INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI

University Microfilms International
A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
313-761-4700     800-521-0600
Characterization of DNA polymorphisms associated with environmentally induced heritable changes in flax

Schneeberger, Richard Gerald, Ph.D.

Case Western Reserve University (Health Sciences), 1992

Copyright ©1992 by Schneeberger, Richard Gerald. All rights reserved.
CHARACTERIZATION OF DNA POLYMORPHISMS ASSOCIATED WITH 
ENVIRONMENTALLY INDUCED HERITABLE CHANGES IN FLAX

by

RICHARD GERALD SCHNEEBERGER

Submitted in partial fulfillment of the requirements for 
the Degree of Doctor of Philosophy

Thesis Advisor: Christopher A. Cullis

Department of Biology
CASE WESTERN RESERVE UNIVERSITY
January, 1992
Copyright © (1992) by
Richard Gerald Schneeberger
We hereby approve the thesis of

Richard G. Schneberger

candidate for the Ph. D.

degree.*

Signed: [Signatures]

(Chairman)

[Signatures]

[Signatures]

[Signature]

Date 9/30/91

*We also certify that written approval has been obtained for any proprietary material contained therein.
CHARACTERIZATION OF DNA POLYMORPHISMS ASSOCIATED WITH ENVIRONMENTALLY INDUCED HERITABLE CHANGES IN FLAX

Abstract

by

RICHARD GERALD SCHNEEBERGER

The flax plant (*Linum usitatissimum*) serves as a model system for the study of factors involved in genome stability. Growth of several inbreeding varieties of fibre flax in defined environments results in the production of changes in the first generation progeny of treated plants. The changes take the form of phenotypic, biochemical and molecular/genetic alterations which are stably inherited by subsequent generations. The stable lines are called genotrophs. The work described in this dissertation focuses on the characterization of DNA polymorphisms and the organization of polymorphic loci in the progenitor line and the genotrophs. The 5S rRNA gene family and associated sequences were characterized to identify DNA polymorphisms. Characterization of the 5S rRNA gene family identified five classes of genes based on genomic organization and DNA sequence homology. The data indicate that the 5S rRNA gene family in flax has high
levels of sequence variability and may contain many pseudogenes. Comparison of the representation of the 5S rRNA genes in the progenitor line and genotroph lines indicated that group 2 5S rRNA genes are reduced in both small and large type genotroph lines. Restriction Fragment Length Polymorphisms (RFLPs) were identified for group 4 5S rRNA genes. Characterization of these polymorphisms indicated that they are due to a DNA rearrangement event. The RFLP pattern identified in four, independently produced, small genotrophs was identical, indicating that similar DNA alterations have occurred in this gene family in association with environmental induction of the small phenotype genotroph. Dissimilar RFLP patterns were found in two large genotrophs, indicating that multiple DNA alteration patterns are possible. Characterization of 5S rRNA gene associated sequences resulted in the identification of an RFLP caused by a transposable element like DNA insertion, termed LIS-1. The LIS-1 insertion is found in all small genotrophs as well as the large genotroph L^H. The data indicate that specific repetitive and low copy DNA rearrangements are associated with environmentally induced heritable changes in flax. The results are discussed in terms of mechanisms for the production of environmentally induced
DNA alterations.
Dedication

This dissertation is dedicated to the memory of Dr. Norman Alldridge, whose friendship, guidance and tireless enthusiasm for teaching and science were a constant source of inspiration to me.
Acknowledgments

There are many people who I am indebted to for their help and support during the course of my dissertation. I would like to thank my family and friends for their support and especially my parents for always allowing me to choose my own way and supporting my decisions and dreams. I would like to thank Chris Cullis for giving me the opportunity to work in his lab and for his guidance and support over the past five years and the rest of the lab, especially Susan Gorman, for their encouragement and constructive criticism. I owe a special thanks to Dr. Tom Mastin for his support and stimulating conversations concerning my work. I am also grateful to the members of my thesis committee, Chuck Rozek, Chris Town, David Setzer, Marcello Jacobs-Lorena, Norman Alldridge and Roy Ritzmann for their constructive comments and input into the dissertation. Last but not least I am indebted to my good friends, the Yeastie Boys, for helping me to maintain some peace of mind in difficult times. I would also like to thank David Threadgill and Terry Magnuson for assistance with the pulsed field electrophoresis studies, Mark Gorman for assistance with genetic linkage analysis and Cyndee Richards for preparing several of the tables.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract</td>
<td>ii</td>
</tr>
<tr>
<td>Dedication</td>
<td>v</td>
</tr>
<tr>
<td>Acknowledgments</td>
<td>vi</td>
</tr>
<tr>
<td>Table of Contents</td>
<td>vii</td>
</tr>
<tr>
<td>List of Figures</td>
<td>x</td>
</tr>
<tr>
<td>List of Tables</td>
<td>xiii</td>
</tr>
<tr>
<td>List of Abbreviations</td>
<td>xiv</td>
</tr>
<tr>
<td><strong>Chapter 1. Introduction</strong></td>
<td></td>
</tr>
<tr>
<td>A. Hereditary Differences Between Plants and Animals</td>
<td>1</td>
</tr>
<tr>
<td>B. Environmentally Induced Heritable Changes in Flax</td>
<td>3</td>
</tr>
<tr>
<td>1. Molecular Changes in Flax Genotrophs</td>
<td>9</td>
</tr>
<tr>
<td>2. 5S rRNA Genes</td>
<td>13</td>
</tr>
<tr>
<td>C. Environmentally Induced Heritable Changes in Other Plants</td>
<td>19</td>
</tr>
<tr>
<td>D. Rationale</td>
<td>21</td>
</tr>
<tr>
<td><strong>Chapter 2. Methods</strong></td>
<td></td>
</tr>
<tr>
<td>A. Plant Material</td>
<td>25</td>
</tr>
<tr>
<td>B. Derivation of Flax Genotrophs</td>
<td>25</td>
</tr>
<tr>
<td>C. Plant DNA Isolation</td>
<td>26</td>
</tr>
<tr>
<td>D. Restriction Enzyme Digests, Electrophoresis, and Transfer of DNA</td>
<td>29</td>
</tr>
<tr>
<td>E. Preparation of Megabase DNA</td>
<td>30</td>
</tr>
<tr>
<td>F. Restriction Enzyme Digests of DNA in Agarose</td>
<td>32</td>
</tr>
<tr>
<td>G. Pulse Field Gel Electrophoresis</td>
<td>33</td>
</tr>
<tr>
<td>H. DNA Probe Preparation, Labeling and Hybridization</td>
<td>34</td>
</tr>
<tr>
<td>I. Recombinant DNA Methods</td>
<td>36</td>
</tr>
<tr>
<td>J. Copy Number Reconstructions</td>
<td>42</td>
</tr>
<tr>
<td>K. DNA Sequencing</td>
<td>43</td>
</tr>
</tbody>
</table>

vii
Chapter 3. Results

A. Molecular Analysis of 5S rRNA Gene Organization in Flax ............................................. 46
   1. Cloning and Restriction Mapping of 5S rRNA Genomic Sequences ............................. 46
   2. Lambda 5S rDNA Clone Homology to pBG13 ......................................................... 55
   3. Sequence Analysis of Group 1 – 5 5S rRNA Genes ......................................................... 63

B. Characterization of Quantitative 5S rDNA Polymorphisms Among Flax Genotrophs ............... 83

C. Characterization of 5S rDNA Restriction Fragment Length Polymorphism Among Flax Genotrophs .......................................................... 88
   1. Identification of 5S rDNA RFLPs ............................................................................. 88
   2. Characterization of 5S-20 5S rDNA ................................................................. 91
   3. The Genomic Organization and Copy Number of 5S-20 5S rDNA ......................... 99
   4. 706 base pair Probe Identifies RFLPs in Flax Genotrophs ........................................ 100
   5. Analysis of Polymorphisms Using pRS20.7 Spacer Probes ........................................... 112
   6. Genetic and Physical Linkage Analysis of Genotroph RFLPs ..................................... 119
   7. Cloning and Characterization of DNA from the Flp-1 locus .................................. 135
   8. Flp-1 Locus Variation in Other Flax Genotypes ......................................................... 145

D. Identification and Characterization of RFLPs Detected by 5S-20 Junction DNA .............. 149
   1. 5S-20 Junction Fragment Detects Low Copy Number RFLPs ...................................... 149
   2. Molecular Cloning of 2010.1 Homologous DNA Fragments from L\(^H\): Characterization of the 2.3 kb RFLP ........................................................... 153
   3. Molecular Cloning of the pZLH2.3 Pl Progenitor Allele ........................................... 163
   4. Reconstruction of the L\(^H\) Allele: Identification of a DNA Insertion Sequence ............... 170

viii
5. Sequence Analysis of Insertion
   Sequence Target Sequence and Junctions. 181
6. Identification LIS-1 RFLPs in Other
   Genotrophs and Flax Cultivars . . . . . 185
7. Genomic Organization of LIS-1 in PI
   and Genotrophs. . . . . . . . . . . 192

Chapter 4. Discussion . . . . . . . . . . . . . . 199

A. Organization of 5S rDNA in Flax: Does a
   Dispersed Organization Facilitate Sequence
   Divergence in a Repetitive Gene Family? . . 200

B. Identification and Characterization of 5S rDNA
   Polymorphisms Between PI and
   Large and Small Genotrophs . . . . . . 208
   1. Quantitative Variation in Flax 5S rDNA. 208
   2. Characterization of Restriction Fragment
      Length Polymorphisms . . . . . . 212
   3. Induction of Small Genotrophs is
      Associated with Specific RFLPs . . . 214

C. Identification and Characterization of Insertion
   Sequence
   Polymorphisms . . . . . . . . . . . . 219

D. Implications of Genotroph Specific Polymorphisms
   on Models for Environmental Induction of
   Heritable Changes in Flax; One Mechanism or
   Many? . . . . . . . . . . . . . . . . . . 222

Bibliography . . . . . . . . . . . . . . . . . . . . 229

Appendix A . . . . . . . . . . . . . . . . . . . . 240
<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Induction of Flax Genotrophs</td>
<td>4</td>
</tr>
<tr>
<td>2.</td>
<td>Restriction Maps of 5S Genomic Clones</td>
<td>49</td>
</tr>
<tr>
<td>3.</td>
<td>EMBL4 5S Clone Homology to pBG13</td>
<td>52</td>
</tr>
<tr>
<td>4.</td>
<td>DNA Blot Analysis of Group 1 - 5 5S rDNA</td>
<td>61</td>
</tr>
<tr>
<td>5.</td>
<td>Sequence comparison of Angiosperm and Flax Group 1 and 2 5S rRNA sequences</td>
<td>75</td>
</tr>
<tr>
<td>6.</td>
<td>Secondary structures of Group 1 - 5 5S rDNA</td>
<td>80-82</td>
</tr>
<tr>
<td>7.</td>
<td>Group 1 Versus Group 2 Quantitative Polymorphism Among Flax Genotrophs</td>
<td>86</td>
</tr>
<tr>
<td>8.</td>
<td>Detection of RFLPs with Group 4 5S rDNA clones 5S-12 and 5S-20</td>
<td>90</td>
</tr>
<tr>
<td>9.</td>
<td>Sequence of pRS20.7 and pRS20.7II and Comparison of the 5S1 and 5S2 Gene Repeats</td>
<td>94</td>
</tr>
<tr>
<td>10.</td>
<td>Restriction Enzyme Site and DNA Probe Map of pRS20.7</td>
<td>97</td>
</tr>
<tr>
<td>11.</td>
<td>pRS20.7 RFLP Patterns Among Flax Genotrophs</td>
<td>103</td>
</tr>
<tr>
<td>12.</td>
<td>Methylation Independent Detection of RFLPs Among Flax Genotrophs</td>
<td>108</td>
</tr>
<tr>
<td>13.</td>
<td>Copy Number Reconstruction of Pl, L^H and S1 EcoRI RFLPs in Flax Genotrophs</td>
<td>111</td>
</tr>
<tr>
<td>14.</td>
<td>High Molecular Weight RFLPs Detected by pRS20.7</td>
<td>114</td>
</tr>
<tr>
<td>15.</td>
<td>DNA Blot Analysis of pRS20.7, 5S1 and 5S2 Spacer Sequences</td>
<td>117</td>
</tr>
<tr>
<td>16.</td>
<td>Linkage Analysis of pRS20.7 RFLPs in L1 x L6 F2 Individuals</td>
<td>121</td>
</tr>
<tr>
<td>17.</td>
<td>Linkage Analysis of High Molecular Weight RFLPs</td>
<td>124</td>
</tr>
</tbody>
</table>
18. Megabase DNA Blot Analysis of Group 4 5S rDNA .......................... 128

19. CHEF Gel Electrophoresis and DNA Blot Analysis of P1, L" and S1 MluI Digested Megabase DNA .......................... 130

20. CHEF Gel Electrophoresis of P1, L" and S1, MluI Digested, Megabase DNA .......................... 132

21. Restriction Site Map and Genomic Organization of p651 5S rDNA .......................... 140

22. DNA Sequence Comparison of pRS20.7 and p651 .......................... 143

23. Hybridization pattern detected by pRS20.7 in EcoRI digested DNA of flax varieties and Linum grandiflorum caeruleum .......................... 147

24. Identification of a Low Copy Number RFLP with the Junction DNA of 5S-20 .... 152

25. DNA Blot Analysis of p2010.1 selected, lambda ZapII, L" EcoRI clones .......................... 156

26. DNA Blot Analysis of pZLH3.9, -3.8, and -2.8 .......................... 159

27. Identification of a P1 EcoRI Fragment + Homologous to pZLH2.3 .......................... 162

28. Comparison of Restriction Enzyme Site Maps of pZLH2.3 and pZPL3.7 .......................... 166

29. Genomic DNA Blot Analysis of EcoRI digested P1, L" and S1 DNAs hybridized with pZPL3.7 .......................... 169

30. Genomic DNA Blot Analysis of EcoRI digested, P1, L" and S1 DNAs hybridized with pZLH6.2 .......................... 173

31. Comparison of the Restriction Enzyme Site Maps of pZPL3.7 and pZLH6.2 .......................... 176
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table I.</th>
<th>Flax 5S Genomic Clone Classification</th>
<th>Page Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table II.</td>
<td>5S rRNA Gene Repeat Pairwise Comparison</td>
<td>67</td>
</tr>
<tr>
<td>Table III.</td>
<td>5S rRNA Gene Coding Region Pairwise Comparison</td>
<td>68</td>
</tr>
<tr>
<td>Table IV.</td>
<td>5S rRNA Spacer Region Pairwise Comparison</td>
<td>69</td>
</tr>
<tr>
<td>Table V.</td>
<td>Copy Numbers of 706 bp Repeat in Comparison to Total 5S rDNA</td>
<td>95</td>
</tr>
</tbody>
</table>
Abbreviations.

bp base pair
CTAB Cetyltrimethylammonium bromide
CHEF Contour-clamped Homogenous Electrical Field
EDTA Ethylene diamine tetra-acetic acid, disodium salt
EtBr Ethidium bromide
FITGE Field Inversion Gel Electrophoresis
kb Kilobase
l liter
Mb Megabase
min minute(s)
milligram
ml milliliter
pfu plaque forming units
rDNA 18S, 5.8S and 25S Ribosomal RNA gene(s).
RFLP Restriction Fragment Length Polymorphism
rpm revolutions per minute
rRNA Ribosomal RNA
SDS Sodium dodecyl sulphate
SSC 0.15M NaCl, 0.015M sodium citrate
TE 10mM Tris-HCl, 0.1mM EDTA pH 8.0
5S rDNA DNA coding for the 5S rRNA and spacer region

xiv
Chapter 1. **Introduction**

A. **Hereditary differences between Plants and Animals.**

The effect of the environment on the phenotype of an organism is largely thought to be physiological with the genome remaining unchanged. A tenet of Mendelian genetics states that, although the environment and the genotype can interact to produce a novel phenotype, this final form is not inherited in the next generation. This principle has been codified in Weismann’s doctrine of the separation of the somatic and germ line tissues (Weismann 1892). According to this doctrine, mutations which occur in the soma of higher eukaryotes are inconsequential to the offspring due to the separation of the germline from the soma early in development. Consequently, the only mutations in the parent which are heritable in the next generation are those which occur in the germline prior to gamete formation. However, this principle of heredity is only valid for those organisms which display a separation of germ and soma. In the plant kingdom it has long been recognized that the germline is not separated from the soma. Therefore Weismann’s Doctrine is not applicable to plant systems (Babcock and Clausen 1927; reviewed in Klekowski 1989).
This fundamental distinction between plants and animals is the consequence of basic differences in development. In animals, after fertilization to form a zygote, development proceeds by the gradual differentiation of cells into all of the tissues and organs of the adult body. The germ line is set aside very early in development and these are the only cells to remain relatively undifferentiated. Therefore, the embryo develops into an essentially complete but immature form of the adult with all tissues and organs present. In contrast, plants develop only representatives of the major organs and tissues (shoot, root, leaf and vascular system) of the adult during embryogenesis and a germ line is not defined until late in development. After germination, cell division and differentiation at sites within the meristems results in continued organ and tissue formation throughout the life of the plant. The cells which make up the apical meristem produce the vegetative structures of the plant in addition to the reproductive structures and gametes. Thus, mutations which occur during vegetative growth in the meristem cell lineages which produce the gametes may be transmitted to the next generation. This phenomenon is well known to horticulturists in the form of "bud sports" which are branches of a plant which produce a
novel fruit. These novelties have been shown in many cases to be due to heritable genetic changes as seeds from these fruits continue to produce and pass on the novelty in a Mendelian fashion. Thus, the primary consequence of the lack of a germ line is that mutations occurring during vegetative growth can be inherited in the next generation. The potential for inheriting somatically derived mutations presents the opportunity for the environment to have more than a physiological influence on the phenotype of a plant.

B. Environmentally Induced Heritable Changes in Flax.

In the mid 1950’s Alan Durrant performed a set of plant nutrition experiments with the inbreeding flax variety, Stormont Cirrus (Durrant 1958, 1962a, 1971). Plants were grown in different fertilizer combinations of Potassium (K), Nitrogen (N) and Phosphate (P) to assess the optimal growth conditions for flax cultivation (Figure 1). The plants treated with the different fertilizer combinations produced the expected large differences in total plant weight and stature due to the different fertilizer environments. However, when the self fertilized progeny of these plants were
Parental Flax

After 1 generation of growth in altered soil the capsule seeds are planted in standard soil.

L Plant  
- smooth seed capsule  
- 10% more DNA, normal ribosomal DNA
  
S Plant  
- hairy seed capsule  
- slightly less total DNA, greatly reduced ribosomal DNA

SEEDS BREED TRUE  
SEEDS BREED TRUE
Figure 1. Induction of Flax Genotrophs (taken from Alberts et al. 1983).
compared in a common environment, he found that some of them no longer resembled the progenitor line. These new plant lines were observed to breed true and maintain the altered phenotypes as a homogeneous genotype. This observation suggested that the growth of plants in the altered fertilizer regimes somehow effected heritable changes in the first generation offspring. Depending on the combinations of fertilizers used a range of phenotypes could be produced in the first generation offspring, with any one environment producing a uniform effect on all individuals (Durrant 1962 a, b). These environmentally induced plant lines were initially classified according to their plant weight at maturity. A wide range of plant weights were observed for the progeny of plants grown in different environments. The extreme types consist of a large type (induced by N), two to three times the plant weight of the original variety, and a small type about one half the weight of the original variety induced by P (Durrant 1958, 1962a, 1971). In between these extremes were a number of intermediate types. Durrant grew the progeny of treated plants in a number of environments including conditions which gave rise to the original differences. He found that the extreme types bred true regardless of growth conditions, while some of the intermediate types
continued to change under the inducing conditions. Thus some environments produced plants which were phenotypically different from the progenitor and were no longer capable of changing. He classified the two extreme types as large \((L)\) and small \((S)\) stable genotrophs. The intermediate lines, which are still capable of change, and the original Stormont Cirrus line, he termed plastic \((P1)\), since these could continue to be induced. Durrant’s and his colleagues’ subsequent work was primarily concerned with the true breeding \(L\) and \(S\) types and their induction from the original Stormont Cirrus line, generally referred to as \(P1\).

In addition to differing in plant weight, the \(L\) and \(S\) genotrophs also differed from \(P1\) in terms of height with \(L\) being taller than \(S\) and both \(L\) and \(S\) being shorter than \(P1\). The large increase in plant weight is largely due to a large increase in basal branching in the \(L\) types which, conversely, is suppressed in \(S\). Another phenotypically altered character is the number of hairs on false capsule septa (Durrant and Nichols 1971). \(P1\) and the \(S\) types are hairy with approximately 60 hairs per septa while the \(L\) types are hairless. An exception to this rule are two genotroph lines induced in 1961, \(L^H\) and \(S^H\), which are hairy and hairless respectively. The inheritance of these phenotypic
traits was studied in crosses between the L and S types. The genotrophs behave as genetically distinct types for all characters tested (Durrant 1971). However, the F1 plants from crosses between L and S are genetically unstable for a number of characters including plant weight, capsule septa hair number, and ribosomal DNA amount (Durrant 1962b, Durrant and Nichols 1971 and Cullis 1979). Increased variation in plant weight was observed in the F1 which is inherited by the F2 families (Durrant 1962b). In addition non-Mendelian segregation ratios were obtained for other characters such as height and acid phosphatase isozymes. The reasons for genetic instability and non-Mendelian segregation are unclear but may be due to genome destabilization in the F1 or to paramutation (Al-Saheal and Larik 1987; Durrant and Nichols 1970). Crosses between the same type of genotroph produced from two different varieties does not result in instability indicating that similar changes were probably induced in the different varieties (Evans, Durrant and Rees 1966). The lack of instability in this case also argues for a genome destabilization in L X S F1 plants. Total DNA content is inherited in an additive manner with the F1 value being the mean of the L and S values. Selection for small and large plants from the F2 segregants indicates a positive correlation
between plant weight and total DNA content (Evans, Durrant and Rees 1966). The polymorphisms in peroxidase isozyme band mobilities are inherited in a Mendelian fashion indicating that not all genotroph traits are unstable (Tyson et al. 1978; Cullis 1979).

The genotrophs also differ from Pl in a number of biochemical characters. The mobility of various isozyme systems was found to differ between the genotrophs and Pl including peroxidase, esterase and acid phosphatase isozymes (Fields and Tyson 1972, 1973a, b; Cullis and Kolodynska 1975). The changes in mobility have been explained in terms of both altered gene regulation and differential, post-translational modification (Cullis and Kolodynska 1975; Fields and Tyson 1972; Fields and Tyson 1983). The mobility shifts in the peroxidase isozymes are likely due to changes in glycosylation processing in the cytoplasm of genotroph cells (Gaudreault and Tyson 1988). An interesting feature of several of the altered characters is that the change is from a dominant to a recessive trait. This indicates that the changes are present on both homologous chromosomes (Durrant and Nichols 1970; Tyson et al. 1978).

1. **Molecular Changes in Flax Genotrophs.**
The first molecular parameter to be studied was total DNA content per nucleus (Evans et al. 1966; Evans 1968a). These studies carried out on apical meristems demonstrated that L has 16% more nuclear DNA than S with Pl having an intermediate amount. The nuclear DNA values were also studied in the apical meristems of plants growing in inducing conditions. These results indicated that during the first five weeks of induction of Pl to produce either L or S, the nuclear DNA values gradually diverged from the Pl value with an increase in L and a decrease in S. At the end of 5 weeks growth in inducing conditions the DNA difference between plants growing under conditions which will result in the L and S plants was similar to that measured in the L and S genotrophs in subsequent generations (Evans et al. 1966). Thus, the 16% total DNA difference observed between L and S genotrophs is directly associated with the divergence seen in the progenitor of the genotrophs grown in inducing conditions. This observation is consistent with the heritable nature of the induced differences which suggested that the environment had produced a change in the nuclear DNA of the plant. The experiment also illustrates an important developmental feature of the process. Since the plant germ line is
not determined until late in development somatic mutations can be transmitted genetically if incorporated into the cell lineages producing the micro- and mega-gametophytes. The period of floral determination in flax is after 5 weeks of development and therefore the transmission of these differences to the offspring is consistent with the developmental time frame for induction of DNA differences.

A number of studies have been directed towards characterizing the DNA changes observed in the genotrophs. The first studies analyzed the proportion of the genome which codes for the large ribosomal RNA genes, the 18S and 25S rDNA, in P1 and the genotrophs. The rDNA in flax is coded for by an essentially homogeneous family of 8.6 kb tandem repeats which are localized to one pair of chromosomes (Goldsbrough and Cullis 1981; Schneeberger et al. 1989). The rDNA copy number was found to be reduced by approximately 1000 copies in all of the S genotrophs whereas the L genotrophs did not differ significantly from P1 (2500 copies; Timmis and Ingle 1973; Cullis 1976). There is some variation in the amount of reduction observed in the S genotrophs with L6 representing the lowest value at 1050 and S1 the highest with 1600. Therefore, although this sequence is reduced in all of the S types
yet studied, the amount of reduction is not identical in different S types. Copy number reductions for the rDNA have also been directly associated with growth in inducing conditions indicating that part of the variation seen in total DNA content can be accounted for by variation in rDNA composition (Cullis and Charlton 1981). It is not known whether the deletion of 1000 copies of rDNA has any effect on the phenotype or any other character in the S genotrophs. An example of a phenotypic effect due to deficiencies at the rDNA locus is the bobbed mutation in Drosophila, which results from deficiencies in rDNA copy number and has pleiotropic effects on growth and development due to inadequate ribosome pools (Tartof 1975). An increase in the number of rDNA genes in a bobbed mutant can rescue this phenotype indicating that a basal number of rRNA genes is required for normal growth and development. Studies on the transcription and accumulation of ribosomes in small genotrophs indicated that ribosome biogenesis is not affected by reduced rDNA copy numbers (Ghogain et al. 1982). Thus, the significance of rDNA reductions in flax genotrophs with respect to their phenotype is not clear.

The composition of the flax genome has been studied by DNA reassociation kinetic experiments (Cullis 1981
a, b). Comparisons of the C₀T curves for L, S and P1 DNAs indicated that changes in the highly repetitive DNA and, to a lesser extent, the middle repetitive and single copy DNA fractions were largely responsible for the observed differences in total DNA content between L and S genotrophs (Cullis 1983). One component of this variation described above is the rDNA, which makes up approximately 1.5% of the genome. Random DNA clones obtained from the most variable C₀T fractions were produced which represented most of the repetitive sequences in the flax genome (Cullis and Cleary 1986). The clones were used to determine the extent of quantitative variation of these sequences among P1 and L and S genotrophs. The results demonstrated that all but one of the sequences varied significantly in the genotrophs compared to P1 (Cullis and Cleary 1986). However, only rDNA and 5S rDNA variation was correlated with a particular genotroph type (S) as described above. Variation among the other sequences was independent of the type of genotroph.

