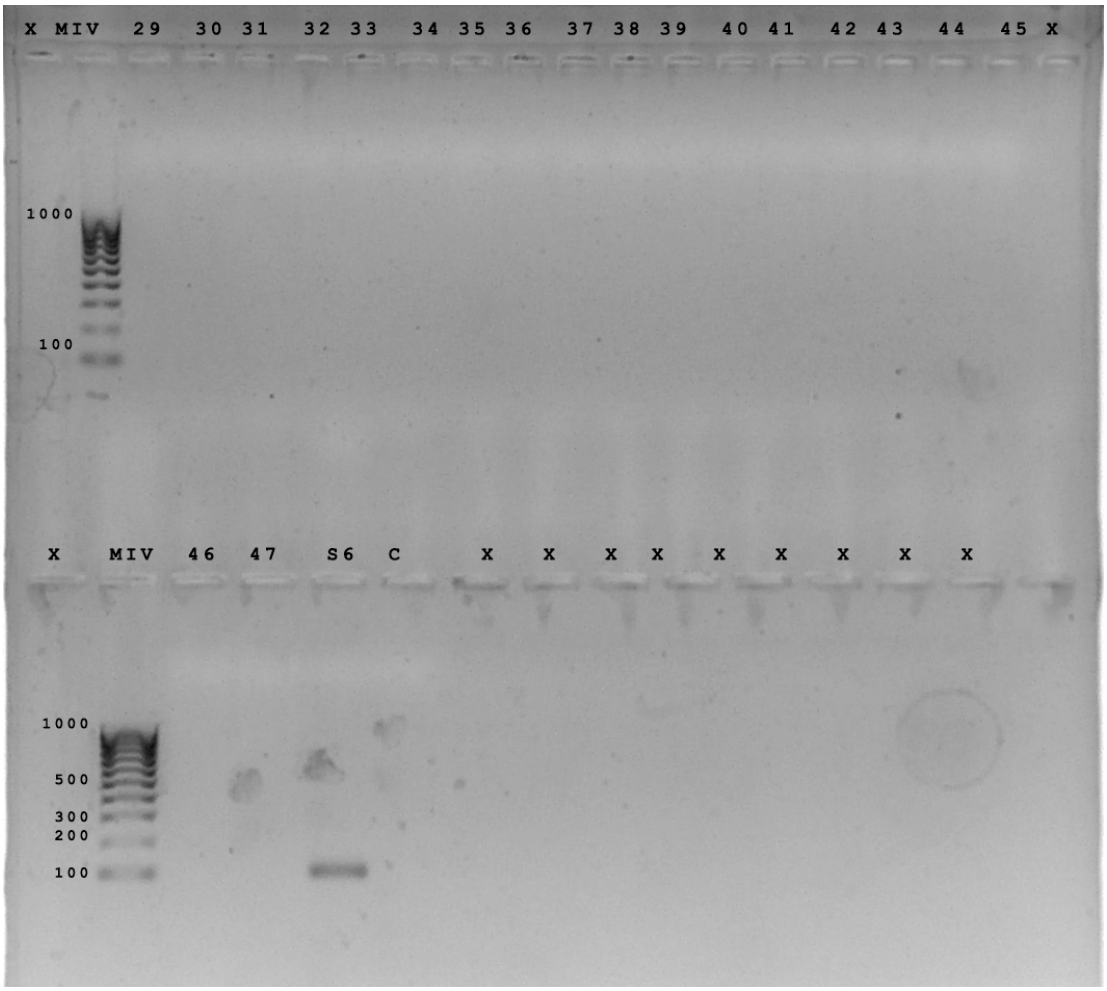


Figure 8.2 I. *Linum* DNAs 29-47 amplified with primer 9, note only the positive control S6 amplified.



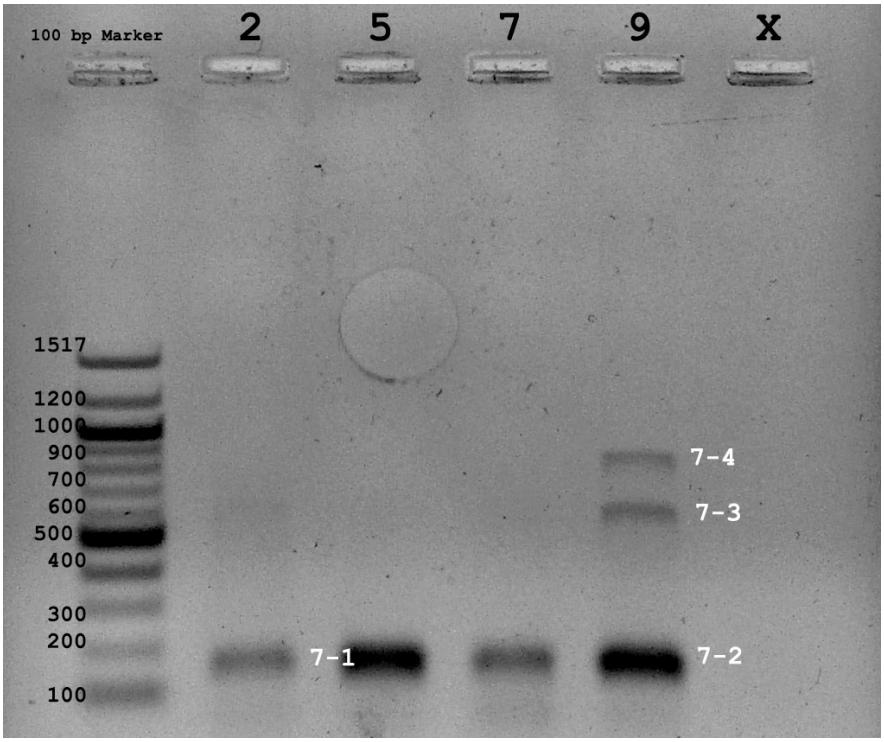
Chapter 9. Results: miRNA primer 7 analysis

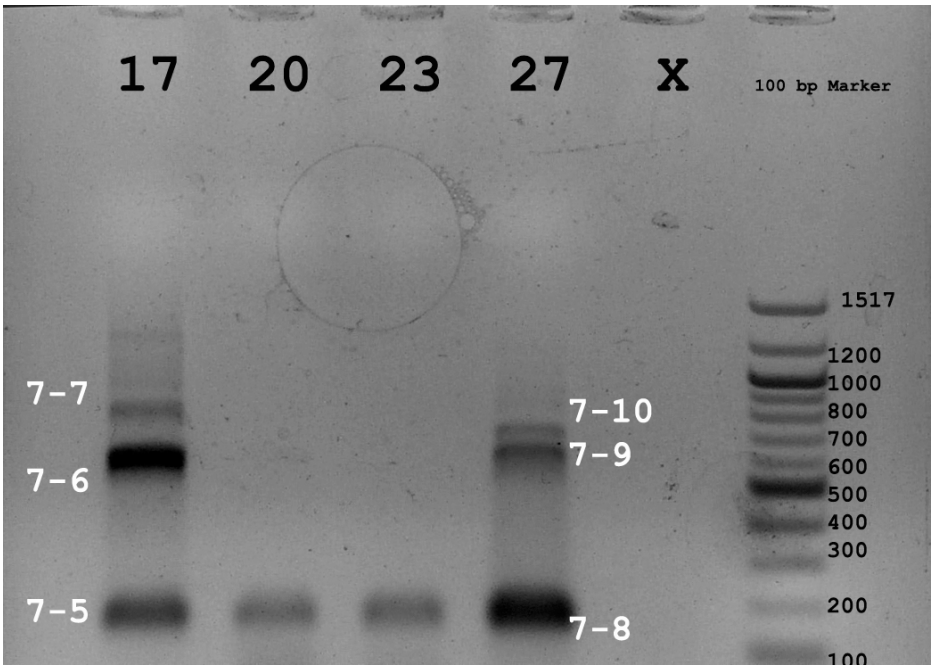
9.1. PCR amplification with select species