2. 5S rRNA Genes.

5S rRNA gene variation has been studied in this
dissertation and the properties of this gene which are relevant to these studies will be reviewed here. The organization and expression of 5S rRNA genes in eukaryotes has been studied in a large number of species. The 5S rRNA gene product is a 120 nucleotide RNA molecule which is an integral component of the ribosome. The eukaryotic gene is thought to be directly related to small RNA molecules in prokaryotes which are also components of ribosomes (Wolters and Erdmann 1986). The function of the 5S rRNA in ribosome biogenesis and protein translation is not clear.

5S rRNA genes in higher eukaryotes are usually present in multiple copies which are separate from the genes encoding the large ribosomal RNAs. Exceptions to this general rule exist; for example, yeast and Dictyostelium, in which the 5S genes are linked to the rDNA (Rubin and Salston 1973; Maizels 1976). The genes are organized in long tandem arrays consisting of the 120 bp gene sequence and a spacer sequence of a characteristic length alternated in a tandem fashion (Long and Dawid 1980). The chromosomal distribution of 5S genes is variable. In Xenopus the 5S rDNA is located at the telomeres of most of the chromosomes (Pardue et al. 1973). In humans the 5S genes are primarily located on the long arm of chromosome 1, in addition to several
other minor locations (Little and Bratten 1989). The of 5S rDNA in angiosperms are generally localized to fewer chromosomal locations than in animals. The genes are localized to one to three pairs of chromosomes in maize, wheat, rye, barley, pea, and *Vicia narbonensis* (Mascia et al 1981; Appels et al. 1980; Ellis et al. 1988; Knalmann and Burger 1986). An exception is the 5S rDNA in flax which is dispersed over many chromosomes (Schneeberger et al. 1989). In the gymnosperm *Pinus radiata* the 5S rRNA genes are dispersed over many chromosomes and in multiple positions per chromosome (Gorman per. comm.). Sequence determinations of 5S rRNA genes and cytoplasmic 5S rRNA from a large number of organisms indicate a high degree of sequence conservation throughout evolution. In general the gene sequence is highly conserved between species over the entire complement of genes. However, the spacer sequence is more tolerant of sequence divergence and can have little or no sequence homology between even closely related species (Long and Dawid 1980). The genes are also known to be highly methylated in a number of species (Goldsbrough et al. 1982; Ellis et al. 1988).

The expression of 5S rRNA genes has been extensively studied in the African clawed toad, *Xenopus*...
laevis. In Xenopus there are two major classes of genes which are developmentally regulated. The oocyte class is comprised of 20,000 copies per haploid genome and is expressed only in developing oocytes where they account for > 95% of the cytoplasmic 5S rRNA. The somatic genes account for 400 copies per haploid genome and are expressed at all stages of development (Wolffe and Brown 1988). The somatic and oocyte gene transcribed sequences differ by six nucleotides. The developmental expression of the oocyte class during oogenesis contributes to providing a large excess of ribosomes which will be needed for early development (Korn 1982).

5S rRNA genes are transcribed by RNA polymerase III (Pol III) and at least 4-5 transcription factors. These include the 5S rRNA gene specific transcription factor TFIIIA; the non specific factors TFIIB and C which are also involved in tRNA and other small RNA gene transcription (Wolffe and Brown 1988). TFIIIA binds internally in the 5S rRNA gene to a 50 base pair sequence termed the internal control region (ICR) (Bogehagen et al. 1980). Deletion and point mutation studies have indicated that the ICR is essential for both transcription and TFIIIA binding in vivo and in vitro (Bogehagen et al. 1980). Sequences in the spacer region of the gene are not thought to play an important
role in 5S rRNA gene transcription. Several theories have been postulated to account for the developmental regulation of oocyte and somatic genes such as transcription factor competition, differential binding of TFIIIA, chromatin effects, and oocyte-somatic tissue specific forms of TFIIIA as well as others (Wolffe and Brown 1988; Blanco et al. 1989).

Information concerning expression of 5S rRNA genes in other systems is limited. Studies in Neurospora crassa have revealed marked differences in comparison to higher eukaryotes. Neurospora has approximately 100 5S genes per haploid genome which are dispersed throughout the genome as single copy sequences (Selker et al. 1981). Transcriptional studies indicate that sequences outside of the gene in the spacer region are important for accurate and efficient transcription including TATA box like sequences (Tyler 1987).

In plants a large amount of information is available concerning organization and sequence but not expression. Studies conducted on rice 5S rRNA gene expression showed that dormant seeds contained one species of 5S RNA while 48 hr germinated seedlings contained at least 2 species of 5S rRNA (Hariharan et al. 1987). Information regarding sequences required for specific transcription are limited to sequence
comparisons rather than functional assays. Comparisons of a number of 5S rRNA sequences indicates that the region corresponding to the ICR is highly conserved in plants (Hemleben and Werts 1988). In flax the 5S rRNA genes (5S rDNA) have been analyzed at the level of genomic organization and sequence composition (Goldsbrough et al. 1981; Goldsbrugh et al. 1982). These studies indicated that flax has a large 5S gene family, with up to 58,500 copies per haploid genome, which is organized in long tandem arrays. The majority of these tandem arrays have a repeating unit of 350 base pairs consisting of a 120 nucleotide transcription unit separated by a 230 base pair spacer. This type of organization is similar to that observed in plant and animal systems previously described (Long and Dawid 1980; Hemleben and Werts 1988). Several studies have been conducted on quantitative 5S rDNA variation in flax genotrophs (Goldsbrough et al. 1981; Cullis and Cleary 1986). These studies indicated that large reductions in 5S rDNA gene copy number were present in $L^H$, $S_1$, $S^h$, $L_6$, $C_1$ and $C_3$ genotrophs. The differences are found in both large and small genotrophs and may therefore be a general feature of many genotroph lines.
C. Environmentally Induced Heritable Changes in other Plants.

A similar phenomenon to that described for flax has also been documented in *Nicotiana rustica* (Hill 1965; Perkins et al 1971). The experimental design paralleled that for flax. A highly inbred variety of *Nicotiana rustica* was grown for a single generation in combinations of N, P and K fertilizers. The self fertilized progeny were then grown in a common environment and were observed to exhibit heritable differences in height and time to flower compared to the parents. A biometrical analysis of these characters in genetic crosses of these lines indicated that the induced changes were due to genetic differences between the induced lines, since segregation of the altered characters was evident (Perkins et al. 1971). The authors concluded that the phenomena found in *N. rustica* and flax are essentially the same. They also note that the ability to undergo environmentally induced heritable change is genotype dependent, since other varieties were unable to be induced. Genotype dependence is also observed in flax where many, but not all fibre varieties are inducible and oil seed varieties are not (Durrant 1971).
Another example of environmentally induced heritable variation is that of paramutation at the R locus in maize (Brink 1956; Mikula 1967). This example is unlike that in flax and *Nicotiana rustica* in that the changes in phenotype are directed at a single gene, the R locus. This gene is involved in kernel pigmentation in the aluerone layer. Paramutation is a phenomenon whereby the ability of the R gene to produce pigment in subsequent generations can be changed by passing the R gene through a heterozygote with its $R^st$ (stippled) allele. When R is removed from the $R/R^st$ combination, a reduction in pigment is observed in the next generation (Brink 1956; Mikula 1967). These changes occur at high frequency (100%), in all of the offspring with the R gene and are stable once removed from the $R^st$ allele (Brink 1956).

This system was used by Mikula to determine if the environment is capable of modifying R gene expression in a heritable fashion (Mikula 1967). The environmental treatments were growth in either a normal light dark cycle or in constant light for a period of four weeks after which the plants were transferred to the field. It was found that growth in constant light produced a heritable reduction in R gene expression. The data also demonstrated that the period just prior to floral
initiation was important for heritable transmission of altered R gene expression. Treatments occurring after this time were ineffective in producing a heritable change. This indicates that a developmental window exists in which changes in R gene expression can be incorporated into the gametes, which is consistent with late germline determination in plants as described in section A.

The exact nature of the paramutation effect is not understood, but it is thought to involve transposable element systems which modify the expression of the R gene (Brink 1960). Whether or not any of the environmentally induced changes described above have any adaptive value has not been tested and is therefore unknown.

D. Rationale

The stability of the gene and genome from generation to generation is an underlying principle to all aspects of genetics. Transgenerational stability provides the basis for the maintenance of genetic integrity and function. Many instances of genetic instability have been described in various organisms which can result in impaired genetic function or the
creation of new genetic combinations. Thus, an understanding of the factors involved in genetic stability is of basic significance to the study of biology. This dissertation is concerned with the study of DNA polymorphisms associated with environmentally induced heritable changes in flax. The environmentally induced genetic instability in flax provides a unique opportunity to study some of the factors involved in genome stability. Although a great deal of information has been obtained concerning the nature of DNA alterations in flax genotrophs, specific mechanisms have not been identified. In addition, the specificity of DNA alterations in association with a particular genotroph phenotype is not clear. Information regarding these two points will further our understanding of the process of environmentally induced heritable change and provide testable models for future studies.

The 5S rRNA gene family was chosen as a model system for studying the nature of DNA alterations in association with environmentally induced heritable changes for several reasons. Previous characterization of this gene family demonstrated that large changes in copy number were associated with environmentally induced heritable changes in a number of flax genotrophs (Goldsbrough et al. 1981; Cullis and Cleary 1986).
These studies provided a molecular characterization of the organization and sequence of 5S rRNA genes including cloned 5S rRNA sequences which could be used as a starting point to conduct detailed studies. In addition the small size of the 5S rRNA gene (ca. 350 bp) was attractive, since the complexities of tandemly arrayed sequences could be addressed more readily than with larger repeating sequences such as the rDNA (8.6 kb).

The genotrophs $L^H$, $S1$ and $S^h$, a large type and two small types respectively, were shown to have a largely reduced 5S rRNA gene copy number in comparison to their progenitor, P1 (Goldsbrough et al. 1981). Large copy number changes are also observed in other genotrophs in which specific blocks of 5S rDNA are preferentially lost in several small genotrophs (Cullis and Cleary 1986). However, with the exception of reduced rDNA copy number, there is no direct correlation between copy number change and the type of genotroph produced from induction experiments. The experiments described in this dissertation were initiated in order to answer the following questions. Firstly, can any 5S rDNA sequences be identified which are preferentially rearranged or altered in representation in flax genotrophs? The uniformity of the genotrophs produced by a particular inducing environment from experiment to experiment
predicts that similar DNA alterations have occurred in all of these individuals. Secondly, can any 5S rDNA-associated sequences, such as junction fragments or interspersed DNA, which are altered in the genotrophs be identified? The identification and characterization of such sequences will provide insights into the nature of the DNA alterations and whether or not these events were random or associated with a particular genotroph phenotype.
Chapter 2. Methods and Materials

A. Plant Material

The inbred flax variety Stormont Cirrus was the progenitor to all the genotroph lines used in this study and is referred to as the plastic line or Pl. All of the genotrophs used in this study have been described elsewhere in detail (Durrant 1971; Cullis 1977; Cullis 1981a).

Plants were grown in a greenhouse in Metromix with fertilization every two weeks using Rapidgro. Light was supplemented to plants grown during the late fall to early spring by sodium lamps.

B. Derivation of Flax Genotrophs

All genotrophs were induced from the flax variety Stormont Cirrus, termed Pl (plastic). L refers to the large genotroph phenotype. S refers to the small genotroph phenotype. However, L6 is also a small genotroph which was derived from a large genotroph, L1 (Cullis 1977). Genotroph characteristics and induction conditions have been previously described in detail (Durrant 1962; Durrant 1971; Cullis 1977). All
genotrophs have been subsequently maintained by growth under standard greenhouse conditions and self fertilization. \(L^H\) and \(S^h\) genotrophs were derived from PI in a separate induction experiment from those described above. The induction conditions and genotrophs have been described (Durrant 1971; Cullis 1977). C3 is a small genotroph obtained from an induction experiment carried out in 1980. The characteristics of C3 have been described (Cullis 1981). Flax varieties were obtained through the USDA-SEA, North Dakota State University. *Linum grandiflorum caeruleum* was obtained from the Hortis Botanicus, Romania. Genetic crosses were performed according to standard procedures.

C. **Plant DNA Isolation**

Plant DNA was isolated from leaves and apical meristems as described by Cullis (1976), except for several flax cultivars described below. Between 1.5 and 15 grams of tissue, depending on the number of plants available, was ground in a mortar and pestle with sand in 1 ml/gram of DNA extraction buffer (3xSSC, 0.1M EDTA pH 8.0, 0.1M sodium diethyldithiocarbamate). The resulting pulp was adjusted to 2% SDS, ground briefly to mix and extracted
for 5 minutes by vigorous swirling in an Erhlemeyer flask with and equal volume of chloroform:iso-amyl-alcohol (24:1, v/v). Organic and aqueous phases were separated by centrifugation in a Sorval SS-34 rotor at 10,000 rpm (8,000 x g) for 10 min. Total nucleic acid was precipitated from the aqueous phase by the addition of 2 volumes of 100% ethanol. The precipitated DNA was spooled and then redissolved in 9.2 ml of extraction buffer and TE buffer, 1:1 (v/v). The solution was adjusted to 1.5 grams/ml CsCl by the addition of 9.0 gm CsCl and 400 ul of 10mg/ml EtBr. The DNA was banded by centrifugation at 45,000 rpm, (174,000 x g) for 36 hrs in a Beckman 50Ti rotor. The banded DNA was collected and extracted 3-4 times with an equal volume of CsCl saturated 1-butanol to remove the EtBr. The DNA was diluted twofold with TE buffer and precipitated with two volumes of 100% ethanol. DNA was recovered by centrifugation washed with 1ml 70% ethanol and dissolved in TE buffer.

DNA from flax cultivars was isolated by a modification of the method described by Murray and Thompson, 1980 (Wagner et al. 1987). Plant tissue (2-10 g) was homogenized in 180 ml of extraction buffer (50 mM Tris-HCl pH 8.0, 5mM EDTA, 720 uM spermine tetrachloride, 1.2 mM spermidine, 350 mM sorbitol, 0.1%
BSA, 10% PEG₃₅₅₀, 0.1% 2-mercaptoethanol) with a
Kinematica CH6010 polytron for 3 intervals of 10 seconds
each at maximum speed with a 10 second pause between
each on ice to cool the homogenate. The homogenate was
filtered through several layers of cheese cloth and one
layer of miracloth into 250 ml polypropylene centrifuge
bottles in an ice bath. Nuclei were pelleted by
centrifugation at 9,000 rpm at 4°C for 15 min. The
supernatant was carefully decanted and the pellet
resuspended in 4 ml chloroplast wash buffer (250 mM
EDTA, 50 mM Tris-HCl pH 8.0, 350 mM sorbitol, 720 uM
spermine tetrachloride, 1.2 mM spermidine, 0.1%
2-mercaptoethanol). The nuclei were lysed at room
temperature for 15 min in Falcon 2059 tubes by addition
of 1/5 volume of 5% sarkosyl. 1/7 volume of 5M NaCl was
added and mixed by inversion followed by 1/4 volume of
8.6% CTAB/0.7M NaCl (final CTAB concentration is 2%).
The mixture was allowed to incubate at 65°C for 12 min
followed by extraction with 5 ml CHCl₃:octanol (24:1).
Extraction is carried out by gentle inversions until a
uniform emulsion forms. The organic and aqueous phases
were separated by centrifugation in a Sorval SS-34 rotor
at 7500 rpm (5000 x g) for 10 min at room temperature.
The aqueous phase was transferred to another tube and
the DNA precipitated by the addition of 2/3 volume of
isopropanol. The precipitated DNA was collected by spooling or centrifugation. The collected DNA was washed for at least 4 hours to overnight in 20 mls of 76% ethanol/10mM NH₄ Acetate. DNA was dissolved and stored in TE buffer.

D. Restriction Enzyme Digests, Electrophoresis and DNA Transfer of DNA

Restriction enzyme digests of genomic DNA contained 2-4 ug of DNA and 10 units of enzyme per ug of DNA unless otherwise noted. Additional enzyme (10 units) was added after 2 hours of digestion and incubated at the recommended temperature for a further 2 hours. Enzymes and reaction buffers were purchased from Boehringer Mannheim Biochemical, United States Biochemical and New England Biolabs. Digested DNAs were electrophoresed on vertical 0.8% agarose gels run in tris-acetate buffer (0.04M tris-HCl, 0.001M Na₄EDTA, 0.005M Na₃acetate, pH 7.9) at 1.5 V/cm for 20 h unless otherwise noted. Denaturation of the DNA in the gels was carried out for 30 min in 1.5M NaCl, 0.5M NaOH, and then neutralized for 30 min in 3.0M NaCl, 0.5M Tris-HCl pH 6.5. The DNA was then transferred to nylon membranes in 20x SSC (1x SSC = 0.15M NaCl, 0.015M Na₃citrate) for
16 hours. Post transfer membranes were rinsed in 2x SSPE buffer (1x SSPE = 3M NaCl, 0.2M NaH$_2$PO$_4$, 0.02M Na$_2$EDTA pH 7.4).

Restriction enzyme digests of cloned DNA (plasmids or phage) were carried out as recommended by the enzyme manufacturer and fractionated on either vertical agarose gels as described above or on horizontal Tris-Borate gels (TBE, 0.089M Tris-HCl, 0.089M Borate, 0.002M EDTA). DNA gel blotting was performed as described above for genomic DNA. Washing temperatures and salt concentrations used in determining approximate homologies of the lambda 5S rDNA clones to pBI13 and for discriminating between related DNA sequences in other gel blot hybridization experiments were calculated according to Meinkoth and Wahl (1984).

E. Preparation of Megabase Size DNA

Megabase DNA was prepared from flax leaf tissue through a modification of the procedure described by Ganal and Tanksley 1989. 2 grams of fresh leaf tissue, (about 40 leaves) with no signs of senescence (generally tissue from plants which are 4 to 6 weeks old is best) was protoplasted by removing the lower epidermal cell layer with a pair of forceps and floating the leaves on
an osmoticum solution consisting of 0.5M manitol, 20mM 2[N-morpholino]-ethansulfonic acid (MES), 1.5 mM CaCl₂, 0.7mM KCl, pH 5.6 (adjusted with KOH) in a Falcon tissue culture dish. The protoplasting media was supplemented with 1% cellulase (Onozuka RS), 0.05% Pectlyase Y-23 (Seishin) 1% BSA fraction V (Sigma). The leaf, osmoticum, enzyme mix was incubated in a Percival environmental chamber on a rotating platform (50 rpm at 24°C). The progress of protoplasting was checked periodically by examination of a 1ul aliquot under a standard light microscope. When most cells (>90%) appeared to be single cells, devoid of cell wall, the protoplast mixture was passed sequentially through 100 um and 60 um sieves to remove unprotoplasted cells and vasculature. The protoplasts were then pelleted by centrifugation in an IEC table top centrifuge for 10 min at 35xg. The pellet was resuspended in 50 ml of fresh osmoticum and repelleted. This washing procedure was repeated once more during which the number of cells was estimated with a hemocytometer. After the final wash the cell pellet was adjusted to a cell concentration of 2x10⁷ cells/ml, warmed to approximately 30 to 35°C and then mixed with an equal volume 1% Incert agarose (FMC) in csmoticum cooled to the same temperature as the protoplasts. The cell agarose solution was mixed gently
and then aliquoted into precooled molds to form blocks. The agarose was allowed to solidify for 15 min. The agarose blocks were then transferred to at least three times the combined block volume of ESP (0.5M EDTA pH 9.3, 1% sarkosyl, 1mg/ml proteinase K) and allowed to incubate overnight at 50°C while rotating in a cell production roller. The ESP solution was replaced after 24 hours and allowed to incubate as above for a further 24 hours. At this point the agarose blocks are translucent and devoid of chlorophyll. The blocks were stored at 4°C in 50 ml falcon tubes wrapped in aluminum foil.

F. Restriction Enzyme Digestion of DNA in Agarose

Agarose blocks were prepared for restriction enzyme digestion by washing up to 15 blocks in 30 ml of 10mM tris-HCl, 10mM EDTA pH 8.0, 1mM PMSF (fresh) at 50°C with rotation for 90 min. This step was repeated once with and once without PMSF. The blocks were then washed twice at 50°C in TE buffer pH8.0. Blocks were then cut in half and prequillibrated in the appropriate digestion buffer supplemented with 50ug/ml BSA (molecular biology grade, Boehringer Mannheim) and 10mM DTT for 30 min at 37°C. The preequilibration buffer was then discarded and
replaced with fresh buffer and 40-50 units of restriction enzyme. Digestion was carried out overnight at 37°C with rotation. Generally additional enzyme (10 units) was added to the digests after 12 hrs. After digestion the blocks were either directly electrophoresed, preequilibrated in 1xTBE and directly electrophoresed or treated with ESP for 2 hours at 50°C and then stored in ESP at 4°C until electrophoresis. Stored blocks were preequilibrated with filter sterilized TBE prior to electrophoresis.

G. Pulsed Field Gel Electrophoresis

Field inversion gel electrophoresis (FIGE) was performed using DNAStar computer software and switching modules (DNAStar, Madison, WI). Gels were made in 0.5 x TBE buffer (1 x = 90 mM Borate, 100 mM Tris-HCl, 1mM EDTA). Gel temperature was constant at 14°C. Prerun time was 10 min. Beginning and ending forward and reverse pulses are indicated in the legends of appropriate figures. Figure legends also indicate applied voltage and length of electrophoresis. After electrophoresis of conventional sized DNA (< 200 kb) the gels were depurinated for five minutes in 0.25M HCl and then denatured, neutralized and transferred to nylon
membranes as described above. Megabase sized DNA gels were not depurinated. Instead the DNA in-gel was nicked by exposure to 302 nm UV light for 5 min to facilitate transfer of the DNA out of the gel.

Contour-Clamped Homogenous Electric Field (CHEF) electrophoresis was carried out using a CHEF-DR II system (BioRad). Electrophoresis was carried out in 0.5 x TEE with buffer cooling to 17°C. Pulse times and length are indicated in the appropriate figure legends. High Molecular Weight DNA standards, yeast chromosome markers (YNN295) and lambda ladders were supplied by Bethesda Research Labs. S.pombe chromosomes were the kind gift of Dr. Terry Magnuson.

H. DNA Probe Preparation, Labeling and Hybridization.

Radiolabeled DNA probes were made from either pure cloned DNA (plasmids or phage) or from DNA fragments isolated from preparative gels. Isolation of DNA from preparative gels was carried out by electroelution of appropriate sized DNA fragments from agarose gel slices into dialysis tubing as described by Maniatis et al. 1982. Alternatively some probes were made directly from
agarose gel fragments containing the fragment of interest. This procedure was carried out according to Feinberg and Vogelstein, 1984. The spacer probes of pRS20.7 shown in Figure 11 were isolated from a preparative restriction enzyme digest by electroelution of the appropriate bands from low melting point agarose. The 209 bp A probe was subsequently blunt end cloned into pBluescript vectors according to conventional procedures (Stratagene, La Jola Ca.). Radioactive DNA probes were generated using the random primer method with [\(^{32}\)P]dCTP or [\(^{32}\)P]dATP supplied by Amersham; 3000 Ci/mmol, (Feinberg and Vogelstein 1984). Filters were prehybridized for 4hr to overnight at 68\(^{\circ}\)C in 5 x SSPE, 0.5% dried milk powder, 1.0% NaDodSO\(_4\), 0.05 mg/ml denatured herring sperm DNA (Sigma) for Zetaprobe membranes. For other Nylon membranes prehybridization was carried out in 5 x SSPE, 0.08% bovine serum albumin, 0.08% Ficoll, 0.08% polyvinylpyrolidone, at 65\(^{\circ}\)C for 4 hours to overnight (Sambrook et al. 1989). Hybridizations were carried out in the same buffer with approximately 10\(^7\) cpm/ml of \(^{32}\)P random labeled DNA (Feinberg and Vogelstein 1984) at 68\(^{\circ}\)C for 16 hours. Filters were washed 20 min each in 0.1x SSC, 0.1%
NaDodSO₄, at 50°C unless indicated otherwise and exposed to Fuji RX film with intensifying screen at -70°C unless stated otherwise.

I. Recombinant DNA Methods

All recombinant DNA manipulations were according to Maniatis et al. (1982). Plasmid clones were generated according to standard techniques using either Bluescript or pBluescript II vectors (Stratagene). Transformation of plasmid clones into E.coli host strains (DH5 alpha, XL-1 Blue unless indicated otherwise) was accomplished by the method of Hannahan, 1985. Transformants were selected by growth on selective media containing 50μg/ml Amp, 0.08% X-gal and 0.04mM IPTG. White colonies were picked and screened by isolating plasmid DNA according to the "miniprep" alkaline lysis method of Birnboim and Doly, 1979 with the following modifications; lysozyme treatment of cells was eliminated; second, after precipitation of the plasmid the DNA was washed with 70% ETOH and further purification by phenol/chloroform extraction was generally not necessary. The alkaline lysis method was also used to isolate plasmid DNA on a large scale (Maniatis et al. 1982).
Phage DNA was isolated by the glycerol step gradient method as described in Maniatis et al., 1982. Recombinant EMBL4 restriction maps were determined through the combined use of single, double and triple restriction enzyme digests in addition to the end labeling procedure described by Rackwitz et al., 1984. All plasmid subclones are in Bluescript or pBluescript II vectors (Stratagene).

Construction of an EMBL4/L^H genomic library was carried out essentially as described by Sambrook et al., 1989. Briefly, purified L^H leaf DNA was partially digested with Sau3A and size selected for 15-20 kb size fragments on a 1.2-5.0 M linear NaCl gradient, centrifuged at 38,000 rpm for 4 hrs, 20^oC in a Beckman SW 40.1 rotor. Size selected fractions were pooled and ligated to BamHI/SalI digested EMBL4 vector DNA at 14^oC for 16 hrs. The resulting recombinants were packaged using Gigapack plus packaging extracts (Stratagene) and the phage introduced into the E. coli host strain VCS-257 for screening (Stratagene) as described by Maniatis et al., (1982). All lambda phage libraries were screened directly in duplicate, without amplification, by the method of Benton and Davis 1977 as described in Maniatis et al. 1982. Schliecher and Schuel nitrocellulose filters were prehybridized in 3 x
SSC, 0.08% bovine serum albumin, 0.08% Ficoll, 0.08% polyvinylpyrrolidone, at 65°C for 4 hours (Denhardt, 1966). Hybridizations were carried out in the same buffer with approximately 10^7 cpm/ml of 32P labeled DNA (Feinberg and Vogelstein 1984) at 65°C for 16 hours. After hybridization the filters were washed 2 x 1 hour in 0.1 x SSC, 0.1% SDS at 50°C. In other cases as stated in the text the filters were post hybridization washed in 0.1 x SSC, 0.1% SDS at high stringency, i.e 65-68 °C.

Construction of lambda ZAPII Pl and L^H sub-genomic libraries was carried out as described by the manufacturer (Stratagene). Briefly, the COS sites of 5ug of undigested ZAPII were ligated in 50mM Tris-HCl pH 7.5, 7mM MgCl_2, 1mM dithiothreitol (DTT), 1mM ATP pH 7.5, and 4 Weiss units of T4 DNA ligase for 1 hour at room temperature followed by 16 hours at 4°C. The T4 DNA ligase was inactivated by heating at 65°C for 15 min and the buffer adjusted to 100mM NaCl, 50mM Tris-HCl pH 7.5, 10mM MgCl_2, 1mM DTT for EcoRI digestion. The COS ligated DNA was digested with 20 units of EcoRI for 2.5 hours at 37°C. After digestion the reaction was stopped by the addition of 1ul of 0.5M EDTA, incubated at 65°C for 15 min and extracted once with phenol/chloroform and once with chloroform. The DNA was the precipitated by
the addition of 2 volumes of 100% ethanol and incubation overnight at \(-20^\circ\text{C}\). The precipitated DNA was recovered by centrifugation at 14,000 rpm for 30 min at \(4^\circ\text{C}\), resuspended in 10 ul 10mM Tris-HCl pH 8.0 and dephosphorylated with calf intestinal alkaline phosphatase (Promega) according to the procedure described by Maniatis et al., 1982. The resulting vector DNA was recovered by precipitation, washed with 70% ethanol and resuspended in 5ul of TE buffer. Prepared lambda ZapII cloning vector was stored at \(-20^\circ\text{C}\). Recovery and digestion were tested by agarose gel electrophoresis with lambda DNA concentration standards (BRL). For construction of the \(L^H\) library 2 ug of \(L^H\) genomic DNA was digested to completion with \(EcoRI\). The DNA was purified away from the restriction enzyme by phenol/chloroform extraction and ethanol precipitation. Ligation of 1 ug of lambda ZAPII arms and 0.2 ug of \(EcoRI\) digested \(L^H\) DNA was carried out in a total volume of 5ul for 1 hour at room temperature followed by 16 hrs at \(4^\circ\text{C}\). The ligated DNA was packaged into phage particles using Gigapack II Gold packaging extracts at \(22^\circ\text{C}\) for 2 hours (Stratagene). The resulting library was stored in 500ul of SM buffer (100mM NaCl, 8.1mM MgSO_4, 50mM Tris-HCl pH 7.5, 2% gelatin) and one drop of chloroform. The library was
titered on PLK-F' host cells. The titer was determined to be $2.3 \times 10^7$ pfu/ml with a background of 100 pfu/ml. Two genome equivalents of phage were infected into PLK-F' and plated on a 24.5 x 24.5 x 0.2 culture dish (Nunc). Plaques were allowed to form from 6 hr to overnight and plaque lifts taken. The filters were screened as described above by hybridization with a $^{32}$P random labeled 706 bp insert of pRS20.7 after prehybridization with 10 ug of sonicated pBluescript to eliminate background. Approximately 60 positives were identified and picked and stored in 50 ul of SM buffer with 1 ul of chloroform.

The construction of ZAPII P1 subgenomic libraries was carried out essentially as described above for L$^H$. except for the following modifications. Total P1 genomic DNA (200ug) was digested with EcoRI to completion and fractionated on a 1.2-5.0 linear NaCl gradient as described above for EMBL4 library construction. Fractions were selected in two size ranges, 2.0-15 kb (P1 12/13) and 5.0-20 kb (P1 11), in order to construct two libraries which would be biased toward larger and smaller size DNA fragments in an effort to facilitate cloning of different sized molecules. In separate reactions 700ng of P1 12/13 DNA and 500ng of P1 11 DNA was ligated to 1ug of \ZAPII
vector arms. The ligations were packaged and titered on PLK-F’ host bacterial cells. The titers of the resulting Pl 11, and Pl, 12/13 libraries were $2.95 \times 10^7$ and $2.5 \times 10^7$ pfu/ml respectively. The Pl 11 library was also plated on Sure bacterial host strain which is recombination deficient (Stratagene). The genotype with respect to recombination deficiency is mcrA, delete(mcrBC-hsdrms-mrr)171, recB, recJ, sbcC, umuC::Tn5 and uvrC (Stratagene). The Pl 12/13 and Pl 11 libraries were screened as above for the $L^H$ library. Thirty five positives were identified and picked for Pl 12/13 and greater than 80 identified for Pl 11 (infected into Plk-F’ host strain). The Pl 11 library infected into Plk-F’ was rescreened with the A spacer probe and 5 strong positives were isolated. The Pl 11 library infected into the SURE bacteria host was also screened with the A spacer probe and 2 strong positives were isolated. The Lambda ZapII insertion vector allows fragments up to 10 kb to be cloned. The cloning experiments performed with this vector did not limit the vector to one EcoRI insert. Therefore multiple EcoRI fragments could be cloned.

Plasmid rescue from lambda Zap II was performed exactly as described by the manufacturer (Stratagene).
Rescued phagemids were infected into 200 ul of mid log phase XL1 Blue host strain resuspended in 10 mM MgSO₄. Confirmation of rescue for was performed on 5 independent colonies by isolation of plasmid DNA followed by EcoRI restriction enzyme digestion and gel electrophoresis to visualize the band sizes.

J. Copy Number Reconstructions

Copy number reconstructions were carried out by diluting pRS20.7 plasmid DNA to 10, 50, 100, 200, 500, 1000, and 2000 copies. The samples were digested with Bam HI. Digestion was terminated by addition of 1 ul of 0.5M EDTA and loading buffer after 1 hour and 1 ug of sonicated herring sperm DNA was added as carrier. These samples were electrophoresed on a vertical 1% agarose gel along with 1 ug Sca I digested samples of P1, L^H, S^h and L. grandiflorum caeruleum DNA. Sca I digests 99% of the pRS20.7 genomic sequence into 353 and 706 bp bands. The gel was subsequently blotted to a nylon membrane and hybridized to random labeled 706 bp insert of plasmid pRS20.7. Autoradiography was performed using preflashed Fuji RX X-ray film. The resultant autoradiograms were scanned on a Shimadzu CS-930 Dual Wavelength TLC Scanner and the peak areas of standards and genomic samples
which were in the linear response of the film were quantitated. In order to make more accurate determinations of copy number changes between the genomic samples, the filters were also hybridized with a cloned sequence shown to be a single copy sequence in Pl, L\textsuperscript{H}, S\textsuperscript{h} and L. grandiflorum caeruleum (p2010.1). The autoradiograms were scanned for this sequence hybridization and the copy numbers normalized to the Pl value to correct for any error in genomic DNA loading. Reconstuctions of RFLP band copy numbers were carried out essentially as above except the genomic DNA was digested with Eco RI.

K. DNA Sequencing

DNA fragments to be sequenced were cloned into the appropriate cloning site of either Bluescript SK (+,-) or pBluescript II SK (+,-) vectors as described above (Stratagene). Single strand DNA was prepared from the plasmids according to the manufacturers instructions and used as template for sequencing reactions. DNA sequence was obtained by using T7 DNA polymerase and the chain terminating dideoxy nucleotides according to the manufacturers instructions (United States Biochemical;
Tabor and Richardson 1987). Primer sites in the plasmid vector were utilized for initiating DNA synthesis. In addition two 5S rRNA gene specific primers were synthesized on an Applied Biosystems 381A DNA synthesiser for sequencing flax 5S rRNA genes. The primer sequences are 5’-CATACCAGCACAATGCACCGG-3’ and 5’-CGGTGCATTAAGCTGTTATG-3’. Several 706 bp repeats from the 5S-20 EMBL 4 genomic clone were subcloned into pBluescript vectors (Stratagene) at the BamHI restriction site according to standard procedures (Sambrook et al. 1989). Single strand DNA was prepared for two clones according to the manufacturer’s instructions (Stratagene). Positives were sequenced using T7 DNA polymerase and chain terminating dideoxy nucleotides (United States Biochemical). Sequence for the opposite strands was obtained by double strand sequencing.

The following list contains the names and origins of the 5S rDNA clones used for sequence analysis in Chapter 3A3.

<table>
<thead>
<tr>
<th>Clone</th>
<th>Size (bp)</th>
<th>Origin</th>
</tr>
</thead>
<tbody>
<tr>
<td>pBG13e</td>
<td>341</td>
<td>Most common repeat in pBG13.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>pBG13 contains; 9 repeats of ca. 341 bp and one 362 bp repeat in</td>
</tr>
<tr>
<td>Clone</td>
<td>Size (bp)</td>
<td>Origin</td>
</tr>
<tr>
<td>--------</td>
<td>-----------</td>
<td>--------</td>
</tr>
<tr>
<td>pRS18</td>
<td>342</td>
<td>EMBL4 5S-18 (Chapter 3A,1)</td>
</tr>
<tr>
<td>pRS13</td>
<td>344</td>
<td>EMBL4 5S-13 (Chapter 3A,1)</td>
</tr>
<tr>
<td>p3</td>
<td>356</td>
<td>EMBL4 5S-3 (Chapter 3A,1)</td>
</tr>
<tr>
<td>pRS20.7</td>
<td>706</td>
<td>EMBL4 5S-20 (Chapter 3A,1)</td>
</tr>
<tr>
<td>20.7 5S1</td>
<td>353</td>
<td>pRS20.7</td>
</tr>
<tr>
<td>20.7 5S2</td>
<td>353</td>
<td>pRS20.7</td>
</tr>
<tr>
<td>p651</td>
<td>651</td>
<td>lambda ZapII Pl RFLP rescued plasmid, p24 (Chapter 3C, 7)</td>
</tr>
<tr>
<td>651 5S1</td>
<td>299</td>
<td>p651</td>
</tr>
<tr>
<td>651 5S2</td>
<td>352</td>
<td>p651</td>
</tr>
<tr>
<td>p6xa2</td>
<td>360</td>
<td>EMBL4 5S-6 (Chapter 3A,1)</td>
</tr>
</tbody>
</table>
Chapter 3. Results

A. Molecular Characterization of 5S rRNA Gene Organization in Flax.


The previous isolation of 5S rRNA genes from flax was based on actinomycin D density gradient centrifugation. This method allows for the enrichment of sequences with high G-C contents from the rest of the genome based on density (Hemleben et al. 1977). 5S rRNA gene sequences tend to have high GC contents and therefore are amenable to isolation by this technique. The drawback of this approach is that sequences of interest which have a lower GC content will not be enriched for by this procedure. In order to obtain a "representative" set of 5S rDNA containing sequences a genomic cloning strategy was utilized. A genomic library in the lambda cloning vector EMBL4 was constructed using total leaf DNA from L^H, which had been partially digested with Sau3A and subsequently size selected for fragments between 15 and 20 kb. The
library was screened with pBG13 and 25 positive clones were identified and selected. Eleven clones which together displayed widely variant signal intensities independent of their relative plaque sizes were plaque purified and restriction mapped. Regions of 5S rDNA homology were determined by Southern blot analysis (Southern 1975) using pBG13 as the 5S rDNA probe. The results of the analyses are shown in Figure 2. The clones are grouped based on restriction site similarity and relative homology to pBG13 (homology data presented below in section 3A2). This classification does not consider junction fragment similarity between the clones. The data are summarized in Table I.

There is considerable heterogeneity among the 5S clones at the restriction site level. The majority of the 5S rRNA genes in flax have previously been shown to be organized in tandem arrays consisting of a 341 and a 362 base pair repeat unit. The 362 bp repeat is a minor class containing a 21 bp duplication in the spacer region. The repeat unit can be defined by a BamHI site near the 5' end of the transcription unit (Goldsbrough et.al., 1982). Initially each clone was digested with BamHI to determine if the clones were composed of BamHI repeat units. Only five of the eleven clones contained 350 bp BamHI restriction fragments with pBG13 homology
Figure 2. Restriction maps of 5S genomic clones.
Groups 1, 2, 3, 4, and 5 refer to classification of 5S clones as described in results section (A,1) and Table I. Blackened areas indicate regions of pBG13 homology with a defined repeat unit. Hatched areas indicate regions of pBG13 homology in the absence of a definable repeat unit. Group 1 and 2 contain 350 bp BamHI 5S DNA homologous fragments (refer to Figure 4 panel A). Group 3 contains 350 bp XhoI repeat unit. Group 4 contains 700, 1000, and 1700 bp BamHI repeat units. Group 5 contains clones without a definable repeat unit. The restriction sites for 350 bp repeat units in groups 1, 2 and 3 are not shown due to space constraints.
Group 1

Group 2

Group 3

Group 4

Group 5

λ clone

5S-18

5S-13

5S-17

5S-91

5S-1

5S-3

5S-20

5S-12

5S-6

5S-11

5S-42

= Bam HI  = Eco RI  = Hind III
**TABLE I**

Flax 5S genomic clone classification

<table>
<thead>
<tr>
<th>Group</th>
<th>Clone Designation</th>
<th>Enzyme/Repeat Unit</th>
<th>Homology to pBG13</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>EMBL4 5S-18</td>
<td><em>BamHI</em>/350 bp</td>
<td>+++</td>
</tr>
<tr>
<td>2</td>
<td>EMBL4 5S-13, -17, -91, -1</td>
<td><em>BamHI</em>/350 bp</td>
<td>++</td>
</tr>
<tr>
<td>3</td>
<td>EMBL4 5S-3</td>
<td><em>XhoI</em>/350 bp</td>
<td>+</td>
</tr>
<tr>
<td>4</td>
<td>EMBL4 5S-20, -12</td>
<td><em>BamHI</em>/700, 1000, 1700 bp</td>
<td>+</td>
</tr>
<tr>
<td>5</td>
<td>EMBL4 5S-11, -42, -6</td>
<td>undefined*</td>
<td>++</td>
</tr>
</tbody>
</table>

*a* Groups are defined based on restriction site and nucleic acid hybridization homology to pBG13.

*b* Clone numbers refer to primary recombinant EMBL4 phage isolated using pBG13 as a probe. See Fig. 2.

*c* Sizes are approximations derived from gel standards.

*d* Homologies are approximations based on initial hybridization intensity and retention of probe, through increasing stringency washes. + Indicates relative homology to pBG13 as determined from Fig. 3. Note that homologies can be due to either pBG13 spacer or 5S rRNA transcription unit sequence.

*e* Digestion with enzymes other than *BamHI* does produce 5S rDNA homologous bands at 350 and 700 bp, in 5S-42 and 5S-11. However there are in addition other bands which are not multimers of 350 or 700 bp and hybridize with the pBG13 insert probe.
Figure 3. EMBL4 5S Clone Homology to pBG13. A;
ethidium bromide stained gel of lambda 5S
rDNAs restricted with BamHI (lanes 2-9, and
11), EcoRI (lane 10), and BamHI/ EcoRI (lane
12). Lane 1 is the 342 bp insert of pBG13.
M is X174 DNA standard. The following
lanes contain the indicated clone digests;
Lane 2, 5S-18; 3, 5S-13; 4, 5S-17; 5, 5S-31;
6, 5S-1; 7, 5S-3; 8, 5S-20; 9, 5S-12; 10,
5S-6; 11, 5S-11; 12, 5S-42. Standard DNA
sizes are from lambda DNA digested with Eco
RI and Eco RI/ HIII. B: The gel shown in A
was Southern blotted and hybridized with a
$^{32}$P random labeled insert of pBG13. The
filter was washed at low stringency 0.1 x
SSC, 0.1 x SDS, 45°C (Tm - 27°C) and
autoradiographed. C: The same filter was
rewashed at high stringency 0.1 x SSC, 0.1 x
SDS at 68°C (Tm - 3°C) and autoradiographed.
D: Autoradiograph of the same filter after
being stripped of pBG13 probe, and reprobed
with the 350 bp Bam HI fragment of 5S13 (Tm
- 3°C). Arrows in A, B, C, and D indicate the
position of 350 bp Bam HI fragments.
(Figure 3, a; lanes 2-6). Each of the five clones contains multiple copies of the BamHI repeat (determined by ethidium fluorescence and partial digestion, Figure 3, A) indicating that these clones are organized in tandem arrays of ca. 350 bp repeating units as predicted from previous studies (Goldsbrough et al. 1981). Since 350 bp is the approximate size of the major repeat unit in flax this figure will be used rather than 342 for the remainder of the text. The remaining clones did not produce 350 bp BamHI repeat units and thus represent sequence variants with respect to pBG13 (Figure, 3, A). Note that there is some size difference between the group 1 and 2 clones (Figure 3, A, compare lanes 1,2 with 3-6). Length heterogeneity in flax 5S rDNA sequences has previously been described (Goldsbrough et al. 1981). Clone 5S-3 lacks a 350 bp BamHI repeat unit. However it does produce a repetitive 350 bp band when digested with XhoI. Two BamHI sites are present within the 5S homologous portion of clone 5S-3 indicating some heterogeneity in this tandem array. Clone 5S-20 contains approximately 7 repeats of a 706 bp 5S rDNA homologous band and may represent a dimer of two 350 bp repeats which has lost a BamHI site and subsequently been amplified. Clone 5S-12 is of particular interest since it is organized as an intradispersed array of 700,
1000, and 1700 bp BamHI restriction fragments containing homology to pBG13. The remaining clones, 5S-11, -42, and -6 do not have identifiable 350/700 bp repeats.

All of the clones contain both 5S rDNA homologous and non-5S rDNA homologous sequences with the exception of clone 5S-12. The non-5S rRNA gene sequences presumably represent junction fragments between 5S rDNA clusters and non-5S rDNA flanking DNA. The genomic organizations of the junction fragments as determined by Southern hybridization are identical to the restriction maps shown in Figure 2 and are therefore not the result of rearrangements occurring during the cloning procedure (Southern 1975, Figure 2). The junction fragments from different lambda clones do not show any common organization based on the restriction maps. The flanking regions have been determined to be mainly low copy number in terms of sequence complexity (5S-18, -13, -17, -91, -1, -3, -20, -6, -42) with 5S-11 representing a repetitive sequence junction fragment (data not shown). Group 1 and 2 clones are similar in that their junction fragments are all low copy number. This observation is not surprising since 40% of the flax genome is single copy DNA. Clone 5S-42 is the only clone which contains 5S sequence flanked by non-5S sequence on either side. Thus, at the level of
restriction site analysis the 5S rRNA genes can be juxtaposed to different types of sequences and special junction sequences do not appear to be present. However, a junction sequence context which is not observable at this level of analysis may be present. The features described above suggest that both sequence heterogeneity and the presence of major and minor sites of 5S rRNA gene clusters contribute to the dispersed nature and relative intensities of the in situ determination of 5S gene localization (Schneebberger, Creissen, and Cullis 1989).

2. **Lambda 5S rDNA Clone Homology to pBG13.**

The extent of sequence diversity revealed by restriction site polymorphisms among the different groups of 5S rDNA presented above was investigated by determining the relative homologies of each of the clones to the original flax 5S rRNA gene clone pBG13 (Goldsbrough et al. 1981). This was done by assaying the stability of hybrid duplexes between the labeled insert of pBG13 and the 5S rDNA genomic clones via Southern blot analysis (Southern 1975). Figure 3, A shows an ethidium bromide stained gel containing each of the 11 genomic clones digested with the enzymes
indicated in the legend to Figure 3. Lane 1 contains the 350 bp Bam HI insert of pBG13 run as the homology control. This fragment was also used as the $^{32}$P labeled probe. Each of the remaining tracks was loaded such that the 350 base pair 5S rDNA homologous restriction fragments from each clone would be present in approximately equal proportion to each other and the pBG13 Bam HI control fragment. Following hybridization the filter was washed at low stringency ($45^\circ$C, 0.1 x SSC, 0.1% SDS, Tm - 27$^\circ$C) and autoradiographed. It is evident that not all of the clones hybridize to pBG13 to the same extent (Figure 3). Since all of the restriction fragments are in approximately equal stoichiometry on the filter, this difference can be attributed to different extents of homology between the 5S rRNA gene homologous portions of the clones and pBG13. The filter was taken through a series of washes of increasing stringency. After each wash the filter was assayed for retention of labeled probe by autoradiography. The difference in homologies is clearly seen when the filter is rewashed at high stringency. Figure 3, C shows the same filter rewashed at 68$^\circ$C in 0.1 x SSC, 0.1 x SDS, Tm - 3$^\circ$C). This level of stringency permits approximately 3% nucleotide mismatch (Meinkoth and Wahl 1984). The results thus far
subdivide the 11 clones into at least 5 groups of pBG13 homologous sequences based on restriction site analysis and hybrid duplex homology. Notably, several clones possess low sequence homology to pBG13, (5S-3, -20, -12, -6, -11 and -42) and may represent pseudogenes. In addition, clone 5S-11 appears to contain sequences which are not uniform in their homology to pBG13. Several bands are no longer detected by the probe, while one 2.4 kb band persists at high stringency (Figure 3, compare B and C, lane 11). Group 1 and 2 clones, while both possessing a 350 bp BamHI repeat unit, are not equally homologous to pBG13. Only 5S-18 hybridizes to pBG13 with equal intensity to pBG13 350 bp control fragment (Figure 3, A). The group 2 clones, 5S-91, -1, -17, and -13, initially hybridize with approximately one third the intensity of group 1. This could be explained simply if the homology between group 1 and group 2 was limited to the 5S rRNA gene transcription unit and did not include the spacer sequence. This situation would be similar to that previously described in Xenopus laevis and wheat. In Xenopus, the oocyte and somatic 5S gene spacers are completely different except for a short conserved sequence near the 5’ and 3’ end of the gene (Peterson et al. 1980). In wheat, two repeat units have been identified which have similar but divergent spacers
(Gerlach and Dyer 1980). To determine if the group 2 clones were more similar to each other than to pBG13 the filter was stripped of probe and reprobed with the 350 bp BamHI fragment of clone 5S-13 (Figure 3, panel D). The comparable levels of hybridization among the group 2 clones indicates they are more similar to each other than to group 1 and thus constitute a class of 5S rRNA genes distinct from pBG13 type sequences (group 1) in flax.

The homology between 5S-20 and 5S-12 was determined by reprobing the same filter with a subclone of the 700 bp 5S homologous portion 5S-20 (pRS20.7). These clones are approximately 90% homologous based on Tm indicating that they are closely related but have diverged significantly from each other. The degree of variation in repeat unit size and sequence homology among flax 5S genes is striking in comparison to other plants. Studies conducted on wheat, yellow lupin, Vigna radiata, Matthiola incana, pea and soybean all indicate that 5S genes are very homogeneous with respect to sequence composition (Gerlach and Dyer 1980; Rafalski et al., 1986; Hemleben and Werts 1988; Ellis et al. 1988; Gottlob-Mchugh et al. 1990). In addition, previous studies conducted with flax suggested a high level of homogeneity for the 5S rDNA (Goldsbrough et al. 1982).
Thus the organization and variation reported here is novel. This heterogeneity presents several problems in relation to proposed mechanisms for the maintenance of homogeneity of tandemly arrayed repetitive sequences which will be discussed in Chapter 4.

The genomic identity of each of the 5S rDNA clones was verified by individually probing \( L^H \) BamHI genomic blots with the 5S rRNA gene homologous portions of each of the clones. After hybridization the filters were washed at high stringency to be able to distinguish between the different groups described above (\( T_m - 2^0 \), chapter 2, H). The organization of the 5S rRNA sequences within the 11 phage clones was consistent with the hybridization profiles shown in Figure 4. For example, digestion of \( L^H \) genomic DNA with XhoI produces a 350 bp band in addition to a 350 bp ladder when hybridized with the 5S rDNA portion of 5S-3 (Figure 4, lane c). The multimers of 350 bp which form the ladder pattern in this lane are either due to loss of the Xho I restriction site or to methylation of the internal cytosine in the recognition site. The major bands detected for group-5 are the same size as the fragments used for the probe in each case (Figure 4, lanes g, h and i).
Figure 4. DNA blot analysis of Group 1 - 5 5S rDNA genomic clones. BamHI $L_H^H$ genomic DNA blot (lanes a,b and d,e,g,h,i), or XhoI DNA blot (lane c). Lane f is 5S-12 cut with BamHI and run on the same gel as a control for 5S-12 hybridization. Individual lanes from the filter were hybridized with (a) pBG13 350 bp insert, (b) 350 bp BamHI fragment of 5S-13, (c) 350 bp XhoI fragment of 5S-3, (d) pRS20.7 insert (700 bp BamHI fragment of 5S-20), (e) pRS1214 (14.0 kb EcoRI insert of 5S-12), (f) as in (e), (g) pRS119.0 (9.0 kb EcoRI fragment of 5S-11, (h) 3.7 kb EcoRI/BamHI fragment containing 5S DNA homology in 5S-42, (i) 3.2 kb EcoRI fragment containing 5S rDNA homology of 5S-6. Washes were carried out as described in methods and materials section D. 3hr exposure time for lanes a and b. 168 hr exposure time for lanes c-e, g-i. 1 hr exposure for lane f. 350 bp indicates the position of group 1, 2, and 3 monomers. Arrows indicate the point of divergence between the group 1 and 2 patterns.
Comparison of the hybridization profiles of group 1 and group 2 clones shows that under high stringency these two classes can be differentiated. 5S-18 produces a hybridization pattern identical to that of pBG13. In addition to hybridizing to a 350 bp monomer band it also hybridizes to multimers of 350 bp. These bands are due to either cytosine methylation at the BamHI site or to loss of the BamHI site resulting in multimers of 350 bp. Incomplete digestion is unlikely as the patterns do not change on further digestion. Although the group 2 clones produce identical hybridization patterns between themselves this pattern differs from that of the group 1 clones (Figure 4). The patterns of hybridization for groups 1 and 2 are broadly similar with respect to relative band intensity and extent of multimers identified. However, the regular spacing of the 5S rDNA ladder is disrupted at the octamer band in group 1 whereas in group 2 the ladder is not interrupted until the 11-mer band. The fact that differences can be observed in banding patterns is consistent with the group-1, group-2 homology differences identified in Figure 3. The hybridization patterns of the rest of the clones are also consistent with the genomic clones.
however the relatively weak hybridization signals indicate that their respective copy numbers are far lower than those of group 1 and 2 (Figure 4).