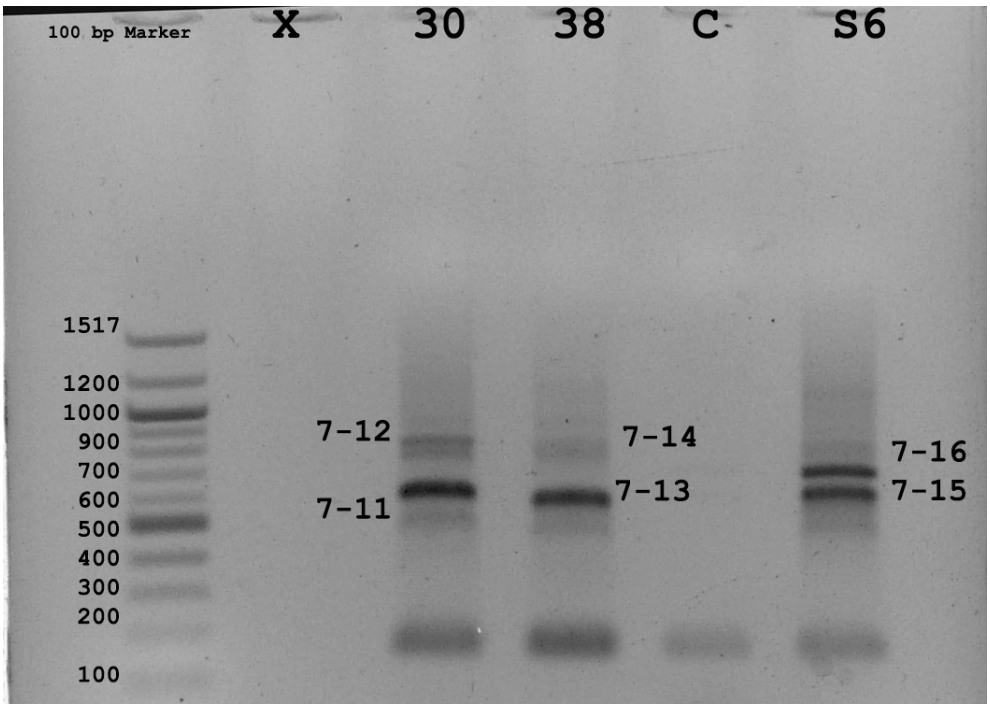
Select Linum species were amplified with miRNA primer 7, in order to do a band cut out for sequencing (Figure 9.1 A-C). The following species were used: 2 (*L. austriacum*), 5 (*L. narbonense*), 7 (*L. perenne*), 9 (*L. perenne*), 17 (*L. perenne*), 20 (*L. flavum*), 23 (*L. hologynum*), 27 (*L. usitatissimum*), 30 (*L. corymibiferum*), 38 (*L. maritimum*), and S6 (*L.*

usitatissimum). After a re-amplification of cut out bands it became apparent that multiple sequences exist within each band and trying a cloning reaction would be better.

Figure 9.1 A-C. *Linum* species 2, 5, 7, 9, 17, 29, 23, 27, 30, 38, and S6 were amplified with miRNA primer 7. Not how both flax varieties (27 and S6 share the same band pattern, as well as both L. perenne species, although 17 has a more intense second band compared to the other perenne 9.




17
20
23
27
X
100 bp Marker
1517
1200
1000
800
700
600
500
400
300
200
100
7-7
7-6
7-5
7-10
7-9
7-8


100 bp Marker
X
30
38
C
S6
1517
1200
1000
900
700
600
500
400
300
200
100
7-12
7-11
7-14
7-13
7-16
7-15

9.2. TOPO TA cloning

To ensure proper cloning, a fresh PCR product was done (Figure 9.2) using species 9, 17, 27, 30, 38, and S6. Once the transformation was done, a plasmid prep was performed on the overnight-incubated broth. From this plasmid prep, a gel was run to ensure there was plasmid present (Figure 9.3). The overnight broth was grown with 2.5 µl and 50µl of ampicillin (but the solution with 2.5µl of ampicillin was used for all colony PCRs. Using 1µl of the plasmid prep solution, a PCR was performed using miRNA primer 7 and the M13 primers included in the TOPO protocol (Figure 9.4). After the success of this gel, colony PCR was done (Figure 9.5 A and B). Note that some of the colonies did amplify two bands, this can be explained by the fact that the colonies were close to each other on the plate, and more than likely an error occurred while picking the colonies and more than one was picked.

Figure 9.5 A. Colony PCR of a 10^3 dilution. 40 colonies were picked. Note the following colonies amplified with the M13 primers: 2, 3, 4, 5, 6, 9, 10, 11, 12, 13, 15, 21, 25, 27, 28, 29, 30, 31, 32, 33, 39, 40. The last two colonies (39 and 40) were blue colonies picked as a control.