3. **Sequence Analysis of Groups 1 - 5 5S rRNA Genes.**

Most 5S rRNA gene families studied to date have reported low levels of intraspecific sequence divergence in the range of 0.3% to 5% between individual members of the same class of genes (Ellis et al. 1988; Fedoroff and Brown 1979; Scowles et al. 1988). Comparison of members of the oocyte and somatic 5S rRNA gene families in *Xenopus laevis* showed high conservation of the coding sequence (ca. 6 bp changes). However the spacer region of the two classes were essentially unique from each other (Peterson et al. 1980). In contrast most plant 5S rRNA genes studied to date do not exceed 3% variation across the gene and spacer within the same species (Goldsbrough et al. 1982; Scowles et al. 1988; Gottlob-McHugh 1989). The general homogeneity of repetitive DNA has been ascribed to homogenization of the sequences by molecular mechanisms collectively termed molecular drive (Dover 1982). Models involving unequal exchange and gene conversion have been proposed
for the maintenance of repetitive gene families, although little experimental data is available to support these theories (Smith 1976; Dover 1982). The hybridization results presented in section 3A2 suggested a high degree of sequence variation between the various groups of flax 5S rDNA (Figure 3). These results are in direct contrast to previous studies in flax which showed a high level of sequence homogeneity among different repeat units of 5S rDNA (0.7% sequence divergence; Goldsbrough et al. 1982). Therefore the DNA sequence for at least one member of each 5S rRNA group was determined to define the range of sequence variation and which groups are likely to be expressed based on comparison to other angiosperm 5S rRNA gene sequences. A list of the groups and clones which were sequenced is presented in Chapter 2K. In addition to the clones described above a clone related to group 4, p651, was also sequenced. Repeat ‘e’ from pBG13 was used for the comparisons described below since it is the most common sequence form in pBG13 and will be referred to as pBG13e in this section (Goldsbrough et al. 1982).

In order to facilitate a meaningful analysis of sequence variation between groups the 5S rRNA gene sequences were compared in pairwise fashion at three levels. First, the coding plus spacer regions (referred
to as gene repeat) of each group were compared to
determine the amount of sequence divergence over the
entire gene repeat. Second, the coding regions were
compared to assess the conservation of the 5S rRNA
coding sequence. The coding regions were compared to the
angiosperm 5S rRNA consensus sequence to determine which
cloned is most homologous to other plant 5S rRNAs.
Third, the spacer regions were compared to determine if
there is any difference in the conservation of the
spacers as opposed to the coding regions. The gene
sequences were set up such that the first 120
nucleotides corresponded to the coding region (as
deduced from comparison to other 5S rRNA sequences),
followed by the spacer region. Since the cloning site
for many of the 5S rDNA repeat subclones occurs in the
coding region (BamHI, SpeI) the coding sequence
represents a circular permutation at the cloning site to
derive a single coding sequence. The entire DNA sequence
of each clone is contained in appendix A. Selected
comparisons are shown to illustrate specific points.

Pairwise comparisons were performed using the
method of Wilbur and Lipman 1983, on the Align program
distributed by DNASTAR software (DNASTAR Madison,
Wisconsin). The program provides a similarity index in
addition to an optimal alignment of the two sequences
being compared. The similarity index is the number of matched bases divided by the sum of matched bases, unmatched bases and the number of gaps (Dayhoff, 1979). The number is expressed as a percentage and is very useful in that it is not length dependent.

The gene repeat comparisons are presented in Table II. The homology values obtained from comparison of pBG13 to the other 8 flax 5S rRNA gene sequences generally reflects the hybridization behavior of these sequences to pBG13e shown in Figure 3. Group 1 clone pRS18 is most similar to pBG13 while homology decreases to a low of 65% in the second gene repeat of clone 20.7 (5S2). The initial grouping of the clones based on the hybridization and repeat structure of the EMBL4 genomic clones is in general agreement with the sequence data presented in Table II. The clones listed in group 5 possess varied homologies to pBG13 based on the results obtained from Figure 3. The sequenced subclone of 5S6 (p6XA2) appears to contain the lowest homology to pBG13 of any of the clones in group 5.

Further consideration of Table II indicates that there are several levels of relationship between the clones. Groups 2 through 5 are more homologous to pBG13 than they are to clone 18 supporting the idea that all of the clones in this sampling of 5S rRNA genes are
<table>
<thead>
<tr>
<th>Group</th>
<th>CLONE</th>
<th>1</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>4</th>
<th>4</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>pBG13e</td>
<td>100</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>pRS18</td>
<td>94.0</td>
<td>100</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>pRS13</td>
<td>79.0</td>
<td>78.4</td>
<td>100</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>p3</td>
<td>72.0</td>
<td>71.4</td>
<td>70.4</td>
<td>100</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>20.7 (5S1)</td>
<td>70.2</td>
<td>66.0</td>
<td>67.0</td>
<td>74.2</td>
<td>100</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>20.7 (5S2)</td>
<td>65.0</td>
<td>64.0</td>
<td>63.0</td>
<td>75.4</td>
<td>77.0</td>
<td>100</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>651 (5S1)</td>
<td>68.3</td>
<td>65.0</td>
<td>65.0</td>
<td>73.2</td>
<td>84.3</td>
<td>82.4</td>
<td>100</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>651 (5S2)</td>
<td>67.4</td>
<td>64.4</td>
<td>67.3</td>
<td>73.0</td>
<td>78.1</td>
<td>88.0</td>
<td>82.0</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>p6XA2</td>
<td>68.0</td>
<td>68.0</td>
<td>66.0</td>
<td>81.0</td>
<td>70.4</td>
<td>74.0</td>
<td>76.0</td>
<td>73.2</td>
<td>100</td>
</tr>
</tbody>
</table>
### TABLE III
5SrRNA Gene Coding Region Pairwise Comparison

<table>
<thead>
<tr>
<th>Group</th>
<th>CLONE</th>
<th>1</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>4</th>
<th>4</th>
<th>4</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>pBG13e</td>
<td>100</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>18</td>
<td>97.0</td>
<td>100</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>13</td>
<td>96.0</td>
<td>93.3</td>
<td>100</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>82.0</td>
<td>76.2</td>
<td>81.0</td>
<td>100</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>20.7 (551)</td>
<td>78.3</td>
<td>75.0</td>
<td>75.0</td>
<td>79.2</td>
<td>100</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>20.7 (552)</td>
<td>74.2</td>
<td>72.0</td>
<td>71.0</td>
<td>78.3</td>
<td>83.0</td>
<td>100</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>651 (551)</td>
<td>78.0</td>
<td>75.0</td>
<td>74.0</td>
<td>82.0</td>
<td>91.0</td>
<td>85.0</td>
<td>100</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>651 (552)</td>
<td>73.1</td>
<td>71.0</td>
<td>70.0</td>
<td>77.0</td>
<td>82.0</td>
<td>93.3</td>
<td>83.3</td>
<td>100</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>6XA2</td>
<td>76.0</td>
<td>76.0</td>
<td>76.0</td>
<td>83.0</td>
<td>71.0</td>
<td>77.0</td>
<td>77.0</td>
<td>73.0</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>Group</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>5</td>
<td>6X1A2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>-------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>pbG13e</td>
<td>100</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>18</td>
<td>92.3</td>
<td>100</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>13</td>
<td>69.2</td>
<td>70.1</td>
<td>100</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>67.3</td>
<td>64.4</td>
<td>65.0</td>
<td>100</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>20.7 (651)</td>
<td>66.0</td>
<td>62.0</td>
<td>62.4</td>
<td>72.0</td>
<td>100</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>20.7 (652)</td>
<td>60.0</td>
<td>58.0</td>
<td>55.0</td>
<td>73.4</td>
<td>100</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>651 (651)</td>
<td>62.0</td>
<td>59.0</td>
<td>65.0</td>
<td>67.0</td>
<td>81.0</td>
<td>100</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>651 (652)</td>
<td>64.0</td>
<td>65.0</td>
<td>60.2</td>
<td>71.2</td>
<td>77.0</td>
<td>86.0</td>
<td>75.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>6X1A2</td>
<td>64.0</td>
<td>64.0</td>
<td>62.0</td>
<td>62.0</td>
<td>71.3</td>
<td>73.0</td>
<td>75.0</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

TABLE IV
5S-rRNA Gene Spacer Region Pairwise Comparison
descended from a pBG13 like sequence. Group 3 through 5 show sequentially reduced homology to group 1 and 2 respectively. However, comparison of Groups-3,4, and 5 to each other indicates that these sequences are more similar to each other than they are to group 1 or 2. This is most pronounced by the 81% similarity of group 3 to group 5 (Table II). This clustering of similarity in 5S rRNA gene groups which are highly diverged from the main sequence could indicate a commonality in their origin or in the way that these variant genes have diverged.

Thus, comparison of groups 1 through 5 at the gene repeat level indicates good agreement with the homology data obtained through hybridization studies described above in section 3A2. The results also suggest that groups 3-5 are more related to each other than to pBG13 and may therefore constitute a level of relationship beyond the groupings so far indicated.

Interspecific comparisons of 5S rRNA genes as well as other repeated genes have historically noted a high degree of sequence conservation in the coding sequence with little to none in the spacer regions (Long and Dawid 1980). Comparatively few studies have addressed the question of intraspecific sequence variation in 5S rRNA genes. The data obtained from those studies
indicated a high degree of sequence conservation between different repeats of 5S rRNA genes (Goldsbrough et al. 1982; Gottlob-Mchugh et al. 1990). However, comparison of individual repeats in *Xenopus laevis* has demonstrated several levels of intraspecific variation. Firstly, members of the oocyte and somatic classes show highly diverged spacer regions which are essentially unique (Peterson et al. 1980). Secondly, repeats from the same class show low sequence variation in the range of several percent (Fedoroff and Brown 1979). In order to precisely define the where sequence variation occurs among flax 5S rDNA, the levels of sequence divergence in each of the two segments of the 5S rRNA gene repeat were investigated by performing pairwise comparisons of the coding and spacer sequences independently. Tables III and IV show the comparisons for the coding region and the spacer region respectively. Comparison of a similarity index in Table III with its cognate in Table IV for almost all of the sequence pairs shows that sequence conservation in the coding region is higher than that in the spacer region. With respect to pBG13 the level of divergence ranges from 5% to 36% in the spacer region as opposed to 3% to 27% in the coding region (pBG13 vs pRS18 and pBG13 vs p6XA2 respectively for each percentage in coding and spacer region).
Therefore even though several of the gene repeats are highly diverged from pBG13 the coding regions are still maintained to a higher degree of homogeneity than the spacer regions. Comparison of the extents of homology between group 1 and 2 sequences demonstrates that the coding region displays the highest similarity to pBG13e and pRS18, while the spacer region is highly diverged. This result is consistent with the ca. 1:3 hybridization ratio between group 1 and 2 (Figure 3). Consideration of these results with those obtained from the hybridization experiment in Figure 3 indicate a good relationship between homology and extent of hybridization at the high stringency level. For example the extent of hybridization between group 1 and 2 at 30°C below the calculated Tm is dramatically reduced. The reduction in hybridization is largely due to the low homology of the group 2 spacer region to group 1 (compare pBG13 to pRS13 in Table IV). These results indicate that the high stringency genomic DNA blots shown in Figure 4 are an accurate representation of the relative abundance of the different groups of 5S rDNA in the flax genome since cross hybridization will be minimal.

As with the gene repeat comparisons described above, both the coding and spacer regions of group 3
through 5 show greater homology between themselves than to groups 1 and 2. The higher degree of spacer homology in clone 6 and clone 3 indicates that these sequences must be closely related. The parent phage clone of p6XA2, 5S6, contains several XhoI restriction fragments with 5S rDNA homology, two of which appear to be a tandem of 356 bp. Thus, p6XA2 may represent a divergence away from group 3. The linkage relations of these clones is not known at this point and therefore it is not known whether these two sequences are located at or near the same chromosomal positions.

The high level of nucleotide divergence between the different groups of flax 5S rDNA suggests that several of the genes may represent pseudogenes. To establish which of the cloned flax 5S rRNA gene sequences most closely resembles those of other plant 5S rRNAs the coding regions of all of the groups were compared to a canonical sequence of angiosperm 5S rRNA sequences derived from both 5S rRNA and 5S rDNA sequence (Erdman and Wolters 1986). For practical purposes only group 1 and 2 sequences are aligned with the angiosperm 5S rRNA canonical sequence in Figure 5 since the other groups are so divergent that they are not candidates for being closely related to other plant 5S rRNA gene sequences (compare groups 3 and 5 to pBGl3 in Table III). All of
Figure 5. Sequence comparison of Angiosperm and Flax Group 1 and 2 5S rRNA sequences. The Angiosperm consensus sequence was compiled from sequences presented in Ellis et al. 1988. Only differences to the angiosperm (ANGIO) consensus sequence are shown. The coding regions for group 1 and clones pRS18 and pRS13 are indicated as 18COD and 13COD respectively. Dash marks indicate single bases, (.) symbols indicate increments of five base pairs and (+) symbols denote increments of 10 base pairs. The alignment was performed on the MULIGN program using the default parameters furnished by DNASTAR (Madison, WI).
ANGI05S: GGGTGCGATCATACCACGACTAATGCAACCAGATCCCATCAAGACTCCGCAGTAAAGCGTG
5SCODING:       a   a
18COD:          a   a
13COD:             t   a   a   a
               ----- --------- ------

       ANG10SS: CTTGGGGGAGATAGTACTAGGATGGGGACCTCTGGGAAGGTCCCTCTCGTGTTGCACCCCC
PBG13:     a   t
18COD:       a   a   t   a   ca
13COD:       a   t   t   a   t
               ----- --------- ------
the coding sequences obtained from pBG13 were identical to each other (Goldsbrough et al. 1982). The 5S rRNA coding sequence present in pBG13 is most homologous to the consensus angiosperm sequence containing only 6 bp mismatches. Due to this homology and high representation in the genome pBG13 will be assumed to be the progenitor of the other flax 5S rRNA gene sequences in the flax genome. All of the differences between pBG13 and the angiosperm sequence are conserved in groups 1 and 2 and are therefore specific to flax. The differences within group 1 and between group 1 and 2 thus represent more recent events than the speciation of flax, consistent with regarding pBG13 as the progenitor of the other flax 5S rRNA genes. The coding sequence from pRS18 contains 4 base pair changes with respect to pBG13 and the angiosperm sequence, several of which have not been identified in any of the other angiosperm sequences (Figure 5, positions 85 and 97; Goldsbrough et al. 1982; Ellis et al. 1988). Thus, although this clone is very similar to pBG13 it may represent a relatively recent divergence event or a new variant in this group. Group 2 contains 5 base pair changes with respect to pBG13 and the angiosperm sequence. Since groups 1 and 2 represent the majority of 5S rRNA sequences present in flax (See Figure 4 above) these two groups would appear
to be the best candidates for contributing the most to the production of cytoplasmic 5S rRNA.

An investigation of the secondary structures of the flax 5S rRNA genes was conducted in order to gain further insight into which sequences are likely to represent functional genes. The secondary structure of 5S rRNA genes is thought to be essentially universal (Vandenbergh et al. 1984). Several models of 5S rRNA secondary structure have been advanced based on comparisons of over 237 different sequences and helix versus single strand secondary and tertiary structure determinations (Erdmann and Wolters 1986). From these studies it has been accepted that the structure of the 5S rRNA is most highly conserved while base sequence conservation, although also high, is secondary to structure (Romby et al. 1990; Erdman and Wolters 1988). Based on these assumptions Erdman and Wolters 1988, have considered 5S rRNAs which demonstrate violations of the common 5S rRNA secondary structure obeyed by a close relative as pseudogenes. Several of the sequences considered as plant pseudogenes were examined and their structures compared to those of flax. Most of the mutations affect the stem regions of the 5S rRNA molecule. In Lupinus luteus 1 and 2 the mutations disrupt three of the five stems by one base pair
mismatch and one by two base pair mismatches. In the *Triticum aestivum* pseudogenes at most three of the five stem regions are interrupted by single base pair mismatches. The putative secondary structures of each of the 5S rRNA coding sequences determined in this study are shown in Figure 6. All of the sequences shown except for pBG13 and pRS18 display mismatches in at least 3 of the stem regions. Inspection of the structures for groups 3 through 5 indicates a large number of base pair mismatches in the stem regions which would presumably have an effect on the structure of the molecule. Studies on the effect of mutations in the stem regions of *Xenopus laevis* 5S rRNA have demonstrated that the integrity of helices B and D are important for TFIIB binding (Romaniuk et al. 1989). The loop region between helix A and B has been termed the ‘hinge’ region. Mutations in the hinge region have been shown to destabilize the tertiary structure of the 5S rRNA molecule such that binding of TFIIB is greatly reduced (Romby et al. 1990). Inspection of the secondary structures in Figure 6 shows that the three nucleotides (A15, G66, and U109) thought to be responsible for the overall organization of the two helical regions are mutated in group 3 (G66 to C66), group 4 5S1 (G66 to C66 and U109 to A109), group 4 5S2 (G66 to U66 and U109 to
Figure 6. Secondary structures of Group 1 - 5 5S rDNA.
Secondary structures are based on the model of Erdmann and Wolters 1988. Differences to
the pBG13 sequence are shown in lower case.
Large, bold, capital letters indicate stem
regions A - E'. Differences from pBG13
representing insertions are denoted with an
arrow. Lines connecting bases indicate
phosphodiester bonds between these bases.
G109) and group 5 (G66 to C66). Following the classification of Erdman and Wolters 1988, all of the flax 5S rRNA genes except the pBG13 and pRS18 sequences would require pseudogene assignments. However, these classifications must be viewed as tentative due to the lack of information regarding RNA expression and structural constraints on 5S rRNA gene function (Nazar et al. 1990).

B. Characterization of Quantitative 5S rRNA Polymorphism among Flax Genotrophs.

The organization and representation of group 1 - 5 5S rDNA was studied in an effort to identify sequences which would provide molecular markers for environmental induction of heritable changes. This was accomplished by comparing the molecular organization of both 5S rDNA and junction sequences of each group in the progenitor line PI and in large and small genotrophs via DNA gel blot analysis. In the course of these studies both quantitative (copy number) and Restriction Fragment Length Polymorphisms (RFLPs) were identified for 5S rDNA. Analysis of the junction sequences revealed RFLPs for one of the junction sequences. This section describes the characterization of 5S rDNA quantitative
polymorphisms among the genotrophs. Results sections 3C and 3D describe the results obtained for characterization of RFLPs in 5S rDNA and junction sequences respectively.

The 5S rDNA portion of each of the clones identified in section 3A was isolated in pure form away from the junction sequences by either plasmid sub-cloning or gel purifying the fragment of interest as described in Chapter 2H. The 5S rDNA of each phage clone was then used as a hybridization probe to DNA gel blots containing restriction digests of P1, large and small genotrophs. The filters were washed at Tm - 3° to eliminate cross hybridization between the different groups of 5S rDNA as described in results section 3A2. Probes which identified differences between P1 and at least two genotrophs were screened against the entire set of genotrophs available in the lab. The patterns identified for groups 3 - 5 did not show any quantitative polymorphisms. The signal intensities were similar to those in L^H genomic DNA shown in Figure 4. The relative amount of group 1 5S rDNA hybridization detected for the genotrophs was approximately equal to P1, except for Cl which showed ca. 27% reduction in signal intensity compared to P1 (Figure 7B). In contrast many differences were observed in the relative
Figure 7. Group 1 versus Group 2 Quantitative Polymorphisms among Flax Genotrophs. BamHI restriction digests of flax P1 and genotroph DNA (1 ug) were electrophoresed, blotted and hybridized first with group 2 5S rDNA (ca. 350 bp insert of pRS13) and autoradiographed (Panel A). The probe was then removed, and the filter rehybridized with group 1 5S rDNA, (ca. 350 bp insert of pBG13), and autoradiographed (Panel B). Arrow indicates ca. 350 bp band.
amount of hybridization detected between genotrophs using the group 2 probe pRS13 (Figure 7A).
Densitometric scans indicate the signal intensity of \( L^H \) is approximately one third (35\%) that of P1 (Figure 7B, compare lanes 1 and 2). In order to determine if copy number reduction is a common feature of genotroph lines the experiment was extended to include lines S1, C3, C2, C1, and L6. The lines S1, C3, C2, C1 and L6 also demonstrate reduced hybridization (Figure 7A, compare lane 1 with 3, 4, 5, 6, and 7). The results from Figure 7 indicate that group 2 sequences are reduced in copy number in many of the genotroph lines. In contrast group 1 sequences tend to be relatively equal among the genotrophs with one exception, C1. Previous studies on copy number differences had demonstrated large reductions of 5S rDNA in the genotrophs \( L^H \) and S1 when using labeled 5S rRNA as a probe (Goldsbrough et al. 1981). The results also show large reductions in 5S rDNA copy number in genotrophs \( L^H \) and S1 for group 2 5S rDNA but not for group 1. This may suggest that the two classes of sequences are differentially modulated with respect to copy number in the genotroph lines. However, since the original \( L^H \) and S1 DNA samples were not available for use in these experiments it is possible that these alterations have occurred since the
environmental induction event. Further studies are required to determine if the reductions are the result of the induction process.

C. Characterization of 5S rDNA Restriction Fragment

   Length Polymorphism among Flax Genotrophs and Pl.

1. Identification of 5S rDNA RFLPs

   The analysis described in section 3B was also employed to determine if any of the 5S rDNA sequences identified RFLPs among the genotrophs and Pl. Group 4 clones 5S-20 and 5S-12 were the only clones to identify RFLPs between the progenitor line Pl and large and small genotrophs. 5S-12 and -20 5S rDNA are approximately 90% homologous based on melting temperature experiments described in 3A2. The homology differences between the clones are evident from the presence of several different size classes of 5S rDNA in 5S-12 (ca. 700, 1000, 1700 bp, Figure 2). The 1 kb band of 5S-12 clearly differentiates the two clones in high stringency genomic DNA gel blots shown in Figure 3, lanes d and e. The hybridization patterns obtained at Tm - 2°C using the 5S rDNA homologous portions of 5S-12 and -20 as probes to EcoRI digested genomic DNA are shown in Figure 8.
Figure 8. Detection of RFLPs with Group 4 5S rDNA clones 5S-12 and 5S-20. P1, \( L^H \) and S1 genomic DNA was digested with EcoRI, electrophoresed, blotted and hybridized with \(^{32}P\) random labeled probes of: the 5S rDNA portions of 5S-20 (A); and 5S-12 (B) and autoradiographed. DNA size markers are from lambda DNA digested with EcoRI and EcoRI/HindIII.
Both probes identify similar polymorphic bands which at lower stringency are more pronounced with 5S-12. The high stringency experiment shown in Figure 8 was performed to determine if one of these clones was more closely related to the RFLPs since 5S-12 and 5S-20 are ca. 90% homologous. EcoRI digested DNA was chosen since it identifies more RFLPs than BamHI, thus allowing for a more accurate determination of which clone is more related to the RFLPs. The results shown in Figure 8 clearly indicate that 5S-20 is more closely related to the DNA fragments comprising the RFLPs than 5S-12. Although 5S-12 still identifies several of the polymorphic bands, the hybridization is much reduced compared to 5S-20. In addition, polymorphic bands at 6 and 9 kb are no longer visible. Based on these results 5S-20 5S rDNA was chosen for further studies.

2. Characterization of 5S-20 5S rDNA.

The substantial homology differences between 5S-12 and 5S-20 prompted an investigation into the homogeneity of the 706 bp repeats present in 5S-20. If substantial heterogeneity existed between the seven 706 bp repeats in 5S-20, the possibility existed that a subset of these repeats might be more useful for identifying RFLPs,
rather than using the entire set or one randomly selected subclone of 5S-20. In addition, the DNA sequence would permit determination of the sites of variation with respect to pBG13 which might contribute to the observed polymorphisms. Therefore, several 706 bp repeats were subcloned from the parent phage clone 5S-20 into the plasmid vector pBluescript (Stratagene) and two randomly selected clones (pRS20.7 and pRS20.7II) were sequenced as described in chapter 2K (Figure 9). The clone used for all subsequent experiments is referred to as pRS20.7. The 5S rDNA of 5S-20 was found by hybrid duplex stability to be weakly homologous to the previously described flax 5S rRNA gene clone pBG13 (chapter 3A2, Figure 3; Goldsborough et al. 1982). Comparison of the pRS20.7 and pBG13 DNA sequences demonstrates a high degree of divergence in both the spacer region and the 5S rRNA transcription unit with an overall homology of approximately 75% (Table II, III and IV, chapter 3A3). Restriction mapping and sequence analysis identified pRS20.7 as a tandemly arranged 5S rRNA gene dimer (Figure 3, Figure 9, panel A and Figure 10). Several features of this clone are noteworthy. First, the 706 bp repeat sequence contains two 5S rRNA gene transcription units designated 5S1 and 5S2. The 5S1 and 5S2 transcription units and spacer regions
Figure 9. Sequence of pRS20.7, pRS20.7II and Comparison of the 5S1 and 5S2 Gene Repeats.

Panel A: Nucleotide sequence of the 706 bp insert of pRS20.7. Differences between this sequence and a second 706 bp repeat (pRS20.7II) isolated from 5S-20 are shown above the pRS20.7 sequence. Spacer sequence is shown in lower case letters and 5S rRNA coding sequence (based on comparison to the plant consensus sequence in Erdmann and Wolters 1986) is shown in upper case letters. The (^) symbol indicates the insertion of the base appearing above the pRS20.7 sequence. The asterisk at base 353 defines the position where 5S1 and 5S2 repeats were defined for comparison in panel B. Panel B: Sequence comparison of the two 5S rRNA gene repeats present in pRS20.7. Differences between 5S2 and 5S1 are shown below the 5S1 sequence. (−) indicate gaps introduced to optimize the overlap of homologous bases.
### A.

<table>
<thead>
<tr>
<th></th>
<th>20</th>
<th></th>
<th>40</th>
<th>60</th>
<th>80</th>
<th>100</th>
</tr>
</thead>
<tbody>
<tr>
<td>t</td>
<td>a</td>
<td>a</td>
<td>g</td>
<td>g</td>
<td>g</td>
<td></td>
</tr>
<tr>
<td>1 gttgataatccctgtggaaac cgagaccaagatgtagtaaa aagtaatcgaacagagttcctaa aatagtttgtgtaatatctaatgagcata</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>100 tttgcatgagaacgcagacga GTGAGGTCAATACCGATGATGTA ATTCATCGGTATCCCATACG ACCTGTAAGGTTAAAACGTACT TGCCCAAGACTAGTACCTACG</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>200 ATGGGTTGACCTTTTATGCTCT TCCCGGAGTACACCCCTTTtttttttttataaagttatatgcctaa tttttaccccaagtgcaaat gaaacggttgtgaactatctcagaa</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>300 cttcgctactgttgtgtaa gtatagtacagggggaac tgggttccgggtgttttccgt gcgcNcagaaatgagacgac aagatatggtatagcgtaag</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>400 gaaatgtagaaagaagag gggggagttagtaagagaca atttatagtttatgaggac gcggagctggtgAATGAGAT CATACCAACACTAATTCATTT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>500 GGATCTGTCAATACCTCTGTA AATTAAAACGTACTTTGGTиг GTGTAAGTACTCGATGGGIG ACCTTITGATCCTATCCCG ATTCGACACCTtcACCTtttttttc</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>600 ttttttttttttttttacctttgacgtatataaggaatg tttaactaaacaacatccitc taatacaatcatggtaatga cctggggaaccggtgatgc</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### B.

<table>
<thead>
<tr>
<th></th>
<th>20</th>
<th></th>
<th>40</th>
<th>60</th>
<th>80</th>
<th>100</th>
</tr>
</thead>
<tbody>
<tr>
<td>tt</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5S1 gttgataatccctgtggaaac cgagaccaagatgtagtaaa aagtaatcgaacagagttcctaa aatagtttgtgtaatatctaatgagcata</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5S2 gc aca at a t t g c g a t g a g g</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5S1 tgtgcatgagaacgcagacga GTGAGGTCAATACCGATGATGTA ATTCATCGGTATCCCATACG ACCTGTAAGGTTAAAACGTACT TGCCCAAGACTAGTACCTACG</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5S2 a c t t A A AC C T TTTG TA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5S1 ATGGGTTGACCTTTTATGCTCT TCC-CGAGTACACCCCTTTtttttttttttataaagttatatgcctaa tttttaccccaagtgcaaat gaaacggttgtgaactatctcagaa</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5S2 A T T G A C t t c tttt t g c t a t a a</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5S1 cttcgctactgttgtgtaa gtatagtacagggggaac tgggttccgggtgttttccgt gcgcNcagaaatgagacgac aagatatggtatagcgtaag</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5S2 t c aca ag ct g g ac t a a</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### C.

...
Table V. Copy numbers of 706 bp repeat in comparison to total 5S rDNA.

<table>
<thead>
<tr>
<th>Plant Line/Species</th>
<th>Copy # 5S rDNA $^a$ 2C</th>
<th>copy # pRS20.7 $^b$ 2C</th>
</tr>
</thead>
<tbody>
<tr>
<td>PI</td>
<td>117,000</td>
<td>1000</td>
</tr>
<tr>
<td>H</td>
<td>49,600</td>
<td>960</td>
</tr>
<tr>
<td>L</td>
<td>52,800</td>
<td>1060</td>
</tr>
<tr>
<td>S1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>L. grandiflorum caeruleum</td>
<td>ND</td>
<td>220</td>
</tr>
</tbody>
</table>


$^b$ : Mean of two determinations. The standard error for these experiments is less than 5% (RIVIN, CULLIS AND WALBOT 1986).
Figure 10. Restriction Enzyme Site and DNA Probe Map of pRS20.7. Filled and hatched boxes represent 5S RNA transcription units 1 and 2 respectively. The 706 bp bracket designates one repeat unit as defined by BamHI which was cloned from 5S-20. The open boxes below show the sub-probes used in hybridization experiments (probes described in section 3C2. The interrupted lines on either side of the 706 bp repeat indicate that this sequence is organized in tandem arrays in the original genomic clone, 5S-20. Restriction enzyme sites: T, TaqI; R, RsaI; B, BamHI; A, AluI; M, MboII.
separating the genes are significantly diverged from one another as shown by their restriction maps in Figure 10. Sequence comparisons of both 5S1 and 5S2 transcription units and spacers; show 76% sequence homology (Figure 9 panel B, table III and IV). This type of heterodimer repeat has not been identified in any of the other 5 groups of flax 5S rRNA genes studied to date (Goldsbrough et al. 1982). Normally 5S rDNA sequences found in tandem arrays are nearly homogenous with the exception of micro-sequence heterogeneity (Long and Dawid 1980). This was previously shown in flax for pBG13 (group 1) 5S rRNA gene sequences (Goldsbrough, Ellis and Lomonossoff 1982). However, different copies of the 5S rDNA repeats present in 5S-20 are very homogeneous with respect to restriction enzyme sites. In addition the two sequenced copies of the 5S-20, 706 bp repeat array (pRS20.7 and pRS20.7II; 5S-20 contains a total of 7 repeats) differ in only eight positions (Figure 9, panel A). This level of variation (1.1%) is similar to that described for other tandemly arranged 5S rDNA as described in chapter 3A3. Thus, the dimer repeat would appear to be a unit which is being maintained in a homogeneous tandem array in a concerted fashion, rather than a set of individually diverging 353 bp units. This situation may be structurally similar to
that described for the *Xenopus laevis* oocyte type 5S rDNA repeat which contains a pseudogene alternated with a functional gene in a tandem array (Miller et al. 1978). On the basis of these results it was concluded that the 7 repeats of 5S-20 are essentially identical and therefore equally suitable for detecting the RFLPs under the hybridization conditions employed for this study. No differences were observed between the hybridization patterns obtained with either pRS20.7 or pRS20.7II on genomic DNA Blots.

3. The Genomic Organization and Copy Number of 5S-20 5S rDNA.

The genomic organization of this 5S rRNA gene family was examined further in PI (Stormont Cirrus), genotrophs, and other flax varieties and *Linum* species. The DNA probes used in this study are shown in Figure 10. The 209 bp A spacer probe is specific to the spacer sequence preceding the 5S2 transcription unit. Probe A starts 16 bp downstream of the 5S1 transcription unit and ends 2 bp upstream of the 5S2 transcription unit. Probe B is a 201 bp fragment specific to the spacer sequence preceding 5S1. Probe B starts at the first bp past 5S2 transcription unit and ends 16 bp upstream of
the 5S1 transcription unit. Probe C contains all of probe B in addition to 40 bp of 5S2 transcription unit sequence. Probe C starts 40 bp upstream of probe B and ends in the same position as probe B. The 706 bp BamHI fragment represents a single repeat from 5S-20.

Genomic reconstructions hybridized at high stringency (Tm - 2°C) indicate that the pRS20.7 5S rDNA sequence is present at approximately 500 copies per Pl haploid genome (Table V; no hybridization to any other group of flax 5S rRNA gene is seen at this stringency; chapter 3A2, Figure 3). The results from DNA blot experiments and the restriction map of 5S-20 indicate that this sequence is organized largely in tandem arrays of 706 bp BamHI fragments, containing the same repeat unit as shown in Figure 9 and 10. Thus, this 5S rRNA gene group appears share characteristics of other tandemly arrayed genes despite the marked divergence of the 5S1 and 5S2 repeats from each other and from other flax 5S rRNA gene groups. However, a survey of hybridization patterns produced with other restriction enzymes revealed a considerable amount of sequence variation. The variation is particularly apparent with enzymes which do not have sites within the 706 bp repeat isolated from 5S-20. The hybridization pattern of the 706 bp probe to genomic DNA gel blots of flax genotroph
and P1 DNAs digested with three different enzymes is shown in Figure 11; panel A, B and C show the hybridization pattern detected in \textit{EcoRI}, \textit{DraI} and \textit{EcoRV} digests respectively. A large number of bands are generated with these enzymes indicating a high degree of sequence variation in the 706 bp repeat family. However, many of the bands do not correspond to multimers of 706 or 353 bp (one half of the 706 bp dimer repeat). These results are interpreted as evidence of large amounts of length and sequence heterogeneity among the members of this group of 5S rDNA. Group 4 clone 5S-12 also displays length variation and is approximately 90\% homologous to pRS20.7 (Figure 2 and 3). This suggests that flax group 4 5S rDNA sequences are exceptionally heterogeneous in comparison to group 1 and group 2 flax 5S rDNA and are not as homogeneous as would be suggested from the analysis of the sequences and restriction sites present in 5S-20 alone (Figure 9).

4. 706 bp probe Identifies RFLPs in Flax Genotrophs.

Comparison of the 706 bp probe banding patterns of P1, L and S genotrophs by genomic DNA gel blot analysis detected polymorphisms with nine different restriction enzymes (\textit{EcoRI}, \textit{DraI}, \textit{EcoRV}, \textit{ScaI}, \textit{SpeI}, \textit{BclI}, \textit{BglI},
Figure 11. pRS20.7 RFLP Patterns among Flax Genotrophs. DNA blot hybridization patterns detected by pRS20.7 in EcoRI (A), DraI (B) and EcoRV (C) digests of flax Pl and genotroph DNAs. Lane 1, S\textsuperscript{h}; lane 2, S1; lane 3, S6; lane 4, C3; lane 5, L6; lane 6, Pl; lane 7, L\textsuperscript{H}; lane 8, L1. Panel A does not show S\textsuperscript{h} (see figure 6, lane 5). Arrows indicate RFLPs. Bracketed areas designate multiple RFLPs. DNA standard sizes are in kilobases. DNA standards for panel A are on the left side. DNA standards for panels B and C are on the right side. Band ‘a’ is 3-5 fold more intense in L\textsuperscript{H} (lane 7) than in small genotrophs (lanes 1-5). Asterisk denotes a potential single copy band in Pl.
HindIII and BstNI). Comparison of the EcoRI, DraI, and EcoRV restriction fragment patterns detected by the 706 bp probe between Pl (Figure 11, lane 6, panel A, B, and C) and both large (lanes 7 & 8, panel A, B and C) and small genotrophs (lanes 1-5, panel A, B and C) shows distinct RFLPs. A large number of bands are detected in each lane, a subset of which are polymorphic. Large and small genotroph patterns differ from one another as well as from Pl. The RFLP patterns are stable and have been reproduced with several independent DNA samples. The stability of the RFLP patterns in stable genotrophs is shown in families such as S1 and S6 which are separated by many generations but have identical patterns (Figure 11, lane 2 and 3, panel A, B and C; Cullis 1977).

An identical hybridization pattern is detected for each restriction enzyme in four small genotrophs, produced from separate, independent environmental induction experiments (Figure 11, compare lanes 1-5, panels A, B, & C). This RFLP pattern will be referred to as the S pattern. The S pattern has been identified in all small genotrophs available in the lab. This observation suggests that common sequence alterations have occurred in each of the small genotrophs. An identical S pattern was found in DNA isolated from four individual C3 plants, one of which is shown in Figure
11. Seed for each of these C3 plants was individually collected from separate, self fertilized Pl plants after environmental induction (Cullis 1981). The largest number of polymorphic bands are detected in the S pattern for each enzyme. The EcoRI, DraI and EcoRV blots show seven, ten and nine polymorphic bands (below 15 kb) respectively between small genotrophs and Pl (compare Figure 11 lanes 1-5 with 6). In contrast to the situation described above for the S pattern, non-identical RFLP patterns were detected in two independently induced large genotrophs, Ll and L^H. Comparison of L^H and Pl shows four, two and three polymorphisms respectively in panels A, B and C. Only one polymorphic band is detected for Ll in the DraI digest, (Figure 11, arrowed band on the right side of panel B). The EcoRI and EcoRV hybridization patterns are the same for Ll and Pl.

In each panel several polymorphisms appear to be identical in the S and L^H RFLP patterns. However, close inspection indicates that these bands are not identical. For example, band ‘a’ in Figure 11, panel A is consistently 3-5 fold more intense in L^H than in the S pattern. The difference in intensity could indicate a higher copy number of the same fragment in L^H, or a different fragment with more homology to the probe. The
difference in band intensity is clearly seen when using a sub-probe of the 706 bp repeat as shown in Figure 15 (compare A with 706 patterns; also discussed below). In each case the polymorphisms in the genotrophs represent both novel bands and deleted bands with respect to Pl. The band intensities of many of the RFLPs differ from one another within and between lanes and do not reflect a linear increase in fragment size. RFLP generation as a result of loss or gain of restriction enzyme sites would predict the loss and appearance of the same number of bands between Pl and the genotrophs. However, this type of predictable band shift is not observed from the calculated band sizes and numbers. This suggests that the RFLPs are not due to point mutation(s) resulting in the loss or gain of restriction enzyme sites. This inference is supported by the ability to detect RFLPs with every enzyme tested to date. The possibility of DNA modification contributing to the identification of RFLPs was tested by digesting equal amounts of Pl, \( L^H \), and S1 DNAs with the isoschizomer enzyme pair EcoRII and BstNI which recognize the sequence \( CC(A/T)GG \) (New England Biolabs). EcoRII is inhibited by methylation of the internal cytosine residue whereas BstNI is not affected by methylation at either cytosine residue. The EtBr stained gel shown in Figure 12A shows clearly that
Figure 12. Methylation Independent Detection of RFLPs Among Flax Genotrophs. Panel A: Ethidium bromide stained gel of P1 (lanes 1 & 2), L\textsuperscript{H} (lanes 3 & 4) and S1 (lanes 5 & 6) digested with EcoRII (lanes 1,3 & 5) or BstNI (lanes 2,4 & 6). Lane M is lambda HindIII digested DNA marker. Panel B: Autoradiograph of DNA gel blot from panel A hybridized with the A spacer probe of pRS20.7 (Figure 10).
a large fraction of the cytosines found in this sequence are methylated in the flax genome resulting in reduced cleavage compared to BstNI. The DNA from this gel was transferred to a nylon membrane and hybridized with a $^{32}$P labeled 706 bp repeat probe. The result shown in Figure 12B indicates that the probe identifies RFLPs in a methylation independent manner.

The intensities of similarly sized bands between lanes in Figure 11 are different, indicating either a difference in 706 bp sequence homology or copy number representation (Figure 11, panel B bracketed area). Copy number reconstructions were performed to determine the copy number of the 706 bp sequence in polymorphic bands detected by the 706 bp probe in an EcoRI digest of genomic DNA. The results indicate that the majority of the polymorphic bands are not single copy in the genome, with each band representing several copies (Figure 13). However, due to the heterogeneity of group 4 sequences it is difficult to determine the exact number of copies represented by each band. Several of the high molecular weight bands may correspond to single copy bands (Figure 11, starred arrow in panel A).

The large degree of hybridization to DNA fragments above 15 kilobases obscures identification of
Figure 13. Copy number reconstruction of Pl, L^H and S1 EcoRI RFLPs in Flax Genotrophs.

Autoradiograph shows one of three exposures used for determination of RFLP copy number in Pl, L^H and S1. The number of copies of the 706 bp insert of pRS20.7 are: lane 1, 1740; lane 2, 700; lane 3, 350; lane 4, 175; lane 5, 70; lane 6, 35; lane 7, 17.5; lane 8, 4.0. Experimental design described in chapter 2J.
amplification in response to methotrexate selection in polymorphisms in this region using conventional gel electrophoresis. The hybridization pattern observed with EcoRI digested genomic DNA fractionated by field inversion gel electrophoresis (FIGE) is shown in Figure 14. A number of additional polymorphisms are detected in the high molecular weight DNA between P1 and the genctrophs (Figure 14 compare lane 3 with lanes 1 & 2). Several intense bands appear in IH which have no obvious counterparts in P1. These bands may represent DNA amplification events. Chromosomal amplification events in other systems (e.g. dihydrofolate reductase gene animal cells) are thought to result from DNA overreplication and integration into the chromosome (Schimke et al. 1986). Such a process here would result in the production of novel bands and/or bands with increased copy number relative to unamplified bands.


The spacer probes were found to detect a subset of the 706 bp probe RFLPS as well as additional sequence polymorphisms. The dimer organization of the 706 bp repeat shown in Figures 9 and 10 is unusual in that most
Figure 14. High Molecular Weight RFLPs detected by pRS20.7. Hybridization pattern detected by pRS20.7 in DNA gel blot of EcoRI digested $S^h$ (lane 1), $L^H$ (lane 2), and P1 (lane 3) DNAs separated by FIGE. DNA standard sizes are in kilobases.
5S rDNA tandem arrays contain nearly homologous repeats (Long and Dawid 1980). An exception to this generalization is the *Xenopus laevis* oocyte type 5S rDNA which contains a pseudogene alternated with a functional gene in each repeat (Korn 1982). In order to determine if the 706 bp repeat is representative of all members of the flax group 4 5S rDNA family including those which show polymorphism, spacer probes were isolated and used to hybridize EcoRI genomic blots of Pl, L^H_1 and S^h DNAs (Figure 15). Clearly, the patterns obtained with the different spacer probes are non-identical to that of the entire 706 bp probe. Both the A and B spacer probes identify a subset of the bands shown by the 706 bp repeat. Although the specific activities of the probes were similar in all of the experiments shown in Figure 15, the A spacer appears to be represented in only one half of the bands detected by the entire 706 bp probe (Figure 15, compare A and 706). In addition the representation of the B spacer in S^h is reduced in comparison to Pl and L^H_1 (Figure 15, B). These results suggest that changes in the sequence copy number of group 4 5S rDNA family members has occurred during the induction process.

The A and B spacer probes distinguish different polymorphic bands as well as common, invariant bands.
Figure 15. DNA Gel Blot Analysis of pRS20.7, 5S1 and 5S2 Spacer Sequences. Autoradiograms of EcoRI digested Pl, L\(^H\), and S\(^h\) total genomic DNA blots hybridized with the indicated pRS20.7 derived spacer probes. See Figure 10 for probe location on the 706 bp repeat restriction map and text for description. Each set of three lanes represents Pl, L\(^H\) and S\(^h\) DNAs, respectively, as shown for probe A. The position of bands which are differentially detected by the A and B spacer probes with respect to the 706 bp probe are indicated by arrows.
However, the polymorphisms all share the 5S2 transcription unit sequence since probe C (which contains 40 bp of 5S2 in addition to the A spacer) identifies all of the polymorphic bands (Figure 15 panel C). The conclusion drawn from these data is that the 706 bp repeat does not always occur with the A and B spacers in the same organization as shown in Figure 10. Comparison of the autoradiographs obtained with the A and B spacer probes indicates the spacers are separate from each other in some cases as they each identify bands exclusive of the other spacer (Figure 15, A and B). This result could be due to the high degree of sequence divergence in group 4 flax 5S rRNA genes resulting in a population of sequences which can be differentiated by their spacer sequence. The spacer probes also specifically differentiate between bands which appear to be identical in the 706 bp EcoRI band pattern (Figure 15, compare arrowed bands in 706 with those in A and B). This latter result suggests two important features of the observed polymorphisms. First, the 706 bp repeat family is highly diverged, as the polymorphic bands identified by the 706 bp probe are not homogenous with respect to the sequence presented in Figure 9. Second, the differences in band intensities of several of the RFLPs suggest complex changes in
either the sequence or organization of these bands (discussed below).


In order to determine whether the observed polymorphisms are in multiple genomic locations or at a single locus the RFLP segregation pattern was followed in L1 x L6 genetic cross. DNA from 23 F2 individuals was isolated and digested separately with EcoRI, DraI and EcoRV. The resulting Southern blots with all 23 individuals represented for each enzyme were then probed with the 706 bp fragment (Figure 16, panel A) and the A spacer probe (Figure 16, panel B) independently to determine the linkage relations of the RFLPs (Southern 1975). Representative blots using conventional gel electrophoresis are shown in Figure 16, panel A and B (Panel A and B represent all of the 23 F2 individuals). Only parental and heterozygote type patterns (containing all of the polymorphic bands) for all probes are observed, indicating that the RFLP patterns detected by both probes segregate as a single unit. This classification was independent of the enzyme used to digest the F2 DNAs (EcoRI, EcoRV and DraI). The
Figure 16. Linkage Analysis of pRS20.7 RFLPs in L1 x L6 F2 individuals. A; Autoradiograms of DraI digested total leaf genomic DNA from L1, L6 and F2 progeny from an L1xL6 cross hybridized with pRS20.7. B; Autoradiograms of EcoRV digested L1, L6 and L1xL6 F2 individual total leaf genomic DNA hybridized with the A spacer probe. C; Autoradiograms of EcoRI digested total leaf genomic DNA from L1, L6 and L1xL6 F2 individual total leaf genomic DNA hybridized with the pRS20.7. F2 individual numbers appear below the lanes for identification.
segregation pattern for both probes is identical indicating that the polymorphisms detected by each are linked to each other. The probability that the RFLPs are linked was tested by $^2x$ analysis of each of the polymorphisms independently for a 1:2:1 segregation ratio. The test was performed in this way due to the low expected numbers for most of the classes of a five gene (RFLP), independent segregation test. A highly significant fit to a 1:2:1 (5-13-5) segregation ratio for each probe was obtained, indicating that the RFLP patterns identified by the 706 bp probe and the A spacer probe follow Mendelian single locus segregation ($p > 0.99$). Therefore, the rearrangements must be all occurring at a single chromosomal locus. The lack of recombination between loci represented by any of the polymorphic bands indicates that either the RFLPs are physically close together or that recombination in this region was suppressed.

Linkage analysis was extended to the high molecular weight polymorphisms. EcoRI digested genomic DNA of the same 23 F2 individuals was fractionated by FIGE and hybridized with the A spacer probe (Figure 17). Complete linkage was also observed for high molecular weight polymorphisms which was identical to that obtained for the entire 706 bp probe (arrowed bands in
Figure 17. Linkage Analysis of High Molecular Weight RFLPs. Autoradiogram of EcoRI digested total leaf genomic DNA from L1, L6 and L1xL6 F2 individual total leaf genomic DNA separated by FIGE and hybridized with the A spacer probe (see Figure 10). The arrows identify two similarly sized restriction fragments, one of which is a segregating RFLP (lower band).
Figure 17, p > 0.99 probability that RFLPs are linked by $^{2}X)$. The segregation patterns for the high molecular weight polymorphisms are identical to those identified for the other probes indicating that these RFLPs are also linked to the same chromosomal region. Thus, all RFLPs detected by the pRS20.7 DNA sequence are linked to a specific chromosomal region. This locus has been named Flp-1, for Flax polymorphic locus 1.

Plant height and weight measurements of induced plants in comparison to the progenitor P1 are used as discriminating values for classifying genotrophs as large or small (Durrant 1962). Analysis of the 23 individuals described above indicates that plant height is not significantly correlated with segregation of an L or S RFLP pattern (p > 0.5). However, a larger sample size may be required to determine if height and other genotroph traits are linked to Flp-1.

The genetic data provided above demonstrated that the RFLPs are linked to a single chromosomal region. However, the linkage of the non-polymorphic bands could not be determined from this analysis. If most of the group-4 5S rDNA are located at this locus in a contiguous fashion, then the physical distance should be in the range of 350 kb based on a copy number of 500 and a repeat size of 706 bp (table V). The physical
identification of the Flp-1 locus should then be possible utilizing pulse field gel electrophoresis provided a restriction enzyme which cuts outside of the locus can be identified. To this end, enzymes which have rare recognition sequences in plant DNA and also do not have sites in the group 4 sequences were used to digest P1 megabase DNA samples (prepared as described in chapter 2E, F and G). After CHEF gel electrophoresis and DNA gel blotting a survey of the pRS20.7 hybridization products indicated that MluI produced the most intact single band (Figure 18 and 19). However, the size of the fragment was too large to be sized accurately under the conditions used for the gel shown in Figure 19. In addition, DNA fragments of multiple sizes can migrate in this region of the gel. In order to accurately size the band(s), the experiment was repeated with the electrophoresis conditions designed to separate S. pombe chromosomes which are ca. 3.5, 4.6 and 5.7 megabases in size (Bio Rad). The EtBr stained gel and autoradiograph upon hybridization with pRS20.7 are shown in Figure 20, A and B respectively. Under these conditions the MluI band is now resolved into a single band of similar size in P1, L^H and S1. The size is approximately 2.0 megabases. Hybridization of p651 (another group 4 5S rRNA gene) also detects the same
Figure 18. Megabase DNA Blot Analysis of Group 4 5S rDNA. Panel A; Pl megabase DNA was digested with the enzymes indicated and electrophoresed on a 0.8% Pulse Field Agarose (Boehringer Mannheim Biochemicals) CHEF gel apparatus for 22.5 hr at 14°. Switch times were 80 s for 17 hr and 30 s for 5.5 hr. YM is a Saccharomyces cerevisiae chromosomal DNA molecular weight marker and lambda is a concatamerized lambda DNA molecular weight marker. DNA sizes are given in megabases (Mb). Panel B; Autoradiogram of the DNA gel blot of Panel A hybridized with pRS20.7 706 bp insert. Panel C; hybridization of Panel A DNA gel blot with the A spacer probe. Undigested Pl DNA is a size control.
Figure 19. CHEF Gel Electrophoresis and DNA Blot
Analysis of P1, L^H and S1 MluI Digested Megabase DNA. Panel A; EtBr stained gel of P1, L^H and S1 megabase DNAs were digested with MluI and electrophoresed on a 0.8% Pulse Field Agarose (Boehringer Mannheim) CHEF gel for 20 hr at 14^0 with a switch time of 70 s. YM is a Saccharomyces cerevisiae chromosomal DNA molecular weight marker with sizes given in megabases. Panel B; Autoradiogram of DNA gel blot of gel in panel A hybridized with pRS20.7 706 bp probe. UD is undigested P1 DNA.
Figure 20. CHEF Gel Electrophoresis of Pl, L^H and S1 MluI Restriction Digested Megabase DNA
Separated in the .5 to 5.0 Megabase Size Range. Panel A; EtBr stained gel of Pl, L^H and S1 MluI restriction digested megabase DNA electrophoresed on a 0.5% PFG CHEF gel for 72 hr at 50 V, 14^0 with a 30 min switch time. Panel B; Autoradiogram of DNA gel blot of gel in panel A hybridized with 706 bp probe (Figure 11). Panel C; Autoradiogram of DNA gel blot of gel in panel A hybridized with 651 bp insert of p651 plasmid. YM is a Saccharomyces cerevisiae chromosomal DNA molecular weight marker with sizes given in megabases. Sp is a Schizosaccharomyces pombe chromosomal molecular weight marker (Bio Rad). Undigested Pl DNA is a size control.
size fragment suggesting that all group 4 sequences are present at one chromosomal locus (Figure 20, C). This result places certain constraints on the type of DNA rearrangements which could account for the RFLPs. The results suggest the alterations to the Flp-1 locus are limited to this region and are therefore unlikely to include mechanisms which would result in the movement of group 4 sequences to other parts of the genome such as transposition, gross chromosomal inversion or illegitimate recombination events between nonhomologous chromosomes since these would have been visible as changes in the size and/or number of fragments in comparison of P1, L^H and S1. Conversely mechanisms which do not disrupt the continuity of the locus, such as small chromosomal inversions, deletion and amplification, or replication slippage would be compatible with the above data (Schimke et al. 1986; Scowles et al. 1988). Further characterization of this locus will be required to differentiate between the possible mechanisms.