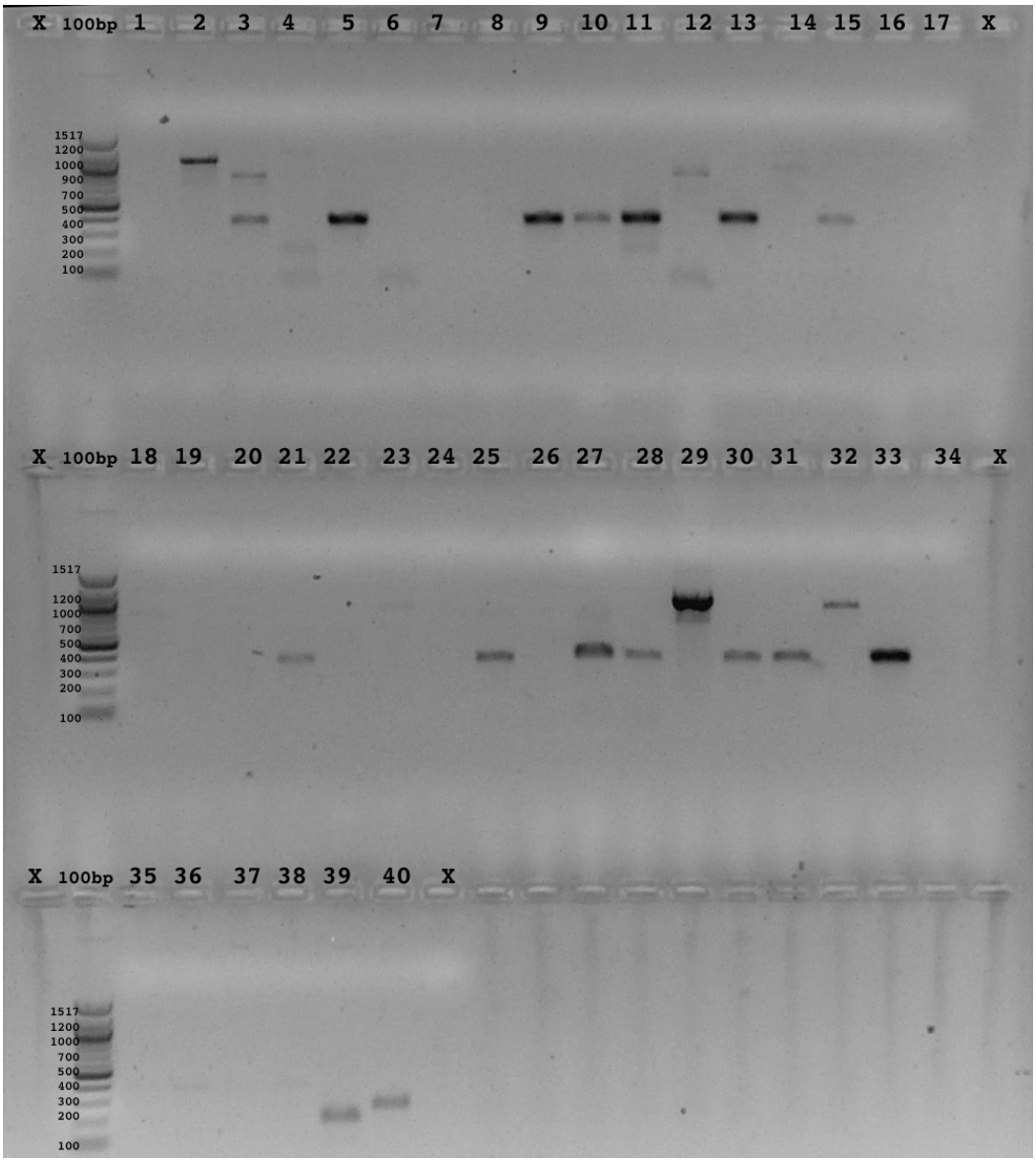
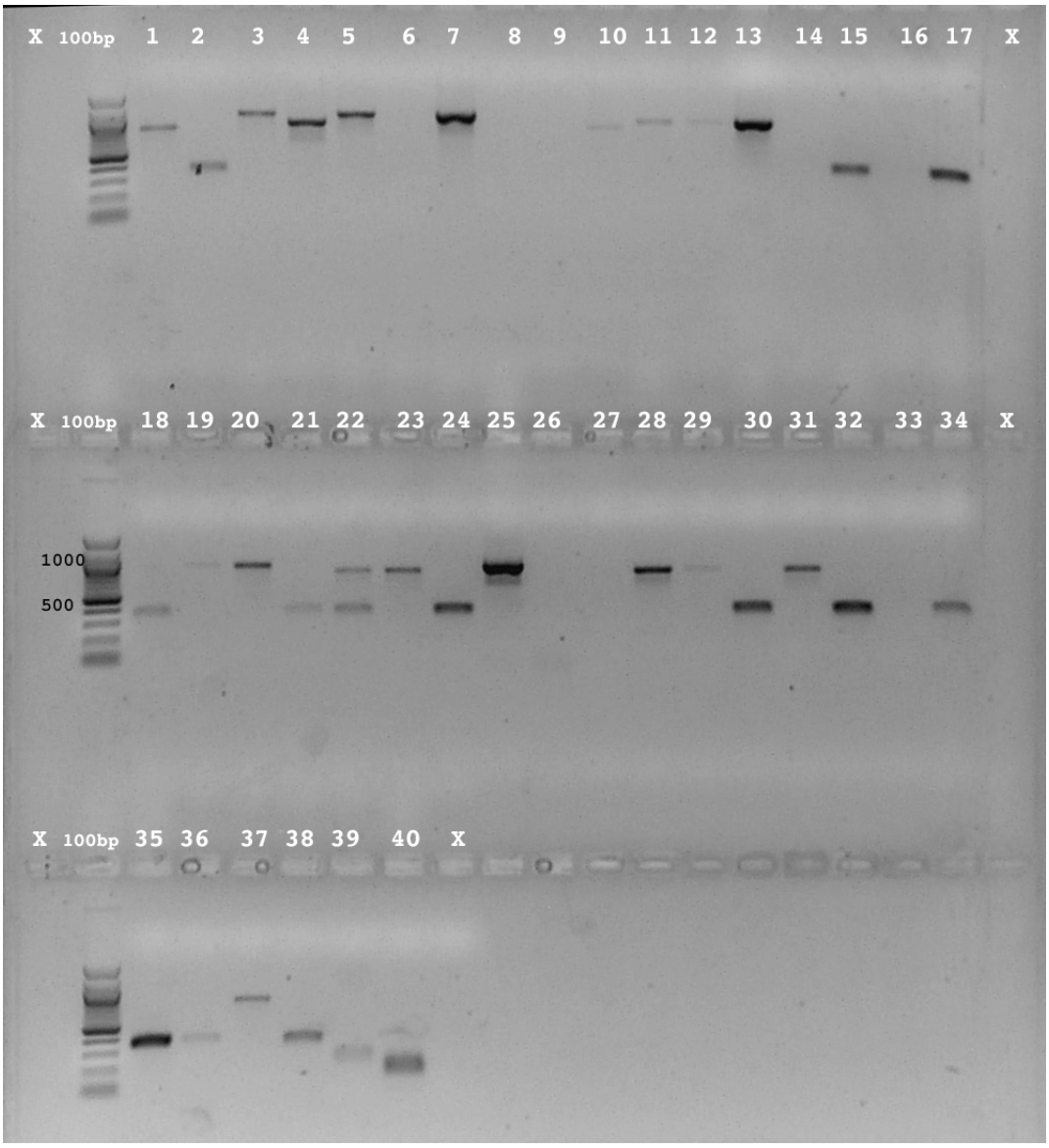


Figure 9.5 B. Colony PCR of a 10^4 dilution. 40 colonies were picked. Note the following colonies amplified with primer M13: 1, 2, 3, 4, 4, 5, 7, 10, 11, 12, 13, 15, 16, 17, 18, 20, 21, 22, 23, 24, 25, 26, 28, 30, 31, 32, 34, 35, 37, 38, 39, 40.



Chapter 10. Discussion

From these experiments, it is clear there is much molecular diversity in the genus *Linum*. The results obtained will allow the grouping to be made of species that based on the data, are most similar to each other and to flax (*L. usitatissimum*). To simplify this discussion, all the species will be categorized by the number assigned to them and listed in Table 1.1. The flax actin gene is expected to be conserved through the *Linum* genus, these results show otherwise. From those that amplified and those that did not, an immediate grouping can be made between species that share homology with the flax actin gene and those that do not (Table 10.1).

Table 10.1. *Linum* species that amplified with the flax actin gene

Amplified with flax actin gene	Did not amplify
4, 5, 6, 9, 10, 16, 17, 20, 27, 30, 31, 38, 39, 45, 49	1, 2, 7, 8, 11, 12, 13, 14, 17, 22, 23, 29, 32, 33, 34, 35, 36, 37, 41, 42, 43, 44, 46, 48

From the data collected with High A T primer 8, further groupings are made with the pattern of amplification found in primer 8. Those that amplified a 200bp band, those few that amplified a 250 or 225bp band, and those that amplified a faint band at 200bp where the groupings made under this section. Putting the under the original grouping of those that amplified the flax actin gene and those that did not (Table 10.2).

Table 10.2. *Linum* species grouping based on actin data and A&T primer 8 data

Amplified with flax actin gene			Did not amplify flax actin gene		
200bp	250/ 225bp	Faint 200bp	200bp	250/ 225bp	Faint 200bp
49, 9, 10, 20, 27, 30, 49	5, 16, 17	31, 45	1, 2, 7, 8, 12, 14, 22, 23, 34, 35, 36, 37, 40, 41, 42	13, 43	29, 32, 33, 44, 46

Data collected from High A and T primer 17 can also add to this grouping. Since not all species amplified with primer 17, these species are grouped according to whether or not they amplified.

Table 10.3. Table showing which species amplified with A&T primer 17, those highlighted in green amplified.

Amplified with flax actin gene			Did not amplify flax actin gene		
P8: 200bp	250/ 225bp	Faint 200bp	P8: 200bp	250/ 225bp	Faint 200bp
4, 9, 10, 20, 27, 30, 49	5, 16, 17	31, 38, 39, 45	1, 2, 7, 8, 12, 14, 22, 23, 34, 35, 36, 37, 40, 41, 42	13, 43	29, 32, 33, 44, 46

Using the miRNA primer 7 we can continue by grouping together like sequences that had similar patterns of band expression.

Table 10.4. Table grouping those species with the same pattern of expression when amplified with primer 7. Flax is highlighted in yellow. Those that amplified with actin are highlighted in pink, those that amplified with primer 17 are highlighted in green, and those that amplified with both are highlighted in orange.