Consideration of the data in Figure 18 provided additional information concerning the Flp-1 locus. Several of the enzymes produced very large pRS20.7 homologous DNA fragments which migrated near the top of the gel (KspI and SfiI). The enzyme NotI produced the
largest fragment which is indistinguishable from uncut DNA.  PvuII and RsrII produced a low and high molecular weight smear of hybridization respectively for both probes (A spacer probe and pRS20.7) with several distinct bands. The last three enzymes, Clai, SmaI and PacI, produce distinct bands for each probe. However, there was a clear distinction in the number and intensity of bands produced by Clai, SmaI and PacI. The A spacer probe identifies two prominent bands in the PacI lane in addition to several other minor bands. In contrast Clai and SmaI generate six to four prominent bands in addition to minor bands and a substantial amount of smeared hybridization. A summation of the bands in the Clai and PacI lanes (>5.0 Mb and 1.9 Mb respectively) leaves a large discrepancy in the amount of DNA homologous to the probe, significantly increasing the size estimate for the Flp-1 locus obtained in the experiment above. However, one interpretation of these results is that different methylation patterns of the Flp-1 locus within the plant results in the generation of additional bands in the Clai and SmaI lanes due to inhibition of the restriction enzyme. This also results in reduced intensity in comparison to the PacI lane which is not reported to be sensitive to methylation (recognition sequence is composed entirely of A/T base
pairs; New England Biolabs). Similar observations have
been reported in mammalian systems (D. Threadgill per. 
comm.). The presence of methylation at this locus is
consistent with the results of Figure 12 showing a high
degree of cytosine modification. Assuming that the PacI
digest is not methylation sensitive, the Flp-1 locus
size estimate of 2.0Mb is similar to the 1.9 Mb size
obtained by summing the bands in the PacI lane.

Comparison of the hybridization pattern obtained
with the A spacer probe and the 706 bp insert of pRS20.7
to the same DNA blot in Figure 18 shows differences in
the number of bands identified in the PacI lane. A band
identified by the 706 bp probe is missing from the A
probe profile. In addition two of the ClaI bands are
also missing or reduced in intensity in the panel B
compared to panel C. These results suggest that a large
region of the Flp-1 locus lacks the A spacer sequence.
This hypothesis correlates well with the results of
Figure 15 showing that the A spacer is only present in a
subset of the 706 bp repeat family.

7. Cloning and Characterization of DNA from the Flp-1
locus.
The results presented in chapter 3C, 1-8 have provided data on the organization, linkage and physical structure of the Flp-1 locus. The results have defined more clearly the types of mechanisms which may be involved in the generation of the DNA polymorphisms. Specifically, the DNA alterations appear to be confined to a localized region of approximately 1.6 Mb on a single chromosome pair. The presence of new bands and absence of other bands in the genotroph RFLP pattern suggests that deletion or amplification events may have occurred. In order to ask more specific questions about the nature of the DNA rearrangement(s) the isolation of individual RFLP fragments by molecular cloning was pursued. The premise for undertaking this study was that the RFLP fragment(s) would provide a unique molecular probe which could then be used in standard DNA gel blotting experiments to characterize the nature of the rearrangement. This would then allow the formulation of a testable hypothesis concerning the molecular mechanism(s) responsible for the production of the RFLPs. The ability to design a specific test regarding the environmental induction of heritable changes in flax is essential to understanding the process. To this end, EcoRI subgenomic DNA libraries were prepared for PI and $L^H$ genotrophs in a lambda ZAP
II cloning vector as described in chapter 2I. Pl and L^H DNAs were chosen since they contained the highest number of RFLP bands between 2 and 10 kb (the size limit of the cloning vector is 10kb). Screening of the L^H and Pl libraries with the 706 bp insert of pRS20.7 produced a large number of positive plaques. A total of 7 out of 60 L^H and 19 of 35 Pl positives were plaque purified and the insert rescued in the form of a pBluescript plasmid for restriction analysis. The two criteria used for establishing the presence of an RFLP clone were EcoRI fragment size and homology to the 706 bp probe. Therefore all of the clones were restriction digested with EcoRI, electrophoresed, blotted and hybridized with the 706 bp probe. A problem concerning the isolation of the rescued plasmid DNAs was encountered during this experiment. Miniprep analysis was carried out on 5 independent colonies from each rescued clone. However, the band size(s) observed for a particular clone often differed between the 5 isolates. This indicated that either the phage used for rescue was not pure, or that something was occurring during rescue to alter the integrity of the DNA inserts. Subsequently, lambda DNAs were isolated from the clones of interest to verify the integrity of the inserts. The reason for this problem is not clear, but it persisted even in recombinant
deficient genetic backgrounds (SURE). This may suggest that the problem lies in the rescue of the pBluescript plasmid.

One P1 clone was isolated, pZ6, which had an insert size of about 4 kb. This fragment size was also absent from the small genotroph EcoRI RFLP patterns in Figure 11. All of the other clones isolated from these experiments corresponded to nonpolymorphic fragments and were not characterized further. Before characterizing pZ6 further, the A spacer probe of pRS20.7 was used to rescreen the P1 11 library which was size selected for DNA fragments in the range of 3 - 10 kb (chapter 2I). The A spacer probe was used to increase the probability of cloning an RFLP since it detects only a subset of the group 4 family (Figure 15). In addition the P1 11 library was infected into two different Escherichia coli host strains, Plk-F′ (also used for the LH and P1 12/13 libraries) and the SURE strain. The SURE strain is recombination-deficient as well as mcrA,B,C and mrr-deficient which increases the representation of methylated DNA sequences (Greener, 1990; Kretz and Short 1989). The P1 11/Plk-F′ library produced 5 strong positives which were purified and analyzed as above. These contained only non-polymorphic DNA fragments. Similarly, the P1 11/SURE library screen produced two
strong positives which were both non-polymorphic. Therefore, before considering alternative cloning strategies, the pZ6 clone was characterized further to determine its usefulness for studying the RFLPs.

Restriction enzyme digests of pZ6 indicated that the clone was composed of 6 repeating units of about 650 bp as defined by SpeI and did not contain sites for BamHI (Figure 21,A). One of the repeats was subcloned into the SpeI site of pBluescript (p651) and sequenced to determine the level of homology to pRS20.7. The comparison of p651 and pRS20.7 sequences showed that p651 is 86% homologous over the entire length of the clone and is organized in a similar fashion to pRS20.7 with two 5S rDNA repeats (Figures 21 and 22). The substantial degree of divergence and difference in size are consistent with the results presented above suggesting that group 4 5S rDNA is highly variable and not subject to sequence homogenization typical of repetitive DNA sequences (Dover 1982). The difference in size is due to a 55 bp deletion in the 5S1 spacer region of p651. The region of deletion in p651 contains a number of purine tracts consisting of a guanine residue followed by a variable numbers of adenine residues (Figure 22, underlined sequence). The purine tracts form a series of small direct repeats which could
Figure 21. Restriction Site Map and Genomic Organization of p651 5S rDNA. Panel A: Organization and restriction site map of p24. E = EcoRI cloning site, S = SpeI. Solid and open boxes indicate 5S1 and 5S2 Transcription units. Triangle denotes site of deletion in 5S1. Panel B: Autoradiogram of EcoRI digested P1, L^H and S1 DNA gel blot hybridized with 4.0 kb insert of pZ6. Molecular weight sizes are in kb.
A

\[ \text{551 bp} \]

pZ6

\[ \begin{align*}
E & \quad S & \quad S & \quad S & \quad S & \quad S & \quad S & \quad S & \quad S & \quad S & \quad E \\
\end{align*} \]

p651

\[ \begin{align*}
5S1 & \quad 5S2 & \quad 5S1 \\
\end{align*} \]

55 bp Deletion

B

\begin{align*}
\text{PI} & \quad \text{L}^H & \quad \text{S1} \\
\text{kb} & \quad -23.1 & \quad -23.1 & \quad -9.4 & \quad -6.5 & \quad -5.1 & \quad -3.5 & \quad 2.0 \\
\end{align*}
Figure 22. DNA Sequence Comparison of pRS20.7 (207063) and p651 (RC1SPE9). The DNA sequences are compared from the SpeI site of p651. This required the movement of bases 1,47 to the end of pRS20.7 since it was cloned at the BamHI site. Dashed lines indicate deleted base pairs with respect to the other sequence. Underlined bases indicate purine tracks. Alignment performed with the ALIGN program distributed by DNASTAR, Madison WI.
participate in replication slippage (Scowles et al. 1988). This could result in the deletion of various size segments of the spacer depending on how much slippage occurs.

The insert of p651 was used as a hybridization probe to a DNA gel blot of EcoRI-digested P1, L^H and S1 DNA to determine if this clone uniquely identified the 4kb P1 band and its corresponding RFLP in S1. The result shown in Figure 21 B indicates that this clone is sufficiently homologous to the rest of the group 4 5S rDNA that it cannot detect itself uniquely at high stringency (Tm - 2^0). This indicates that at least some of the RFLP fragments detected in Figure 11 are composed of repeating units of pRS20.7 related sequences. Therefore the RFLPs are not likely to represent useful probes for studying the exact mechanism of DNA change, since their fate cannot be followed as unique DNA fragments under the experimental conditions employed in this study. This, however, does not diminish their utility as markers for environmentally induced heritable change in flax. Understanding the mechanisms involved in variation at this locus which is associated with environmental induction will await further long range characterization of the Flp-1 locus and flanking sequences.
8. **Flp-1 Locus Variation in Other Flax Genotypes.**

The 706 bp repeat was used to probe genomic blots of flax and linseed varieties and another Linum species, *Linum grandiflorum caeruleum* (Figure 23). In no case is the hybridization pattern observed in these lines identical to that in the genotrophs (Figure 23, compare lanes 1-4 with lanes 5-7). The RFLP pattern detected in *L. grandiflorum caeruleum* lacks several of the invariable bands detected in the genotrophs as well as showing non-genotroph polymorphisms. In addition, the overall copy number of the pRS20.7 sequence is reduced to approximately 110 copies/1C as determined by copy number reconstructions (compare lanes 1 and 7, Figure 23; Table V). RFLPs are also detected in nine other flax varieties. Of these, only two, Dakota and Abyssian Brown, appear identical for EcoRI (Figure 23, lanes 2 and 4). The others (Royal, Williston Brown, Kenya, Victory, Koto, Leona, and Barnes) show RFLPs in comparison to Pl (Barnes is shown in Figure 23, lane 2; others not shown). Several of the polymorphisms detected in Barnes resemble those present in $S^h$. These results suggest that restriction fragment length
Figure 23. Hybridization pattern detected by pRS20.7 in EcoRI digested DNA of flax varieties and Linum grandiflorum caeruleum. Lanes 5, 6, and 7 show the pattern detected in S\textsuperscript{h}, L\textsuperscript{H} and Pl respectively for comparison. Lane 1 is Linum grandiflorum caeruleum. Lane 2, 3, and 4 are Dakota, Barnes and Abyssian Brown respectively.
polymorphism at this 5S rDNA locus is high in comparison to other flax 5S rRNA gene families and the induced polymorphism is similar to that which exists naturally.
D. Identification and Characterization of RFLPs Detected by 5S-20 Junction DNA.

1. 5S-20 Junction Fragment p2010.1 Detects Low Copy Number RFLPs.

In order to determine if the 5S rDNA junction sequences of the EMBL4 clones in Figure 2 detected polymorphism between the progenitor line and the genetrophs, a DNA blot hybridization approach similar to that used in sections 3B and C was employed. The non-5S rDNA junction sequences of the EMBL4 clones, described in section 3A1, were used as hybridization probes to DNA blots of P1, large and small genetroph DNAs. These studies were conducted with the three enzymes used to map the $L^H$ clones in Figure 2 (BamHI, EcoRI and HindIII). At this level of analysis the junction sequences did not show any RFLPs or any identifiable quantitative polymorphisms, with the exception of 5S-20. The hybridization patterns obtained with the remaining junction fragments in P1 and all genetrophs were consistent with the $L^H$ restriction maps shown in Figure 2. Analysis of these clones was not pursued further since polymorphisms were not identified for the junction
fragments. The initial hybridization pattern obtained with the 10 kb junction DNA of 5S-20 did not show any differences between P1 and genotroph EcoRI DNA blots (not shown). The 1.0 kb EcoRI fragment, which borders the EMBL4 right cloning vector arm, was subsequently used as a purified, plasmid hybridization probe (p2010.1) to verify the genomic size of this fragment which is defined by the next chromosomal EcoRI site down from the right Sau3A cloning site in the 5S-20 EMBL4 clone (Figure 24; the left terminal EcoRI site in 5S-20 is donated by EMBL4 and is just to the left of the Sau3A cloning site; Sambrook et al. 1989). This DNA probe detects a strongly hybridizing 3.8 kb band in EcoRI DNA blots. The 3.8 kb fragment represents the distance from the right, nonvector, EcoRI site in 5S-20 to the next EcoRI site on the chromosome (Figure 24). Extended exposure of this DNA blot revealed that p2010.1 also detects two weakly hybridizing bands of 2.8 and 2.3 kb. The 2.3 kb EcoRI band defines a low copy RFLP which is present in L^H and S1, but absent from P1 (Figure 24). The association of the 2.3 kb RFLP with those identified by the 5S rDNA of 5S-20 suggested a relationship between the two RFLPs. If the 5S-20 junction DNA represented one end of the F1p-1 locus or an internal low copy segment, then the RFLP identified by the 1.0 kb probe
Figure 24. Identification of a Low Copy Number RFLP with the Junction DNA of 5S-20. The EMBL4 5S-20 clone map shows EcoRI, HindIII and BamHI restriction enzyme sites a white flag, black triangle flag and black flag respectively. The Sau3A cloning site is shown at the left end of the map. The black portion of the map indicates 5S rDNA (706 bp repeat). The hatched area shows the extent of the p2010.1 hybridization probe. The position of the 3.8 kb EcoRI fragment detected by p2010.1 is indicated below the 5S-20 map. The arrows on the EcoRI genomic DNA blot denote the bands detected by p2010.1.
\( \lambda \text{ 5S-20} \)

- Sau3A

- PI \( ^M \) S1

- 3.8 kb

- 2.8 kb

- 2.3 kb RFLP
may be due to rearrangements occurring in the sequences flanking or internal to this locus. This type of DNA rearrangement would be useful in determining the type of DNA alterations responsible for the polymorphisms observed at this locus. In order to determine if there is a relationship between the Flp-1 and p2010.1 polymorphisms, the 2.3 kb RFLP identified by p2010.1 was studied further by isolation of the section of the genome making up the RFLP.

2. Molecular Cloning of p2010.1 Homologous DNA

Fragments from \( L^H \): Characterization of the 2.3 kb RFLP.

Isolation of the EcoRI fragments detected by p2010.1 was accomplished by screening an \( L^H \), lambda ZapII library (described in 2I and 3C7) with the gel-purified insert of p2010.1. A total of 9 positive plaques were identified from a screen of two genome equivalents of lambda ZapII recombinants (Clarke and Carbon 1976; 250,000 plaques). The positives showed strong and weak signals as would be expected from the hybridization pattern in Figure 24. The 9 positives were plaque purified and phage DNA isolated for restriction enzyme analysis before plasmids were rescued.
due to the difficulties encountered in section 3C7. Digestion of the phage DNAs with EcoRI produced one 3.8 kb band for clones 1 and 8 and two or more bands for clones 3, 4, 5, 6, 7, 9 and 10. The presence of multiple bands is due to the nonselective cloning strategy of lambda ZapII (see section 2I). The DNA gels were blotted to a nylon membrane and hybridized with the labeled insert of p2010.1. The resulting autoradiograph shows that all of the EcoRI fragments detected by p2010.1 in a genomic DNA blot (Figure 24) are represented in addition to a 3.9 kb fragment (clone 7; Figure 25). Clones 1, 8 and 10 contain a strongly hybridizing 3.8 kb band which is homologous to the 3.8 kb band identified by p2010.1 in Figure 24 (Figure 26, A). Clone 9 contains a weakly hybridizing 2.8 kb EcoRI fragment in addition to 3 other nonhybridizing bands. Finally, the 2.3 kb EcoRI fragment is represented by clones 4, 5 and 6, each of which also contain smaller nonhybridizing bands. The inserts of the recombinant lambda ZapII clones were rescued in the form of pBluescript plasmids as described in Chapter 2I. Miniprep DNA analysis of 5 independent colonies for each clone was consistent with the results obtained with the purified phage DNAs. Therefore, in contrast to the results obtained in section 3C, 7, the lambda ZapII
Figure 25. DNA Blot Analysis of p2010.1 selected, Lambda ZapII, L^H, EcoRI Clones. The numbers above the lanes indicate clone designations. Lambda ZapII DNA isolations from each clone were digested with EcoRI, electrophoresed and blotted. The DNA blot was hybridized with p2010.1. Hybridized bands are indicated by arrows and size in kilobases.
rescue protocol did not generate any aberrant plasmids. The 2.3, 2.8 and 3.9 kb p2010.1 homologous EcoRI bands were subcloned from the other unrelated DNA fragments in the primary clones. Subcloning was not necessary for the 3.8 kb fragment since clones 1 and 8 contained only the 3.8 kb fragment. The resulting plasmids were named pZLH2.3, pZLH2.8 pZLH3.8 and pZLH3.9. The nomenclature refers to the original cloning vector, lambda ZapII, the source DNA for cloning, L^H, and the size of the insert, 2.3 kb, etc. This nomenclature is followed for the rest of this section with reference to the other clones generated during the course of this study.

Hybridization of the clones representing nonpolymorphic bands in Figure 24 (2.8 and 3.8), in addition to the 3.9 kb clone, to EcoRI DNA blots of Pl, L^H, and S1 DNAs at Tm - 19^0, resulted in identification of the appropriately sized bands for each of the clones (Figure 26). The 2.8 and 3.9 kb bands detect in addition to their cognate band size, many less intense bands which form a dispersed hybridization signal, especially for pZLH3.9 (Figure 26, B and C). This hybridization is presumably due to the presence of some repetitive sequence in both clones. The 3.9 kb band, represented by clone 7, was not visible in Figure 24 due
Figure 26. Genomic DNA Blot Analysis of pZLH3.9, -3.8 and -2.8. EcoRI genomic DNA blots of Pl, LH and S1 DNAs were hybridized individually with labeled inserts of the plasmids pZLH3.9, -3.8 and -2.8. The EcoRI band corresponding to the size of the clone is indicated by an arrow and the size in kilobases.
to the intense hybridization of the 3.8 kb band. The pZLH3.8 hybridization pattern shows much reduced hybridization to the 2.8 and 2.3 kb EcoRI bands in comparison to p2010.1 (Compare Figure 24 to Figure 26, B). This indicates that the homology responsible for the detection of the 2.3 and 2.8 kb EcoRI fragments resides predominantly in the p2010.1 portion of pZLH3.8 (Figure 24).

The genomic organization of pZLH2.3 which represents the RFLP identified by p2010.1, was investigated by EcoRI DNA blot analysis of Pl, L^H and S1 DNAs. When the entire 2.3 kb is used as a probe at Tm - 19^0 two predominant 2.3 kb bands in L^H and S1 are detected in addition to a highly dispersed hybridization signal (Figure 27). Comparison of the Pl lanes with the L^H and S1 lanes shows that a 2.3 kb band is not present in Pl. However, a 3.7 kb band is detected in Pl which is absent from the L^H and S1 lanes (Figure 27). The 3.7 kb band is further defined after raising the stringency of the post hybridization wash to Tm - 1^0 (Figure 27, B). The presence of the 3.7 kb band in Pl and its absence from L^H and S1 strongly suggests that the 3.7 kb band is the progenitor of the 2.3 kb genotroph RFLP. The 3.7 kb band in Pl also shows reduced signal intensity compared to the 2.3 kb bands in L^H and S1,
Figure 27. Identification of a Pl, EcoRI Fragment Homologous to the pZLH2.3 RFLP clone. Panel A; Autoradiograph of an EcoRI Genomic DNA blot hybridized with the insert of pZLH2.3 and washed at low stringency (Tm = 19^0). Panel B; The filter from panel A rewashed at high stringency (Tm = 1^0). The major bands are indicated by arrows and size in kilobases.
suggesting that the 3.7 kb P1 band and pZLH2.3 do not share complete homology with each other. The results suggest that the RFLPs present in $L^H$ and S1 are due to a DNA alteration which has produced a band shift as well as a change in DNA homology between the progenitor and genotroph DNAs. This type of polymorphism could be generated in a number of different ways; the loss and gain of EcoRI restriction sites in the region of the progenitor 3.7 kb EcoRI fragment could generate a new fragment with only partial homology to the original 3.7 kb band. Alternatively, recombination, deletion or insertion of foreign DNA could result in the observed changes in band size and homology. In order to differentiate between these possibilities and determine the nature of the polymorphism, the isolation of the 3.7 kb P1 band was pursued.


The P1, 3.7 kb EcoRI fragment identified by pZLH2.3 in Figure 27 was isolated from the P1 12/13 library described in Chapters 2I and 3C, 7. The library was hybridized with a gel purified, $^{32}$P labeled 1.2 kb SspI - EcoRI fragment of pZLH2.3 (derived from the left EcoRI site in Figure 28). The library filters were washed at
high stringency (Tm − 1°) and two positive plaques were identified (A and B). The two positives were plaque purified and the inserts rescued as pBluescript plasmids directly, since no problems were encountered with the isolation of clones described in section 3D, 2.

Restriction enzyme digests of the two clones with EcoRI indicated that clone A contained a single 3.7 kb insert while clone B contained 3.7 kb and 2.6 kb insert sizes. DNA blot analysis of this gel indicated that the 3.7 kb band in each clone hybridized to pZLH2.3 but not the 2.6 kb fragment in clone B (not shown). Clone A, pZPl3.7, was characterized further as it contained only one insert.

In order to determine the extents of homology between pZLH2.3 and pZPl3.7, restriction enzyme site maps for each clone were generated and then compared. Comparison of the maps in Figure 28 indicates that pZLH23 and pZPl3.7 contain similar restriction enzyme sites over an approximately 900 bp region from the left terminal EcoRI site to just prior to the SspI site in pZPl3.7 (Figure 28). In contrast the right sides of the clones do not show any similarity in restriction enzyme sites. This suggests that either a recombination, deletion or insertion event has occurred since an explanation based on the loss and gain of restriction
Figure 28. Comparison of Restriction Enzyme Site Maps of pZLH2.3 and pZPL3.7. Restriction enzyme sites are indicated below the figure. The hatched box indicates pZPL3.7, non-homologous DNA. The probes used in Figure 37 are indicated above the pZLH2.3 map. The sequenced regions of pZLH2.3 and pZPL3.7 are indicated by arrows below the maps. The number of nucleotides determined are indicated at the end of the arrow. The point of insertion of LIS-1 and the number of nucleotides from the left EcoRI site, in pZLH2.3, is shown at the border of the open and hatched boxes.
enzyme sites is not consistent with the structure of the pZLH2.3 map in comparison to pZPl3.7. The latter explanation predicts that the left end EcoRI site in pZLH2.3 would be displaced farther down the chromosome to the left and the right end EcoRI site be placed near the point were the two clones diverge. However, the restriction site maps in Figure 28 do not reflect this organization.

The identity of pZPl3.7 to the 3.7 kb EcoRI band detected by pZLH2.3 was tested by hybridizing pZPl3.7 to a duplicate DNA blot of the one used in Figure 27 (Figure 29). The probe detects a 3.7 kb band in PI as well as several other weaker bands which are also present in $L^H$ and S1. A weak 2.3 kb and a strong 6.2 kb band are also identified by pZPl3.7 in $L^H$ and S1. The 2.3 kb band coincides with the size of pZLH2.3. The identity of pZLH2.3 and the 2.3 kb band in Figure 29 is strengthened by the reversed ratio of 2.3 to 3.7 kb band hybridization observed with pZPl3.7 in comparison to the ratio observed when pZLH2.3 is used as the probe (Figure 27). This reduced level of hybridization is consistent with the lack of restriction site similarity of pZLH2.3 and pZPl3.7 over more than 2/3 of the pZLH2.3 restriction site map (Figure 28). The identification of the 6.2 kb fragment in $L^H$ and S1 suggested that this
Figure 29. Genomic DNA Blot Analysis of EcoRI Digested P1, L^H, and S1 DNAs hybridized with pZP13.7. Bands identified by the probe are indicated by arrows with the size given in kilobases. The starred arrow indicates the position of a 5.8 kb band which was subsequently cloned in section 3D4.
fragment contains the region of pZPl3.7 not represented in pZLH2.3. This prediction was tested by isolation of the 6.2 kb EcoRI fragment from the L^H lambda ZapII library using pZPl3.7 as a hybridization probe.