7, 8, 9, 10, 15, 12
4, 3, 21
1, 17, 31, 32, 33, 36
5, 27
22, 23

From the data in these tables conclusions are made in respect to those species most closely related to flax in the molecular qualities studied, they include: 4 (*L. grandiflorum caeruleum*), 10 (*L. perene*), and 30 (*L. corymibiferum*), because theses species amplified with the actin primer, the same pattern as flax in primer 8, and also with primer 17. The next set close to flax would be 5 (*L. narbonense*), 17 (*L. perenne),* 38 (*L. maritimum*), and 39 (*L. mucronatum*), because they amplified with actin, primer 8 (but with a different pattern) and primer 17. The next set would be 9 (*L. perenne*), 20 (*L. flavum*), 49 (*L. strictum subsp. Strichum*), 16 (*L. maritimum*), 31 (*L. elegans*), and 45 (*L. tenufifolium*) since they amplified with actin, and primer 8, although did not amplify with primer 17. The next set includes 2 (*L. austriacm*), 7 (*L. perenne*), 8 (*L. perenne*), 12 (*L. austriacum*), 14 (*L. grandiflorum reubrum*), 34 (*L. grandiflorum desf.*), and 37 (*L. marginale*) because they did not amplify with the actin primers, but they did amplify with 8 and 17. Finally are the species that only amplified with 8. They include: 1 (*L. alpinum*), 22 (*L. altaicum*),

23 (*L. hologynum*), 35 (*L. hirsutum*), 36 (*L. lewisii*), 40 (*L. perenne*), 41 (*L. sp.*), 42 (*L. stelleroides*), 13 (*L. bienne*), 43 (*L. strictum*), 29 (*L. campanulatum*), 32 (*L. flavum* UK), 33 (*L. flavum* GER), 44 (*L. suffruticosum*), and 46 (*L. thracicum*). All of these are in order of bp amplified in primer 8; those that amplified at 200bp are first because they share the same amplification size as flax with primer 8.

The actin gene, which is expected to be highly conserved in these species, was determined to not amplify all species studied, showing to the vast molecular differences see in these species. Using the high A and T primers (8 and 17) it was determined that this region may be conserved to some exist to some extent in the genus although not in nearly the same amount as flax (where it comprises 15% of the total genome), since one primer did not amplify in all species and in the same intensities or sizes as flax. In order to make concrete assumptions about this region, it is definitely necessary to run more High A and T primers.

LIS-1 is inserted in a target region between two known genes in flax. Looking at primers that amplified this region showed little homology in the genus to this region, showing that evolution of this region may be a more recent occurrence, or that this region is molecularly different in these species despite being a region between two genes

Potential miRNA sites were also studied in these *Linum* species. Amplifications with primers 1-6 and 8-9 yielded virtually no amplifications, showing these regions are poorly conserved. One miRNA primer did amplify with all the species. Primer 7 yielded unusual patters of amplification with multiple bands of different sizes and intensities, this was especially useful in determining which species could be grouped together based on expression pattern (Table 10.4).

Chapter 11. Future Directions

To further complete this data set, it will be necessary to finish the TOPO TA cloning reaction to determine the homology of those miRNA primer 7 bands in each *Linum* species using it as another means of determining like families and completing the data from miRNA primer 7. Continuing PCR analysis with more data from the High A and T primers would also be useful and discovering how this light satellite A and T region is represented in these different *Linum* species. This can be done by amplifying more primers, choosing those that are determined to be highly repetitive from the Bethune/Bison BLAST.

Appendix A: Actin Sequences

>A-4 L
GAACGGCTTCATCGACCACAGGAAGTGCTTCTAAAGCTCCAGGAATTGGGAT
GCT

>A-4 R
CTCGGGAGATTGAGGCCGGATTCGTCGTACTCAGACTTCGAGATCCACATCT
GCTG

>A-5 L
GAACGGCTTCATCGACCACAGGAAGTGCTTCTAAAGCTCCAGGAATTGGGAT
GCT

>A-5 R
CTCGGGAGATTGAGGCCGGATTCGTCGTACTCAGACTTCGAGATCCACATCT
GCTG

>A-6 L
GAACGGCTTCATCGACCACAGGAAGTGCTTCTAAAGCTCCAGGAATTGGGAT
GCT

>A-6 R
CTCGGGAGATTGAGGCCGGATTCGTCGTACTCAGACTTCGAGATCCACATCT
GCTG

>A-9 L
GAACGGCTTCATCGACCACAGGAAGTGCTTCTAAAGCTCCAGGAATTGGGAT
GCT

>A-9 R
CTCGGGAGATTGAGGCCGGATTCGTCGTACTCAGACTTCGAGATCCACATCT
GCAG

>A-10 L
GAACGGCTTCATCGACCACAGGAAGTGCTTCTAAAGCTCCAGGAATTGGGAT
GCT

>A-10 R
CTCGGGAGATTGAGGCCGGATTCGTCGTACTCAGACTTCGAGATCCACATCT
GCTG

>A-16 L
GAACGGCTTCATCGACCACAGGAAGTGCTTCTAAAGCTCCAGGAATTGGGAT
GCT

>A-16 R
CTCGGGAGATTGAGGCCGGATTCGTCGTACTCAGACTTCGAGATCCACATCT
GCTG

>A-17 L
GAACGGCTTCATCGACCACAGGAAGTGCTTCTAAAGCTCCAGGAATTGGGAT
GCT

>A-17 R
CTCGGGAGATTGAGGCCGGATTCGTCGTACTCAGACTTCGAGATCCACATCT
GCTG

>A-20 L
GAACGGCTTCATCGACCACAGGAAGTGCTTCTAAAGCTCCAGGAATTGGGAT
GCT

>A-20 R
CTCGGGAGATTGAGGCCGGATTCGTCGTACTCAGACTTCGAGATCCACATCT
GCTG

>A-27 L
GAACGGCTTCATCGACCACAGGAAGTGCTTCTAAAGCTCCAGGAATTGGGAT
GCT

>A-27 R
CTCGGGAGATTGAGGCCGGATTCGTCGTACTCAGACTTCGAGATCCACATCT
GCTG

>A-30 L
GAACGGCTTCATCGACCACAGGAAGTGCTTCTAAAGCTCCAGGAATTGGGAT
GCT

>A-30 R
CTCGGGAGATTGAGGCCGGATTCGTCGTACTCAGACTTCGAGATCCACATCT
GCTG

>A-31 L
GAACGGCTTCATCGACCACAGGAAGTGCTTCTAAAGCTCCAGGAATTGGGAT
GCT

>A-31 R
CTCGGGAGATTGAGGCCGGATTCGTCGTACTCAGACTTCGAGATCCACATCT
GCTG

>A-38 L
NN

>A-38 R
CTCGGGAGATTGAGGCCGGATTCGTCGTACTCAGACTTCGAGATCCACATCT
GCTG

>A-39 L
GAACGGCTTCATCGACCACAGGAAGTGCTTCTAAAGCTCCAGGAATTGGGAT
GCT

>A-39 R
CTCGGGAGATTGAGGCCGGATTCGTCGTACTCAGACTTCGAGATCCACATCT
GCTG

>A-49 L
GAACGGCTTCATCGACCACAGGAAGTGCTTCTAAAGCTCCAGGAATTGGGAT
GCT

>A-49 R
CTCGGGAGATTGAGGCCGGATTCGTCGTACACAGACTTCGAGATCCACATCT
GCTG

>A-S6 L
GAACGGCTTCATCGACCACAGGAAGTGCTTCTAAAGCTCCAGGAATTGGGAT
GCT

>A-S6 R
CTCGGGAGATTGAGGCCGGATTCGTCGTACTCAGACTTCGAGATCCACATCT
GCTG

Appendix B: High A and T Primer 8 sequences

Primer 8 388 L (GAC AAA TTA TTC ACC GGC AGA) **R** (AAG CAC TAA TCG GGC ATT TC)

AAAGAACACAATCCTAAATCAAATAAATAAAATTAATCAAATAAAGAAAAT
TACAAAAGAAAAGAAAAAAAGTTATAATGCTAAAAGCTTCAATTTCAAATCA
TTAAAAAGGATTTTGCAGCTATTAAAAATACGAAATGCCCGATTAGTGCTTA
A

>8-4 R

CTTTTTATGATTTGAATTGAAGCTTTTAGCATTATAACTTTTTTTCTTTTCTTTT
GTAATTTTCTTTATTTGATTAATTTTATTTATTTGATTTAGGATTGTGTTTCTTT
TATCAGTCAAATCTTTCATTTTGCTTTCTGCCGGTGAATAATTTGTCA

>8-6L
GAAGATTTCGACTGATAAATGAAACACAATCCTCAATCAAATAAATAAAATT
AATCAAATAAAGAAAATTACAAAGGAAAAGAAAAAAAGTTATAATGCTAAA
GACTTCCATTTCAAATCATTAAAAAGGATTTTGCAGCTATTAAAAATGCGAA
ATGCCCGATTAGTGCTTAAG