The 6.2 kb EcoRI fragment identified by pZPl3.7 in Figure 29 was isolated from the lambda ZapII, L^H library used to isolate pZLH2.3 in section 3D, 2. The library was screened by hybridization with a ^32P labeled insert of pZPL3.7. Six positive clones were isolated, plaque purified and rescued as pBluescript plasmids. Restriction enzyme analysis with EcoRI and gel electrophoresis showed the presence of 6.2 kb fragments in clones 1 and 3, which also hybridized strongly to pZPl3.7 upon DNA blot analysis. Clones 2 and 5 contained 2.3 kb bands which also hybridized strongly to pZPl3.7. Clones 2 and 5 were subsequently shown to be identical to pZLH2.3 by hybridization and restriction mapping (not shown). Therefore, the 2.3 kb band identified by pZPl3.7 in Figure 29 is the equivalent of pZLH2.3. The remaining clones, 4 and 6 contained a 5.8 kb band which hybridized weakly to pZPl3.7. This
fragment presumably represents the 5.8 kb band identified in $L^H$ and S1 in Figure 29 (starred band). Clone 3, PZLH6.2 was analyzed further as it contained only the 6.2 kb fragment. Hybridization of the entire 6.2 kb insert of pZLH6.2 to EcoRI genomic DNA blots of P1, $L^H$ and S1 DNAs detected the 3.7 kb band in P1, and 6.2 kb bands in $L^H$ and S1 at high stringency (Tm $-1^0$; Figure 30, B). In addition to these bands a 5.8 kb band was identified as well as a number of less intense bands. Lower stringency hybridizations resulted in the identification of a large number of bands with weak signal intensity (Figure 30, A). These fragments form a dispersed hybridization signal which covers a wide range of molecular weights; 1.0 to 30 kb. This result will be returned to below. PZLH6.2 was restriction mapped with the same enzymes used for pZLH2.3 and pZPL3.7 to determine the regions of similarity between the clones. Comparison of the pZLH6.2 map with pZPL3.7 indicates that the 6.2 kb fragment overlaps with pZPL3.7 over an approximately 2.8 kb region (Figure 31). This is the same region which is not contained in pZLH2.3. PZLH6.2 also contains 3.0 kb of nonoverlapping sequence which starts in the region were pZLH2.3 and pZPL3.7 diverge (Figure 31 and 32). To verify that the regions of similarity in pZLH6.2 and pZPL3.7 are homologous, an 800
Figure 30. Genomic DNA Blot Analysis of EcoRI digested, P1, L^H and S1 DNAs hybridized with pZLH6.2.
Panel A: Hybridization with entire 6.2 kb insert of pZLH6.2 with a low stringency wash Tm – 19^0. Panel B: Filter from panel A, rewashed at Tm – 1^0. Panel C: Hybridization of an 800 bp NsiI/NsiI probe from pZLH6.2. Position of the probe is shown in Figure 32.
bp region of pZLH6.2 (Figure 31), which overlaps with pZPL3.7, was used as a hybridization probe to a duplicate genomic DNA blot of that in Figure 30, A and B, at low stringency (Tm = 19°C). The probe detects the 6.2 kb and 3.7 kb EcoRI bands as expected, in addition to the 5.8 kb band. However, the dispersed hybridization signal detected with the entire 6.2 kb insert is not detected. The experiments described in section 3D, 7 suggest that the dispersed hybridization signal is due to the non-pZPL3.7 homologous DNA in pZLH6.2.

Reconstruction of the LH allele with reference to the pZLH2.3 and pPL3.7 maps is depicted in Figure 32. The reconstructed RFLP suggests two methods for its production. The first is by the interruption of the 3.7 kb Pl EcoRI band by an insertion sequence of approximately 5.3 kb. The second is by a recombination event which splits the 3.7 kb EcoRI fragment into two fragments which are now juxtaposed to unrelated DNA segments. To determine which of these two possibilities is correct, the expected and observed DNA fragment sizes, based on the presence of an insertion sequence, were compared in a BamHI DNA blot, hybridized with pZLH2.3. The expected fragment sizes are 2.6 kb for Pl and 7.2 kb for LH and S1 based on the map in Figure 32.
Figure 31. Comparison of the Restriction Enzyme Site Maps of pZPl3.7 and pZLH6.2. Restriction enzyme sites are indicated below the figure. The hatched box indicates pZPl3.7, non-homologous DNA. The probes used in Figures 31 and 38 are indicated above the pZLH6.2 map. The sequenced region of pZLH6.2 is indicated by an arrow below the map and the number of nucleotides (nt) determined at the end of the arrow.
Figure 32. Reconstruction of $L^H$ and Pl pZPL3.7 homologous alleles: Composite Restriction Site and Clone Maps. Restriction enzyme sites are indicated below the figure. The hatched box indicates pZPL3.7, non-homologous insertion sequence DNA. The extents of the clones pZLH2.3, pZLH6.2 and pZPL3.7 are indicated above and below the respective clones. The insertion point of LIS-1 is shown by the two lines extending from the ends of the insertion sequence to pZPL3.7.
Figure 33. Genomic DNA Blot verification of $L^H$ and P1 Restriction Site Maps. A BamHI genomic DNA blot of P1, $L^H$, S1, C1, C2 and $S^h$ was hybridized with the entire insert of pZIH2.3 and subsequently washed at high stringency ($T_m - 10^0$). The position and sizes of the observed bands are indicated in kilobases.
The observed pZLH2.3 homologous fragment sizes, shown in Figure 33, are consistent with this organization and the restriction site maps of the $L^H$ and P1 alleles in Figures 31 and 32. DNA blot analysis with EcoRV also produces the expected restriction fragment sizes (not shown). Therefore, the conclusion is that the RFLP originally identified by p2010.1 is due to a DNA insertion sequence which will be referred to as LIS-1 (Linum Insertion Sequence-1).

The relationship between p2010.1 and the RFLP is not clear since the flanking sequence to p2010.1 contained in pZLH3.8 does not appear to contain any of the rearrangement described in this section (Figure 24). Linkage analysis was used to establish if the Flp-1 locus and LIS-1 RFLPs are genetically linked to each other. The L1 X L6 F2 population used in section 3C, 5 was analyzed for segregation of the LIS-1 RFLP. The data indicates that Flp-1 and the LIS-1 loci are not linked (recombination = 45\% \(+/- 10\%)\).

5. **Sequence Analysis of the Insertion Sequence Target Site and Junctions.**
Insertion sequences such as the one described in this section have commonly been shown to represent transposable element (TE) related sequences (Voytas and Ausubel 1988; Harberd et al. 1987; Martienssen and Baulcombe 1989). Many transposable elements are distinguished by the presence of terminal inverted repeats at either end of the element and duplication of a characteristic number of nucleotides at the site of insertion referred to as the target site (Shapiro 1983). The DNA sequence at the borders of the insertion sequence was determined and compared to the sequence of pZPl3.7 in the region of the insertion. The regions of pZLH2.3, pZLH6.2 and pZPl3.7 which were sequenced are shown in Figure 28 and 31. This analysis allowed definition of the site of insertion at 611 bp from the left end of pZPl3.7 (Figure 28). Comparison of the Pl insertion site with the sequence flanking the insertion sequence’s right and left borders indicated that a 3 bp sequence, TCC, was duplicated (Figure 34). The TCC sequence to the right of the poly T tract is maintained at the pZLH6.2 side of the insertion as well as at the pZLH2.3 proximal side. Duplications as small as 3 bp have been identified for other plant transposable elements such as Tam-1 (Bonas et al. 1984). There are also several sequence deviations in the region
Figure 34. DNA Sequence Analysis of Insertion Sequence Target Site and Termini. The $L^H$ and Pl sequences are indicated to the left of the sequences. The target site is underlined in both Pl and $L^H$ (duplicated in $L^H$). The terminal 19 bases of the left and right ends of the element are shown vertically above the site of insertion. Watson-Crick base pairs between the termini are depicted as black dots. Direct and inverted TCC repeats are shown as right and left ended arrows respectively.

$H$ 5' GGGAGGTATTGGAGAGAC TCC TCC GGT

L 3' TTTTTTTTTTTTCA TATC GGT

Pl 5' GGGAGGTATTGGAGAGAC TCC TTTTTTTTTTTCTGTATATTG

3'
surrounding the target site (Figure 34). These deviations may be due to unrelated point mutations or to aberrations caused by the putative activity of LIS-1.

Comparison of the terminal 19 bp of each end of the insertion sequence did not identify any inverted repeats. Therefore, although the target site shows a duplication of sequence, inverted repeats are not present at the termini of the insertion. Cross hybridization experiments indicate a small degree of homology between the terminal regions of the insertion (not shown). This may indicate that there are homologous regions at the termini of the element. However, further sequence analysis is required to clarify this point.

6. Identification of LIS-1 RFLPs in other Genotrophs and Flax Cultivars.

The prevalence of LIS-1 related sequences in other flax genotrophs and cultivars was investigated through DNA blot hybridization with pZLH2.3. Low and high stringency hybridization patterns detected by pZLH2.3 in P1 and genotroph DNAs are shown in Figure 35, A and B. At low stringency the probe detects predominant bands at 2.3 and 3.7 kb in addition to a large number of less
intense bands in all of the genotrophs. The 2.3 and 3.7 kb bands represent pZLH2.3 and pZPl3.7 and are the only species to remain hybridized to the probe at high stringency ($T_m - 1^0$; Figure 35, B). The presence of the 2.3 kb band indicates that LIS-1 is inserted in the 3.7 kb EcoRI band represented by pZPl3.7. The striking result from this experiment is that all of the small genotrophs and $L^H$ contain the LIS-1 sequence inserted in the 3.7 kb EcoRI fragment; L1, C1 and C2 do not. This result parallels that observed with the pRS20.7 RFLPs except that the $L^H$ pattern is the same as the small genotrophs and L1 does not have the LIS-1 insertion. In addition the insertion point would appear to be the same since the 2.3 kb band size appears identical to $L^H$ and S1 (Figure 35, B). If LIS-1 were inserted elsewhere in the 3.7 kb fragment, then the band hybridizing to pZLH2.3 would be either larger or smaller depending on the placement of LIS-1 in pZPl3.7. This result suggests that LIS-1 has inserted into a very similar location in 5 independent cases ($L^H$, $S^h$, S1, C3 and L6). In addition all of the genotrophs and P1 are homozygous for either the 3.7 kb progenitor band or the LIS-1 insertion. The historical derivation of the genotrophs suggests that LIS-1 has inserted into this locus in association with environmental induction. The
Figure 35. EcoRI DNA Blot Analysis of P1 and Genotroph DNAs for the Presence of LIS-1. The blot was hybridized with the 2.3 kb insert of pZLH2.3 and washed at low stringency (Panel A) and subsequently at high stringency (Panel B). The lanes corresponding to P1 and genotroph DNAs are indicated above the lanes. The positions of the 2.3 kb and 3.7 kb bands are shown with arrows and size. The high molecular weight RFLP is shown by an additional arrow in Panel A. Molecular weight standards are from HindIII digest of lambda DNA.
B

PI  L^H  L1  L3  L6  S1  S3  S6  S^H  C3  C2  C1  kb

←3.7

←2.3
Figure 36. DNA Blot Analysis of Flax Cultivars for the presence of LIS-1 sequences. An EcoRI genomic DNA blot of flax cultivars was hybridized with the 2.3 kb insert of pZLH2.3 and washed at low stringency (Tm \(-19^\circ\))

Lane 1, Hollandia; lane 2, Stormont Cirrus; lane 3, Liral Monarch; lane 4, Liral Prince; lane 5, Rembrandt; lane 6, Stormont Motley; lane 7, Stormont Gossamer. Molecular weight markers are from the HindIII digest of lambda DNA. The position of the 2.3 kb and 3.7 kb bands is indicated on the left side in addition to the position of a high molecular weight RFLP.
homozygous nature of this polymorphism is similar to the results obtained in section 3C for 5S rDNA polymorphisms. The LIS-1 insertion has not been detected in any of the laboratory P1 stocks used for induction experiments. Although most of the genotroph lines have been propagated for several generations after induction, the DNA used for genotroph C3 represents the first generation progeny of induced plants. Therefore, the insertion event must have occurred close to the time of induction and resulted in a homozygous condition. Close inspection of the low stringency DNA blot in Figure 35 reveals the presence of another polymorphic band. This fragment, at approximately 20 kb (arrowed band in Figure 35), is present in P1, L'H, S'h and C2 but missing from L1, L3, L6, S1, S3, S6 and C3 (undetermined in C1). Therefore, all genotrophs are polymorphic for sequences related to LIS-1.

One feature of transposable elements which has been useful in their study is reversion of the insertion through excision of the element (Shapiro 1983). This usually results in the reestablishment of the wildtype preinsertion fragment size (Martin et al. 1985). Close inspection of Figures 29, 30, and 35, B, reveals the presence of a 3.7 kb band in S1 which is a fraction of the intensity of the 2.3 kb band in S1 and the 3.7 kb
band in P1. This fragment could represent a somatic reversion event which recreates the 3.7 kb EcoRI fragment by excision of LIS-1. The possibility of contamination with P1 DNA would seem unlikely since this same DNA does not show P1 band patterns with pRS20.7 in Figure 12. Further experiments will be required to determine if this insertion sequence is a mobile DNA element.

The LIS-1 sequence is also present in other flax cultivars as shown in Figure 16. However, analysis of 5 Hollandia individuals for LIS-1 indicate that they are homozygous for either the 3.7 kb band or the insertion profile (not shown).

7. Genomic Organization of LIS-1 in P1 and Genotrophs.

The genomic organization of LIS-1 was investigated to determine if multiple copies of the insertion sequence are present in the flax genome and the region of LIS-1 which has homology to p2010.1. The DNA blot studies of pZLH2.3 shown in Figure 28 identified the 2.3 and 3.7 kb bands at high stringency and many less intense bands at low stringency (Tm - 19°). Hybridizations to duplicate EcoRI DNA blots of P1, LH and S1 with subfragments of pZLH2.3 and pZLH6.2 were performed to identify LIS-1
sequences only (probes shown in Figures 28 and 31) . Figures 37 and 38 show the respective autoradiographs obtained with these probes. The general result obtained from all of the probes is that LIS-1 sequences are either not present in P1 or have an altered sequence context such that they are not identified by the probes at high stringency (Tm - 10; Figure 37 and 38, A.2 and B.2). For example, all 4 probes, which cover 90% of LIS-1, show little to no detectable hybridization in the P1 lanes at high stringency. Several weakly hybridizing bands are indicated by arrows in Figures 37 and 38. These bands appear to be present in both P1 and genotrophs. In contrast to the species identified at high stringency, the low stringency hybridizations show multiple to many bands depending on the probe. The 280 bp EcoRV/EcoRV probe (probe-1) detects a 3.8 and 2.8 kb EcoRI band in addition to the 2.3 kb band (Figure 37, B.1) and also displays the strongest hybridization to p2010.1 (not shown; Figure 24). Therefore, the 3.8 and 2.8 kb fragments identified by probe-1 probably represent hybridization to pZLH3.8 and pZLH2.8 as a result of its homology to p2010.1. Although undetectable in Figure 37, B.2, probe-1 also detects these bands to a small extent at high stringency. Unlike probe-1, probes 2, 3 and 4 do not identify the
3.8 and 2.8 kb bands. This result is consistent with the homology of p2010.1 residing in the 280 bp probe-1 region of LIS-1. Probe-3 also detects a number of weak bands in addition to the 3.7 Pl and 6.2 kb L\(^H\) bands at both low and high stringency. The 500 bp SspI/SspI probe (probe-2) detects the 2.3 kb band and a large number of less intense, highly dispersed bands of equal intensity in Pl, L\(^H\) and S1. The high stringency blot results in loss of the dispersed signal and retention of the 2.3 kb band in L\(^H\) and S1 only. Similar results are obtained with probes-3 and -4 from pZLH6.2 (Figure 38). Interestingly, the 600 bp NsiI/NsiI probe-4 also detects a dispersed hybridization signal similar to the one identified with probe-2 (Compare Figures 37 A.1 and 38 B.2). The position of probe-4 on the LIS-1 restriction map is similar to the position of probe-2 at the other end of the element. The loss of the dispersed hybridization signal at high stringency for probes-2 and -4 suggests that the hybridization is due to a degenerate repetitive sequence in LIS-1. Sequence analysis will be required to determine if the similar hybridization profiles of probes-2 and -4 are due to sequence similarity. This possibility would be significant in terms of locating transposable element like symmetry at the ends of LIS-1.
Figure 37. Genomic Organization of LIS-1 using pZLH2.3 derived probes. Probes 1 and 2 shown in Figure 29 were hybridized to duplicate EcoRI genomic DNA blots of P1, L\(^H\) and S1 DNAs separately. Panel A, 1 and 2, shows the autoradiographs obtained with probe-2 at low and high stringency respectively. Panel B, 1 and 2, shows the autoradiographs obtained with probe-1 at low and high stringency respectively. The position of the 2.3 kb band is indicated as well as the 2.8 and 3.8 kb bands in panel B.
Figure 38. Genomic Organization of LIS-1 using pZLH6.2 derived probes. Probes 3 and 4, shown in Figure 32, were hybridized to duplicate EcoRI genomic DNA blots of P1, L⁻¹ and S1 DNAs separately. Panel A, 1 and 2, shows the autoradiographs obtained with probe-3 at low and high stringency respectively. Panel B, 1 and 2, shows the autoradiographs obtained with probe-4 at low and high stringency respectively. The position of the 6.2 band is indicated as well as the position of a 3.7 kb band in panel A.
Chapter 4. **Discussion**

At the time this study began, little information was available concerning the specific mechanisms involved in DNA alterations associated with environmentally induced heritable changes in flax. In addition, no low copy DNA polymorphisms had yet been described. The similar phenotypes of plants induced under the same conditions suggests that specific DNA alterations had occurred in each of these individuals. Results obtained by other workers showed that several isozyme polymorphisms were specific to L or S flax genotrophs (Fields and Tyson 1973b; Cullis 1979). Molecular studies have shown that total DNA content changes are directly associated with induction of heritable changes (Evans et al. 1966). In addition, repetitive sequence alterations have also been directly associated with environmental induction (Cullis and Charlton 1981). However, the mechanisms whereby the alterations are produced have not been elucidated. This study was initiated to determine if specific DNA alterations are associated with particular genotroph phenotypes and to attempt to determine the mechanism(s) by which the changes were produced. These studies are important for understanding and developing testable
models for how the environment affects genome stability. The first section discusses the characterization of 5S rDNA genomic clones which were subsequently used to assay flax genotrophs for DNA polymorphisms. The following sections discuss the results of characterization of 5S rDNA and low copy DNA rearrangements, respectively.

A. **Organization of 5S rDNA in flax: Does a dispersed organization facilitate sequence divergence in a repetitive gene family?**

The results from the analysis of flax 5S rRNA gene genomic clones in section 3A, 1-3, show similarities and differences to those from other species with respect to sequence, organization and evolution. Previous studies on the organization of flax 5S rDNA indicated that this class of genes was organized similarly to other eukaryotes with a gene-spacer tandem arrangement and little sequence variation between repeats (Goldsbrough et al. 1981; Goldsbrough et al. 1982). The restriction maps of the EMBL4 clones in Figure 2, in addition to in situ localizations performed in parallel with these studies, indicate that the 5S rDNA in flax is
distributed over many chromosomes (Schneeberger et al. 1989). Failure to recover clones of complete tandem arrays of group 1 or 2 type sequence may be due to clone instability, methylation, or the lack of many such large arrays in the genome. All but one of the genomic clones (5S-12) contained non-5S rDNA junction fragments in addition to 5S rDNA, which is consistent with a dispersed organization. This result is in contrast to all other angiosperms studied to date, in which the 5S rDNA is localized to one to three pairs of chromosomes (Mascia et al. 1981; Appels et al. 1980; Ellis et al. 1988; Knalmann and Burger 1986). The only other plant to show a dispersed chromosomal organization is the gymnosperm, *Pinus radiata* (Gorman, personal communication).

The homology determinations shown in Figure 3 and organization of the genes in the different genomic isolates allowed classification of the clones into 5 groups which are summarized in Table I. Each group can be identified specifically on genomic DNA blots by hybridization at high stringency (Figure 4). The sequence determinations of at least one gene repeat from each group are in good agreement with the relative homologies of the clones obtained in Figure 3 (Table I). The genomic organization and sequence homology results
demonstrate that there are two major classes of 5S rRNA in flax, group 1 and 2, which differ predominantly in the spacer sequence and are almost identical in length and 5S rRNA coding sequence (Figures 2, 3 and 4; 342 and 341 bp respectively; Table IV). The relative hybridization intensities shown in Figure 4 demonstrate that group 1 and 2 clones make up the majority of the 5S rRNA genes in the flax genome represented by this sampling of clones. The data described here are consistent with that previously described for flax (Goldsbrough et al. 1981; Goldsbrugh et al. 1982). Comparison of the 5S rRNA sequences of groups 1 and 2 to the consensus angiosperm sequence indicates that the pBG13 sequence is most closely related to other plant 5S rRNAs. The pBG13 sequence is therefore assumed to be the progenitor of the other 5S rRNA sequences in the flax genome. In addition to groups 1 and 2, there are three other groups which show reduced homology to group 1 in both the gene and spacer sequence (groups 3-5; Tables III and IV). These groups also have low copy numbers as evidenced by the much longer exposure required to visualize these groups in comparison to groups 1 and 2 (58 times longer exposure; Figure 4). Group 4 5S rDNA is represented by approximately 500 copies per haploid genome (Table V). Groups 3 - 5 have
distinctive types of genomic organization. Group 3 is organized in tandem arrays defined by XhoI, of similar size to group 1 and 2 (Table I). Group 3 also shares the greatest degree of spacer and gene region homology with pBG13 (Tables II, III and IV). Group 4 is also organized in tandem arrays. However, the family is composed of two alternating repeats of 353 bp, which are 76% identical (Figures 9 and 10). Sequence comparisons to another group 4 member, p651, and the results presented in section 3C, indicate that group 4 5S rDNA contains a high degree of sequence and repeat length divergence (Figure 22; also discussed below in 4B, 2). Group 5 is a collection of 5S rDNAs which do not have a characteristic repeat structure and have low homology to pBG13 (Figure 2; Table I). In view of the differences in sequence and structure shown by groups 3, 4 and 5, and their low representation in the genome, these clones may represent pseudogenes which are not under any selective pressure. This would explain their divergence away from the main body of 5S RNA genes represented by groups 1 and 2. The predicted 5S rRNA secondary structures of group 3 - 5 5S rDNA is consistent with the suggestion that they are pseudogenes (Figure 6). Based on comparisons to mutagenesis experiments in Xenopus, which have defined base pairs important for 5S
rRNA structure, TFIIB binding, and canonical 5S rRNA structures, groups 3 - 5 show multiple helix destabilizing mutations as well as mutations in the hinge region (Romaniuk et al. 1989; Romby et al. 1990; Erdmann and Wolters 1988; Figure 6). These mutations would presumably affect the structure and function of the molecule in a negative fashion. However, to further distinguish genes from pseudogenes, it will be necessary to determine which genes are expressed and if there is a heterogeneous population of 5S RNA sequences present in the cytoplasm of flax cells. Approximately seven types of 5S RNA sequences have been identified in Neurospora crassa ribosomes (Selker et al. 1985).

Much of the interest in repetitive genes lies in the observation that the spacer region and the coding region of tandemly arranged genes display different forms of sequence evolution (Brown et al. 1972; Dover 1982). In general, the gene sequence is highly conserved within and between species, while the spacer sequence is only conserved within species. Sequence evolution of this type has been termed concerted or horizontal evolution (Zimmer et al 1980; Brown et al. 1972). This form of evolution requires a mechanism to maintain the homogeneity of repetitive genes and prevent the random accumulation and spread of mutations.
throughout the population. 5S rRNA gene organization has been useful in studying this phenomenon in that the genes are relatively easy to isolate and the size of the gene is generally small, permitting rapid sequence determinations.

Studies within *Xenopus* species have demonstrated that the major oocyte and somatic gene classes have different spacer sequences which are conserved throughout members of the same class, but not between classes. In addition, comparison of the somatic and oocyte genes of *X. laevis* to *X. borealis* indicates that the 5S rRNA sequence is conserved to a high degree between classes and species, but the spacer regions are divergent (Fedoroff and Brown 1978; Fedoroff 1979). The somatic genes are located at the telomeres of most chromosomes and the oocyte genes are located on one chromosome (Pardue et al. 1973). In wheat, 2 5S rRNA gene loci have been identified (Appels et al. 1980; Reddy and Appels 1989). Sequence determinations and in situ experiments indicate that these two loci have independent spacer evolution but concerted gene evolution (Reddy and Appels 1989; Scowles et al. 1988). These results typify those obtained in other eukaryotes which show that while there may be several classes of a tandemly repeated gene with divergent spacers, the gene
sequences are very highly conserved within and between species (Long and Dawid 1980; Fedoroff 1979).

As stated above, the dispersed organization of 5S genes in flax appears to be unique among angiosperms so far investigated. In addition, studies of intraspecific 5S rRNA gene variation in other organisms have not described levels of divergence comparable to those in this study, especially where the coding region is concerned (Scowles et al. 1988; Fedoroff and Brown 1978; Gottlob-McHugh et al. 1990; Erdmann and Wolters 1988). The lack of group 1 – 2 coding region homology in groups 3 – 5 indicates that these 5S genes are not evolving in a concerted fashion with the majority of the 5S rDNA (Table III). Thus the organization and sequence divergence of 5S rDNA in flax presents a problem in terms of mechanisms for maintenance of repetitive gene homogeneity. The most favored mechanisms for the maintenance of homogeneity of repetitive genes are unequal crossover, deletion and reamplification, and gene conversion (Nagylaki and Petes 1982; Dover 1982). A simple model based on homologous but uneven crossing over such as that described by Ohta (1976) and Smith (1976) is plausible for systems in which the repetitive sequences are present in one to several large arrays such as in all other angiosperms. In flax this model
may operate at a multi-locus level where only sequences on homologous chromosome positions are corrected with each other. In this scheme sequences which have become separated from main functional arrays to different chromosomal locations can no longer participate in direct homologous exchange. Here sequences may be subject to increased accumulation of mutations and thus diverge at a rapid rate. Precise chromosomal locations for individual clones would need to be identified to test this model. However, as noted above for *Xenopus*, dispersion of the 5S rDNA on many chromosomes does not necessarily equate with sequence divergence. Alternatively, the identification of a relatively large number of pseudogene classes in flax may be due to the large size of the gene family (>50,000 per 1N) and the cell’s inability to homogenize all sequences.

In spite of the high degree of sequence and organization divergence of groups 3-5 from the majority of flax 5S rDNA, these groups still show higher coding region homology to pBG13 than the spacer region (Table III and IV). This observation suggests that either the 5S rRNA gene sequence is more resistant to mutation than the spacer; or that selection for different variants during evolution has resulted in the production of a number of different sets of 5S rDNA sequences.
Therefore, the maintenance of sequence homology between repeated genes is likely to depend on multiple factors, one of which may include chromosomal distribution.