>8-6 R
TTTTTATGATTTGAATTGAAGTCTTTAGCATTATAACTTTTTTTCTTTTCTTTTG
TAATTTTCTTTATTTGATTAATTTTATTTATTTGATTTAGGATTGTGTTTCTTTT
ATCAGTCAAATCTTTCATTTTTCTTTCTGCCTCTTTCATTTTTCTTTCTGCCGGT
GAATAATTTGTCA

>8-13 L
TCCTGGTCAATAAAAAAAATTAATCAAATAAAGAAAATTAAAAAAGAAAAG
AAAAAAAGTTATAATGCTAAAAGCTTCAATTTCAAATTCGTTAAAAAGGATT
TTGCAGCTATTAAAAATAGGAAATGCCCGATTAGGGCTTAGAA

>8-13 R

TTTATGATTTGAATTGAGCTTTTAGCATTATAACTTTTTTTCTTTTCTTGTTGTA
ATTTTCTTTATTTGATTAATTTTTTTTATTTGATTTAGGATTGTG

>8-14 L
TAAATCAAATAAATAAAATTAATCAAATAAAGAAAATTACAAAAGAAAAGA
AAAAAGGTTATAATGCTAAAAGCTTCAATTTAAAATCATTAAAAAGGATTTT
GCAGCTATTAAAAATACGAAATGCCCGATTAGTGCTTAA

>8-14 R
ATTTTAATTGAAGCTTTTAGCATTATAACCTTTTTTCTTTTCTTTTGTAATTTTC
TTTATTTGATTAATTTTATTTATTTGATTTAGGATTGTGTTTCTTTTATCAGTCA
AATCTTTCATTTTTCTTTCTGCCGGTGAATAATTTGTCA

>8-16 L
CACAATCCTAAATCAAATAAAAAAAATTAATCAAATAAAGAAAATTAAAAA
AGAAAAGAAAAAAAGTTATAATGCTAAAAGCTTCAATTTCAAATCATTAAAA
AGGATTTTGCAGCTATTAAAAATACGAAATGCCCGATTAGTGCTTA

>8-16 R
TCCTTTTTATGATTTGAATTGAAGCTTTTAGCATTATAACTTTTTTTCTTTTCTT
TTTTAATTTTCTTTATTTGATTAATTTTTTTTATTTGATTTAGGATTGTGTTTTT
TTTATCAGTCAAATCTTTGATTTTTCTTTCTGCCGGTGAATAATTTGTCAA

>8-32 L
GATAAAGAACCCATCCTAATCAATAAATAAAATTAATCTTTAAAGAAATTAC
TAGAGAAAAGAAAAAAAGTTATAATGCTGAAAGCTGCAATGGCTAAGGTTC
AAAAAGGAGGATGCAGCTTTTAAAAATACTACGGCGGGGTTTATTACTTCAT
AA

8-32 R
NN

>8-33 L
AAAGAACCCGTCCTAATCAAAAAATAAAATTAATCAATTAAGGAAATTACAA
TTGAAGGAAAAAAAGTTGTAATGCTAAAAGCTTCAGTTTCAAAGCTGAGAAG
GGATTAGGCGGCTATTAAAAATACGGACAGCGGGGGGGGTGTTTGTTTGTAA

>8-33 R
TATGATTTGAATTGAAGCTTTTAGCATTATAACTTTTTTTCTTTTCTTTTGTAA
TTTTCTTTATTTGATTAATTTTATTTATTTGATTTAGGAAGGGGGT

>8-34 L
AACACAATCCTAAATCAAATAAATAAAATTAATCAAATAAAGAAAATTACAA
AAGAAAAGAAAAAAGGTTATAATGCTAAAAGCTTCAATTTAAAATCATTAAA
AAGGATTTTGCAGCTATTAAAAATACGAAATGCCCGATTAGTGCTTA

>8-34 R
CCTTTTTATGATTTTAATTGAAGCTTTTAGCATTATAACCTTTTTTCTTTTCTTT
TGTAATTTTCTTTATTTGATTAATTTTATTTATTTGATTTAGGATTGTGTTTCTT
TTATCAGTCAAATCTTTCATTTTTCTTTCTGCCGGTGAATAATTTGTCA

>8-37 L
TGATAAAGAACACAATCCTAAATCAAATAAATAAAATTAATCAAATAAAGA
AAATTACAAAAGAAAAGAAAAAAAGTTATAATGCTAAAAGCTTCAATTTCAA
ATCATTAAAAAGGATTTTGCAGCTATTAAAAATACGAAATGCCCGATTAGTG
CTTA

>8-37 R
TTTTATGATTTGAATTGAAGCTTTTAGCATTATAACTTTTTTTCTTTTCTTTTGT
AATTTTCTTTATTTGATTAATTTTATTTATTTGATTTAGGATTGTGTTTCTTTTA
TCAGTCAAATCTTTCATTTTGCTTTCTGCCGGTGAATAATTTGTCAA

>8-38 L
GACTGATAAGAACACAATCCTAAATCAAATAAATAAAATTAATCAAATAAAG
AAAATTACAAAAGAAAAGAAAAAAAGTTATAATGCTAAAAGCTTCAATTTCA
AATCATTAAAAAGGATTTTGCAGCTATTAAAAATACGAAATGCCCGATTAGT
GCTTA

>8-38 R
TTTTTATGATTTGAATTGAAGCTTTTAGCATTATAACTTTTTTTCTTTTCTTTTG
TAATTTTCTTTATTTGATTAATTTTATTTATTTGATTTAGGATTGTGTTTCTTTT
ATCAGTCAAATCTTTCATTTTGCTTTCTGCCGGTGAATAATTT

>8-39 L
AATCCTAGTCAAATAAAAAAAATTAATCAAATAAAGAAAATAAAAAAAGAA
AAGAAAAAAAGTTATAATGCTAAAAGCTTCAATTTCAAATCGTTAAAAAGGA
TTTTAGCAGCTATTAAAAATACGAAATGCCCGATTAGTGCTTAA

>8-39 R
TTTTTATGATTTGAATTGAAGCTTTTAGCATTATAACTTTTTTTCTTTTCTTTTT
TAATTTTCTTTATTTGATTCTTTTTTTTTATTTGATTTAGGATTGTGTTTTTTTT
ATCAGTCAAATCTTTCATTTTTCTTTCTGCCGGTGAATAATTTGTCACTGAGGT
CAATGTTATCC

Appendix C: Primer 8 Left and Right Alignments

```
CLUSTAL 2.1 multiple sequence alignment


8-13L_L_bienne                  ------------------------------TCCTGGTCAA-TAAAAAAAA
8-39L_L_mucronatum              ----------------------------AATCCTAGTCAAATAAAAAAAA
8-6L_L_perenne                  GAAGATTTCGACTGATAAATGAAACACAATCCTCAATCAAATAAATAAAA
8-16L_L_maritimum_FR            ------------------------CACAATCCTAAATCAAATAAAAAAAA
8-14L_L_grandiflorum_reubrum    -------------------------------TAAATCAAATAAATAAAA
8-34L_L_gradiflorum             ----------------------AACACAATCCTAAATCAAATAAATAAAA
8-4L_L_grandiflorum_caeruleum   ------------------AAAGAACACAATCCTAAATCAAATAAATAAAA
8-37L_L_marginale               --------------TGATAAAGAACACAATCCTAAATCAAATAAATAAAA
8-32L_L_flavum_UK               ------------------GATAAAGAACCCATCCTAATCAA-TAAATAAAA
8-33L_L_flavum_GER              --------------------AAAGAACCCGTCCTAATCAA-AAAATAAAA
                                                                 ****  *** ****

8-13L_L_bienne                  TTAATCAAATAAAGAAAATTAAAAAAGAAAAGAAAAAAAGTTATAATGCT
8-39L_L_mucronatum              TTAATCAAATAAAGAAAATAAAAAAAGAAAAGAAAAAAAGTTATAATGCT
8-6L_L_perenne                  TTAATCAAATAAAGAAAATTACAAAGGAAAAGAAAAAAAGTTATAATGCT
8-16L_L_maritimum_FR            TTAATCAAATAAAGAAAATTAAAAAAGAAAAGAAAAAAAGTTATAATGCT
8-14L_L_grandiflorum_reubrum    TTAATCAAATAAAGAAAATTACAAAAGAAAAGAAAAAAAGGTTATAATGCT
8-34L_L_gradiflorum             TTAATCAAATAAAGAAAATTACAAAAGAAAAGAAAAAAAGGTTATAATGCT
8-4L_L_grandiflorum_caeruleum   TTAATCAAATAAAGAAAATTACAAAAGAAAAGAAAAAAAGTTATAATGCT
8-37L_L_marginale               TTAATCAAATAAAGAAAATTACAAAAGAAAAGAAAAAAAGTTATAATGCT
8-32L_L_flavum_UK               TTAATCT-TTAAAGAAA-TTACTAGAGAAAAGAAAAAAAGTTATAATGCT
8-33L_L_flavum_GER              TTAATCAATTAAGGAAA-TTACAATTGAAG-GAAAAAAAGTTGTAATGCT
                                ******   *** **** * *  *  ***  ******* *** *******

8-13L_L_bienne                  AAAAGCTTCAATTTCAAATTCGTTAAAAAGGATTTT-GCAGCTATTAAAA
8-39L_L_mucronatum              AAAAGCTTCAATTTCAAAT-CGTTAAAAAGGATTTTAGCAGCTATTAAAA
8-6L_L_perenne                  AAAGACTTCCATTTCAAAT-CATTAAAAAGGATTTT-GCAGCTATTAAAA
8-16L_L_maritimum_FR            AAAAGCTTCAATTTCAAAT-CATTAAAAAGGATTTT-GCAGCTATTAAAA
8-14L_L_grandiflorum_reubrum    AAAAGCTTCAATTTAAAAT-CATTAAAAAGGATTTT-GCAGCTATTAAAA
8-34L_L_gradiflorum             AAAAGCTTCAATTTAAAAT-CATTAAAAAGGATTTT-GCAGCTATTAAAA
8-4L_L_grandiflorum_caeruleum   AAAAGCTTCAATTTCAAAT-CATTAAAAAGGATTTT-GCAGCTATTAAAA
8-37L_L_marginale               AAAAGCTTCAATTTCAAAT-CATTAAAAAGGATTTT-GCAGCTATTAAAA
8-32L_L_flavum_UK               GAAAGCTGCAATGGCTAAG--GTTCAAAAAGGAGGATGCAGCTTTTAAAA
8-33L_L_flavum_GER              AAAAGCTTCAGTT-TCAAA--GCTGAGAAGGGATTAGGCGGCTATTAAAA
                                 **  ** *  *    **     * * ** *      ** *** ******

8-13L_L_bienne                  ATAGGAAATGCCCGATTAGGGCTTAGAA----
8-39L_L_mucronatum              ATACGAAATGCCCGATTAGTGCTTAA------
8-6L_L_perenne                  ATGCGAAATGCCCGATTAGTGCTTAAG-----
8-16L_L_maritimum_FR            ATACGAAATGCCCGATTAGTGCTTA-------
8-14L_L_grandiflorum_reubrum    ATACGAAATGCCCGATTAGTGCTTAA------
8-34L_L_gradiflorum             ATACGAAATGCCCGATTAGTGCTTA-------
8-4L_L_grandiflorum_caeruleum   ATACGAAATGCCCGATTAGTGCTTAA------
8-37L_L_marginale               ATACGAAATGCCCGATTAGTGCTTA-------
8-32L_L_flavum_UK               ATACT-ACGGCGGGGTT--TATTACTTCATAA
8-33L_L_flavum_GER              ATACGGACAGCGGGGGGGGTGTTTGTTTGTAA
                                **    *  **  *       *
```

```
CLUSTAL 2.1 multiple sequence alignment


8-14R_L_grandiflorum_reubrum       -----------ATTTTAATTGAAGCTTTTAGCATTATAACCTTTTTTCTT
8-34R_L_gradiflorum                -CCTTTTTATGATTTTAATTGAAGCTTTTAGCATTATAACCTTTTTTCTT
8-16R_L_maritimum_FR               TCCTTTTTATGATTTGAATTGAAGCTTTTAGCATTATAACTTTTTTTCTT
8-39R_Linum_mucronatum             ---TTTTTATGATTTGAATTGAAGCTTTTAGCATTATAACTTTTTTTCTT
8-37R_L_marginale                  ----TTTTATGATTTGAATTGAAGCTTTTAGCATTATAACTTTTTTTCTT
8-4R_L_grandiflorum_caeruleum      --CTTTTTATGATTTGAATTGAAGCTTTTAGCATTATAACTTTTTTTCTT
8-38R_Linum_mucronatum             ---TTTTTATGATTTGAATTGAAGCTTTTAGCATTATAACTTTTTTTCTT
8-13R_L_bienne                     -----TTTATGATTTGAATTGA-GCTTTTAGCATTATAACTTTTTTTCTT
8-33R_L_flavum_GER                 -------TATGATTTGAATTGAAGCTTTTAGCATTATAACTTTTTTTCTT
8-6R_L_perenne                     ---TTTTTATGATTTGAATTGAAGTCTTTAGCATTATAACTTTTTTTCTT
                                              **** ****** *  ********************* *********

8-14R_L_grandiflorum_reubrum       TTCTT-TTGTAATTTTCTTTATTTGATTAATTTTATTTATTTGATTTAGG
8-34R_L_gradiflorum                TTCTT-TTGTAATTTTCTTTATTTGATTAATTTTATTTATTTGATTTAGG
8-16R_L_maritimum_FR               TTCTT-TTTTAATTTTCTTTATTTGATTAATTTTTTTTATTTGATTTAGG
8-39R_Linum_mucronatum             TTCTT-TTTTAATTTTCTTTATTTGATTCTTTTTTTTTATTTGATTTAGG
8-37R_L_marginale                  TTCTT-TTGTAATTTTCTTTATTTGATTAATTTTATTTATTTGATTTAGG
8-4R_L_grandiflorum_caeruleum      TTCTT-TTGTAATTTTCTTTATTTGATTAATTTTATTTATTTGATTTAGG
8-38R_Linum_mucronatum             TTCTT-TTGTAATTTTCTTTATTTGATTAATTTTATTTATTTGATTTAGG
8-13R_L_bienne                     TTCTTGTTGTAATTTTCTTTATTTGATTAATTTTTTTTATTTGATTTAGG
8-33R_L_flavum_GER                 TTCTT-TTGTAATTTTCTTTATTTGATTAATTTTATTTATTTGATTTAGG
8-6R_L_perenne                     TTCTT-TTGTAATTTTCTTTATTTGATTAATTTTATTTATTTGATTTAGG
                                   ***** ** ******************  **** ***************

8-14R_L_grandiflorum_reubrum       ATTGTGTTTCTTTTATCAGTCAAATCTTTCATTTTTCTTTCTGCCGGTGA
8-34R_L_gradiflorum                ATTGTGTTTCTTTTATCAGTCAAATCTTTCATTTTTCTTTCTGCCGGTGA
8-16R_L_maritimum_FR               ATTGTGTTTTTTTTATCAGTCAAATCTTTGATTTTTCTTTCTGCCGGTGA
8-39R_Linum_mucronatum             ATTGTGTTTTTTTTATCAGTCAAATCTTTCATTTTTCTTTCTGCCGGTGA
8-37R_L_marginale                  ATTGTGTTTCTTTTATCAGTCAAATCTTTCATTTTGCTTTCTGCCGGTGA
8-4R_L_grandiflorum_caeruleum      ATTGTGTTTCTTTTATCAGTCAAATCTTTCATTTTGCTTTCTGCCGGTGA
8-38R_Linum_mucronatum             ATTGTGTTTCTTTTATCAGTCAAATCTTTCATTTTGCTTTCTGCCGGTGA
8-13R_L_bienne                     ATTGTG------------------------------------------
8-33R_L_flavum_GER                 AAGGGGGT----------------------------------------
8-6R_L_perenne                     ATTGTGTTTCTTTTATCAGTCAAATCTTTCATTTTTCTTTCTGCCTCTTT
                                   *  * *

8-14R_L_grandiflorum_reubrum       ATAATTTGTCA----------------------
8-34R_L_gradiflorum                ATAATTTGTCA----------------------
8-16R_L_maritimum_FR               ATAATTTGTCAA---------------------
8-39R_Linum_mucronatum             ATAATTTGTCACTGAGGTCAATGTTATCC---
8-37R_L_marginale                  ATAATTTGTCAA---------------------
8-4R_L_grandiflorum_caeruleum      ATAATTTGTCA----------------------
8-38R_Linum_mucronatum             ATAATTT--------------------------
8-13R_L_bienne                     ---------------------------------
8-33R_L_flavum_GER                 ---------------------------------
8-6R_L_perenne                     CATTTTTCTTTCTGCCGGTGAATAATTTGTCA
```