B. Identification and characterization of 5S rDNA polymorphisms between P1 and genotroph DNAs.

1. Quantitative Variation in Flax 5S rDNA.

The main goal of this study was to isolate and characterize 5S rDNA associated polymorphisms which could be used to ask questions concerning environmentally induced heritable changes in flax. In a survey of groups 1 – 5 5S rDNA, quantitative variation was only found for groups 1 and 2 (Figure 7). The results of quantitative DNA blot analysis indicate that group 2 sequences are reduced in L^H, S1, S6, C3, C2, C1, and L6 by approximately one third in comparison to P1 (Figure 7, A). In contrast, group 1 5S rDNA appears to be slightly reduced in S1, C3, C2 and C1 and unchanged in the other genotrophs. Previous studies of 5S rDNA copy number in P1 and flax genotrophs were carried out using labeled cytoplasmic 5S rRNA and pBG13 as hybridization probes (Goldsbrough et al. 1981; Cullis
and Cleary 1986, respectively). The study employing labeled 5S rRNA from P1 tissue culture cells showed approximately 50% reductions in $L^H$ and S1, and an 86% reduction in $S^h$. The study using pBG13 produced different results. In this case $L^H$ showed a slight increase over P1 and S1 showed a 73% reduction ($S^h$ was not assayed). The relative hybridization intensities obtained for Group 1 5S rDNA (i.e. pBG13) are broadly consistent with those obtained previously with pBG13 (Figure 7, B; Cullis and Cleary 1986). The results obtained with group 2 are more consistent with those obtained using labeled cytoplasmic 5S rRNA, although the amount of S1 hybridization appears less than that previously obtained (Goldsbrough et al. 1981). This result may suggest that group 2 sequences are expressed. The discrepancies in the results between experiments could be due to statistical variations in experimental procedure and replicate experiments would need to be performed to define more significant numbers. However, comparison of the identical DNA blots in Figure 7, A and B, indicates a clear reduction in group 2 5S rDNA in all of the genotrophs tested. The differences obtained in the copy number determinations described previously and those in this report could be due to experimental differences. First, this study employed high stringency
hybridization to distinguish between groups 1 and 2; the former studies were carried out at lower stringency (Tm - 150'). Therefore the previous studies may have been assaying variations in other, as yet uncharacterized, 5S rDNA genes. Second, the tissue culture derived probe is of unknown sequence complexity. The probe may have contained several different classes of 5S rRNA. This would be similar to the effect of using reduced stringency as stated above, resulting in the detection of variation in multiple groups of 5S rDNA. Further experimentation will be required to clarify these issues.

The low representation of group 2 5S rDNA in all of the genotrophs tested in Figure 7, A suggests that this quantitative polymorphism may be associated with environmental induction of heritable changes. The DNA used for genotroph C3 is from the first generation after induction. Therefore the reduction in 5S rDNA must have occurred close to the time of induction. However, stochastic fluctuations in the copy numbers of repetitive genes due to unequal crossing over events have been documented for genes such as rDNA (Tartof 1975). This type of mechanism could have generated the observed reductions. Experiments assaying the copy number of group 2 sequences need to be conducted in Pl
plants undergoing induction and the first generation progeny, to definitively determine if the environment has a heritable effect on 5S rDNA redundancy.

2. Characterization of 5S rDNA Restriction Fragment Length Polymorphisms.

Examination of the 5S rDNA portions of the 11 genomic clones isolated in section 3, A1, revealed that clones 5S-12 and 5S-20 produced RFLP’s between P1 and genotroph DNAs (Figure 8). High stringency hybridizations indicated that the RFLPs were more closely related to clone 5S-20 than to 5S-12 (Figure 9). The 706 bp 5S rDNA repeats found in 5S-20 display some interesting features not typical of most 5S gene families. The 706 bp repeat exists as a heterodimer composed of two 353 bp repeats which are only 76% identical (Figures 9 and 10). This form of heterodimer repeat is unusual in that most 5S gene tandem arrays are composed of nearly identical repeats. A similar organization has been described for Xenopus laevis oocyte genes, in which the repeat unit contains a divergent pseudogene juxtaposed to a functional gene and spacer sequence (Miller et al. 1978; Fedoroff 1979). The two 353 bp repeats which compose the 706 bp repeat
predominantly differ in single base pair changes with possible one base pair insertions in each repeat (Figure 9).

Another interesting feature of group 4 5S rDNA is the high level of sequence heterogeneity. The amount of sequence divergence in the transcription unit is greater than any observed within plant 5S rRNA sequences described to date (Erdmann and Wolters 1986). The degree of heterogeneity displayed by this group of 5S rRNA genes is clearly seen in the DNA gel blots shown in Figure 11. Typically, tandemly arrayed repetitive genes, such as rDNA and 5S rRNA genes, are maintained in a concerted fashion and do not show large amounts of restriction enzyme site polymorphism (Long and Dawid 1980). A similar situation is observed for pRS20.7 homologous sequences with enzymes such as BamHI which is highly conserved in plant 5S rRNA genes (Figure 4, lane d). However, other enzymes such as EcoRI, DraI, and EcoRV, produce fragments which do not fall into a multimer repeat size (Figure 11, panels A, B and C). This indicates the pRS20.7 gene family is heterogeneous with respect to sequence and repeat sizes and may not be subject to the sequence homogenization characteristic of other tandemly arrayed genes. This interpretation is strengthened by the analysis of pRS20.7 spacer sequences
shown in Figure 15 (Dover 1982). Comparison of the hybridization patterns of the A and B spacer probes illustrates differences in representation and organization, with each spacer detecting bands exclusive of the other spacer sequence. This indicates that the arrangement of the A and B spacers displayed in Figures 9 and 10 is not completely representative of the entire gene family. Analysis of another group 4 family member, p651, demonstrates that length variation is partly due to deletions in the 5S1 spacer sequence (Figure 22). Clone p651 is also organized in a fashion similar to pRS20.7, except that this repeat lacks BamHI sites. Numerous spacer deletions have been shown to cause length variation in wheat 5S rRNA genes (Scowles et al. 1988). However, in wheat extensive sequence divergence between repeats is not evident between members of the same class. Comparison of the DNA sequences of pRS20.7 and p651 shows 85% homology (Figures 21 and 22). Here, the sequence divergence takes the form of single base pair mutations as well as the 55 bp deletion in the 5S1 spacer. These results taken together show that group 4 flax 5S rRNA genes are highly variable.

Analysis of the predicted 5S rRNA secondary structure for both 5S1 and 5S2 of pRS20.7 and p651
indicates poor conservation of the stem loop domains thought to be important for 5S rRNA structure (Figure 6; Erdmann and Wolters 1986). This result in addition to the high degree of divergence of pRS20.7 and p651 from the majority of flax 5S rRNA genes (as represented by pBG13) strongly suggests that group 4 may represent a pseudogene class. As discussed in section 4A, this may have implications for the lack of conservation of this group of 5S rDNA sequences with respect to others in the genome (Dover 1982).

3. **Induction of Small Genotrophs is Associated with Specific RFLPs.**

A defined set of RFLPs was observed when the 706 bp probe hybridization pattern for P1 was compared with that for the large and small genotrophs. An identical pattern was observed for all restriction enzymes tested on four small genotrophs (Figure 11). This result is of particular significance for several reasons. The four small genotrophs analyzed were induced in different environmental induction experiments (Durrant 1962; Cullis 1977; Cullis 1981). In addition, all four share common characteristics including reduced plant weight, plant height, DNA content, copy number of rDNA genes,
and an altered peroxidase isozyme band pattern (Cullis 1977; Cullis 1981). Thus, a specific DNA RFLP pattern is associated with the induction of small genotrophs.

That an induction event is required to generate these polymorphisms is evidenced by the derivation of the small genotroph L6 from a large genotroph L1. It has been observed that maintenance of stability for some genotroph lines requires growth under defined stabilizing conditions (Durrant 1962). Under other conditions several genotroph lines have been observed to still be plastic and capable of heritable change (Jorder et al. 1975). L6, a small genotroph, was derived from the large genotroph L1 after growth outdoors for six generations (Cullis 1977). DNA extracted from L3 plants contains the same pRS20.7 polymorphisms detected in L1 (data not shown for 5S rDNA; see Figure 36 for LIS-1 RFLP). These plants had retained the large genotroph phenotype. However, there was a reduction in the total nuclear DNA content of L3 plants (Durrant and Jones 1971; Jorder et al. 1975). After three more generations of growth outdoors the resulting L6 plants changed to a small phenotype as defined by lower DNA content, lower rDNA amount, reduced plant weight and height, concomitant with the acquisition of the small genotroph pRS20.7 RFLP pattern (Cullis 1977; Figure 3,
lane 5). An example of the stability of the small genotroph pattern is shown in a plant series derived from S1. Genotrophs S3 and S6 are sublines separated from S1 by three and six generations (respectively) of growth outdoors. All three show the same small RFLP pattern (compare S1 and S6, Figure 3 lanes 2 and 3, S3 not shown). This indicates that pRS20.7 polymorphisms are rapidly generated in Stormont Cirrus only in response to an induction event in which plant stature is altered. The instability of the large pattern and the stability of the small pattern also suggests that different states of genomic plasticity/stability exist which may be differentially affected by environmental conditions. The detection of RFLPs which are identical in the small genotrophs suggests that this DNA alteration is specific to the induction of small genotrophs. In addition, RFLPs are detected in all small and large genotrophs indicating that the 706 bp probe may represent a valuable diagnostic marker for stress-induced heritable changes in flax.

Although the exact nature of the RFLPs is unknown, the results indicate that a DNA rearrangement, such as sequence rearrangement, deletion or amplification may be responsible for the altered band patterns. It is unlikely that the RFLP pattern is due to point mutations
since the patterns are observed with many restriction enzymes (Figure 11). The polymorphic pattern is not lost when genomic DNAs are digested with methylation insensitive enzymes such as DraI (Figure 11, panel C). Methylation independent identification of RFLPs was formally demonstrated by use of the isoschizomer restriction enzyme pair EcoRII/BstNI (Figure 12). In addition, due to the enzymes used in this study, both specific adenine and cytosine methylation would be required to explain the polymorphisms as the result of DNA modification. If DNA modification is responsible, then the modification is regulated and heritable for many generations. This situation would be similar to an imprinting mechanism except that sexual transmission does not affect the RFLP patterns (Kermicle 1978). The results of FIGE analysis of high molecular weight polymorphisms indicate that DNA amplification and/or recombination may be responsible for the new bands in LH.

Linkage data from L X S F2s show that all of the RFLPs are tightly linked (Figures 16 and 17). Analysis of 23 F2 individuals did not show any recombination of polymorphic bands. Therefore, the polymorphisms constitute a single chromosome locus. Due to the localization of the polymorphisms to this region, the
locus has been named Flax polymorphic locus 1 (Flp-1). Since segregation of the RFLPs is not observed, the rearrangements must be confined to this chromosomal region. This conclusion is supported by experiments which addressed the physical size of the locus (section 3, C6). Megabase DNA blot analysis showed that all of the group 4 sequences are contiguous in a single MluI restriction fragment of approximately 2.0 Mb (Figure 19-21). This size could comfortably contain the approximately 500 copies (equal to 350 kb) of group 4 5S rDNA which includes both polymorphic and nonpolymorphic sequences.

The organization of the group 4 sequences in the Flp-1 locus is unknown at present. However, comparison of the hybridization patterns identified in PacI digests of P1 DNA with the A spacer probe and pRS20.7 are informative (Figure 19, B and C). Hybridizations with pRS20.7 detect a 210 kb fragment which is not identified by the A spacer probe. As discussed above, the A spacer is present in a subset of pRS20.7 repeats (Figure 15). The megabase DNA results suggest that this subset of group 4 5S rDNA may form a contiguous length of DNA of approximately 210 kb (Figure 19, B and C). This type of arrangement would be in contrast to a dispersal of the A spacer sequence at random among the other repeats in the
locus. A dispersed organization would not permit exclusive detection of sequences containing only the A spacer. While further experimentation is required to substantiate this type of organization, the results of the spacer probe and megabase DNA analysis clearly show a complex organization for group 4 5S rDNA. Further analysis of the higher order structure of the Flp-1 locus may provide insights to the nature of the DNA rearrangements occurring at this locus.

C. Identification and Characterization of Insertion Sequence Polymorphisms.

The previous section discussed alterations in 5S rDNA which were associated with the induction of L and S genotrophs. Although a great deal has been learned about these rearrangements, the exact mechanism awaits clarification. The experiments discussed in this section were initiated in order to isolate low copy DNA polymorphisms which could be used to determine a specific mechanism for their production. Identification of a specific mechanism would allow the design of straight-forward tests for its involvement with environmental induction of heritable changes. To this end, the low copy, non-5S rDNA junction fragments of the
11 EMBL4 clones isolated in section 3, A1, were used to try and identify RFLPs. Of the 11 clones, only the 5S-20 flanking sequence identified a low copy number RFLP. The results described in section 3D clearly indicate that the RFLP originally identified by the flanking sequence of 5S-20 (p2010.1) is due to the presence of a 5.6 kb insertion sequence named LIS-1 (Linum Insertion Sequence - 1; Figure 33). LIS-1 was found to be inserted into a 3.7 kb Pl fragment in the genotrophs L^H, L6, S1, S3, S6, S^h and C3. Insertion sequences have been commonly found to represent transposable DNA elements (TEs) (Shapiro 1983). LIS-1 shares several properties with other transposable elements. The first is target site duplication. The sequence TCC is duplicated at the site of insertion and is present at the right and left borders of the insertion (Figure 35). Three base pair duplications have also been identified for Tam1 in *Antirrhinum majus* and a soybean insertion sequence (Bonas et al. 1984; Vodkin et al 1983). Another characteristic of many, but not all transposable elements is the presence of short 5 - 20 bp inverted repeats at the termini of the element (Shapiro 1983). These DNA motifs are thought to be important for transposition, and their deletion can result in a defective element (Freeling
1984). Sequence determination of 19 base pairs at each terminus of LIS-1 did not reveal any typical terminal inverted repeats in this sampling of bases. There are several inverted repeats of the sequence TCC (GAA) present in the right most terminus of LIS-1 in addition to the surrounding P1 target sequence. The significance of these sequences is not known at this point, but they may represent relics of terminal inverted repeats. A criterion commonly used as evidence of transposition is the recovery of revertant alleles which have excised the element (Freeling 1984; Martin et al. 1985). Observation of small amounts of a 3.7 kb fragment in S1 lines may be evidence of transposition of LIS-1 resulting in production of the P1 sized fragment (Figures 30, 31 and 35). The origin of the 3.7 kb fragment in S1 does not appear to be due to contamination of DNA samples since P1 group 4 5S rDNA bands are not observed in S1 DNA samples (Figure 11). Thus, although further experimentation is required, LIS-1 appears to resemble other eukaryotic TEs. The observation that LIS-1 appears to be associated with both large and small genotrophs is similar to the results obtained for the group-4 5S rDNA. However, in this case LIS-1 is also present at the same position in $L^H$. In addition, a 20 kb RFLP is also detected by LIS-1
at low stringency in the genotrophs L1, L3, L6, S1, S3, S6 and C3 (Figure 37). As described above, the derivation of the small genotroph L6 strongly suggests that the appearance of the LIS-1 insertion is associated with environmental induction. In addition, four separate C3 individuals have been tested which are all first generation progeny of induced plants. Each of these C3 individuals has the LIS-1 insertion. Therefore, the insertion must have occurred close to the time of induction. Thus, the results obtained from these studies indicate that LIS-1 detects low copy RFLPs which are associated with environmentally induced flax genotrophs, the first such low copy DNA polymorphism to be described.

D. Implications of Genotroph Specific Polymorphisms on Models for Environmental Induction of Heritable Changes in Flax; One Mechanism or Many?

The mechanism(s) that are responsible for the described DNA alterations are not clear. Large changes in the DNA content and copy number of repetitive sequences have been documented in flax genotrophs, suggesting deletion and amplification of large amounts of DNA (Cullis and Cleary 1986). Specific deletions of 5S rDNA involving
the loss of a cluster 5S rDNA repeats resistant to cleavage by TaqI has also been associated with a reduction in 5S rDNA copy number (Cullis and Cleary 1986). Thus, based on previous studies and the nature of the 5S rDNA RFLPs described above, it seems likely that a DNA rearrangement is responsible for the observed polymorphisms. The spacer regions of the 706 bp repeat differentiate between bands which appear to be identical when the entire repeat is used as a probe (Figure 15). This result may indicate that these bands represent a heterogeneous population of fragments of a specific size which contain different spacer regions. Differential modulation of these fragments as a result of deletion, amplification or copy correction (gene conversion) may result in different representations in the genotrophs. Copy number determinations for the entire 706 bp of Pl, L^H and S^H did not show a conclusive change (Table V). However, the results presented in Figure 14 suggest that copy number alterations have occurred in this gene family. Comparison of the MluI, megabase DNA hybridization patterns obtained in Pl, L^H and S1, with pRS20.7 and p651, indicate that the RFLPs observed in conventional DNA blots do not involve large scale changes which would change the size or the number of MluI restriction fragments. The observation of a small
chromosomal region encompassing the 5S rDNA RFLPs and invariant group 4 sequences places certain restrictions on the type of DNA alterations which can be envisioned for the production of the RFLPs. In effect, the mechanism must not result in the relocation of DNA sequences outside of the 2.0 Mb MluI fragment or to other nonhomologous chromosomes. Therefore processes such as transposition, recombination or chromosomal restructuring which would disrupt the integrity of the locus are unlikely. Alternatively mechanisms such as deletion, amplification, unequal crossing over, and gene conversion within in the limits of the locus are distinct possibilities. Further experimentation is required to determine the nature of these complex DNA alterations.

The results from hybridizations to other Linum species and flax varieties show that similar types of DNA polymorphisms exist in these lines and that the RFLPs are not the result of an artifact in the induction experiments. The mechanism(s) by which these polymorphisms arose is unknown, but may be due to the occurrence of "induction events" similar to those in Stormont Cirrus. Identification of specific "fingerprint" like polymorphisms in flax species and varieties suggests that this probe may also be useful in
breeding or mapping experiments. While a specific mechanism was not determined for the 5S rDNA RFLPs, analysis of low copy DNA polymorphisms indicates that a transposition event is responsible (Figure 33). The identification of a transposable element-like insertion, associated with flax genotrophs, is of particular interest. Transposable elements have been shown to be activated by various environmental and cellular stresses (McClintock, 1984; Carpenter et al. 1987; Freeling 1984). Furthermore, transposable elements have been shown to cause chromosomal breakage and high rates of mutation (McClintock 1984; Freeling 1984). The molecular changes associated with environmental induction of heritable changes in flax bear resemblance to many of the changes produced by transposable elements. Future studies of LIS-1 should allow determination of the involvement of this sequence in environmentally induced heritable changes.

The possibility of selection of genetic variants at the whole plant level for the production of the 5S rDNA and LIS-1 RFLPs is unlikely, since all plants undergoing induction survive, reproduce and their progeny are alike (Durrant 1962; Cullis 1977). Furthermore the S pattern and the LIS-1 RFLP are detected in four individual C3 plants which are each first generation progeny of
different, simultaneously induced parent P1 plants. Thus, the changes are occurring in all plants undergoing induction and are not due to selection of preexisting variation in P1. Since Stormont Cirrus is self pollinating and the RFLPs are homozygous, the changes may be occurring in all of the cells of a meristem before the development of gametes such that both micro and megagametophytes contain the altered DNA. Alternatively, there may be a selection for cells which contain specific alterations during the course of growth under inducing conditions. This somatic selection again results in a homogeneous meristem before gametogenesis. Since the cells undergoing induction are diploid and there is no evidence for heterozygosity in P1, the alterations must be occurring on both homologues before gametogenesis. This can occur either by independent changes in both homologous chromosomes, chromosome loss, or gene conversion. Thirdly, gametophytic selection may result in preferential transmission of the altered genotypes. Previous data indicates that both nuclear DNA content and rDNA gene copy numbers change during growth under inducing conditions (Evans, Durrant and Rees 1966; Cullis and Charlton 1981). Experiments directed at determination of the RFLP pattern during growth under inducing conditions will clarify these possibilities.
The appearance of homozygous RFLPs is consistent with the changes in peroxidase band patterns. In this case, the changes in relative mobilities of anionic peroxidase isozymes is controlled by dominant and recessive alleles in L and S, respectively, which show simple 3:1 Mendelian inheritance (Tyson et al. 1978). The specificity and reproducibility of the observed RFLP patterns could possibly indicate a limited repertoire of potential rearrangements. However, the group 4 5S rDNA patterns for the large genotrophs show that more than one type of rearrangement can occur.

How the environmental conditions are responsible for this very specific set of polymorphisms is not yet known. One possibility is that the new arrangement is advantageous in the new environment, but the basis for this "adaptation" is not known. Generation of variation in this manner would clearly be advantageous to inbreeding plants with limited genetic variability. On the other hand, the genotypes of the plastic varieties may be unstable under certain environmental conditions. Instabilities such as this are common with traits resulting from insertion sequences (Peterson 1988). Recent evidence indicates that such instabilities may result in "adaptive" variation in prokaryotes (Hall 1988). However, to date there is no
evidence to suggest that any of the environmentally induced characters is directly advantageous to the genotrophs. Physical characterization of the pRS20.7 and LIS-1 RFLPs and surrounding sequences will aid in determining the type of alteration as well as where and when in the development of the plant the changes are occurring during induction. In addition, studies directed at the expression of genes in the vicinity of the rearrangements will help clarify whether any adaptive advantage accrues through these alterations. These experiments will be important in gaining an understanding of environmentally induced heritable change in plants.
BIBLIOGRAPHY


Kretz, P.L. and J.M. Short. _1989._ Gigapack\textsuperscript{TM} II: Restriction free (hsd\textsuperscript{−}, mcrA\textsuperscript{−}, mcrB\textsuperscript{−}, mrr\textsuperscript{−}) packaging extracts. Stratagies 2: 25-26.


Martin, C., R. Carpenter, H. Sommer, H. Saedler and
E.S. Coen. 1985. Molecular analysis of instability in flower pigmentation of Antirrhinum majus, following isolation of the pallida locus by transposon tagging. EMBO J. 4: 1625-1630.


Romaniuk, P.J., I. Leal de Stevenson and Qimin You. 1989. The specificity of the RNA binding activity of


Appendix A. 5S rRNA Gene Sequences.

pRS18 Sequence

1 GGGTGCATCATAACCAGACTAATGCAACCGGATCCATCGAGACTCCGAA
GTAAAACGTCTGGGCCAGACTAGTACTAAGATAGGTGACCTCTTGGAG
AGTCTGCATGACCCCCCTCTCCTTTGCAATTTTTTCGCCGGGACGTTG
TGACACAATATTAGCGAGTTTACATCGTTTATATATTACATCGAATTAAAC
AGCGAAGCCGGTGGGGGAAGGTTTTACGTCGACGAGGGGAGAAGGCA
AAATGTAATAAAAAGAAAGGAAAGCATGAAATACAAAGGGAGAGAATG
TGAGGCGGGGTGTTCAAATAGCCGTGATCGTGAGACGTTGGAC 342

pRS13 Sequence.

1 GGGTGCATCATAACCAGACTAATGCAACCGGATCCATCGAGACTCCGAA
GATAAACGTCTGGGCCAGAGTACTAAGATTGGTACCTCTTGGAG
AGTCTGCATGACCCCCCTCTCCTCTACTACAAAAATCTTTTTGGCCCGTGCT
CGTCACTCCTTCAACGATATAAAAATCATGTATATTCTCTACATTACGAC
GTAGGGGACAGCCAGACGTGGTTAAGTCCTCCGAGGGTTCCGGAAAGGGGC
CAATCAGGTGTTCAAGATAAAAAACGGGAGAAGGCAAGGAGGAGAAGGAG
AGTGCAGGGTTGTTGATATACACATGAGGGTGGACGATGGAC 344

pRS13 Sequence.

1 GGGTGAGATCATACCGCCGACCTAATCCACCCGTCATCCTACTCCCGAA
CATAAAAGTCCTGGCGAGGAGTCTACTAAGATGGTACCTCTTGGATC
CGTCCCTTTATCCACCTTCTCTTGTATTATAATTTTTATTTATTTTATTTT
TTCTCGAGGTTGATGGAAATATTTTCAAGCAACATCATCTACGAGAAA
TTTCGCAAATAGCCGCGGGTTAGGCGGATGCGGGTTTTGTTTGCAGACTCCG
CGGAAAAACAGCGCAAGCATAGTAAGGATACGAAAGGAGGAGTTGAAAT
AAAGGGGAGGATTAGTGAAAAGACTAGTTAATGTACATATGAGGCTGAGGA
GCAGAA 356

p651, 5S1 Sequence.

1 GAGTGAGGTCTATAACCGACGACTAATCCACCCGATTTCCATCTCATCTTTTGA
GTTAAACGTCTGGGCCAGACTAGTACTGATGGTACCTCTTGGATC
TATCCTCCGAGATACACCTTTTATCTATTCTTTTGTATTTTTTTAC
CCTAAGGGTAAAGTGGACGTATTCATAAACAACCIATCTTATGAGAATCT
GAGTTATGAATTGCCGAAACCGGTGCTGGGGTTAGGTATGCTCGTG
AGGAAGTAAGGCAAGAACATTTCATGTTACTTATGAGCACCGAGAGCGGAT 299
p651, 5S2 Sequence.

1 GAGGGAGATCACACCAACACACTAATTCATCGGATCTCGTCATCCTCTTTAA
   ATGGAAGCTCTTGGGTATGGTACTACGATGAGTTACCTTTTTGATC
   CATTCTCCTTGACACCTCTCTTTTATCTCTTTTTTGGATATTTTTAACC
   CGAAGGGATGTGAATTGAAACGGTTAAGTAAAAACATCTCTCATCGGAACTT
   ACGTAATGAATCCGGGAGGAGCCATGAGTGCGATTAGTTGTTCTCTCGG
   GAACAGAGACCAGATGCTAGTAAAAAGTAGATCGAAAATGAGTGAAAAAGAA
   GGGCGGAAATAGTGATAAATACATTTCGTCATGACATGACATGACAAACGAC
   AC 352

p6XA2 Sequence.

1 TGGTGAGATCACACCAACACACTAATTCATCGGATCTCGTCATCCTCTTTAA
   CAGATACGACTTGGCTGAGGATGACTAAAGAATAGGGTACCTATTTCA
   CGTCTCTCTGTGGACCCCTTCTCATCTATTTTTGATGCTATTTTTTT
   CTCGAGCTCGAGGTGAAATGGAACATTTAACCTAGAAAAACATCTCTCTAGCA
   GAACCTACGGAAATAAGCCCGGGGAGGAGGAGGTAGGAGTTTACCCTACGGC
   CCAACGGAAACCAGACCAAGACACTGGAAAAGTAGAAGAAAAAAGACCAA
   AAAGGAAAAGGAGGAAGTCTGAAAGACTAGTTAATCTACATATAACCCTG
   AAGAAGCACAA 360