Appendix D: Target, and Bienne sequences

>Target LR
GGGTTTCAGAACTGTAACGAATGAGTAAGGGGGGAAGAGGAGTTGAACAAATTAGTGAA
GCAAAGCAAGGAGCTGGGAGGAATCAGAGCTTACTGGTTTGAAACTCGTCAAGGAAAGA
CATTGGTTGAGATAATGGCGTCTTTGAATTTGATTTCTTCTGTTATCCGCCCCCTTCTTC
AGTTCTGCTCAACGTGATGACCCTTCCCGCCACCGATTTAGGCAACTCTGGCGTATCTGC
AATTGGGGATTGAAAGGGAGGTATTGGAGAGACTCCTTTTTTTTTCCTTGTATATTGGTG
GTTAGTTAAGAGGAGCGATTATAATCGTAATCGGTGGGTACAGGGATGGAACGAAATTT
ATCACCGAAATAGAAATAGTGATGATTATGGTTAATAACCGCCACAAAGGTGGATGAAT
AATGGATATGCCATAACCATCCCCGAACCTCCTCGTTGCCAAGTCTAAATCCAAGCTGTT
ACAAGATAAACTTTGTATTATTTTCACACCATATAATATACTACATACATTATTATAAA
AAAAAGTTATTTATGAGGGCAATATAATACTTTAACTTTTATTTAATCAAATATAACTC
TTTATTTTGAGTTAATATAGCAGAATAATTTTTATAAAATCCAAAATATTTTATAAGTA
TCACCTTTTCAAATAATATATTATTTTATTTTAAAATACTACATATTACAAAATAACTC
ATAAAACACATCAAAATTAATCTCATTGAACTAAAAATTAAAAAAAACAAATAAAATTT
GCGAAGGAAACATGATTTAAAGTTTAAACTGTAGAGACATGATCACAAAATCTAAAGAA
AAAATAAAATTAAGAAGAGAGAAAAATGAAAGAAAAATATCACAGAAAAAGAAAGAAG
AGCGAGTAGGAAACACGTGCGGCTGTGGGCTCCGTAGCCGTCGAATCAATGGCTGAGATT
TAGTTTCAGTGGTTGTTTGGGATTCGAAATCGACGCAGCTTCCGTTTCCAATTTCAGTTT
TACAATACGAAGACTTTCGCCCCGTTTCCCATTTCTCTCCTTCTTCTTCTTCCCCGTTTTG
CTCTCTCTTTCTGTAACTCTCTCTCTCTCTCTTTCTCTCTCTCTCACTCCACCAACTGAAA
TAACCCACCTCCACGCGGCCATTAACCGAAGCAAATCCAAACCATTTCCCTTGTTTCTCCC
CTTTGAATGGCTTTCTTCCGGAGCTAAAAAACAAAACCCGGCCGCCACCGTTGCGCACCG
CCGCCGTGTTGAATGGGAAGGTACATGGCCAAGACAAAGAGAAACGCCAGCTGCCGGCA
TGGACCTCTCCCTCGATCCGTACCAGAGCCAAAAACCCTAGCTCTCCTCCGCCTTGAACA
GCAGCAAACCACCCCGAAATTGCATGCCTCTCCTCTCTCTTATCCTCCGCCAGATGCTGCT
GCTCCTGCCTCCGCCGGATATCTGCAGCTTCGGAGCCGCCGTCTTGAGAAGCCTCCCGTTG
TGTTGTACGGCTCTAACCGGCGACAGTCCAGGCAGGCTGATGATGTTGCTGACCAAGTGG
CTGGTGGTGGAAACCCTAACCCTAGCCC
>21
GGGTTTCAGAACTGTAACGAATGAGTAAGGGGGGAAGAGGAGTTGAACAAATTAGTGAA
GCAAAGCAAGGAGCTGGGAGGAATCAGAGCTTACTGGTTTGAAACTCGTCAAGGAAAGA
CATTGGTTGAGATAATGGCGTCTTTGAATTTGATTTCTTCTGTTATCCGCCCCCTTCTTC
AGTTCTGCTCAACGTGATGACCCTTCCCGCCACCGATTTAGGCAACTCTGGCGTATCTGC
AATTGGGGATTGAAAGGGAGGTATTGTAGAGACTCCTTTTTTTTTCTTGTATATTGGTG
ATTAGTTAGGAGGAGCGATTATAATCGTAGCCGGGGGATATGGGGATGGAACGAAATCT
ATCACCGAAATAGAAATGGTGATGATTATGGTTAATAACCGTCACAAAGGTGGATGAAT
AATGGATATGCCATACCCATCCCCGAACCTCCTCGTTGTCAAGCCTAAATCCGACCTGTT
ACAAGATAAACTTTATATTATTTTCACACCATATAATATATTACATACATTATTATAAA
AAAAAGTTATTTATGAGGACAATATAATACTTTAACTTTTATTTAATCAAATATAACTC
ATTATTTTGAGTTAATATAGCAGAATAATTTCTATAAAATCCAAAATATTTTATAAGTA
TCATCTTTTCAAATAATATATTATTTTATTTTAAAATACTACATATACAAAATAACTCA
TAAAAAACACCAAAATTAATCTCATTGAACTAAAAATTAAAAAAAACAAATAAAATTTG
CGAAGGAAACATGATTTAAAGTTTAAACTGTAGAGACATGATCACAAAATCTAAAGAAA

AATAAAATTAAGAAGAGAGAAAAATGAAAGAAAAATATCACAGAAAAAGAAAGAAGAG
CGAGTAGGAAACACGTGCGGCTGTGGGCCTCCGAAGCCCGTCGCATCAATGGCTGAGATT
TAGTTTCAGTGGTTGTTTGGGATTCGAAATCCGACGCAGCTTCCCGTTTCCAATTTCAGT
TTTACAAATACGAAGACTTCGCCCCCCGTTTCCCATCTCTCCTTCTTCTTCTTCCCCGTTT
TGCTCTCTCTTTCTGTAACTCTCTCTCTCTCTCTCTCTCACTCCACCAACTGAAATAACCC
ACCTCCACGCGGCCATTAACCGAAGCAAATCCAAACCATTTCCCTTGTTTCTCCCCTTTG
AATGGCTTTCTTCCGGAGCTAAAAAACAAAACCCGGCCGCCACCGTTGCGCACCGCCGCC
GTGTTGAATGGGAAGGTACATGGCCAAGACAAAGAGAACCGGCCACGTCGCCGGCATGG
ACCTCTCCCTCGGCGGCGTCCGTACCAGAGCCAAAACCCTAGCTCTCCTCCGCCTTGAACA
GCAGCAAACCACCCCGAAATTGCATGCCTCTCCTCTCTCTTGTCCTCCGCCAGATGCTGCT
GCTCCTGCCTCCGCCGGATATCTGCAGCTTCGGAGCCGCCGTCTTGAGAAGCCTCCCGTTG
TGTTGTACGGCTCTAACCGGCGACAGTCCAGGCAGGCTGATGATGTTGCTGACCAAGTGG
CTGGTGGTGGAAACCCTAACCCTAGCC
>8
GGGTTTCAGAACTGTAACGAATGAGTAAGGGGGGAAGAGGAGTTGAACAAATTAGTGAA
GCAAAGCAAGGAGCTGGGAGGAATCAGAGCTTACTGGTTTGAAACTCGTCAAGGAAAGA
CATTGGTTGAGATAATGGCGTCTTTGAATTTGATTTCTTCTGTTATCCGCCCCCTTCTTC
AGTTCTGCTCAACGTGATGACCCTTCCCGCCACCGATTTAGGCAACTCTGGCGTATCTGC
AATTGGGGATTGAAAGGGAGGTATTGTAGAGACTCCTTTTTTTTTCTTGTATATTGGTG
ATTAGTTAGGAGGAGCGATTATAATCGTAACCGGTGGATATAGGGATGGAACGAAATCT
ATCACCGAAATAGAAATGGTGATGATTATGGTTAATAACCGTCACAAAGGTGGATGAAT
AATGGATATGCCATAACCATCCCCGAACCTCCTCGTTGTCAAGCCTAAATCCGACCTGTT
ACAAGATAAACTTTGTATTATTTTCACACCATATAATATATTACATACATTATTATAAA
AAAAAAGTTATTTATGAGGACAATATAATACTTTAACTTTTATTTAATCAAATATAACT
CATTATTTTGAGTTAATATAGCAGAATAATTTCTATAAAATCCAAAATATTTTATAAGT
ATCATCTTTTCAAATAATATATTATTTTATTTTAAAATACTACATATACAAAATAACTC
ATAAAAAACACCAAAATTAATCTCATTGAACTAAAAATTAAAAAAAACAAATAAAATTT
GCGAAGGAAACATGATTTAAAGTTTAAACTGTAGAGACATGATCACAAAATCTAAAGAA
AAATAAAATTAAGAAGAGAGAAAAATGAAAGAAAAATATCACAGAAAAAGAAAGAAGA
GCGAGTAGGAAACACGTGCGGCTGTGGGCCTCCGAAGCCCGTCGCATCAATGGCTGAGAT
TTAGTTTCAGTGGTTGTTTGGGATTCGAAATCCGACGCAGCTTCCCGTTTCCAATTTCAG
TTTTACAAATACGAAGACTTCGCCCCCCGTTTCCCATCTCTCCTTCTTCTTCTTCCCCGTT
TTGCTCTCTCTTTCTGTAACTCTCTCTCTCTCTCTCTCTCACTCCACCAACTGAAATAACC
CACCTCCACGCGGCCATTAACCGAAGCAAATCCAAACCATTTCCCTTGTTTCTCCCCTTTG
AATGGCTTTCTTCCGGAGCTAAAAAACAAAACCCGGCCGCCACCGTTGCGCACCGCCGCC
GTGTTGAATGGGAAGGTACATGGCCAAGACAAAGAGAACCGGCCACGTCGCCGGCATGG
ACCTCTCCCTCGGCGGCGTCCGTACCAGAGCCAAAACCCTAGCTCTCCTCCGCCTTGAACA
GCAGCAAACCACCCCGAAATTGCATGCCTCTCCTCTCTCTTGTCCTCCGCCAGATGCTGCT
GCTCCTGCCTCCGCCGGATATCTGCAGCTTCGGAGCCGCCGTCTTGAGAAGCCTCCCGTTG
TGTTGTACGGCTCTAACCGGCGACAGTCCAGGCAGGCTGATGATGTTGCTGACCAAGTGG
CTGGTGGTGGAAACCCTAACCCTAGCC

Appendix E: Target 3' sequence from species 27 or Flax transformed 37

6-B Target L 3’ (Ft37) L

GCACTCTGGCGAATCTGAAATTGGGGATTGAAAGGGAGGTCTTGGAGGAGTCCGGATCT
GGAGCCAGGAAATCAAATTCTGGATCTGATTCGCATTGGCACAACCCATCCGAATCCTCC
CAAAAATGGTACCCTTGATCTAACTCCTGAAATTTTGGGGGCATGTAAATAGTCTCCTTC
CTTCTTCATCTTCCATCTTCAAAGGT

6-B Target L 3’ (FT37) 3 R

GGCGTTAGATCAAGGGTACCATTTTTTGGACGATTCAGATGGGTTGTGCCATTCCGAATC
AGATCCAGAATTTGATTTCCTGGCTCCAGATCCGGACTCCTCCAAGACCTCCCTTTCAAT
CCCCAATTTCAGATTCGCCAGAGTTGCCTAAATCGGTGGCGGGAAGGGTCATCACATTGA
GCAGAACTGAAGAAGGGGGA

References

Background on Nucleic Acid Dot Plots. (1998, June 20). *arbl.cvmbs.colostate.edu*. Retrieved February 16, 2011, from http://www.vivo.colostate.edu/molkit/dnadot/bkg.html

Bennetzen, J. L., Ma, J., & Devos, K. M. (2005). Mechanisms of recent genomic size variation in flowering plants. *Annals of Botany*, *95*, 127-132.

Chen, Y., Lowenfeld, R., & Cullis, C. A. (2009). An environmentally induced adaptive (?) insertion event in flax. *International Journal of Genetics and Molecular Biology* , *1*(3), 038-047.

Chen, Y., Schneeberger, R. G., & Cullis, C. A. (2005). A site-specific insertion sequence in flax genotrophs induced by environment. *New Phytologist*, *167*, 171-180.

Cullis, C. A. (1980). DNA sequence organisation in the flax genome. *Biochimica Et Biophysica Acta*, *652*, 1-15.

Cullis, C. A., & Kolodynska, K. (1975). Variation in the isozymes of flax (Linum usitatissimum) genotrophs. *Biochemical Genetics*, *13*, 687-697.

Cullis, C. A. (1977). Molecular aspects of environmental induction of heritable changes in flax. *Heredity*, *38*(2), 129-154.

Cullis, C. A. (2004). *Plant genomics and proteomics*. San Francisco: John Wiley & Sons.

Cullis, C. A. (2005). Mechanisms and control of rapid genomic changes in flax. *Annals of Botany*, *95*, 201-206.

Diederichsen, A., & Richards, K. (2003). Cultivated flax and the genus Linum L.: Taxonomy and germplasm conservation . *Flax: The genus Linum* (pp. 22-54). New York: Taylor & Francis Ltd.

Durrant, A. (1962). The environmental induction of heritable change in Linum. *Heredity*, *17*, 27-61.

Evans, G. M. (1968). Nuclear changes in flax. *Heredity*, *23*, 25-38.

Kellogg, E. A., & Bennetzen, J. L. (2004). The evolution of nuclear genome structure in seed plants. *American Journal of Botany*, *91*(10), 1709-1725.

Nucleic Acid Dot Plots. (1998, June 24). *arbl.cvmbs.colostate.edu*. Retrieved February 20, 2011, from http://www.vivo.colostate.edu/molkit/dnadot/

Oh, T. J., & Cullis, C. A. (2003). Labile DNA sequences in flax identified by combined sample representational difference analysis (csRNA). *Plant Molecular Biology*, *52*, 527-536.

Project Description. (n.d.). *University of Texas*. Retrieved August 15, 2010, from www.sbs.utexas.edu/simpsonlab/josh%20files/mcdilldig.pdf

Schnable, P., Ware, D., Fulton, R., Stein, J., Wei, F., Pasternak, S., et al. (2009). The B73 maize genome: complexity, diversity, and dynamics. *Science*, *326*, 1112-1115.

Springer, N., Jia, Y., Wu, W., Richmond, T., Kitzman, J., Rosenbaum, H., et al. (2009). Maize inbreds exhibit high levels of copy number variation (CNV) and presence/absence variation (PAV) in genome content. *PLoS Genetics*, *5*(11), 1-17.

Tyson, H., & Fields, M. A. (1982). Molecular-weight and net charge of peroxidase isozymes in F1 hybrids between L and S flax genotrophs. *Biochemical Genetics*, *20*, 919-927.

VersionU. (2006). TOPO TA Cloning User Manual . *Invitrogen*, *25-0184*, ii-27. Retrieved February 5, 2011, from www.invitrogen.com

Yermanos, D. M. (1957). Cytogenetic relationships among linum species. *Flax Institute of the United States* (pp. 1-2). Riverside, Ca: Committee on Co-ordination of Flax research.